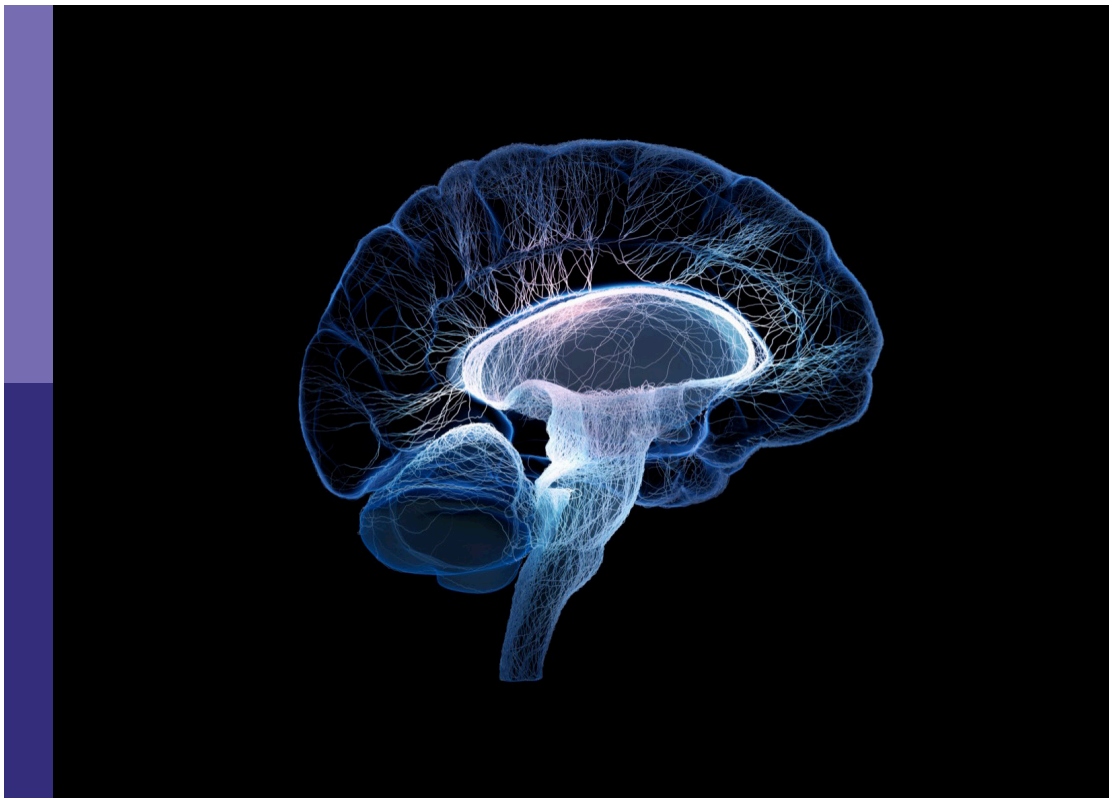
frontiers | Research Topics

# Listening with two ears – new insights and perspectives in binaural research

**Edited by**
Huiming Zhang, Yi Zhou and Lina Reiss

**Published in**
Frontiers in Neuroscience
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-3982-8
DOI 10.3389/978-2-8325-3982-8

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Listening with two ears – new insights and perspectives in binaural research

**Topic editors**

Huiming Zhang — University of Windsor, Canada

Yi Zhou — Arizona State University, United States

Lina Reiss — Oregon Health and Science University, United States

**Citation**

Zhang, H., Zhou, Y., Reiss, L., eds. (2023). *Listening with two ears – new insights and perspectives in binaural research*. Lausanne: Frontiers Media SA.

doi: 10.3389/978-2-8325-3982-8

# Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 07 November 2023
DOI 10.3389/fnins.2023.1323330



OPEN ACCESS

EDITED AND REVIEWED BY
Isabelle Peretz,
Montreal University, Canada

*CORRESPONDENCE
Huiming Zhang
✉ hzhang@uwindsor.ca

RECEIVED 17 October 2023
ACCEPTED 27 October 2023
PUBLISHED 07 November 2023

CITATION
Zhang H, Zhou Y and Reiss L (2023) Editorial: Listening with two ears – new insights and perspectives in binaural research.
*Front. Neurosci.* 17:1323330.
doi: 10.3389/fnins.2023.1323330



# Editorial: Listening with two ears – new insights and perspectives in binaural research

Huiming Zhang[1]*, Yi Zhou[2] and Lina Reiss[3]

[1]Department of Biomedical Sciences, University of Windsor, Windsor, ON, Canada, [2]College of Health Solutions, Arizona State University, Tempe, AZ, United States, [3]Department of Otolaryngology, Oregon Health and Science University, Portland, OR, United States

KEYWORDS

**binaural integration, sound localization, spatial release from masking, binaural fusion, auditory plasticity, auditory scene analysis, hearing loss, cochlear implant**

Editorial on the Research Topic
Listening with two ears – new insights and perspectives in binaural research

While advantages of seeing with two eyes (i.e., binocular vision) were noted many centuries ago by ancient Greek scholars including Klaudios Ptolemaios (c. 100–c. 178CE), those of hearing with two ears (i.e., binaural hearing) were not reported until the end of the 18th century (Wells, 1792; Venturi, 1796, 1802). Great strides were made in the study of binaural hearing after the "Duplex Theory" of sound localization (Strutt, 1907), i.e., the involvement of both the interaural-level and the interaural-time difference (ILD and ITD), was established at the beginning of the last century. Major discoveries provided insight into some important aspects of binaural hearing including neural bases of sound localization (e.g., Jeffress, 1948; Goldberg and Brown, 1969; Colburn and Durlach, 1978; Durlach and Colburn, 1978; Moiseff and Konishi, 1981; Yin and Chan, 1990; Blauert, 1996). These early studies paved the way for addressing a wide range of questions related to functions and mechanisms of binaural hearing. Among these questions is how spatial cues can be used to aid in the detection of a sound in a noisy environment. Other important questions include how speech perception is dependent on the integration of temporal and spectral acoustic information received by the two ears, and how binaural hearing can be shaped by auditory experience. Recently, significant progress has been made in understanding disorders in binaural hearing, i.e., abnormal conditions related to alterations of central binaural integration rather than peripheral cochlear damage. Some of these latest findings are highlighted in the eighteen original research articles published on the present Research Topic.

## 1. Preview of studies on the present Research Topic

### 1.1. Binaural hearing in normal systems: spatial release from masking

One benefit of binaural hearing is that it aids in the recognition of a sound. In a natural acoustic environment, the detection and perception of a sound of interest can be masked by a background noise (Gelfand, 2004). A spatial separation between the two sounds can reduce the effect of masking, resulting in spatial release from masking (SRM) (Plomp and Mimpen, 1981; Saberi et al., 1991). A related phenomenon is the binaural masking-level difference

(BMLD), in which the detection of a sound is improved when the phases of the sound at the two ears become different from those of a masker (Licklider, 1948).

SRM and BMLD were investigated in five studies in this Research Topic. Asim et al. demonstrated in the rat midbrain that neurophysiological responses of an ensemble of neurons to a sound could be suppressed by a preceding sound and the effect was only mildly dependent on local excitation/inhibition. Such a suppressive effect could be reduced by a spatial separation between the sounds, which was reminiscent of SRM. Fan et al. measured responses to diotic and dichotic tone-in-noise stimuli from individual neurons in the midbrain of awake rabbits and revealed that BMLD was related more to interaural correlation between sounds at the two ears than to ITD or ILD. Using a modeling/simulation approach, Smith et al. trained an artificial neural network to yield a BMLD performance that matched the performance of human listeners. Functions of inner nodes of the model resembled interaural correlation functions observed in animal neurophysiological studies, suggesting that BMLD is dependent on interaural correlation.

Previous investigations of masking and SRM have been conducted only under anechoic conditions and have not considered stimulus statistics. Biberger and Ewert extended such investigations to more complex environments by examining how factors such as room reverberation affected detection and quality perception of a target sound in the presence of colocalized or spatially separated maskers. Encke and Dietz characterized the interaural statistics of tone-in-noise stimuli, providing a basis for future studies of the relationship between these statistics and SRM.

## 1.2. Sound localization in abnormal/disordered systems

Understanding how sound localization is affected by hearing loss and other disorders can not only help develop clinical approaches to deal with such problems, but also provide insights into mechanisms underlying normal binaural hearing. Four studies in this Research Topic examined how sound localization was affected by aging, stroke, tinnitus, and replacement of natural acoustic stimulation by electrical stimulation generated by cochlear implants (CIs).

Previous studies have reported worsening of sound localization abilities in aging populations (see Russell, 2022 for review). In this Research Topic, Eddins et al. used electroencephalography to demonstrate that the processing of ITD was more heavily dependent on the activation of the contralateral than the ipsilateral auditory cortex. This asymmetry along with across-hemisphere differences in response waveform over specific time windows was reduced with age, which may be among the factors affecting the sensitivity to ITD in older adults. Dietze et al. found that lesions of specific brain regions caused by ischemic stroke impaired sound lateralization, with the impairment manifested in different ways depending on lesion sites. Specifically, brainstem lesions caused compressed and distorted response choices in lateralization, thalamic lesions led to a shift of perceived auditory space, and cortical lesions resulted in strong effects on lateralization of stimuli contralateral to the lesion. Long et al.s' study on sound-localization abilities in listeners with tinnitus showed that tinnitus percepts could affect localization of tones but not words. Future work is needed to determine the structure(s) within the auditory pathway that is/are responsible for such interference.

The acuity of sound-source localization, especially that based on ITD cues, is known to be significantly reduced in individuals with bilateral CIs (see Laback et al., 2015 for review). Müller et al. investigated this phenomenon using neurophysiological recordings and mathematical modeling/simulation. They revealed that sensitivities of neurons in the lateral superior olivary nucleus (LSO) to ITD were dependent on the temporal precision of spiking of inputs to the LSO from lower brainstem structures. In comparison to neural inputs to the LSO driven by acoustic stimulation, those driven by electrical stimulation (e.g., generated by CIs) exhibited hyper precision and low jitter, which led to reduced sensitivity to ITD in olivary neurons. This finding suggests that localization ability based on ITD can be improved by introducing jitter into stimulation generated by CIs.

## 1.3. Dependence of speech perception on binaural integration in normal and impaired auditory systems

A notable gap in literature exists regarding how speech perception depends on the integration of acoustic (including spectral) cues received by the two ears. Six studies in this Research Topic investigated effects of perturbation of this integration on speech perception.

Two studies used simulation to create asymmetries of inputs in normal-hearing listeners. Yoon and Morgan revealed that consonant recognition was possible even if large amounts of spectral information were missing at individual ears, as long as complementary information could be integrated across ears. This finding suggests that effective bimodal hearing (i.e., with one ear having a CI while the contralateral ear having acoustic hearing) can be achieved when the implanted ear is provided with information within a frequency range that complements rather than overlaps that of the contralateral ear. Anderson et al. used a vocoder to simulate CI processing and manipulate the dynamic range of speech at each ear to create a "better ear" and a "poorer ear". Decreasing the dynamic range in one ear led to increased binaural interference for single words, whereas for dichotic double word presentations, this manipulation led to increased word fusion and blending. These findings suggest that increased binaural fusion due to dynamic range asymmetry can result in abnormal fusion and interference.

Abnormal fusion does occur in listeners with hearing loss and can lead to binaural interference as well as difficulties with speech understanding in a noisy environment. Oh et al. demonstrated in listeners with hearing aids that there was significant inter-subject variation in binaural pitch fusion, i.e., fusion of sounds with different pitches across ears. Broad binaural fusion was correlated with a reduced ability to use voice fundamental frequency differences in speech recognition

in the presence of background talkers. This correlation was also observed in normal-hearing listeners, suggesting that underlying mechanisms are of central rather than peripheral origin. Burg et al. examined listening effort in users of bilateral CIs. They found that listening effort increased when a poorer ear was used in addition to a better ear, suggesting negative consequences of binaural integration when asymmetries in hearing are present between ears.

Two other studies developed new methods with potential application for future studies of binaural integration. Dolhopiatenko and Nogueira demonstrated that decoding of selective auditory attention could be obtained in bimodal CI users using electroencephalography signals, despite the presence of stimulus artifacts from the CI in these signals. Chou et al. developed an algorithm based on a biologically inspired network to process both special and directional acoustic information driven by the two ears. This algorithm is able to segregate sounds based on spatial and spectral information and may also have applications in the development of hearing devices or software. Methods used in both studies provide researchers an opportunity to explore how binaural integration contributes to neural processing.

### 1.4. Brain plasticity: auditory training and the influence of auditory experience

Another emerging area of research is about how binaural hearing and underlying mechanisms are shaped by auditory experience. Nisha et al. showed in listeners with hearing loss that auditory training using stimuli delivered in a virtual acoustic space improved spatial acuity of sound localization. Ding et al. examined the detection of a binaural gap, i.e., a period without correlation between acoustic signals received by the two ears, in listeners with normal hearing. Performance was correlated with the sensitivity to temporal fine structure of monaural acoustic stimulation, and this correlation was reduced by musical training. Sanchez Jimenez et al. used the ferret as a model system to study plastic changes in sound-localization behaviors following unilateral conductive hearing loss. They found that training facilitated recovery of sound localization abilities. Recovery could generalize to more naturalistic listening conditions, so long as the target sounds provided sufficient spatial information.

## 2. Significance and future directions

The current Research Topic explored some exciting directions in the field of binaural hearing using both normal and disordered/clinically relevant systems. These studies provide new knowledge about functions and underlying mechanisms of some established binaural phenomena. They also show how binaural hearing can be shaped by auditory experience and provide new applications of electrophysiological tools and computational models. Despite these advances, many important questions remain to be answered. For instance, how does the brain use spatial along with temporal and spectral cues to stream and group information to form cohesive individual acoustic images? Conversely, how is this information used to segregate multiple acoustic images, as in the cocktail party effect? A multidisciplinary approach is needed to address these questions and help understand how the auditory scene is analyzed by the brain. Human psychoacoustical and animal behavioral experiments can improve our understanding of binaural hearing at the functional level. Neurophysiological recordings along with neurostimulation, and neuropharmcological or molecular/genetic manipulation conducted in normal and disordered systems may reveal key binaural components through gain-of-function and loss-of-function analyses. Mathematical models will be critical for simulating binaural components not easily measured/manipulated using experimental techniques. Taken together, multiple approaches integrated across studies as well as within studies will pave the way for future advances in the study of binaural hearing.

## Author contributions

HZ: Conceptualization, Writing—original draft, Writing—review & editing. YZ: Conceptualization, Writing—original draft, Writing—review & editing. LR: Conceptualization, Writing—original draft, Writing—review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This research is supported by a Discovery Grant from the Natural Science and Engineering Research Council (NSERC) of Canada to HZ and grants from the U.S. National Institutes of Health - National Institute of Deafness and Other Communication Disorders (NIH - NIDCD) to YZ (Grant No. RO1 DC019278) and LR (Grant No. RO1 DC013307).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References


Blauert, J. (1996). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press.

Colburn, H. S., and Durlach, N. I. (1978). "Models of binaural interaction: handbook of perception," in eds E. C. Carterette and M. P. Friedman (New York, NY: Elsevier Inc.), 467–518.

Durlach, N. I., and Colburn, H. S. (1978). "Binaural phenomena," in *Handbook of Perception*, eds E. C. Carterette and M. P. Friedman (New York, NY: Elsevier Inc.), 365–466.

Gelfand, S. A. (2004). *Hearing – An Introduction to Psychological and Physiological Acoustics 4th Edn*. New York, NY: Marcel Dekker.

Goldberg, J. M., and Brown, P. B. (1969). Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: some physiological mechanisms of sound localization. *J. Neurophysiol.* 32, 613–636. doi: 10.1152/jn.1969.32.4.613

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41, 35–39. doi: 10.1037/h0061495

Laback, B., Egger, K., and Majdak, P. (2015). Perception and coding of interaural time differences with bilateral cochlear implants. *Hear. Res.* 322, 138–150. doi: 10.1016/j.heares.2014.10.004

Licklider, J. C. R. (1948). The influence of interaural phase relations upon the masking of speech by white noise. *J. Acoust. Soc. Am.* 20, 150–159. doi: 10.1121/1.1906358

Moiseff, A., and Konishi, M. (1981). Neuronal and behavioral sensitivity to binaural time differences in the owl. *J. Neurosci.* 1, 40–48. doi: 10.1523/JNEUROSCI.01-01-00040.1981

Plomp, R., and Mimpen, A. M. (1981). Effect of the orientation of the speaker's head and the azimuth of a noise source on the speech reception threshold for sentences. *Acustica* 48, 325–328.

Russell, M. K. (2022). Age and auditory spatial perception in humans: review of behavioural firings and suggestions for future research. *Front. Psychol.* 13, 831670. doi: 10.3389/fpsyg.2022.831670

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., and Perrott, D. R. (1991). Free-field release from masking. *J. Acoust. Soc. Am.* 90, 1355–1370. doi: 10.1121/1.401927

Strutt, J. (1907). On our perception of the direction of a source of sound. *Philos. Mag.* 13, 214–323. doi: 10.1080/14786440709463595

Venturi, J. B. (1796). Considérations sur la connaissance de l'étendue que nous donne le sens de l'ouïe. *Mag. Encycl. J. Sci. Lett. Arts* 3, 29–37.

Venturi, J. B. (1802). Betrachtungen über die Erkenntniss der Entfernung, die wir durch dasWerkzeug des Gehörserhalten. *Archiv. Physiol.* 5, 383–392.

Wells, W. C. (1792). *An Essay Upon Single Vision With Two Eyes: Together With Experiments and Observations on Several Other Subjects in Optics*. London: Cadell.

Yin, T. C., and Chan, J. C. (1990). Interaural time sensitivity in medial superior olive of cat. *J. Neurophysiol.* 64, 465–488. doi: 10.1152/jn.1990.64.2.465

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 01 September 2022
DOI 10.3389/fnins.2022.957012



OPEN ACCESS

EDITED BY
Huiming Zhang,
University of Windsor, Canada

REVIEWED BY
Anuprasad Sreenivasan,
Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), India
Michael Akeroyd,
University of Nottingham, United Kingdom

*CORRESPONDENCE
Yu Ding
dingyuzero@mail.tsinghua.edu.cn
Chunmei Cao
caocm@tsinghua.edu.cn

†These authors have contributed equally to this work and share first authorship

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 30 May 2022
ACCEPTED 11 August 2022
PUBLISHED 01 September 2022

CITATION
Ding Y, Lei M and Cao C (2022) The relationship between interaural delay in binaural gap detection and sensitivity to temporal fine structure in young adults with or without musical training experience. *Front. Neurosci.* 16:957012. doi: 10.3389/fnins.2022.957012



# The relationship between interaural delay in binaural gap detection and sensitivity to temporal fine structure in young adults with or without musical training experience

Yu Ding[1*†], Ming Lei[2†] and Chunmei Cao[1*]

[1]Division of Sports Science and Physical Education, Tsinghua University, Beijing, China, [2]Laboratory of Artificial Intelligence and Cognition, School of Tourism Sciences, Beijing International Studies University, Beijing, China

Humans can detect the presence of a break in interaural correlation (BIC, also called binaural gap) even if a large interaural time delay (ITD) is introduced, which is important for detecting, recognizing, and localizing sounds in everyday environments. To investigate the relationship between interaural delay in binaural gap detection and the sensitivity of temporal fine structure (TFS), 40 young college students with normal hearing took the BIC delay threshold test, the TFS1 test (the test of monaural TFS sensitivity), and the TFS-AF test (the test of binaural TFS sensitivity). All participants were asked whether they had any musical training experience in their childhood. Results showed that the BIC delay threshold was significantly correlated with the TFS1 test ($r = -0.426$, $p = 0.006$), but not with the TFS-AF performance ($r = -0.005$, $p = 0.997$). The correlation between BIC delay threshold and monaural TFS sensitivity was observed in the non-music training group ($r = -0.508$, $p = 0.010$), but not in the music training group ($r = -0.290$, $p = 0.295$). These findings suggest that the interaural delay in binaural gap detection is related to the monaural sensitivity of TFS, this significant correlation was mainly found in young adults without musical training experience.

KEYWORDS

**temporal fine structure, binaural gap, break in interaural correlation, interaural delay, primitive auditory memory**

## Introduction

One of the benefits of having two ears is that binaural spatial cues can be obtained, as the time and intensity of the signal reaching both ears vary depending on the location of the sound source (Blauert, 1997; Schnupp et al., 2011). Extracting and integrating binaural information not only provides a basis for sound localization but is also crucial

for target detection and speech recognition in complex environments (Bronkhorst, 2000; Darwin, 2006; Eramudugolla et al., 2008; Maddox and Shinn-Cunningham, 2012). Binaural information processing involves both the binaural calculation of the similarity of the acoustic details (mostly temporal fine structure, TFS) between the two ears (Huang et al., 2009; Li et al., 2009) and the monaural/binaural sensitivity (Moore, 2014) of the TFS signal.

Usually, speech signals are decomposed into several narrowband signals in the cochlea, and these narrowband signals can be considered as a relatively slow variation in amplitude over time (envelope, ENV) and the rapid oscillations with a rate close to the center frequency of the band (TFS) (Moore and Sek, 2009). There are already many measurement methods for TFS sensitivity. Among them, monaural sensitivity to TFS can be measured with the "TFS1" test (Moore and Sek, 2009; Sęk and Moore, 2012), while binaural sensitivity to TFS can be assessed using the "TFS-AF" (TFS-adaptive frequency) test (Füllgrabe et al., 2017). TFS1 and TFS-AF tests are mature behavioral measurement methods and have been used in many studies (Füllgrabe et al., 2018; Tarnowska et al., 2019).

Listeners are very sensitive to the dynamic changes in interaural correlation, such as detecting a dynamic break in interaural correlation (BIC, also called BIAC or binaural gap, a brief drop of interaural correlation from 1 to 0 and then return to 1) in a steady-state noise, showing the marked ability to temporally resolve fast changes in interaural configurations (Akeroyd and Summerfield, 1999; Boehnke et al., 2002). Introducing a change in interaural correlation does not change the monaural energy and spectrum in the sound signals, but changes the loudness (Moore, 2003) and dichotic repetition pitch (Bilsen and Goldstein, 1974) of the signals. A study based on the frequency-following responses (FFRs) of the rat auditory midbrain found that introducing a BIC causes more reduction in $FFR_{TFS}$ than in $FFR_{Env}$ (Wang and Li, 2015), and an earlier study also showed that the ENV is not as important as the TFS in determining the BIC detection (Boehnke et al., 2002).

Furthermore, even if a large interaural time delay (ITD) is introduced, humans can still detect the presence of BIC (Huang et al., 2009; Li et al., 2013; Liu et al., 2016). The past studies associated with judging sidedness showed that laterality cues can be discriminated at large ITD, which indirectly measured the ability to detect interaural correlated sounds (Mossop and Culling, 1998). Results of early studies have suggested that the representation of the TFS may persist for up to 9–15 ms (Cherry, 1954; Blodgett et al., 1956; Langford and Jeffress, 1964; Mossop and Culling, 1998). The preservation of the sensitivity to the BIC even when a large ITD is introduced indicates that the TFS information of noise is maintained for the time of the ITD (Huang et al., 2008). Measuring the ITD when the BIC is detectable can provide a way of investigating the temporal storage of acoustic details, which is called the "BIC delay threshold" test (Huang et al., 2009). This faithful auditory storage of TFS has been recognized as the early point in the chain of the transient auditory memory system and termed primitive auditory memory (PAM) (Li et al., 2013; Kong et al., 2015; Liu et al., 2016).

The relationship between interaural delay in binaural gap detection and sensitivity to TFS may vary in populations with different characteristics. A recent study (age range 21–65 years) found that the BIC delay threshold and TFS-AF tests were significantly correlated in a tinnitus group but not in a normal group, since binaural integration may be more difficult due to overt/covert hearing loss with aging and tinnitus (Ding et al., 2022). However, both BIC delay threshold and TFS-AF tests are binaural-based tests that are likely to be affected by monaural coding of TFS information before binaural interaction (Whiteford et al., 2017). Furthermore, many young participants had musical training in their childhood, and music training is related to both monaural sensitivity (Mishra et al., 2015) and binaural sensitivity (Bianchi et al., 2019) of TFS. Therefore, it is unclear whether the BIC delay threshold is associated with monaural/binaural sensitivity of TFS in young adults with or without musical training experience. This research focuses on the relationship between the BIC delay threshold and the monaural/binaural sensitivity of TFS, investigating whether the performance of the BIC delay threshold is related to the performance of the TFS1 or TFS-AF test, considering childhood musical training experience as a potential influencing condition.

# Materials and methods

## Participants

Forty university students (21 males and 19 females; mean age = 22.93 between 18 and 30 years) with normal hearing participated in this study. To estimate the required number of participants, we used the results from the first 20 participants rather than any independent estimate from the literature or a pilot. We noted that the correlation coefficient for them between the BIC delay threshold and TFS1 scores was 0.42. Entering this into G-power gave 39 as the number of participants required to maintain this value of correlation in the whole data set for $\alpha = 0.05$, power = 0.8, and two tails (Faul et al., 2009). Their pure-tone thresholds were no more than 20 dB hearing level (HL) between 0.125 and 8 kHz (ANSI-S3.6, 2004) in each ear, and the threshold difference between the two ears in each frequency was less than 15 dB HL. All the participants gave their written consent to participate in the study and were paid a modest stipend for their participation. The study was approved by the Tsinghua University Ethics Committee.

All participants were asked whether they had any musical training experience in their childhood. The specific problems were stated as follows: Did you receive musical training (including professional instrumental or vocal training) in your

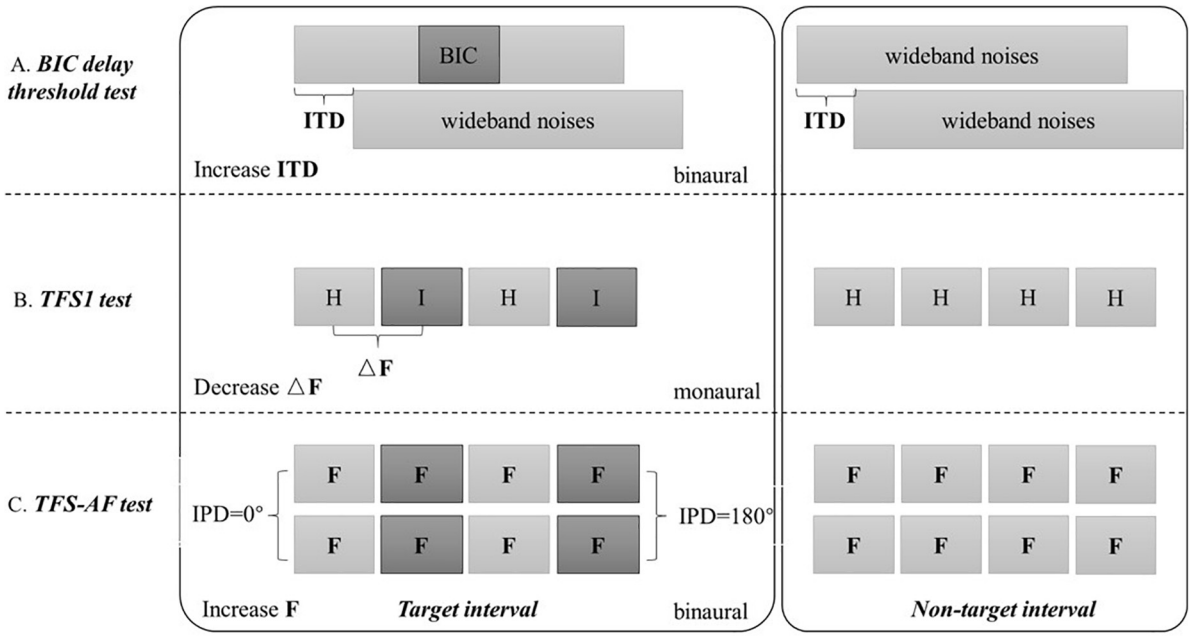



FIGURE 1
Schematic diagram of three test methods. The BIC delay threshold test, TFS1 test, and TFS-AF test use the maximum ITD, minimum ΔF/F0, and maximum F as thresholds, respectively.

childhood? What kind of music training did you have? When did you start musical training? How long did your music training last? Among them, 25 participants did not receive any musical training, and the other15 participants received musical training in their childhood (including 7 piano trainees, 3 guzheng trainees, 1 loner trainee, 1 electronic organ trainee, 1 harmonica trainee, and 2 vocal trainees). All musical trainees began before the age of 13 and the mean duration of their musical training was 5.93 ± 4.41 years.

## Apparatus and stimuli

All tests were carried out in a soundproof room where environmental noise was less than 29 dB SPL. All acoustic signals were calibrated by a sound-level meter (AUDit and System 824, Larson Davis, Provo, UT, United States), delivered by the Creative Sound Blaster (Sound Blaster X-Fi Surround 5.1 Pro, Creative Technology Ltd., Singapore), and presented to participants over two earpieces of Sennheiser HD 650 headphones. For the BIC delay threshold test, we performed a direct calibration on the generated noise. For TFS1 and TFS-AF tests, these two softwares have built-in calibration routines (Sęk and Moore, 2012), and we followed its procedure to calibrate.

All participants were tested for pure-tone hearing threshold first (125–8,000 Hz). The order of the three tests was randomized among participants. Before each test, there would be a practice phase to ensure that participants understood the experimental task (details of the practice phase are described below).

For the BIC delay threshold, 2,000 ms Gaussian wideband noises (including 30 ms rise-fall time, 60 dB SPL) were synthesized using the "randn()" function in the MATLAB (the Math Works Inc., Natick, MA, United States) at the sampling rate of 48 kHz with 16-bit resolution. The generated signals were then lowpass filtered at 10 kHz.

Two special software packages were used in this study to perform the testing of TFS1 (Sęk and Moore, 2012) and TFS-AF (Füllgrabe et al., 2017), which can be downloaded from the Internet.[1] Most of the parameters use the default settings.

## Break in interaural correlation delay threshold test

For consistency and reproducibility, the parameters and procedures of the BIC delay threshold test have been described in detail in multiple previous studies (Huang et al., 2009; Li et al., 2009, 2013; Lei and Ding, 2021; Ding et al., 2022). There were two kinds of signals; in one presentation, the left-headphone noise was an exact copy of the right-headphone noise. In the other presentation, the temporal middle of the left-headphone noise was substituted with a randomly selected independent

1 https://www.psychol.cam.ac.uk/hearing

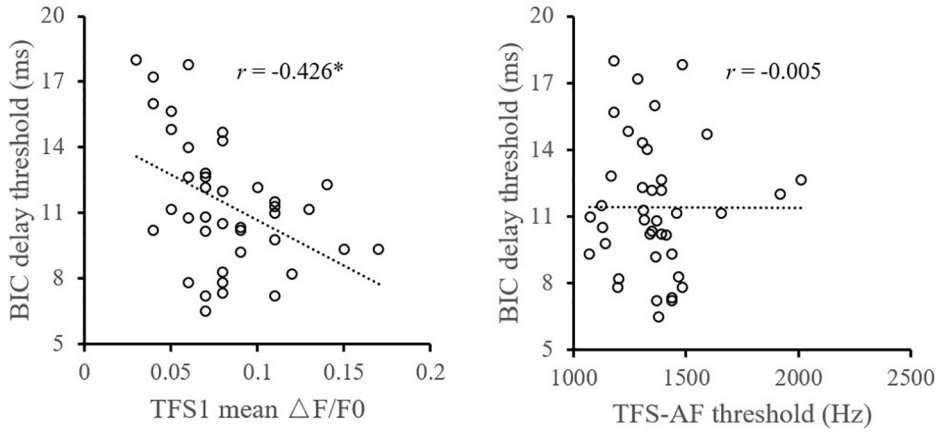



FIGURE 2
Illustration of the correlation analysis of the BIC delay threshold and the TFS1 and TFS-AF test. *Significant effect after Bonferroni's correction, $p < 0.025$.

noise fragment with a fixed duration of 200 ms before filtering, introducing a brief break of interaural correlation, from 1 to 0 and then returning to 1. In the practice phase, participants became familiarized with binaurally presented noise either with or without the BIC. The task was to identify which of the two presentations contained the BIC. The offset-to-onset interval between the two presentations was 500 ms.

The longest ITD for BIC detection was measured using a three-up-one-down paradigm (Levitt, 1971): The ITD increased following three consecutive correct choices of the presentation containing the BIC, and decreased following one incorrect choice. The initial step size was 16 ms, which was altered by a factor of 0.5 with each reversal of direction until the minimum size of 1 ms was reached, and the longest ITD was defined as the mean ITD for the last six reversals. Visual feedback was given after each trial to indicate whether the choice was correct or not.

## Test for monaural/binaural sensitivity of temporal fine structure

TFS1 (Moore and Sek, 2009) and TFS-AF (Füllgrabe et al., 2017) each used methods described in the references. The TFS1 test involved discrimination of a harmonic complex tone (H, with a fundamental frequency, F0) and an inharmonic tone (I, all harmonics shifted upwards by ΔF). The task was a two-interval forced-choice task, and each interval contains four bursts of sound (HIHI or HHHH), the participants were required to discriminate harmonic complex tones and corresponding "frequency-shifted" tones by clicking on the appropriate box on the screen. The fundamental frequency was 200 Hz, the center frequency was 1,800 Hz, and the width of the passband was equal to F0 as the recommended value (Hopkins and Moore, 2011). The signal sound intensity was 60 dB SPL, the noise sound intensity was 45 dB SPL, and the initial change frequency was 100 Hz as the default settings (Sęk and Moore, 2012).

In the TFS-AF test, two consecutive intervals were presented on each trial, separated by 500 ms. Each interval contained four consecutive 400 ms tones, separated by 100 ms. In one interval, the IPD of all tones was always 0° (the standard), while tones with IPD = 0° are perceived as emanating from close to the center of the head. In the other interval, the first and third tones were the same while the second and fourth tones differed in their IPD by φ (the target). Participants were asked to indicate which of the two intervals contained a sequence of tones that appeared "Moving" within the head. The initial frequency for the TFS-AF test was 200 Hz, the sound intensity of the left and right ears was 30 dB SL (sensation level), and the phase difference (φ) was set to 180°.

The TFS1 and TFS-AF used the two-down-one-up (or two-up-one-down) procedure to estimate the "threshold" corresponding to 70.7% correct. It should be noted that the BIC delay threshold test used the three-up–one-down procedure, which was to be consistent with past studies and facilitate horizontal comparison with the results of past studies.

All the results were automatically output by the software after the test. The principle of the three test methods used in this research is shown in **Figure 1**.

# Results

## Test scores of participants

The results showed that the longest ITD for the BIC detection varied between 6.5 and 18.0 ms across 40 participants (mean = 11.4 ms, SD = 3.0 ms). A previous study showed

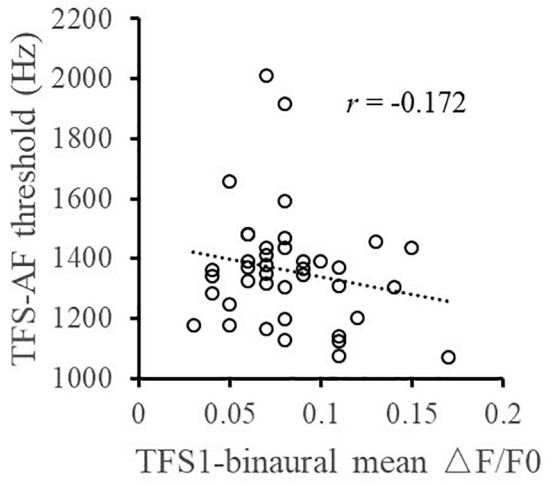



FIGURE 3
Illustration of the correlation analysis of the TFS1 and TFS-AF test.

that the BIC delay threshold is related to the frequency of the noise. For narrow-band noise, the BIC delay threshold decreases with the increase of the center frequency. The BIC delay threshold is approximately 12 ms for narrowband noise (center frequency = 200 Hz) and approximately 10 ms for wideband noise (Li et al., 2013). For monaural sensitivity, the results of TFS1 showed that the relative frequency shift threshold was between 0.017 and 0.221 (mean = 0.087, SD = 0.042) for left ear, and from 0.037 to 0.152 (mean = 0.076, SD = 0.031) for right ear. The mean monaural sensitivity of both ears ranged from 0.030 to 0.170 (mean = 0.0,815, SD = 0.031). A previous study (center frequency = 2,000 Hz, F0 = 222 Hz) showed that the relative frequency shift threshold for musicians was around 0.07–0.11, and slightly higher for non-musicians, around 0.11–0.17 (Mishra et al., 2015). For binaural sensitivity, the results showed that the TFS-AF threshold varied between 1,070.6 and 2,010.0 Hz (mean = 1,359.7 Hz, SD = 193.4 Hz). A previous study showed that the threshold for TFS-AF (180°) was approximately between 1,000 and 2,000 Hz, with a mean of 1,382 Hz (Füllgrabe et al., 2017). All the above results were not far from the scope of previous reports, and all three tests varied remarkably across participants. K-S (Kolmogorov-Smirnov) tests showed that there is no evidence that any of test indicators data differ from normal distribution (for TFS1: $p = 0.442$; for TFS-AF: $p = 0.237$; for BIC: $p = 0.884$).

## The relationship of temporal fine structure sensitivities with break in interaural correlation delay threshold

Pearson correlation analysis (**Figure 2**) showed that the TFS1 score was significantly correlated with the BIC delay threshold. The monaural TFS sensitivity averaged across ears of TFS1 was significantly correlated with the BIC delay threshold ($r = -0.426$, $p = 0.006$), but there was no evidence of a significant correlation between the BIC delay threshold and TFS-AF performance ($r = -0.005$, $p = 0.997$). In addition, this study did not observe significant correlation between TFS1 and TFS-AF ($r = -0.172$, $p = 0.289$) (**Figure 3**).

Furthermore, the relationship between the TFS sensitivity of the left and right ears and the BIC delay threshold was investigated. Pearson correlation analysis (**Figure 4**) showed that the TFS1 score of the left and right ears was significantly correlated with the BIC delay threshold (for left ear: $r = -0.367$, $p = 0.020$; for right ear: $r = -0.358$, $p = 0.023$). **Figure 5** shows

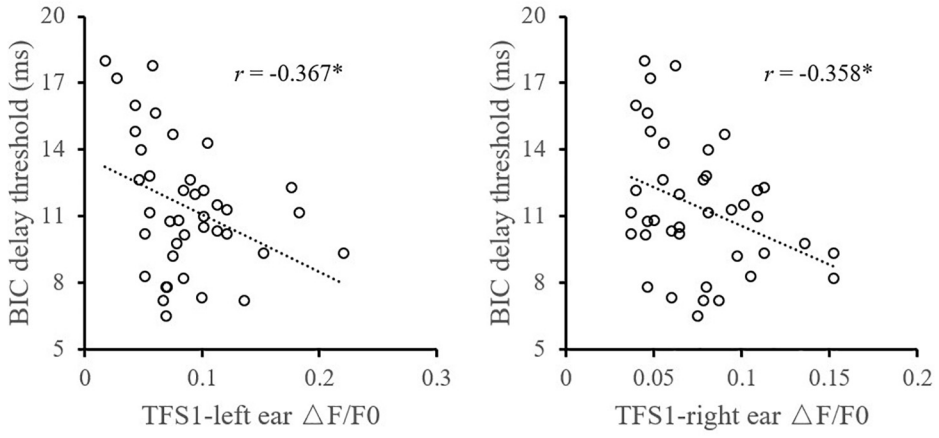



FIGURE 4
Illustration of the correlation analysis of the BIC delay threshold and the TFS1 and TFS-AF test. *Significant effect after Bonferroni's correction, $p < 0.025$.

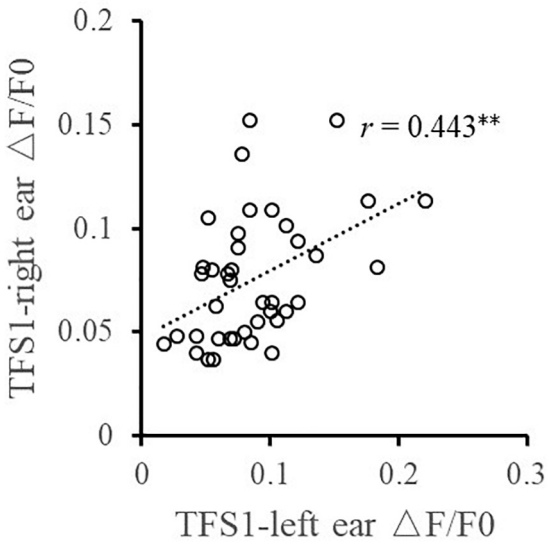



FIGURE 5
Illustration of the correlation analysis of the TFS1 scores of the left and right ears. $^{**}p < 0.01$.

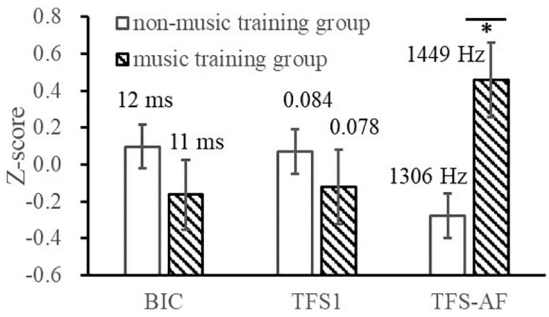



FIGURE 6
Illustration of the BIC delay threshold and TFS sensitivities of the music training group and the non-music training group. $^{*}p < 0.05$.

that there was a significant correlation between TFS1 scores of the left and right ears ($r = 0.443$, $p = 0.004$).

## The effect of music training

This study investigated whether the music training would affect the measurement results (and their relationship) of the BIC delay threshold and TFS sensitivity tests (**Figure 6**). A 3 (Test indicators: BIC delay threshold, TFS1 binaural mean, TFS-AF threshold) × 2 (music training experience: music training group, non-music training group) within-subject repeated measures ANOVA showed that the interaction between the two factors was significant [$F(2, 76) = 5.729$, $p = 0.005$, $\eta_p^2 = 0.131$] and the main effect of the music training experience was significant [$F(1, 38) = 5.623$, $p = 0.023$, $\eta_p^2 = 0.129$]. The independent sample $t$-test showed that the music training group had better TFS-AF scores [1,448.7 Hz for the music training group and 1,306.2 Hz for the non-music training group, $t(38) = 2.386$, $p = 0.022$]. Note that for the TFS-AF and BIC delay threshold tests, higher scores are better, while for the TFS1 test, lower scores represent higher sensitivity. Therefore, it indicates that for musical training experience, both monaural and binaural showed a trend toward better sensitivity to TFS, while the BIC delay threshold did not. It is also important to note that this study did not specifically recruit music majors, but only considered and recorded the effects of music training in normal participant recruitment. Such surveys lack necessary information, such as music level and daily training duration, so this grouping is insufficient compared to the definition of musicians in previous studies and leads to a reduction in statistical power. Insufficient statistical power means that there is a greater chance of making Type 2 errors (β), and some effects may not be detected. Therefore, some interpretations of the results need to be conservative. We used the G-power software to calculate the t-tests achieved power (*post hoc*) of the BIC delay threshold, TFS1, and TFS-AF, which were 0.09, 0.17, and 0.59, respectively (Faul et al., 2009). It should be noted that a reduction in statistical power may affect the reproducibility of this part of the results.

Considering the influence of music training experience, the relationship between TFS1 and BIC was compared between the music training group and those without any music training experience, respectively (**Figure 7**). For participants without any musical training experience, the monaural TFS sensitivity averaged across ears of TFS1 is significantly correlated with the BIC delay threshold ($r = -0.508$, $p = 0.010$). For participants with musical training experience, there was no evidence of a significant correlation between the BIC delay threshold and TFS1 performance ($r = -0.290$, $p = 0.295$).

# Discussion

The results of this study found that, in young adults, the maximal ITD of detecting binaural BIC was significantly correlated with monaural TFS sensitivity, but not with binaural TFS sensitivity. The correlation between BIC delay threshold and monaural TFS sensitivity was observed in both left and right ears, but this correlation was not found in the participants with musical training experience.

Binaural information integration is crucial for speech recognition in complex scenes. In previous studies, the BIC delay threshold was considered an effective method to measure transient the auditory storage capacity of acoustic details (Li et al., 2013; Kong et al., 2015; Liu et al., 2016). However, by measuring and comparing the effects of interaural delay and

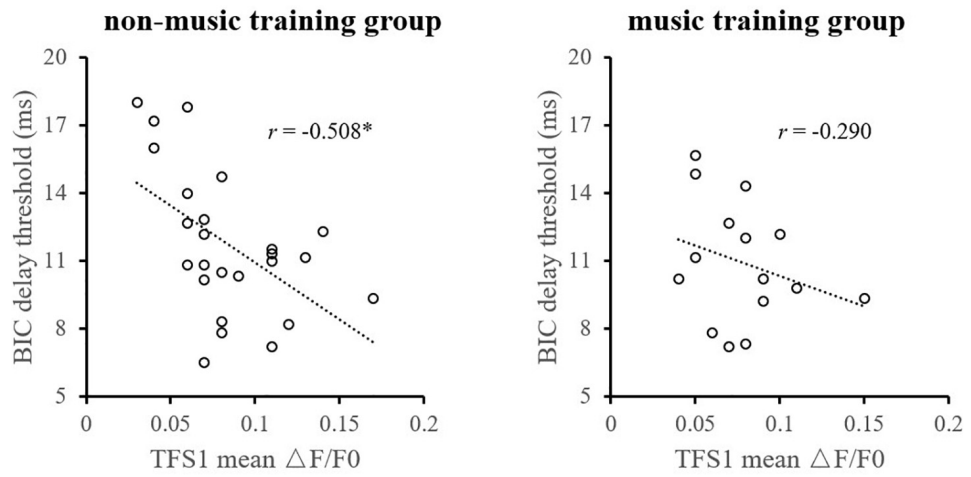



FIGURE 7
Illustration of the correlation analysis of the BIC delay threshold and the TFS1 and TFS-AF test. *Significant effect after Bonferroni's correction, $p < 0.025$.

interaural correlation in a group of participants, a previous study discovered a linear relationship between the changes in interaural correlation and interaural delay required to produce an equivalent decline of sensitivity to the BIC: an increment of 1 ms in BIC delay threshold is equivalent to a reduction of about 0.07 in interaural correlation (Kong et al., 2015). Furthermore, BIC detection involves not only the binaural calculation of the similarity of the TFS signals between the two ears but also the monaural coding of the TFS signal (Lei and Ding, 2021). Although introducing a change in interaural correlation does not alter the monaural energy spectrum of the sound signals, it changes dichotic repetition pitch (Bilsen and Goldstein, 1974) and the loudness (Moore, 2003) of the noise. Therefore, the detection ability of BIC is related to the sensitivity of pitch and loudness, while the TFS1 test reflects the pitch sensitivity to a certain extent, which may be one of the reasons why the two tests are related. In summary, the BIC delay threshold test primarily examines the ability to temporally store sound details, but it also reflects sensitivity to changes in interaural correlation and is associated with many monaural sensitivities.

Music training is related to both monaural sensitivity (Mishra et al., 2015) and binaural sensitivity (Bianchi et al., 2019) of TFS. Studies have found that compared to non-musicians, musicians have a superior ability to discriminate complex sounds based on their TFS, and this ability is unaffected by contralateral stimulation or ear of presentation (Tarnowska et al., 2019). Our study faced the problem of being underpowered (sample sizes: 15 with training, 25 without) but showed similar trends. Studies on BIC testing for musicians are lacking, and no significant results were observed in this study. BIC delay threshold and TFS1 test scores were only significantly correlated in the non-music training group, which may be due to the different effects of music training on those abilities, such as improving TFS sensitivity. This suggests the importance of background checks on participants in auditory-related research, considering that there may be many people who have received musical training in their childhood and that even non-professional training may have an impact on the test results.

## Summary

Overall, the measurements did not show any significant link between the BIC delay threshold and binaural TFS sensitivity, though we note the experimental power was low. However, test scores showed that the BIC delay threshold was significantly correlated with monaural TFS sensitivity. The significant correlation between the BIC delay threshold and monaural TFS sensitivity was mainly found in young adults without musical training experience.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Tsinghua University Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

YD and ML: conception, design, acquisition of data, analysis of data, interpretation of data, and writing – original draft and review and editing. CC: conception, design, funding acquisition, and writing—review and editing. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Planning subject for the 13th Five-Year Plan of National Education Sciences of China, a National Level Project (BLA190216), and the "Shuimu Tsinghua Scholar" Project (2020SM055).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Akeroyd, M. A., and Summerfield, A. Q. (1999). A binaural analog of gap detection. *J. Acoust. Soc. Am.* 105, 2807–2820. doi: 10.1121/1.426897

Bianchi, F., Carney, L. H., Dau, T., and Santurette, S. (2019). Effects of musical training and hearing loss on fundamental frequency discrimination and temporal fine structure processing: Psychophysics and modeling. *J. Assoc. Res. Otolaryngol.* 20, 263–277. doi: 10.1007/s10162-018-00710-2

Bilsen, F. A., and Goldstein, J. L. (1974). Pitch of dichotically delayed noise and its possible spectral basis. *J. Acoust. Soc. Am.* 55, 292–296. doi: 10.1121/1.1914500

Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT press. doi: 10.7551/mitpress/6391.001.0001

Blodgett, H. C., Wilbanks, W., and Jeffress, L. A. (1956). Effect of large interaural time differences upon the judgment of sidedness. *J. Acoust. Soc. Am.* 28, 639–643. doi: 10.1121/1.1908430

Boehnke, S. E., Hall, S. E., and Marquardt, T. (2002). Detection of static and dynamic changes in interaural correlation. *J. Acoust. Soc. Am.* 112, 1617–1626. doi: 10.1121/1.1504857

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acust.* 86, 117–128.

Cherry, E. C. (1954). Some further experiments on the recognition of speech with one and two ears. *J. Acoust. Soc. Am.* 26, 554–559. doi: 10.1121/1.1907373

Darwin, C. J. (2006). Contributions of binaural information to the separation of different sound sources: Contribuciones de la información binaural en la separación de diferentes fuentes sonoras. *Int. J. Audiol.* 45, 20–24. doi: 10.1080/14992020600782592

Ding, Y., Liang, Y., Cao, C., Zhang, Y., and Hu, M. (2022). Relationships among temporal fine structure sensitivity, transient storage capacity, and ultra-high frequency hearing thresholds in tinnitus patients and normal adults of different ages. *Front. Aging Neurosci.* 14:869708.

Eramudugolla, R., McAnally, K. I., Martin, R. L., Irvine, D. R., and Mattingley, J. B. (2008). The role of spatial location in auditory search. *Hear. Res.* 238, 139–146. doi: 10.1016/j.heares.2007.10.004

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Füllgrabe, C., Harland, A. J., Sêk, A. P., and Moore, B. C. (2017). Development of a method for determining binaural sensitivity to temporal fine structure. *Int. J. Audiol.* 56, 926–935. doi: 10.1080/14992027.2017.1366078

Füllgrabe, C., Sêk, A. P., and Moore, B. C. (2018). Senescent changes in sensitivity to binaural temporal fine structure. *Trends Hear.* 22, 2331216518788224. doi: 10.1177/2331216518788224

Hopkins, K., and Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J. Acoust. Soc. Am.* 130, 334–349. doi: 10.1121/1.3585848

Huang, Y., Huang, Q., Chen, X., Wu, X., and Li, L. (2009). Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *J. Exp. Psychol. Hum. Percept. Perform.* 35:1618. doi: 10.1037/a0015791

Huang, Y., Kong, L., Fan, S., Wu, X., and Li, L. (2008). Both frequency and interaural delay affect event-related potential responses to binaural gap. *Neuroreport* 19, 1673–1678. doi: 10.1097/WNR.0b013e32831576c7

Kong, L., Xie, Z., Lu, L., Qu, T., Wu, X., Yan, J., et al. (2015). Similar impacts of the interaural delay and interaural correlation on binaural gap detection. *PLoS One* 10:e0126342. doi: 10.1371/journal.pone.0126342

Langford, T. L., and Jeffress, L. A. (1964). Effect of noise crosscorrelation on binaural signal detection. *J. Acoust. Soc. Am.* 36, 1455–1458. doi: 10.1121/1.1919224

Lei, M., and Ding, Y. (2021). Interaural delay modulates the prepulse inhibition of the startle reflex induced by binaural gap in humans. *JASA Express Lett.* 1:064401. doi: 10.1121/10.0005110

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Li, H., Kong, L., Wu, X., and Li, L. (2013). Primitive auditory memory is correlated with spatial unmasking that is based on direct-reflection integration. *PLoS One* 8:e63106. doi: 10.1371/journal.pone.0063106

Li, L., Huang, J., Wu, X., Qi, J. G., and Schneider, B. A. (2009). The effects of aging and interaural delay on the detection of a break in the interaural correlation between two sounds. *Ear Hear.* 30, 273–286. doi: 10.1097/AUD.0b013e318198703d

Liu, Z., Wang, Q., You, Y., Yin P, Ding H, Bao X., et al. (2016). The role of the temporal pole in modulating primitive auditory memory. *Neurosci. Lett.* 619, 196-202. doi: 10.1016/j.neulet.2016.03.025

Maddox, R. K., and Shinn-Cunningham, B. G. (2012). Influence of task-relevant and task-irrelevant feature continuity on selective auditory attention. *J. Assoc. Res. Otolaryngol.* 13, 119–129. doi: 10.1007/s10162-011-0299-7

Mishra, S. K., Panda, M. R., and Raj, S. (2015). Influence of musical training on sensitivity to temporal fine structure. *Int. J. Audiol.* 54, 220–226. doi: 10.3109/14992027.2014.969411

Moore, B. C. (2014). *Auditory Processing of Temporal Fine Structure: Effects of Age and Hearing Loss*. Singapore: World Scientific. doi: 10.1142/9064

Moore, B. C. J. (2003). Temporal integration and context effects in hearing. *J. Phonet.* 31, 563–574. doi: 10.1016/S0095-4470(03)00011-1

Moore, B. C., and Sek, A. (2009). Development of a fast method for determining sensitivity to temporal fine structure. *Int. J. Audiol.* 48, 161–171. doi: 10.1080/14992020802475235

Mossop, J. E., and Culling, J. F. (1998). Lateralization of large interaural delays. *J. Acoust. Soc. Am.* 104, 1574–1579. doi: 10.1121/1.424369

Schnupp, J., Nelken, I., and King, A. (2011). *Auditory Neuroscience: Making Sense of Sound*. Cambridge, MA: MIT press. doi: 10.7551/mitpress/7942.001.0001

Sęk, A., and Moore, B. C. (2012). Implementation of two tests for measuring sensitivity to temporal fine structure. *Int. J. Audiol.* 51, 58–63.

Tarnowska, E., Wicher, A., and Moore, B. C. (2019). The effect of musicianship, contralateral noise, and ear of presentation on the detection of changes in temporal fine structure. *J. Acoust. Soc. Am.* 146:1. doi: 10.1121/1.5114820

Wang, Q., and Li, L. (2015). Auditory midbrain representation of a break in interaural correlation. *J. Neurophysiol.* 114, 2258–2264. doi: 10.1152/jn.00645.2015

Whiteford, K. L., Kreft, H. A., and Oxenham, A. J. (2017). Assessing the role of place and timing cues in coding frequency and amplitude modulation as a function of age. *J. Assoc. Res. Otolaryngol.* 18, 619–633. doi: 10.1007/s10162-017-0624-x

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 14 October 2022
DOI 10.3389/fnins.2022.1004071



OPEN ACCESS

EDITED BY
Yi Zhou,
Arizona State University, United States

REVIEWED BY
Yan Gai,
Saint Louis University, United States
John C. Middlebrooks,
University of California, Irvine,
United States

*CORRESPONDENCE
Kamal Sen
kamalsen@bu.edu

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 26 July 2022
ACCEPTED 28 September 2022
PUBLISHED 14 October 2022

CITATION
Chou KF, Boyd AD, Best V, Colburn HS and Sen K (2022) A biologically oriented algorithm for spatial sound segregation.
*Front. Neurosci.* 16:1004071.
doi: 10.3389/fnins.2022.1004071



# A biologically oriented algorithm for spatial sound segregation

Kenny F. Chou[1], Alexander D. Boyd[1], Virginia Best[2], H. Steven Colburn[1] and Kamal Sen[1]*

[1]Department of Biomedical Engineering, Boston University, Boston, MA, United States, [2]Department of Speech, Language and Hearing Sciences, Boston University, Boston, MA, United States

Listening in an acoustically cluttered scene remains a difficult task for both machines and hearing-impaired listeners. Normal-hearing listeners accomplish this task with relative ease by segregating the scene into its constituent sound sources, then selecting and attending to a target source. An assistive listening device that mimics the biological mechanisms underlying this behavior may provide an effective solution for those with difficulty listening in acoustically cluttered environments (e.g., a cocktail party). Here, we present a binaural sound segregation algorithm based on a hierarchical network model of the auditory system. In the algorithm, binaural sound inputs first drive populations of neurons tuned to specific spatial locations and frequencies. The spiking response of neurons in the output layer are then reconstructed into audible waveforms *via* a novel reconstruction method. We evaluate the performance of the algorithm with a speech-on-speech intelligibility task in normal-hearing listeners. This two-microphone-input algorithm is shown to provide listeners with perceptual benefit similar to that of a 16-microphone acoustic beamformer. These results demonstrate the promise of this biologically inspired algorithm for enhancing selective listening in challenging multi-talker scenes.

KEYWORDS

**multitalker speech perception, sound (audio) processing, sound segregation, cocktail party problem, binaural hearing, spatial listening, hearing loss**

## Introduction

Attending to a single conversation partner in the presence of multiple distracting talkers (i.e., the Cocktail Party Problem, CPP) is a complicated and difficult task for machines and humans (Haykin and Chen, 2005; McDermott, 2009; Qian et al., 2018). While some listeners can accomplish this task with relative ease, other groups of listeners report great difficulty—such as those with sensorineural hearing loss (Kochkin, 2000, 2007; Shinn-Cunningham and Best, 2008), cochlear implant users (Bernstein et al., 2016; Goupell et al., 2016, 2018; Litovsky et al., 2017), subgroups of children (Dhamani et al., 2013), persons with aphasia (Villard and Kidd, 2019) and adults with "hidden

hearing loss" (Pichora-Fuller et al., 2017; Shinn-Cunningham, 2017; Parthasarathy et al., 2019). At a cocktail party, talkers are distributed in space. Listeners use spatial cues (i.e., interaural timing and level differences, or ITDs and ILDs, respectively) for sound localization. Additionally, normal-hearing listeners appear to make use of spatial cues in addition to a variety of other talker-related cues, to perceptually segregate the competing talkers and attend to the one of most interest. Indeed, spatial listening has been shown to provide enormous benefit to listeners in cocktail-party scenarios (Litovsky, 2012; Rennies and Kidd, 2018).

Sound processing algorithms can be designed with the distinct goals of sound localization or spatial sound segregation. Specifically, spatial processing plays a key role in several sound segregation algorithms that aim to help hearing-impaired listeners overcome the CPP. For example, acoustic beamforming techniques utilize multiple microphones to selectively enhance signals from a desired direction (Gannot et al., 2017; Chiariotti et al., 2019), and are often employed in hearing aids (Greenberg and Zurek, 2001; Chung, 2004; Doclo et al., 2010; Picou et al., 2014; Launer et al., 2016). Machine learning approaches such as clustering using Gaussian mixture models (MESSL) (Mandel et al., 2010) and deep neural networks (DNN) (Wang et al., 2014), among others, also make use of ITDs and ILDs to localize the target sound.

The ability of human listeners with normal hearing to solve the CPP is quite remarkable. Many animals, too, appear to have robust solutions to their own versions of the CPP (Bee and Micheyl, 2008). Unlike beamformers, which benefit from using microphone arrays, humans and animals require only two inputs—the left and right ear. These listeners are also able to solve the CPP in novel and unpredictable settings, a challenge for algorithms that rely on supervised learning (Bentsen et al., 2018; Wang and Chen, 2018). This raises the idea that spatially selective algorithms may benefit from incorporating insights from the human and/or animal brain. From a practical standpoint, biological processing, which is based on neural spikes, also has practical advantages that make it uniquely suited for always-on, portable devices such as hearing aids. Spike-based processing is computationally efficient and can be implemented with higher temporal resolution than algorithms operating on sampled waveforms (Ghosh-Dastidar and Adeli, 2009), especially when implemented on specialized neuromorphic hardware (Roy et al., 2019).

We recently proposed a biologically inspired algorithm for sound processing. The primary goal of this algorithm was to use spatial cues to perform sound segregation and selection, not sound localization. In this algorithm, sound mixtures were segregated by spatially selective model neurons, and selection was achieved by selective integration *via* a cortical network model (Chou et al., 2019). For the tested conditions, which included a frontal target talker and two symmetrically placed masker talkers, the algorithm showed segregation performance similar to MESSL and DNN, and provided proof-of-concept for a biologically based speech processing algorithm. However, the algorithm operated in the spiking domain, and employed a linear decoding algorithm to recover the target speech (Mesgarani et al., 2009), which resulted in low objective speech intelligibility. Like many typical beamformers, the algorithm also did not preserve binaural cues in the output, which can be particularly problematic in multitalker mixtures (Best et al., 2017; Wang et al., 2020). These drawbacks limited its practical use for applications in hearing-assistive devices and machine hearing.

In this study, we present a new biologically oriented sound segregation algorithm (BOSSA) that overcomes specific limitations of our previous algorithm. We introduce a time-frequency mask estimation method for decoding processed neural spikes that improves the quality of recovered target speech compared to the current standard approach (Mesgarani et al., 2009). We compared the proposed two-channel algorithm to a 16-microphone super-directional beamformer, using both objective measures and human psychophysics, and showed equivalent performance. Our algorithm overcomes some of the challenges faced by current state-of-the-art technologies, and provides an alternative, biologically based approach to the CPP.

## Algorithm design and implementation

The proposed BOSSA algorithm contains three modules (**Figure 1**) that together generate neural output patterns that are inputs to the target-reconstruction stage. The first module resembles peripheral filtering by the cochlea. The second module performs spatial segregation by constructing model neurons sensitive to specific spatial cues in narrow frequency bands. Ensembles of neurons then encode sounds that share the same spatial cues. In the third module, the spiking activity of output neurons are decoded into intelligible waveforms using a novel reconstruction approach. All modules are implemented in MATLAB (MathWorks, Natick, MA, United States).

### Peripheral filtering

Left and right channels of the input audio are filtered with a gammatone equivalent-rectangular-bandwidth (ERB) filterbank, implemented using the auditory toolbox in MATLAB (Slaney, 1998). The bandwidths were calculated using $ERB = \left[\left(f_c/Q\right)^x + b^x\right]^{\frac{1}{x}}$ with parameters Q = 9.26449 (Glasberg and Moore, 1990), minimum bandwidth (b) = 24.7 Hz, order (x) = 1. The filterbank used here has 64 channels with center frequencies ranging from $f_1 = 200\ Hz$ to $f_{64} = 20\ kHz$. The filterbank outputs are two sets of 64

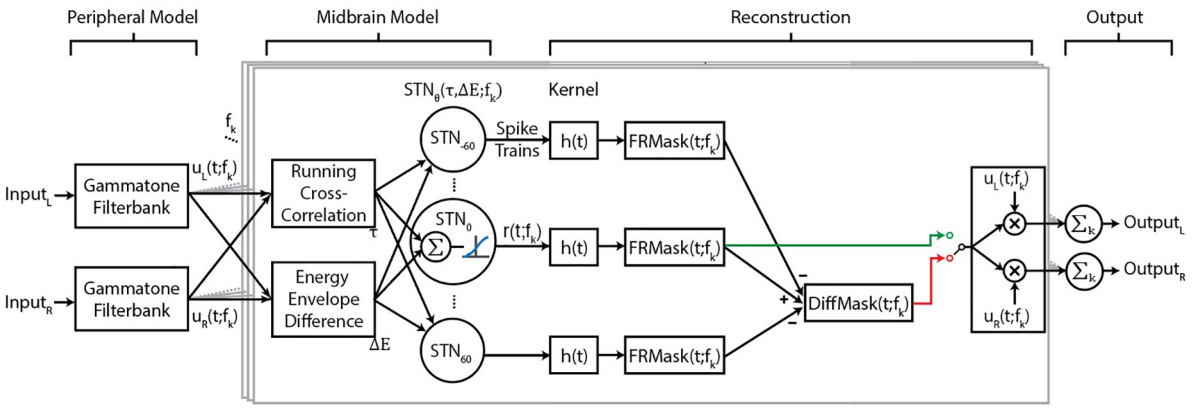



FIGURE 1
Flow diagram of the proposed algorithm. Central boxes, outlined in gray, show processing for a single-frequency band. The functions $u_L$ (t; $f_k$) and $u_L$ (t; $f_k$) are the narrowband signals of the left and right input channels for each frequency channel, and $f_k$ denotes the $k^{th}$ frequency channel. The midbrain model is based on spatially tuned neurons (STNs), where each STN has a "best" ITD and ILD, denoted τ and ΔE, respectively. The best ITD and ILD values of a neuron depend on the direction θ and frequency $f_k$ to which the STN is tuned. h (t) represents the reconstruction kernel that converts spike trains to waveforms. We implemented two masks, FRMask (green line) and DiffMask (red line), either of which could be used for reconstruction, as indicated by the switch, The implementation of DiffMask in our analysis involves five sets of STNs, where θ ∈ {0, ± 30, ± 60}; however, other implementations of the model may involve different sets of θ.

channels of narrowband signals, $u_L(t;f_k)$ and $u_R(t;f_k)$, corresponding to the left and right channels, respectively.

## Midbrain model

First, binaural cues of input signals are extracted based on a model of the barn-owl inferior colliculus (Fischer et al., 2009). ITD was calculated as a short-time running cross correlation between the energy-normalized $u_L(t;f_k)$ and $u_R(t;f_k)$ and ILD as the energy envelope difference between $u_L(t;f_k)$ and $u_R(t;f_k)$. Gain modulation steps matching those used in Fischer et al. (2009) were applied to the filterbank outputs such that the inputs to the cross correlation calculation, ($u_L(t;f_k)$ and $u_R(t;f_k)$), varied as a linear function of stimulus level. Further gain control applied during the cross correlation calculation in conjunction with a logarithmic energy envelope calculation resulted in an approximately stimulus level invariant ILD representation. For a detailed description of the mathematical operations and their physiological basis, we refer interested readers to Fischer et al. (2009).

We then constructed sets of spatially tuned neurons (STNs), where each set consists of 64 neurons tuned to $f_k$ of the previous module. The 64 neurons in each set are sensitive to the same specific direction θ in the horizontal plane ($STN_\theta$, **Figure 1**), and each neuron has specific parameters $\tau(\theta;f_k)$ and $\Delta E(\theta,f_k)$, corresponding to the ITD and ILD for that specific θ. Each neuron's preferred time-lag τ was calculated using the Woodworth formulation (Woodworth, 1938), with the approximation that ITDs are independent of frequency. Preliminary studies found that using frequency-dependent ITD values, calculated as described by Fischer et al. (2009) or the ones described by Aaronson and Hartmann (2014), provided no benefit in terms of objective measures of algorithm performance. On the other hand, $\Delta E$ is frequency-dependent, and was derived by calculating the ILD of a narrow band noise placed at various azimuths. Directionality of the narrow band noise was imparted by convolving with Head Related Transfer Functions (HRTFs) of the Knowles Electronic Manikin for Acoustic Research (KEMAR) (Burkhard and Sachs, 1975; Algazi et al., 2001).

The responses of model neurons were then calculated as follows. If the stimulus energy envelope difference was within a preset range of the neuron's preferred $\Delta E$, then that energy-envelope difference was weighted by the energy envelope of either $u_L(t;f_k)$ or $u_R(t;f_k)$. The ITD and ILD components were combined additively at the subthreshold level and then transformed *via* a sigmoidal input-output non-linearity (i.e., an activation function) to obtain an instantaneous firing rate. Finally, a Poisson spiking generator was used to generate spike trains for each neuron [$r_\theta(t;f_k)$, **Figure 1**]. This sequence of operations is expected to produce a multiplicative spiking response to ITD and ILD in each model neuron as explained in Fischer et al. (2009). These steps, including the activation function, were kept identical for all frequency channels. Parameters for the input-output nonlinearity were modified from a step-function to a sigmoidal function to increase the dynamic range of the model neurons' firing rates.

The model can be implemented with any number and configuration of STNs. For illustrations of spatial tuning curves in **Figure 2A**, nine sets of STNs were constructed where θ ∈ {0°, ± 30°, ± 45°, ± 60°, ± 90°}. The ILDs used in generating the neuron spatial tuning curves are shown in

Figure 2B, where each line represents $\Delta E(\theta, f_k)$ for a set of $STN_\theta$. All other results were obtained by constructing five sets of STNs, where $\theta \in \{0°, \pm 30°, \pm 60°\}$.

## Stimulus reconstruction

The stimulus reconstruction module decodes ensembles of neural spikes into audible waveforms, using an approach similar to ideal time-frequency mask estimation (Wang, 2005). The concept of time-frequency masks can be summarized as follows: for a time-frequency representation of an audio mixture (e.g., spectrogram) consisting of a target and interferers, one can evaluate each element (i.e., time-frequency tile) of such a representation and determine whether the energy present is dominated by the target or the masker. If the target sound dominates, a value of unity (1) is assigned to that time-frequency tile, and zero (0) otherwise. This process creates an ideal binary mask. Alternatively, assigning the ratio of energies of the target to total energies in a time-frequency tile yields the ideal ratio mask (Srinivasan et al., 2006). One can then estimate the target sound by applying the mask to the sound mixture *via* element-wise multiplication. This process has been shown to recover the target with high fidelity in various types of noise (Wang, 2005). A key idea to both binary and ratio masks is the application of a gain factor to each time-frequency tile of a signal. In the proposed BOSSA algorithm we adopt a similar approach but calculate the gain factor for each time-frequency tile based solely on user-defined knowledge of the target location, as explained below.

The spiking responses from the spatially tuned neurons, $r(t; f_k)$, were convolved with a kernel, $h(t)$, to calculate a smoothed, firing-rate-like measure. We set the kernel to be an alpha function: $h(t) = te^{-t/\tau_h}$, a common function involved in modeling neural dynamics. We used a value of $\tau_h$ = 20 ms (see section Model Parameters) and the kernel was restricted to a length of 100 ms.

The same kernel was convolved with the spike trains of each frequency channel independently. The resulting firing rates of each set of STNs were treated as a non-binary time-frequency mask:

$$FRMask(t; f_k) = r(t; f_k) * h(t)$$

where * denotes convolution. We note that the FRMask is akin to a smoothed version of the firing rate. Thus, in theory, FRMask could be directly derived from the firing rate (without the need for spikes). However, the midbrain model can be used as a front-end to spiking network models, where the calculation of spikes is necessary (Chou et al., 2019). Thus, we kept this more versatile implementation.

The mask was then applied (i.e., point-multiplied) to the left and right channels of the original sound mixture. Then, we summed (without weighting) each frequency channel of the FRMask-filtered signal to obtain an audible, segregated waveform. We designated this result as $\hat{S}$.

$$\hat{S}_j = \sum_k FRMask(t; f_k) \cdot u_j(t; f_k), j \in \{L, R\}$$

This procedure resulted in a binaural signal and retained the natural spatial cues of the sound sources.

To reduce spatial leakage, we calculated a DiffMask by calculating FRMasks for each $STN_\theta$, then subtracting scaled versions of the off-center $STN_\theta$ from $STN_0$, followed by rectification:

$$DiffMask = Max(FRMask_0 - a\Sigma FRMask_\theta, 0)$$

where $\theta \in [\pm 30°, \pm 60°]$ corresponds to the location of maskers in our experimental stimuli (see section "Psychophysical Experiment"). In this operation, each FRMask was first scaled to [0,1]. The scaling factor $a$ was chosen to be 0.5 (see section "Model Parameters") and was fixed across all frequencies and spatial channels to reduce the amount of computational complexity in the algorithm.

## Model parameters

Although a behavioral measure of algorithm performance using human psychophysics is the gold standard, such experiments are too time-consuming to explore model parameter variations. For practical reasons, most model parameters were fixed to biologically plausible values. We explored variations in the time-constant of the alpha function kernel ($\tau_h$), and the scaling factor for DiffMask ($a$). We chose the specific values of these parameters using an iterative process by trying a range of values, quantifying algorithm performance using the Short Time Objective Intelligibility (STOI) measure (Taal et al., 2010), and choosing parameters that produced the highest average STOI. STOI is an approximation of speech intelligibility, and ranges between 0 and 1. We do not claim that this approach produces an optimal set of parameters for reconstruction. However, objective measures combined with our behavioral results indicate that the parameter values we chose generated good reconstructions.

# Algorithm performance

## Spatial tuning characteristics

Spatial tuning responses of STNs were important predictors of the model's segregation performance. We define "spatial tuning curves" as the spiking activity of STNs as a function of stimulus location. To construct spatial tuning curves, white Gaussian noise was convolved with anechoic KEMAR HRTFs, then presented to the algorithm. Figure 2 shows the responses of

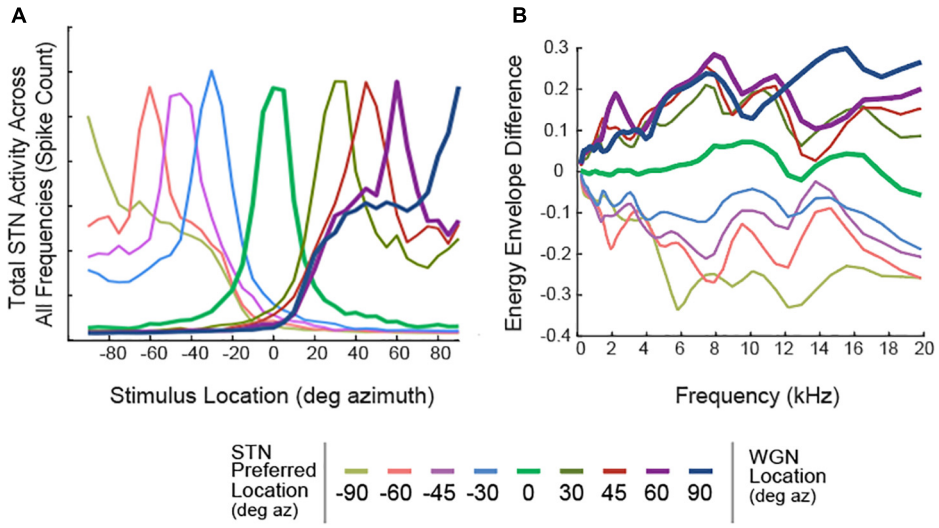



FIGURE 2
Spatial Tuning Characteristics of STNs. **(A)** Spatial Tuning Curves of STNs, represented as total neural activity in response to a White Gaussian Noise (WGN). Each colored line represents the total response from a set of STNs. Tuning curves of STNs tuned to 0°, 60°, and 90° azimuth are bolded. **(B)** Preferred ILDs of STNs, calculated using WGNs placed at various locations along the azimuthal plane.

STNs combined across frequency channels. Ideally, STNs would only respond to stimuli from one specific direction. However, Figure 2 shows that all STNs also respond to off-target locations. For example, STNs tuned to 0° azimuth (Figure 2A, green curve) respond to stimuli at ±30° azimuth and even have a non-zero response to stimuli at ±90° azimuth. We refer to this property as "spatial leakage," which occurs due to overlap in the bandpass filters as well as the fact that a given binaural cue can occur for stimuli from multiple locations (Figure 2B) and thus contain some ambiguity (Brainard et al., 1992).

## Spatial leakage

Leakage across spatial channels limits the performance of the algorithm, especially when multiple sound sources are present. To demonstrate, two randomly selected sentences were presented individually to the model from 0° azimuth, 90° azimuth, or simultaneously from both locations. The responses of three set of STNs, tuned to 90°, 45°, and 0°, are shown as spike-rasters in Figure 3. Each row within a raster plot represents the spiking response from the neuron tuned to that particular frequency channel. Due to spatial leakage, all STNs respond to the single sentence placed at 0° or 90° (Figures 3A,B). When both sentences are present, ITDs and ILDs interact to produce complicated STN response patterns (Figure 3C). Spatial leakage limits the ability of STNs to respond to a single talker, since any one spatial channel contains information from other spatial channels. Lateral inhibition was designed to address the issue of spatial leakage by suppressing neural activation by off-target sound streams.

## DiffMask

The DiffMask operation was inspired by lateral inhibition observed in biological networks. This operation was applied to the spatial tuning curves of 0° STNs to illustrate its sharpening effect on spatial tuning. Figure 4A shows the tuning curves prior to the DiffMask operation. Some neurons within the 0° STNs were activated by stimuli from as far away as 90° (see side peaks). After the DiffMask operation, spiking activity elicited by far-away stimuli was silenced, and side-peaks were suppressed considerably (Figure 4B). Using a subset of STNs during the DiffMask operation, such as those tuned to ±30° (Figure 4C) or ±60° (Figure 4D), did not suppress side-peaks as effectively as if both ±30° and ±60° were used.

# Psychophysical experiment

A psychophysical experiment was conducted to quantify the perceptual benefit provided by the algorithm for listeners with normal hearing. The performance of FRMask and DiffMask was compared against a 16-microphone super-directional beamformer, called BEAMAR (Kidd et al., 2015; Best et al., 2017). BEAMAR attenuates off-center sounds by combining the weighted output of 16 omni-directional microphones into a single channel, using an optimal-directivity algorithm (Stadler and Rabinowitz, 1993). BEAMAR does not process frequencies below 1 kHz in order to retain natural spatial cues in that frequency region. The combination of beamforming at high frequencies and natural binaural signals at low frequencies has

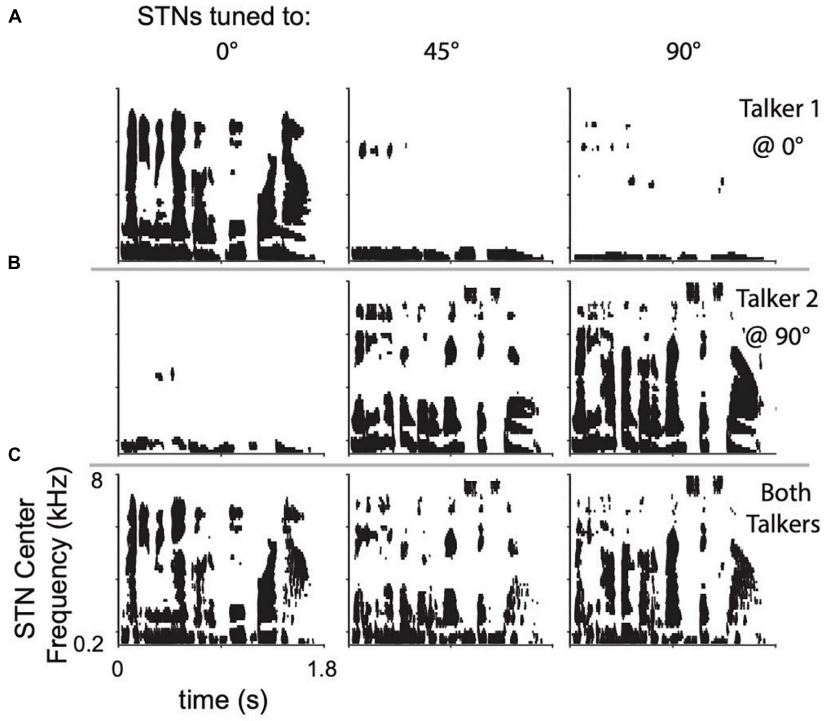



FIGURE 3
Raster plots of STN responses to **(A)** top row, a single sentence placed at 0° azimuth, **(B)** center row, a different sentence placed at 90° azimuth, and **(C)** bottom row, both sentences present at their respective locations. Columns show the STN responses when tuned to the location indicated.

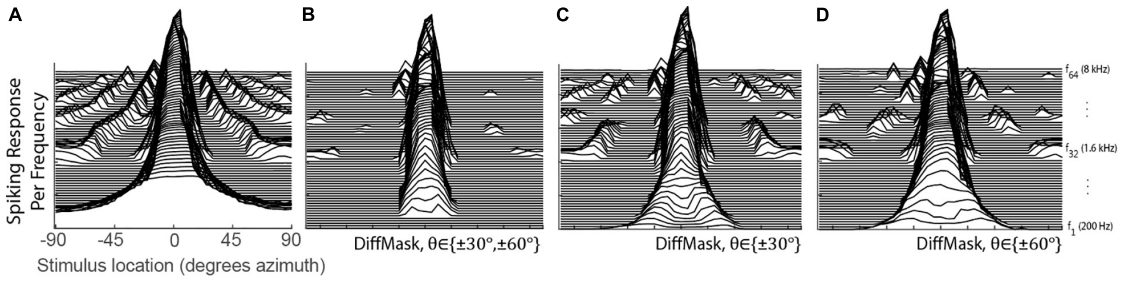



FIGURE 4
Spatial Tuning of the 0° STNs for before **(A)** and after **(B–D)** the DiffMask operation. Each line represents the spatial tuning curve of a single frequency-specific neuron within the set of STNs ranging from 200 to 8 kHz with ERB spacing. STN of the neuron tuned to the lowest frequency is placed on the bottom of the plots. STNs involved in the DiffMask operation are denoted in each subplot.

been shown to provide a significant benefit to both normal-hearing and hearing-impaired listeners attending to a target speech sentence in a multi-talker mixture (Best et al., 2017).

## Participants

Participants in this study were eleven young normal-hearing listeners, ages 18–32. All listeners had symmetrical audiogram measurements between 0.25 and 8 kHz with hearing thresholds within 20 dB HL. Participants were paid for their participation and gave written informed consent. All procedures were approved by the Boston University Institutional Review Board (protocol 1301E).

## Stimuli

Five-word sentences were constructed from a corpus of monosyllabic words (Kidd et al., 2008), with the form [name-verb-number-adjective-noun] (e.g., "Sue found three red hats"). The corpus contains eight words in each of the five categories.

Each word in each sentence was spoken by a different female talker, randomly chosen from a set of eight female talkers, without repetition. During each trial, a target sentence was mixed with four masker sentences, all constructed in the same manner. Words from the target and masker sentences were time-aligned, so that the words from each category shared the same onset. The design of these stimuli was intended to reduce the availability of voice and timing-related cues, and as such increase the listener's use of spatial information to solve the task.

The five sentences were simulated to originate from five spatial locations: 0°, ±30°, and ±60° azimuth, by convolving each sentence with anechoic KEMAR HRTFs. The target sentence was always located at 0° azimuth. The four maskers were presented at 55 dB SPL from ±30°, and ±60° azimuth. The level of the target was varied to achieve target-to-masker ratios (TMRs) of –5, 0, and 5 dB.

Stimuli were processed using one of three methods: BEAMAR, FRMask, and DiffMask. A control condition was also included, in which stimuli were spatialized using KEMAR HRTFs to convey "natural" cues but were otherwise unprocessed.

## Procedures

Three blocks were presented for each of the four conditions, with each block containing five trials at each of the three TMRs (15 total trials per block). This resulted in 15 trials per TMR for each of the four processing conditions, and a total of 180 trials across all conditions. The order of presentation of TMRs within a block, and the order of blocks for each participant, were chosen at random. The experiment took approximately 1 h to complete.

Stimuli were controlled in MATLAB and presented *via* a real time processor and headphone driver (RP2.1 & HB7, Tucker Davis Technologies, Alachua, FL, United States) through a pair of headphones (Sennheiser HD265 Linear). The sound system was calibrated at the headphones with a sound meter (type 2250; Brüel & Kjær, Nærum, Denmark). Participants were seated in a double-walled sound-treated booth. A computer monitor inside the booth displayed a graphical user interface containing a grid of 40 words (five columns of eight words, each column corresponding to one position of the five word sentence). For each trial, participants were presented a sentence mixture and were instructed to listen for the target sentence located directly in front. They responded with a mouse by choosing one word from each column on the grid.

## Analysis

Each participant's performance was evaluated by calculating the percentage of correctly answered keywords across all trials for a given condition. Psychometric functions were generated by plotting the percent correct as a function of TMR and fitting a logistic function to those data. Speech reception thresholds (SRTs), which are the TMRs corresponding to 50% correct, were extracted from each function using the psignifit toolbox (Schütt et al., 2016). Differences in SRTs between the natural condition and each of the processing conditions was taken to be the "benefit" provided by that processing method. Statistical analysis was done in Python using the statsmodels package (Seabold and Perktold, 2010).

# Results

**Figure 5A** shows the percentage of correct responses for each TMR and processing method. A two-way repeated-measures ANOVA found a significant interaction between processing method and TMR on performance [$F_{(6,60)} = 6.97$, $p < 0.001$]. *Post hoc* pairwise comparisons using Tukey's HSD test found significant differences between the natural condition and each of the three processing methods for all three TMRs ($p < 0.001$), suggesting that subjects significantly benefitted from listening to processed speech across all TMRs. At +5-dB TMR, performance was equivalent under all three processing conditions. However, at –5-dB and 0-dB TMR, performance was better for DiffMask than FRMask, and similar for DiffMask and BEAMAR. **Figure 5B** presents the same results in terms of SRTs, and **Figure 5C** shows the benefit (in dB) of each processing method relative to the natural condition. A one-way repeated measures ANOVA followed by Tukey's multiple pairwise comparison showed that all three algorithms provided significant benefit to listeners ($p < 0.001$). Benefits provided by BEAMAR and DiffMask were not significantly different ($p = 0.66$). Out of the eleven listeners, two achieved the lowest SRT and gained the most benefit from BEAMAR, while nine achieved the lowest SRT and gained the most benefit from DiffMask.

# Discussion

Extensive research has been devoted to developing a solution for the CPP [for review, see Qian et al. (2018)], and many approaches benefit from using multiple microphones. For example, the performance of methods using independent component analysis degrades quickly as the number of sources exceeds the number of microphones (Hyvärinen et al., 2001). In acoustic beamforming, performance of the beamformer can be significantly improved by increasing the number of microphones used (Greenberg and Zurek, 2001; Greenberg et al., 2003). Although traditional beamformers produce a single-channel output, which cannot carry binaural information, a variety of spatial-cue preservation strategies have been proposed to overcome this limitation (Doclo et al., 2010;

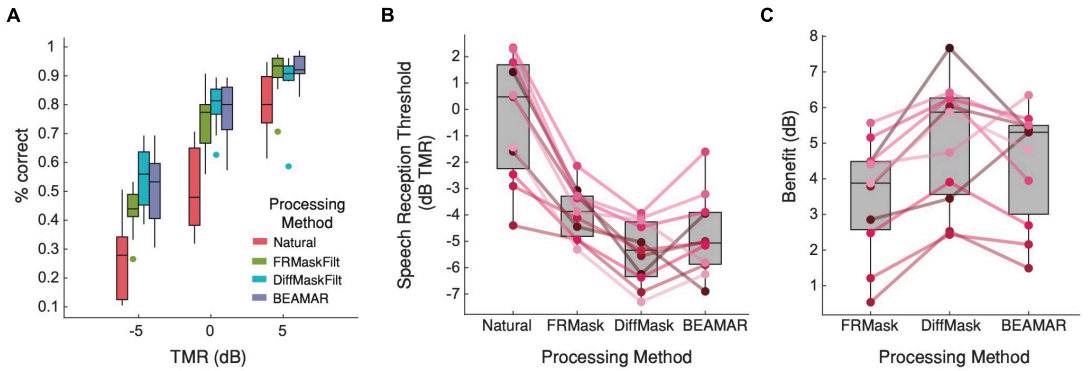



FIGURE 5
Behavioral evaluation results. **(A)** Average percent correct for each processing condition as a function of TMR. $n$ = 11. Higher is better. **(B)** Average and individual subject speech reception threshold for each processing method. Lower is better. Solid lines connect the SRTs for each participant. **(C)** Average and individual subject perceptual benefits relative to the natural condition. Higher is better.

Best et al., 2017; Wang et al., 2020). Here we demonstrated that equivalent performance to a highly optimized beamformer (such as BEAMAR) may be possible using a biologically inspired algorithm that uses only two microphones placed in the ears. Our biologically oriented sound segregation (BOSSA) model provided a substantial benefit in a challenging cocktail party listening situation, and this benefit was larger than that provided by BEAMAR in the majority of our young, normal hearing participants. While this is a promising result, further work is needed to examine the benefits of BOSSA under a wider variety of scenarios and in other groups of listeners. Comparisons to other two-microphone solutions such as binaural beamformers (Doclo et al., 2010; Best et al., 2015), as well as deep-learning solutions that operate on two or even a single microphone (Roman et al., 2003; Healy et al., 2013), would also be interesting.

Spiking neural networks traditionally do not have applications in audio processing due to the lack of a method that produces intelligible, high-quality reconstructions. The "optimal prior" method of reconstruction is often used to obtain reconstructions from physiologically recoded neural responses (Bialek et al., 1991; Stanley et al., 1999; Mesgarani et al., 2009; Mesgarani and Chang, 2012), but produces single-audio-channel reconstructions of poor quality and intelligibility (Chou et al., 2019). The optimal prior method computes a linear filter between a training stimulus and the response of neuron ensembles, and filter needs to be re-trained if the underlying network changes. In contrast, the mask-based reconstruction method used in this study estimates time-frequency masks from spike trains. It is able to obtain reconstructions with much higher intelligibility and preserves spatial cues, all without the need for training. These properties enable rapid development of spiking neural network models for audio-related applications.

Within the biologically plausible algorithms we tested, the difference in performance between FRMask and DiffMask is noteworthy and interesting. The spatial tuning plots (**Figure 2**) quantify the tuning of a given spatial channel to a single sound as it is moved around the lateral spatial field which are reasonably well-tuned. Moreover, **Figures 3A,B**, for example, illustrate the response of the 0° channel to sounds presented at 0° and 90°. In this case, the 0° channel responded primarily to the frontal sound. By themselves, these plots do not suggest problems with spatial tuning and leakage. However, in our psychophysical experiments, we presented a target sound at 0° with four competing maskers from ±30° and ±60°, a far more challenging scenario. In such a scenario, spatial leakage is more significant, and refining/improving spatial tuning improves sound segregation, as demonstrated in the improvement with DiffMask over FRMask.

It is also worth noting that our algorithms were based on processing in the barn owl midbrain which contains a topographic map of space, whereas, in mammals, no such topographic map has been found. Despite this difference, the spatially tuned responses of neurons in the model could be leveraged to improve speech segregation performance in humans. This demonstrates that brain inspired algorithms based on non-human model systems can improve human perception and performance.

The work presented here represents a preliminary evaluation of the BOSSA model, and it identified a number of issues and limitations that deserve further investigation. While the formulation of DiffMask can sharpen the spatial tuning of the STNs, neurons tuned to frequencies below 300 Hz were completely silenced for the stimuli we tested (**Figure 4B**). Low spatial acuity in this frequency range results in a similar response at on and off target STNs. The off-target response scaling and summation that forms DiffMask then results in a complete subtraction of on-target activity below 300 Hz. Additionally, some side peaks still persist even after the DiffMask operation, implying that spatial leakage was not fully addressed. Different formulations of the DiffMask may address these shortcomings.

Moreover, our DiffMask implementation used a specific number of off-target STNs at specific locations, which were aligned with the locations of makers in our experimental stimuli. Further works is needed to explore how DiffMask can be optimized to support arbitrary target and masker configurations, and how the resolution of the STNs affects model performance. We have avoided using deep-learning approaches in this study in favor of biological interpretability, but such approaches may help guide the optimization of DiffMask and could be very valuable in that respect. Another potential limitation of the algorithm is that it processes each frequency channel independently. While this design choice reduces both the complexity of the algorithm and its computation time, it excludes the possibility for across-frequency processing that could improve performance (Krishnan et al., 2014; Szabó et al., 2016). Finally, animals have been observed to resolve binaural cue ambiguity by having neurons preferentially tune to more reliable spatial cues in different frequency regions (Cazettes et al., 2014). Inspiration could be taken from these observations to improve spatial tuning and overcome spatial leakage. Again, deep-learning based optimization methods may help identify these reliable cues for human listeners and multitalker mixtures.

Future work with the BOSSA model could include both sound segregation and localization by comparing the response of each spatial tuning curve to predict source azimuth, possibly utilizing a denser array of STNs. Another idea we plan to explore in the future is to apply automatic speech recognition systems to optimize the parameters of the algorithm. This optimization can be performed relatively fast before conducting time-consuming psychophysics experiments. During this optimization process we also plan to investigate the effects of varying sound pressure level and source dynamics on BOSSA performance.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by the Boston University Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

KC designed the algorithm under the supervision of KS and HC, conducted the experiment, analyzed the data, and wrote the first version of the manuscript, with editing by AB, VB, HC, and KS. KC and VB designed the psychophysical experiment. AB assisted with additional data analysis. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by R01 DC000100-42 (HC-PI, KS-Co-I) 12/01/2015–7/31/2020 NIH/NIDCD, Title: Binaural Hearing. VB and AB were supported in part by NIH/NIDCD R01 DC015760.

## Acknowledgments

The authors would like to thank Matthew Goupell for reviewing an earlier version of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Aaronson, N. L., and Hartmann, W. M. (2014). Testing, correcting, and extending the Woodworth model for interaural time difference. *J. Acoust. Soc. Am.* 135, 817–823. doi: 10.1121/1.4861243

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001). "The CIPIC HRTF database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No.01TH8575)*. (Piscataway, NJ: IEEE), 99–102.

Bee, M. A., and Micheyl, C. (2008). The cocktail party problem: What is it? How can it be solved? And why should animal behaviorists study it? *J. Comp. Psych.* 122, 235–251. doi: 10.1037/0735-7036.122.3.235

Bentsen, T., May, T., Kressner, A. A., and Dau, T. (2018). The benefit of combining a deep neural network architecture with ideal ratio mask estimation in computational speech segregation to improve speech intelligibility. *PLoS One* 13:e0196924. doi: 10.1371/journal.pone.0196924

Bernstein, J. G. W., Goupell, M. J., Schuchman, G. I., Rivera, A. L., and Brungart, D. S. (2016). Having two ears facilitates the perceptual separation of concurrent talkers for bilateral and single-sided deaf cochlear implantees. *Ear Hear.* 37, 289–302. doi: 10.1097/AUD.0000000000000284

Best, V., Mejia, J., Freeston, K., van Hoesel, R. J., and Dillon, H. (2015). An evaluation of the performance of two binaural beamformers in complex and dynamic multitalker environments. *Int. J. Audiol.* 54, 727–735. doi: 10.3109/14992027.2015.1059502

Best, V., Roverud, E., Mason, C. R., and Kidd, G. Jr. (2017). Examination of a hybrid beamformer that preserves auditory spatial cues. *J. Acoust. Soc. Am.* 142, EL369–EL374. doi: 10.1121/1.5007279

Bialek, W., Rieke, F., de Ruyter van Steveninck, R., and Warland, D. (1991). Reading a neural code. *Science* 252, 1854–1857. doi: 10.1126/science.2063199

Brainard, M. S., Knudsen, E. I., and Esterly, S. D. (1992). Neural derivation of sound source location: Resolution of spatial ambiguities in binaural cues. *J. Acoust. Soc. Am.* 91, 1015–1027. doi: 10.1121/1.402627

Burkhard, M. D., and Sachs, R. M. (1975). Anthropometric manikin for acoustic research. *J. Acoust. Soc. Am.* 58, 214–222. doi: 10.1121/1.380648

Cazettes, F., Fischer, B. J., and Pena, J. L. (2014). Spatial cue reliability drives frequency tuning in the barn Owl's midbrain. *Elife* 3:e04854. doi: 10.7554/eLife.04854

Chiariotti, P., Martarelli, M., and Castellini, P. (2019). Acoustic beamforming for noise source localization – Reviews, methodology and applications. *Mech. Syst. Signal. Process.* 120, 422–448. doi: 10.1016/j.ymssp.2018.09.019

Chou, K. F., Dong, J., Colburn, H. S., and Sen, K. (2019). A physiologically inspired model for solving the cocktail party problem. *J. Assoc. Res. Otolaryngol.* 20, 579–593. doi: 10.1007/s10162-019-00732-4

Chung, K. (2004). Challenges and recent developments in hearing aids: Part I. speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends Amplif.* 8, 83–124. doi: 10.1177/108471380400800302

Dhamani, I., Leung, J., Carlile, S., and Sharma, M. (2013). Switch attention to listen. *Sci Rep* 3:1297. doi: 10.1038/srep01297

Doclo, S., Gannot, S., Moonen, M., and Spriet, A. (2010). "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, eds S. Haykin and K. J. R. Liu (Hoboken, NJ: John Wiley and Sons), 269–302. doi: 10.1002/9780470487068.ch9

Fischer, B. J., Anderson, C. H., and Peña, J. L. (2009). Multiplicative auditory spatial receptive fields created by a hierarchy of population codes. *PLoS One* 4:e8015. doi: 10.1371/journal.pone.0008015

Gannot, S., Vincent, E., Markovich-Golan, S., and Ozerov, A. (2017). A consolidated perspective on multimicrophone speech enhancement and source separation. *IEEE/ACM Trans. Audio Speech Lang. Process.* 25, 692–730. doi: 10.1109/TASLP.2016.2647702

Ghosh-Dastidar, S., and Adeli, H. (2009). Spiking neural networks. *Int. J. Neural Syst.* 19, 295–308. doi: 10.1142/S0129065709002002

Glasberg, B. R., and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear Res.* 47, 103–138. doi: 10.1016/0378-5955(90)90170-T

Goupell, M. J., Kan, A., and Litovsky, R. Y. (2016). Spatial attention in bilateral cochlear-implant users. *J. Acoust. Soc. Am.* 140, 1652–1662. doi: 10.1121/1.4962378

Goupell, M. J., Stakhovskaya, O. A., and Bernstein, J. G. W. (2018). Contralateral interference caused by binaurally presented competing speech in adult bilateral cochlear-implant users. *Ear Hear.* 39, 110–123. doi: 10.1097/AUD.0000000000000470

Greenberg, J. E., and Zurek, P. M. (2001). "Microphone-array hearing aids," in *Microphone Arrays*, eds M. Brandstein and D. Ward (Berlin: Springer), 229–253. doi: 10.1007/978-3-662-04619-7_11

Greenberg, J. E., Desloge, J. G., and Zurek, P. M. (2003). Evaluation of array-processing algorithms for a headband hearing aid. *J. Acoust. Soc. Am.* 113:1646. doi: 10.1121/1.1536624

Haykin, S., and Chen, Z. (2005). The cocktail party problem. *Neural Comput.* 17, 1875–1902. doi: 10.1162/0899766054322964

Healy, E. W., Yoho, S. E., Wang, Y., and Wang, D. (2013). An algorithm to improve speech recognition in noise for hearing-impaired listeners. *J. Acoust. Soc. Am.* 134, 3029–3038. doi: 10.1121/1.4820893

Hyvärinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. New York, NY: Wiley. doi: 10.1002/0471221317

Kidd, G., Best, V., and Mason, C. R. (2008). Listening to every other word: Examining the strength of linkage variables in forming streams of speech. *J. Acoust. Soc. Am.* 124, 3793–3802. doi: 10.1121/1.2998980

Kidd, G., Mason, C. R., Best, V., and Swaminathan, J. (2015). Benefits of acoustic beamforming for solving the cocktail party problem. *Trends Hear.* 19:233121651559338. doi: 10.1177/2331216515593385

Kochkin, S. (2000). MarkeTrak V: "Why my hearing aids are in the drawer" : The consumers ' perspective. *Hear. J.* 53, 34–41. doi: 10.1097/00025572-200002000-00004

Kochkin, S. (2007). MarkeTrak VII: Obstacles to adult non-user adoption of hearing aids. *Hear J.* 60, 24–51. doi: 10.1097/01.HJ.0000285745.08599.7f

Krishnan, L., Elhilali, M., and Shamma, S. (2014). Segregating complex sound sources through temporal coherence. *PLoS Comput. Biol.* 10:e1003985. doi: 10.1371/journal.pcbi.1003985

Launer, S., Zakis, J. A., and Moore, B. C. J. (2016). "Hearing aid signal processing," in *Hearing Aids*, 1st Edn, eds G. R. Popelka, B. C. J. Moore, R. R. Fay, and A. N. Popper (Berlin: Springer International Publishing), 93–130. doi: 10.1007/978-3-319-33036-5_4

Litovsky, R. Y. (2012). Spatial release from masking. *Acoust. Today* 8:18. doi: 10.1121/1.4729575

Litovsky, R. Y., Goupell, M. J., Misurelli, S. M., and Kan, A. (2017). "Hearing with cochlear implants and hearing aids in complex auditory scenes," in *The Auditory System at the Cocktail Party. Springer Handbook of Auditory Research*, Vol. 60, eds J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay (Cham: Springer), 261–291. doi: 10.1007/978-3-319-51662-2_10

Mandel, M. I., Weiss, R. J., and Ellis, D. P. W. (2010). Model-based expectation maximization source separation and localization. *IEEE Trans. Audio Speech Lang. Process.* 18, 382–394. doi: 10.1109/TASL.2009.2029711

McDermott, J. H. (2009). The cocktail party problem. *Curr. Biol.* 19, R1024–R1027. doi: 10.1016/j.cub.2009.09.005

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *J. Neurophysiol.* 102, 3329–3339. doi: 10.1152/jn.91128.2008

Parthasarathy, A., Hancock, K. E., Bennett, K., DeGruttola, V., and Polley, D. B. (2019). Neural signatures of disordered multi-talker speech perception in adults with normal hearing. *bioRxiv [Preprint]* doi: 10.1101/744813

Pichora-Fuller, M. K., Alain, C., and Schneider, B. A. (2017). "Older adults at the cocktail party," in *The Auditory System at the Cocktail Party*, eds J. C. Middlebrooks, J. Z. Simon, A. N. Poppe, and R. R. Fay (Cham: Springer), 227–259. doi: 10.1007/978-3-319-51662-2_9

Picou, E. M., Aspell, E., and Ricketts, T. A. (2014). Potential benefits and limitations of three types of directional processing in hearing aids. *Ear Hear.* 35, 339–352. doi: 10.1097/AUD.0000000000000004

Qian, Y. M., Weng, C., Chang, X., Wang, S., and Yu, D. (2018). Past review, current progress, and challenges ahead on the cocktail party problem. *Front. Inf. Technol. Electron. Eng.* 19:40–63. doi: 10.1631/FITEE.1700814

Rennies, J., and Kidd, G. (2018). Benefit of binaural listening as revealed by speech intelligibility and listening effort. *J. Acoust. Soc. Am.* 144, 2147–2159. doi: 10.1121/1.5057114

Roman, N., Wang, D., and Brown, G. J. (2003). Speech segregation based on sound localization. *J. Acoust. Soc. Am.* 114, 2236–2252. doi: 10.1121/1.1610463

Roy, K., Jaiswal, A., and Panda, P. (2019). Towards spike-based machine intelligence with neuromorphic computing. *Nature* 575, 607–617. doi: 10.1038/s41586-019-1677-2

Schütt, H. H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vis. Res.* 122, 105–123. doi: 10.1016/j.visres.2016.02.002

Seabold, S., and Perktold, J. (2010). "Statsmodels: Econometric and statistical modeling with Python," in *Proceedings of the 9th Python in Science Conference*. (Austin, TX: SciPy Society), 57. doi: 10.25080/Majora-92bf1922-011

Shinn-Cunningham, B. (2017). Cortical and sensory causes of individual differences in selective attention ability among listeners with normal hearing thresholds. *J. Speech Lang. Hear. Res.* 60, 2976–2988. doi: 10.1044/2017_JSLHR-H-17-0080

Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306

Slaney, M. (1998). Auditory toolbox: A Matlab toolbox for auditory modeling work. *Interval Res. Corp Tech. Rep.* 10:1998.

Srinivasan, S., Roman, N., and Wang, D. L. (2006). Binary and ratio time-frequency masks for robust speech recognition. *Speech Commun.* 48, 1486–1501. doi: 10.1016/j.specom.2006.09.003

Stadler, R. W., and Rabinowitz, W. M. (1993). On the potential of fixed arrays for hearing aids. *J. Acoust. Soc. Am.* 94, 1332–1342. doi: 10.1121/1.408161

Stanley, G. B., Li, F. F., and Dan, Y. (1999). Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *J. Neurosci.* 19, 8036–8042. doi: 10.1523/JNEUROSCI.19-18-08036.1999

Szabó, B. T., Denham, S. L., and Winkler, I. (2016). Computational models of auditory scene analysis: A review. *Front. Neurosci.* 10:524. doi: 10.3389/fnins.2016.00524

Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J. (2010). "A short-time objective intelligibility measure for time-frequency weighted noisy speech," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. (Dallas, TX: IEEE), 4214–4217. doi: 10.1109/ICASSP.2010.5495701

Villard, S., and Kidd, G. Jr. (2019). Effects of acquired aphasia on the recognition of speech under energetic and informational masking conditions. *Trends Hear.* 23:2331216519884480. doi: 10.1177/2331216519884480

Wang, D. (2005). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech Separation by Humans and Machines*. (Boston, MA: Kluwer Academic Publishers), 181–197. doi: 10.1007/0-387-22794-6_12

Wang, D., and Chen, J. (2018). Supervised speech separation based on deep learning: An overview. *IEEE/ACM Trans. Audio Speech Lang. Process.* 26, 1702–1726. doi: 10.1109/TASLP.2018.2842159

Wang, L., Best, V., and Shinn-Cunningham, B. G. (2020). Benefits of beamforming with local spatial-cue preservation for speech localization and segregation. *Trends Hear.* 24:233121651989690. doi: 10.1177/2331216519896908

Wang, Y., Narayanan, A., and Wang, D. (2014). On training targets for supervised speech separation. *IEEE/ACM Trans. Speech Lang. Process.* 22, 1849–1858. doi: 10.1109/TASLP.2014.2352935

Woodworth, R. S. (1938). *Experimental Psychology*. New York, NY: Holt.

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpsyg.2022.1009463



OPEN ACCESS

EDITED BY
Lina Reiss,
Oregon Health and Science University,
United States

REVIEWED BY
Jesyin Lai,
St. Jude Children's Research Hospital,
United States
Tim Jürgens,
University of Applied Sciences Lübeck,
Germany

*CORRESPONDENCE
Yang-Soo Yoon
yang-soo_yoon@baylor.edu

SPECIALTY SECTION
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology

RECEIVED 01 August 2022
ACCEPTED 05 October 2022
PUBLISHED 20 October 2022

CITATION
Yoon Y-S and Morgan D (2022) Dichotic spectral integration range for consonant recognition in listeners with normal hearing.
*Front. Psychol.* 13:1009463.
doi: 10.3389/fpsyg.2022.1009463



# Dichotic spectral integration range for consonant recognition in listeners with normal hearing

Yang-Soo Yoon* and Dani Morgan

Laboratory of Translational Auditory Research, Department of Communication Sciences and Disorders, Baylor University, Waco, TX, United States

Dichotic spectral integration range, or DSIR, was measured for consonant recognition with normal-hearing listeners. DSIR is defined as a frequency range needed from 0 to 8,000Hz band in one ear for consonant recognition when low-frequency information of the same consonant was presented to the opposite ear. DSIR was measured under the three signal processing conditions: (1) unprocessed, (2) target: intensified target spectro-temporal regions by 6dB responsible for consonant recognition, and (3) target minus conflicting: intensified target regions minus spectro-temporal regions that increase confusion. Each consonant was low-pass filtered with a cutoff frequency of 250, 500, 750, and 1,000Hz, and then was presented in the left ear or low-frequency (LF) ear. To create dichotic listening, the same consonant was simultaneously presented to the right ear or high-frequency (HF) ear. This was high-pass filtered with an initial cutoff frequency of 7,000Hz, which was adjusted using an adaptive procedure to find the maximum high-pass cutoff for 99.99% correct consonant recognition. Mean DSIRs spanned from 3,198–8,000Hz to 4,668–8,000Hz (i.e., mid-to-high frequencies were unnecessary), depending on low-frequency information in the LF ear. DSIRs narrowed (i.e., required less frequency information) with increasing low-frequency information in the LF ear. However, the mean DSIRs were not significantly affected by the signal processing except at the low-pass cutoff frequency of 250Hz. The individual consonant analyses revealed that /ta/, /da/, /sa/, and /za/ required the smallest DSIR, while /ka/, /ga/, /fa/, and /va/ required the largest DSIRs. DSIRs also narrowed with increasing low-frequency information for the two signal processing conditions except for 250 vs. 1,000Hz under the target-conflicting condition. The results suggest that consonant recognition is possible with large amounts of spectral information missing if complementary spectral information is integrated across ears. DSIR is consonant-specific and relatively consistent, regardless of signal processing. The results will help determine the minimum spectral range needed in one ear for consonant recognition if limited low spectral information is available in the opposite ear.

KEYWORDS

**dichotic hearing, spectral integration, binaural integration, consonant recognition, articulation-index gram**

## Introduction

Normal hearing (NH) listeners receive the same or similar auditory input from each ear, and the input is then sent to the higher auditory system for further processing, such as spectral integration (Ronan et al., 2004; Hall et al., 2008; Fox et al., 2011; Räsänen and Laine, 2013; Grose et al., 2016). However, individuals with hearing loss may receive different spectral information from each ear and are forced to integrate them across the ears, that is, dichotic spectral integration (Tononi, 2010; Kong and Braida, 2011; Yang and Zeng, 2013; Reiss et al., 2014; Obuchi et al., 2015). This dichotic spectral integration occurs when different frequency information is dichotically and simultaneously presented to both ears. The improvement in the performance of listeners with hearing loss as signal bandwidth widens is thought to reflect the ability of the auditory system to integrate information across a wide frequency range in complex sounds (Spehar et al., 2008; Happel et al., 2010). Regardless of hearing status, dichotic spectral integration is important for efficient communication, such as speech perception. However, it is hard to find dichotic spectral integration studies with NH and hearing-impaired listeners. In the present study, a frequency range needed in the right ear for consonant recognition was determined when low-frequency information of the same consonant presented to the right ear is presented in the left ear in NH listeners. In this study, this frequency range was named a "dichotic spectral integration range (DSIR)."

It is known that consonant recognition requires the listener's ability to discriminate specific spectral and temporal acoustic cues such as voicing, the onset of the noise burst, and spectral and temporal transitions (Miller and Nicely, 1955; Stevens and Klatt, 1974; Stevens and Blumstein, 1978; Blumstein and Stevens, 1979, 1980). In contrast, a few studies measured the range of spectral information needed for consonant recognition. Lippmann (1996) measured consonant–vowel–consonant syllable recognition in NH listeners when low-frequency information below 800 Hz was combined with high-frequency information above 4,000 Hz in a monotic listening condition (i.e., different frequency information is simultaneously presented to the same ear). This monotic spectral integration study showed that removing midfrequency consonant information (800–4,000 Hz) did not significantly alter consonant recognition. Ronan et al. (2004) did not determine the spectral integration range but demonstrated a relationship between speech perception and monotic spectral integration in NH listeners. They filtered consonants in two widely separated bands (0–2,100 Hz and 2,100–4,500 Hz) of speech. They observed that consonant enhancement is related to the ability of integrating widely separated two bands.

Some other studies showed that dichotic spectral integration (i.e., different frequency information is dichotically and simultaneously presented to both ears.) facilitates sentence perception (Hall et al., 2008; Grose et al., 2016). Hall and colleagues first determined the bandwidths required for approximately 15–25% correct sentence recognition in quiet and noise conditions in listeners with NH and hearing loss (Hall et al., 2008). They then adaptively varied the bandwidth of filtered sentences centered on low (500 Hz) and high (2,500 Hz) frequencies and measured speech perception when the two bandwidths were presented simultaneously to both ears. NH and hearing-impaired listeners observed higher percent performance (64–94%) with dichotic spectral listening compared to a 30–50% additive combination of information presented in the single-band conditions. Grose et al. (2016) also reported similar results as Hall et al. (2008) study but with middle-aged and older NH listeners.

The ability to integrate spectral information across ears may be affected when useful frequency information for speech perception is manipulated, such as being intensified or eliminated. Allen's group identified specific frequency and time regions for the consonant perception that resulted in an improved consonant recognition, called "target frequency and time regions." They also identified specific frequency and time regions that lead to significant consonant confusions, called "conflicting frequency and time regions" (Li et al., 2010, 2012). Consonant recognition was measured with + 6 dB gain on the target (frequency and time) regions and complete removal of the conflicting (frequency and time) regions for consonants. The results from these four studies indicated that the intensified target and removal of the conflicting regions enhance consonant recognition by a minimum of 3 percentage points to a maximum of 70 percentage points. This type of signal processing with the target and conflicting regions will enhance speech perception in listeners with normal hearing. However, listeners with hearing loss with or without devices may not integrate these regions appropriately across ears due to abnormal binaural spectral integration, i.e., fusion and averaging of information from widely different frequency regions (Reiss et al., 2014). This can lead to interference, as was shown for vowel integration (Fowler et al., 2016). Under this listening condition, some listeners may experience spectral interference rather than spectral integration. For example, Fowler et al. (2016) demonstrated that bimodal patients who had better residual hearing (< 60 dB HL at 250 and 500 Hz) in the hearing aid ear received improved speech perception in quiet when low-to-mid (approximately 440–982 Hz) frequencies in cochlear implant ear were removed. Removing mid-frequency information processed by cochlear implant ear may reduce bimodal interference and/or enhance bimodal integration. It is also possible that the AI-Gram signal processing would result in ear-dominance listening when the target and conflicting regions are processed by one ear with a better performing ear (e.g., cochlear implant ear in bimodal hearing). An ear dominance listening results in information presented to one ear being primarily processed and perceived, while information presented to the opposite ear is less utilized and perceived (Reiss et al., 2016). Under ear-dominance listening, dichotic spectral integration will be less affected with the poorer ear (i.e., hearing aid ear in bimodal hearing). So, the findings of Li et al. studies (Li et al., 2010, 2012) led to the working hypothesis that the DSIR will be significantly reduced if target regions are

intensified while the conflicting regions are removed. It is also hypothesized that DSIRs will be narrowed with increasing low-frequency information in the opposite ear.

In summary, previous studies demonstrate that spectral integration within and between ears is important for speech perception using two broad frequency bands (Ronan et al., 2004; Hall et al., 2008; Räsänen and Laine, 2013; Grose et al., 2016). However, spectral integration may occur at specific narrower frequency bands, and additional spectral integration on other bands may not be critical for speech perception. It is also possible that the spectral integration range is listener specific for speech recognition. For example, individuals with different degrees of hearing loss in one ear may need different ranges of spectral information in the opposite ear for good speech recognition. Another challenging aspect of the previous studies is the use of sentences (Hall et al., 2008; Grose et al., 2016). Sentences are more realistic stimuli compared to tones or nonsense syllables. However, the minimum spectral ranges required for sentence perception would be similar regardless of the use of different sentences. Measuring DSIRs for phonemes (i.e., basic units of sentences) will provide us discrete information which can be effectively used in training machine-learning algorithms. In the present study, DSIRs were determined for individual consonant recognition in the right ear when different amounts of low-frequency information were presented to the left ear in NH listeners. The DSIR measurement was administered under the three signal processing conditions: unprocessed, with target frequency and time regions intensified by +6 dB gain (i.e., target condition), and both the target regions intensified, and conflicting regions removed (i.e., target minus conflicting or target-conflicting condition). The results of the present study will help determine the minimum spectral range needed in one ear for individual consonant recognition if limited low spectral information is available in the opposite ear. The results can also serve as control data for future studies with hearing-impaired listeners and bimodal users.

# Materials and methods

## Subjects

Fourteen NH adults participated (11 females and three males; average age: 24 ± 6.7). A reason for this imbalance of subject gender was that the subjects were mainly recruited from Robbins College of Health and Human Sciences at Baylor University, where female students outnumber male students. All subjects were native speakers of American English. All participants had thresholds better than 20 dB HL (hearing level) for both ears at audiometric frequencies ranging from 250 to 8,000 Hz. Interaural threshold differences were less than 10 dB HL. All procedures were approved by the Baylor University Institutional Review Board (#1253711). The Board has determined that the research agrees with the declaration of Helsinki.

## Stimuli

Stimuli included 14 frequently used American English consonants with the common vowel /ɑ/ (/pa/, /ba/, /ta/, /da/, /ka/, /ga/, /ma/, /na/, /fa/, /va/, /sa/, /za/, /ʃa/, and /ʃa/; Hayden, 1950). Each consonant was produced with a sampling frequency of 44,100 Hz by a single female talker whose average fundamental frequency was 228 Hz. Completely silent parts from both onsets and offsets of consonant syllables were identified on time waveforms and spectrograms and manually removed. The average duration and standard deviation (SD) of consonants was 406.57 ± 102.61 ms. The duration of each consonant is provided in Table 1. To limit the spectral range of consonants to 0–8,000 Hz, each consonant was low-pass filtered with a cutoff frequency of 8,000 Hz (IIR second-order Butterworth with 12 dB/oct roll-off and a zero-phase shift). All stimuli were then normalized to have the same long-term root-mean-square energy (65 dBA sound pressure level or SPL). The stimuli was delivered to both ears *via* circumaural headphones (Sennheiser HDA-200) at the subject's most comfortable level (ranges: 50–70 dB SPL), which was established by the subjects' responses to the 14 unprocessed consonants from the stimuli listed above in quiet according to the Cox loudness rating scale (Cox, 2005).

## Articulation index-gram processing on the target and conflicting frequency and time regions

The same target frequency and time regions of consonants used in our previous study in NH listeners (Yoon, 2021) was

TABLE 1 The target and conflicting frequency and time regions used for the AI-Gram processing (Yoon, 2021).

| Consonants | Duration [ms] | Target frequency [kHz] | Conflicting frequency [kHz] | Target time [ms] |
|---|---|---|---|---|
| /pa/ | 240 | 0.3–7.4 | 1.4–2.0 | 32–62 |
| /ba/ | 331 | 0.3–4.5 | 0.6–2.2 | 7–22 |
| /ta/ | 338 | 3.0–7.4 | 1.6–2.8 | 42–62 |
| /da/ | 240 | 4.0–7.8 | 1.4–2.8 | 38–48 |
| /ka/ | 447 | 1.4–2.0 | 5.0–7.8 | 30–50 |
| /ga/ | 348 | 1.4–2.0 | 3.9–5.0 | 10–30 |
| /ma/ | 350 | 0.5–1.3 | 1.2–1.9 | 25–55 |
| /na/ | 400 | 1.5–2.2 | 0.4–0.9 | 77–127 |
| /fa/ | 548 | 0.6–2.2 | 3.0–7.8 | 141–166 |
| /va/ | 349 | 0.6–1.4 | 1.4–4.4 | 16–46 |
| /sa/ | 501 | 3.9–7.8 | 5.4–7.8 | 80–115 |
| /za/ | 501 | 3.6–7.8 | 3.5–5.4 | 90–120 |
| /ʃa/ | 549 | 2.0–3.7 | 4.0–7.8 | 40–160 |
| /ʒa/ | 550 | 1.9–3.7 | 5.4–7.8 | 15–115 |

Duration of each consonant is also given. The entire range of frequency that was used for the identification of the target frequency regions was 0–8 kHz.

employed in this study. Figure 1 shows timewave forms and power spectrum for the unprocessed, target, and target-conflicting processed /ka/. Arrows indicate the target portion of timewave forms. Dotted rectangles and solid ovals on the power spectrums indicate the target and conflicting regions, respectively. The AI-Gram was originally developed by Li et al. (2010, 2012). The AI-Gram was implemented on the MATLAB platform (The MathWorks, 2017) for our conditions (Yoon, 2021). The AI-Gram construction procedures are explained in detail in Yoon (2021). A full discussion of how consistent the target and conflicting frequency regions are with respect to earlier studies (Li et al., 2010, 2012) can also be found in Yoon (2021).

In brief, using a low-pass and high-pass filtering scheme (IIR second-order Butterworth with −12 dB/oct roll-off and a zero-phase shift for both filters), the target frequency regions were identified. These target regions for each consonant are the frequency regions responsible for significant changes in consonant recognition. For example, /ka/ was presented and its perception scores were considerably improved (from 40 to 90%) when the low-pass filter (LPF) cutoff was moved from 1.4 to 1.5 kHz. So, the lower edge of the target frequency would be 1.4 kHz. When the high-pass filter (HPF) cutoff was moved from 2.0 to 2.1 kHz, the recognition of /ka/ considerably dropped (from 90 to 40%). So, the upper edge of the target frequency would be 2.0 kHz. Therefore, the final target frequency region would be 1.4–2.0 kHz for /ka/. This target frequency region included the spectral region (i.e., 1.4 kHz) that leads to improved consonant perception when LPF cutoff frequency raised from 1.4 kHz to 1.5 kHz but excluded the spectral region (i.e., 2.1 kHz) that leads to a potential deteriorated consonant perception when HPF cutoff frequencies raised from 2.0 to 2.1 kHz. The conflicting frequency regions are the frequency regions that yielded the peak errors of the most confused consonants and 20% less than the peak error from both filtering schemes. For example, when /ʃa/ was presented, the recognition of the confused consonant /sa/ reached 24% when the LPF cutoff was 4.0 kHz and a maximum of 30% when the cutoff was moved from 4.0 to 4.1 kHz (i.e., 24% is 20% below the peak 30% error). Therefore, the lower edge of the conflicting frequency would be 4.0 kHz. When the HPF cutoff was 7.8 kHz, the recognition of the confused consonant /sa/ reached a score of 24% and a maximum of 30% when the cutoff was moved from 7.8 to 7.7 kHz. So, the upper edge of the conflicting frequency would be 7.8 kHz. Thus, the final conflicting frequency range would be 4.0–7.8 kHz for recognition of the consonant /ʃa/. Full descriptions of selection criteria for target and conflicting regions can be found in Yoon (2021).

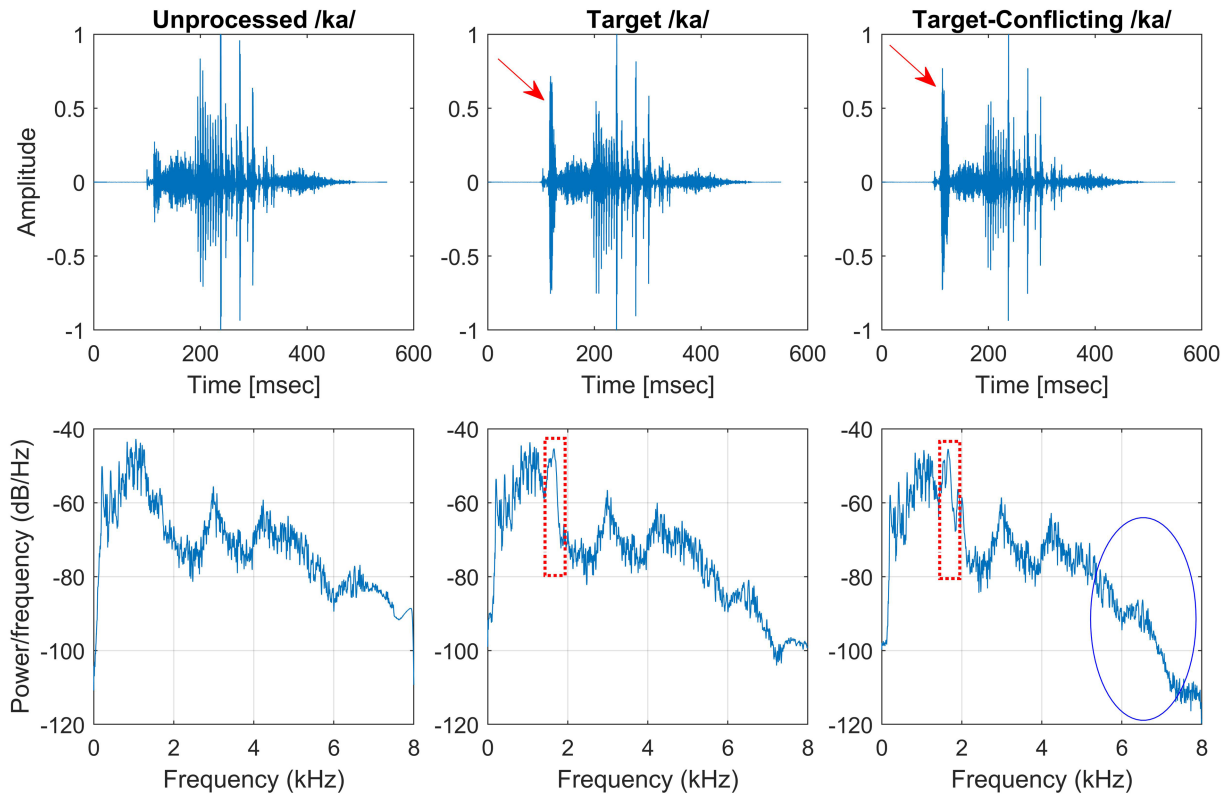



FIGURE 1
Stimulus wave forms (top panels) and power spectra (bottom panels) for the unprocessed (left column), target (middle column), and target-conflicting (right column) conditions forn/ka/. Arrows indicate the target portion of stimulus wave forms. Dotted rectangles and solid oval on the power spectra indicate the target and conflicting regions, respectively.

Analogously, using a truncation approach, the target time regions for each consonant was identified by finding the time segment of the consonant responsible for significant change in consonant recognition. The initial duration of each consonant was 3% of the total duration from the onset (i.e., the remaining 97% of the consonant was truncated out) so that minimal consonant information was presented. The duration of the consonant was increased by 1 ms when a participant's response was incorrect. If perception scores for /ka/ dropped significantly (i.e., from 90 to 40%) when the time-truncation point increased from 30 to 50 ms the onset of the consonant, it suggests that important temporal cues resided within the 30–50 ms time window. Again, these target frequency and time regions used for the current study were obtained from NH listeners in the binaural hearing condition and in quiet (Yoon, 2021). After identifying the target frequency and time regions for each of the 14 consonants using the AI-Gram, a 6-dB gain was applied to those target frequency and time regions for each consonant (i.e., other frequency and time regions for each consonant were intact). The conflicting frequency and time regions were also removed. For the three consonants (/pa/, /ba/, and /za/) with overlapping target and conflicting frequency ranges (Figure 2 in Yoon, 2021), the target frequency ranges were intensified, while the overlapped conflicting frequency ranges were not removed. It should be noted that AI-Gram does not have the ability to apply a 6 dB gain and removal on the exact target and conflicting regions. So, some variations should be expected on the power spectrums for the target and target-conflicting processing conditions, as shown in Figure 1. The completed AI-Gram processing was then verified by five adult NH listeners. The verification procedures can also be found in Yoon (2021). Table 1 lists the resultant target and conflicting frequency and time regions. Note that the target time region in Table 1 indicates a temporal duration of consonants from the onset of the consonant.

## Procedure

The DSIR was binaurally measured in quiet under three signal processing conditions: (1) unprocessed, (2) target: intensified target frequency and time regions responsible for consonant recognition, and (3) target-conflicting: combined intensified target frequency and time regions and removed conflicting frequency and time regions responsible for consonant confusions. Subjects were seated in a single-walled sound-treated booth (Industrial Acoustics Company). Before formal testing, a 30-min familiarization on all 14 consonants was binaurally provided for the target and target-conflicting signal processing conditions in a quiet environment (15-min each). Each consonant was low-pass filtered (IIR fifth-order Butterworth with 30 dB/oct roll-off) in the left ear, with one of the four fixed cutoff frequencies: 250, 500, 750, and 1,000 Hz. Group delay created by filtering was removed by applying zero-phase filtering technique. These cutoff frequencies were purposefully chosen because they are the typical frequencies of residual hearing in individuals who utilize bimodal hearing (Smith-Olinde et al., 2004; Jürgens et al., 2011; Bianchi et al., 2016; Patel and McKinnon, 2018; Varnet et al., 2019; Yoho and Bosen, 2019). Results from these chosen cutoff frequencies can be used for future comparison with data that will be measured in individuals with hearing aids and cochlear implants. In the right ear, the same consonant was presented with an initial HPF cutoff frequency of 7,000 Hz (IIR fifth-order Butterworth with 30 dB/oct roll-off). Zero group delay was achieved by applying a zero-phase filtering on filtered signals. An incorrect response lowered the cutoff frequency in 100-Hz decrements (i.e., the cutoff frequency was reduced from 7,000 to 6,900 Hz). So, low-frequency information was presented to the left ear which was designated as the "low-frequency or LF ear," and the high-frequency information was presented to the right ear which was designated as the "high-frequency or HF ear." Under these LF and HF ear settings, the stimulus was dichotically and simultaneously delivered *via* an audiometer (GSI AudioStar Pro) to Sennheiser HDA-200 circumaural headphones. In fixed block trials, DSIR was determined, using the 15-alternative forced-choice paradigm, along with the additional option of "none of these." With each of the four fixed low-pass filter cutoff frequencies used in the LF ear, each consonant was presented five times for each signal processing, and the order of consonant presentation was fully randomized. The DSIR was determined when the consonant presented was correctly selected three times in a row. These procedures were repeated for the unprocessed, target, and target-conflicting signal processing conditions. No trial-by-trial feedback was provided during the test. The complete test protocol (3 signal processing conditions × 4 LPF cutoff frequencies × 14 consonants × 5 repetitions), including five-minute breaks (at least two breaks per hour and instructed to take breaks as needed) and the consenting process, took approximately 11 h per listener, requiring four separate visits.

## Data analysis

Parametric statistics were used with Sigma Plot (SYSTAT, 2021). Before performing statistical analyses, the normality (Shapiro–Wilk) test and equal variance (Brown-Forsythe) test were performed, and all passed. To determine the main effect of the AI-Gram signal processing and LPF cutoff frequencies on mean DSIRs (Figure 2), a two-way repeated measures analysis of variance (ANOVA) was performed with two within-subject factors: the AI-Gram (Unprocessed, Target, and Target-Conflicting) and LPF cutoff frequency (250, 500, 750, and 1,000 Hz). A two-way repeated ANOVA was also performed with two within-subject factors (i.e., the AI-Gram and each consonant) to determine how DSIR for individual consonants was affected by the AI-Gram signal processing (Figure 3). A two-way repeated ANOVA was performed with two within factors (LPF cutoff frequency and each consonant) to determine how the DSIR of each consonant was affected by the LPF cutoff frequency used in the LF ear (Figure 4). Pearson correlation analyses were conducted

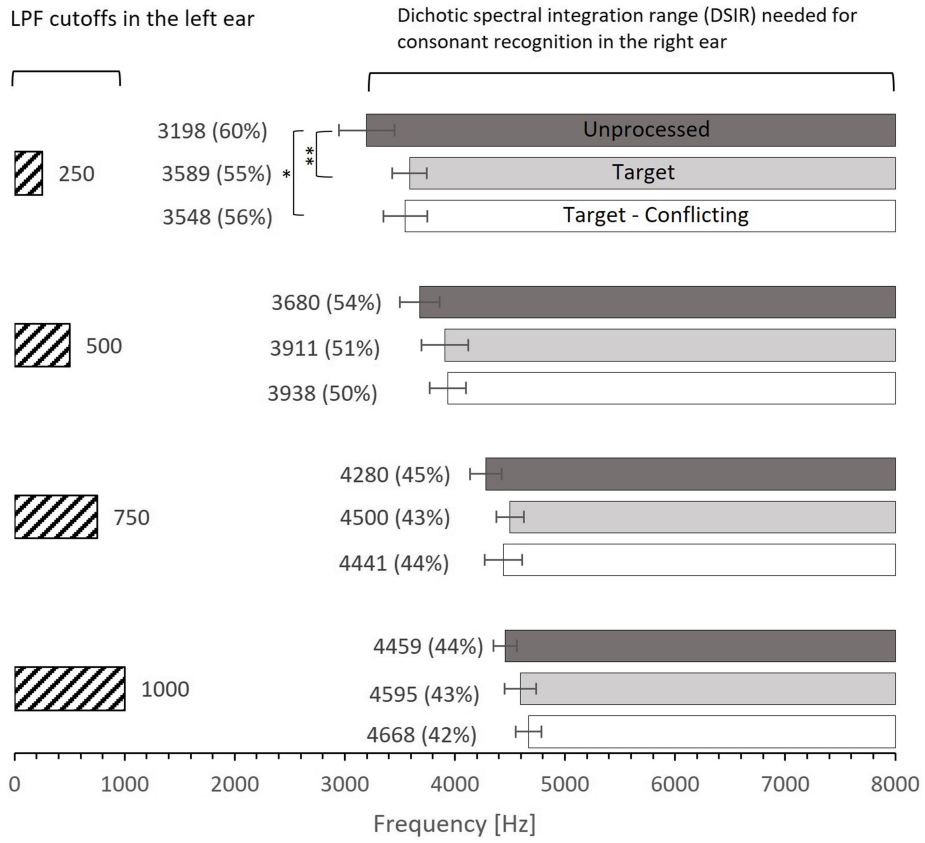



FIGURE 2
Mean dichotic spectral integration range (DSIR) needed for consonant recognition in the HF ear for each LPF cutoff frequency in the LF ear. Dark- and light-filled bars indicate the unprocessed and target conditions, while the open bars indicate the target-conflicting condition. Numbers in the parentheses are the percentages of DSIR out of 0–8,000Hz band required for consonant recognition in the HF ear (e.g., 60% for the unprocessed condition at 250Hz was obtained from 8,000–3,198Hz=4,802Hz, which is 60% of the 0–8,000 band). Error bars indicate standard errors. $^{**}p<0.01$ and $^{*}p<0.05$.

to determine any systematic relationship in the DSIRs with different LPF cutoff frequencies (Figure 5). The results of all statistical analyses were assessed against an alpha level of 0.05 with a two-tailed test. Planned multiple comparisons were performed using an overall alpha level of 0.05 with the Bonferroni correction.

# Results

## Mean DSIR

Figure 2 shows mean DSIR with the standard error for each LPF cutoff frequency used for the LF ear. All DSIRs should be interpreted as lower bound frequencies required for consonant recognition from the 0–8,000 Hz band. For example, DSIR of 3,198 Hz (for 250 Hz cutoff frequency and the unprocessed conditions) means that a frequency range of 3,198–8,000 Hz was required for consonant recognition in the HF ear when low-frequency information below 250 Hz was presented to the LF ear. The numbers in parentheses are the percentages of DSIRs needed for consonant recognition from the 0–8,000 Hz band. For instance, 60% (for 250 Hz cutoff frequency and the unprocessed conditions) means that the DSIR of the 3,198–8,000 Hz covers 60% of the upper portion of the 0–8,000 Hz band. The results show that consonant recognition was achieved with large amounts of spectral information missing. DSIRs narrowed (i.e., required less spectral information) with increasing the LPF cutoff frequency. A two-way repeated measure analysis of variance (ANOVA) showed a significant effect of AI-Gram processing effect, $F(2,78) = 4.28$, $p = 0.02$ and of the LPF cutoff frequency on DSIRs, $F(3,36) = 46.55$, $p < 0.001$. However, no significant interactions were observed between the signal processing and the LPF cutoff frequency, $F(6,78) = 1.29$, $p = 0.32$. All pairwise multiple comparisons across signal processing conditions, with Bonferroni correction for the AI-Gram processing, showed that only two pairs were significant within the cutoff frequency of 250 Hz: unprocessed vs. target ($p = 0.005$) and unprocessed vs. target-conflicting ($p = 0.01$), indicated by asterisks in Figure 2. Across the LPF cutoff frequency, differences between all pairs are significant except for pair 750 Hz vs. 1,000 Hz within all three signal-processing conditions and pair 250 vs. 500 Hz within the target condition. Details are given in Table 2.

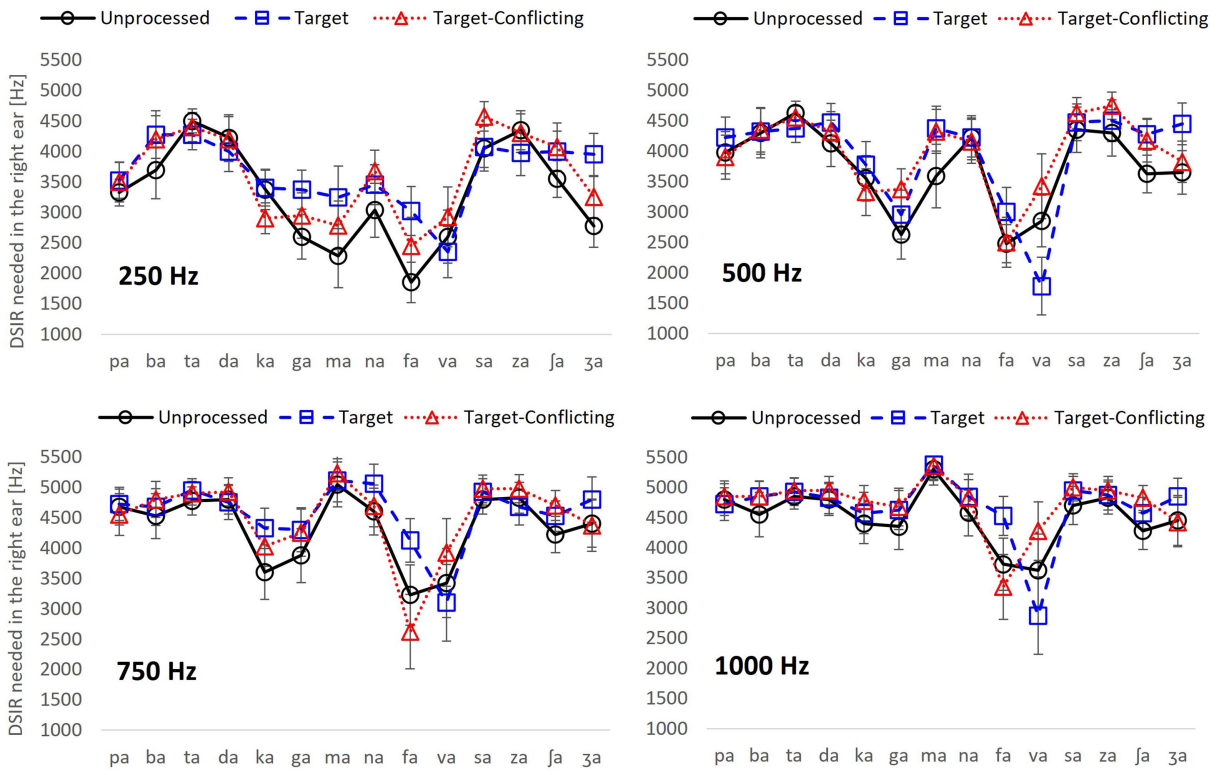



FIGURE 3
Dichotic spectral integration range (DSIR) out of 0–8,000Hz band in the HF ear with the standard errors for individual consonant as a function of signal processing for each LPF cutoff frequency in the LF ear.

## Dichotic spectral integration range for individual consonant

To determine what frequency range is needed for the recognition of each consonant, DSIR per consonant was plotted as a function of the signal processing condition for each LPF cutoff frequency in Figure 3. Two overall findings are that DSIRs are highly consonant-specific, and the patterns of DSIRs are similar between 250 and 500 Hz LPF cutoff frequencies, as well as between 750 and 1,000 Hz LPF cutoff frequencies.

For the LPF cutoff frequency of 250 Hz, a two-way repeated measure ANOVA showed that DSIRs were significantly different across consonants, $F(13,338) = 10.70$, $p < 0.001$ but not across the AI-Gram signal processing, $F(2,338) = 1.91$, $p = 0.17$. Significant interactions were observed, $F(26, 338) = 1.82$, $p = 0.009$. Based on the shapes of the DSIRs, there were two subgroups: five consonants (/ka/, /ga/, /ma/, /fa/, and /va/), requiring wide DSIRs, and the remaining nine consonants requiring relatively narrow DSIRs. This subgrouping was supported by the results of pairwise multiple comparisons with Bonferroni correction (Table 3). These five consonants required significantly wider DSIRs compared to the other nine consonants. For the LPF cutoff frequency of 500 Hz, significant difference in DSIRs were observed across consonants, $F(13,338) = 14.36$, $p < 0.001$, but no significant effect of the AI-Gram signal processing, $F(2,338) = 0.94$, $p = 0.40$. Significant interactions were observed, $F(26, 338) = 2.52$, $p < 0.001$. Observed with the 250 Hz, the same five consonants (/ka/, /ga/, /ma/, /fa/, and /va/) required wider DSIRs in the HF ear than DSIRs for other nine consonants. It should also be noted that DSIRs for the two consonants (/ka/ and /ma/) slightly narrowed, compared to those with the 250 Hz. Table 4 shows the results of pairwise multiple comparisons.

With the LPF cutoff frequency of 750 Hz, each consonant required significantly different DSIRs, $F(13,338) = 6.28$, $p < 0.001$, but AI-Gram signal processing did not affect DSIRs significantly, $F(2,338) = 1.80$, $p = 0.19$. There were significant interactions between the variables, $F(26, 338) = 2.64$, $p < 0.001$. With the LPF cutoff frequency of 1,000 Hz, a significant main effect of the consonant was observed, $F(13,338) = 5.60$, $p < 0.001$, but no significant main effect of the AI-Gram signal processing was observed, $F(2,338) = 1.35$, $p = 0.28$. Significant interactions occurred between the type of consonant and AI-Gram signal processing, $F(26, 338) = 2.95$, $p < 0.001$. The patterns of DSIRs are similar between the 750 Hz and 1,000 Hz cutoff frequencies, as observed in the 250 and 500 Hz LPF cutoff frequency conditions, four consonants (/ka/, /ga/, /fa/, and /va/) still required relatively wider DSIRs in the two higher cutoff frequencies. The two consonants, (/fa/ and /va/) in particular, required wider DSIRs than the other two consonants (/ka/ and /ga/). However, /ma/ then had very narrow DSIRs for LFP of 750 and 1,000 Hz for all signal

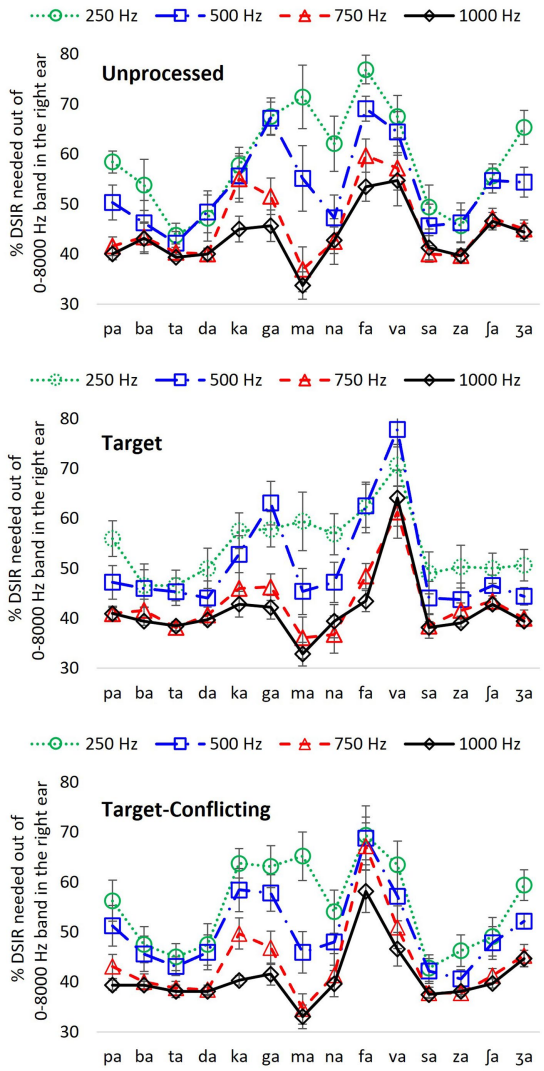



FIGURE 4
Percentage of DSIR out of 0–8,000Hz band in the HF ear for individual consonant recognition as a function of LPF cutoff frequency in the LF ear for each signal processing condition.

processing conditions. The pairwise multiple comparisons supported these findings. Tables 5, 6 present all pairwise multiple comparisons for the 750 Hz and 1,000 Hz cutoff frequencies, respectively.

## Effect of low-frequency information on DSIRs

Figure 3 presents the actual frequency values of DSIRs per consonant for each low-frequency information available in the LF ear. However, it is hard to remember these frequency values and to see the effect of the different low-frequency information on the DSIR metrics. To generate easier metrics, the DSIRs were converted into percentages of the frequency ranges from the 0–8,000 Hz band. As discussed above in the *Mean DSIR* part of the Results section, these percentages of the DSIRs indicate the upper portion of the 0–8,000 Hz band required for consonant recognition. For example, 70% means 70% of the upper portion of the 0–8,000 Hz band, that is, the 2,400–8,000 Hz range. Figure 4 shows the mean percentages of the DSIRs in the HF ear as a function of the LPF cutoff frequency used in the LF ear.

For the unprocessed condition, four consonants (/ta/, /da/, /sa/, and /za/) required less than 50% of the 0–8,000 Hz band, while two consonants (/fa/ and /va/) needed more than 50% regardless of the LPF cutoff frequency. For the remaining nine consonants, the percentage of the DSIRs varied (more than 20% differences), depending on LPF cutoff frequencies. A two-way repeated measure of ANOVA showed significant effects of the LPF cutoff frequency, $F(3,507) = 29.64$, $p < 0.001$ and of the consonant, $F(13,507) = 12.85$, $p < 0.001$. Significant interactions were also observed, $F(39,507) = 4.97$, $p < 0.001$. Pairwise multiple comparisons with a Bonferroni correction were also performed. However, to demonstrate the different overall effects of the LPF cutoff frequency, the mean differences among the LPF cutoff frequencies were reported rather than to present all pairwise multiple comparisons. The analyses showed significant mean differences between any pairs of the LPF cutoff frequencies ($p < 0.001$) except for the pair 750 vs. 1,000 Hz ($p = 1.00$).

Compared to the unprocessed condition, smaller percentages of DSIRs were needed with the target condition. Seven consonants including the three observed in the unprocessed condition (/ba/, /ta/, /da/, /sa/, /za/, /ʃa/, and /ʒa/) needed less than 50% of the 0–8,000 Hz band regardless of the LPF cutoff frequency, while only /va/ needed more than 50%. The remaining six consonants, including the six observed in the unprocessed condition, exhibited more than 20% differences across the LPF cutoff frequencies. There was a significant difference in the percentage of DSIRs across consonants, $F(13,507) = 18.52$, $p < 0.001$ and the LPF cutoff frequency, $F(3,507) = 27.41$, $p < 0.001$. Significant interactions were also observed, $F(39,507) = 3.37$, $p < 0.001$. Significant mean differences were evident in multiple comparisons between any pairs of the low frequencies ($p < 0.001$), except for the pair 250 vs. 500 Hz ($p = 0.12$) and the pair 750 vs. 1,000 Hz ($p = 1.00$).

For the target-conflicting condition, six consonants (/ba/, /ta/, /da/, /sa/, /za/, and /ʃa/) required less than 50%; however, consonant /fa/ needed more than 50% regardless of the LPF cutoff frequency. The remaining seven consonants, including the five observed in the unprocessed and target conditions, exhibited more than 20% differences across the LPF cutoff frequencies. There was a significant difference in the percentage of DSIRs across consonants, $F(13,507) = 15.00$, $p < 0.001$ and across the LPF cutoff frequency, $F(3,507) = 18.66$, $p < 0.001$. Significant interactions were also observed, $F(39,507) = 3.82$, $p < 0.001$. Multiple comparisons showed significant differences between any pairs of the LPF cutoff frequencies ($p < 0.001$), except for pair 250 vs. 500 Hz ($p = 0.12$) and pair 750 vs. 1,000 Hz ($p = 1.00$).

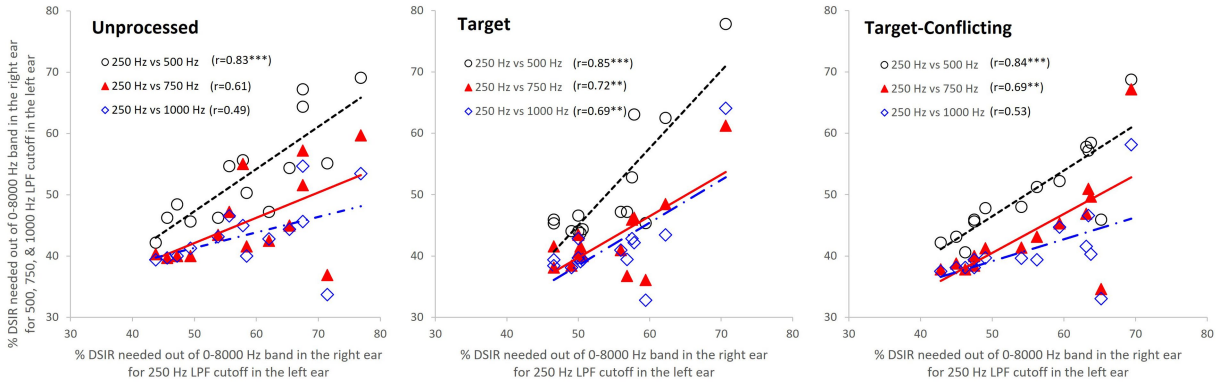



FIGURE 5
Scatter plots among percentages of DSIRs. Each data point represents the average DSIR percentage of each syllable across subjects. The *X*-axis indicates DSIRs for 250Hz LPF cutoff frequency, while the y-axis indicates DSIRs for LPF cutoff frequencies of 500, 750, and 1,000Hz. The open circle is a scatter plot of DSIRs of 250 and 500Hz LPF cutoff frequencies, while the filled triangle is a scatter plot of DSIRs of 250 and 750Hz LPF cutoff frequencies. The open diamond is a scatter plot of DSIRs of 250 and 1,000Hz cutoff frequencies. ***$p$<0.001 and **$p$<0.01.

TABLE 2 Pairwise multiple comparisons for LPF cutoff frequencies in the LF ear.

| | 250 vs. 500 Hz | 250 vs. 750 Hz | 250 vs. 1,000 Hz | 500 vs. 750 Hz | 500 vs. 1,000 Hz | 750 vs. 1,000 Hz |
|---|---|---|---|---|---|---|
| Within unprocessed | ** | *** | *** | *** | *** | ns |
| Within target | ns | *** | *** | *** | *** | ns |
| Within target-conflicting | * | *** | *** | *** | *** | ns |

***$p$ < 0.001; **$p$ < 0.01; *$p$ < 0.05.
ns stands for not significant.

## Interrelationship among percentages of DSIRs

To quantify the relationship between the changes of DSIRs and different LPF cutoff frequencies, Pearson's correlation analyses were conducted. Figure 5 shows scatter plots with $r$ values and regression lines. As a reference, the DSIRs assessed with the LPF cutoff frequency of 250 Hz were on the x-axis and DSIRs assessed with the other three cutoff frequencies were on the y-axis. Since the DSIR data assessed with the 250 Hz cutoff frequency was used three times for the analyses, a Bonferroni corrected $p$ value (i.e., 0.05/3 = 0.017) was used. The overall trends of the analyses show that consonants requiring wide DSIRs in the 250 Hz condition also required wide DSIRs in the 500 Hz condition (and vice versa), but less so in the 750 Hz and 1 kHz conditions. This is consistent across the different AI-gram signal processing. For the unprocessed condition, DSIRs assessed with 250 Hz and 500 Hz (open circles) were significantly correlated, $r(14) = 0.83$, $p = 0.0002$. However, correlation was not significant between 250 Hz and 750 Hz (filled triangles), $r(14) = 0.61$, $p = 0.021$ and between 250 Hz and 1,000 Hz (open diamonds), $r(14) = 0.49$, $p = 0.08$. For the target condition, all three correlations were significant, and $r$ values were higher than the corresponding $r$ values for the unprocessed condition. The DSIRs between 250 and 500 Hz were strongly correlated, $r(14) = 0.85$, $p = 0.0001$. Correlations were also significant between 250 and 750 Hz, $r(14) = 0.72$, $p = 0.003$ and between 250 and 1,000 Hz, $r(14) = 0.69$, $p = 0.005$. For the target-conflicting condition, all three $r$ values were lower than those in the target condition but higher than those in the unprocessed condition. Significant correlations were observed between 250 and 500 Hz, $r(14) = 0.84$, $p = 0.0001$ and between 250 and 750 Hz, $r(14) = 0.69$, $p = 0.006$. However, no significant correlation was observed between 250 and 1,000 Hz, $r(14) = 0.53$, $p = 0.05$.

## Discussion

In this study, frequency ranges needed for consonant recognition in the HF ear were measured when different low-frequency information was simultaneously presented to the LF ear under three signal processing conditions: unprocessed, target, and target-conflicting. The results showed that spectral integration and consonant recognition is possible without midfrequency consonant information. DSIRs were not significantly affected by the two signal processing conditions, except for at the LPF cutoff frequency of 250 Hz in the LF ear. DSIR narrowed significantly with increasing LPF cutoff frequency. Individual consonant analyses showed that four consonants (/ta/, /da/, /sa/, and /za/) required the least amount of spectral information. On the other hand, the four consonants (/ka/, /ga/, /

TABLE 3 Pairwise multiple comparisons among consonants for the LPF cutoff frequency of 250Hz.

| | ka | ga | ma | fa | va |
|---|---|---|---|---|---|
| pa | | | | * | |
| ba | | | *** | *** | *** |
| ta | * | *** | *** | *** | *** |
| da | | | *** | *** | *** |
| na | | | | * | |
| sa | * | * | *** | *** | *** |
| za | * | ** | *** | *** | *** |
| ʃa | | | *** | *** | *** |
| ʒa | | | | ** | |

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.
Consonant pairs with significant difference are only listed.

TABLE 4 Pairwise multiple comparisons among consonants for LPF cutoff frequency of 500Hz.

| | ka | ga | ma | fa | va |
|---|---|---|---|---|---|
| pa | | *** | | *** | *** |
| ba | | *** | | *** | *** |
| ta | * | | *** | *** | *** |
| da | | *** | | *** | *** |
| ma | | | | *** | *** |
| na | | *** | | *** | *** |
| sa | ** | *** | | *** | *** |
| za | ** | *** | | *** | *** |
| ʃa | | *** | | *** | *** |
| ʒa | | ** | | *** | *** |

***indicates $p < 0.001$; **$p < 0.01$; *$p < 0.05$.
Consonant pairs with significant difference are only listed.

TABLE 5 Pairwise multiple comparisons among consonants for LPF cutoff frequency of 750Hz.

| | ka | ga | ma | fa | va |
|---|---|---|---|---|---|
| pa | * | | | *** | *** |
| ba | * | | | ** | ** |
| ta | * | | ** | ** | ** |
| da | ** | | | *** | *** |
| ma | | | | ** | ** |
| sa | ** | | | *** | *** |
| za | ** | | | *** | *** |
| ʒa | | | | * | * |

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.
Consonant pairs with significant difference are only listed.

fa/ and /va/) required the widest amount of spectral information. The trends for these nine consonants were consistent, regardless of signal processing and the amount of low-frequency information available in the LF ear. The recognition of the remaining six consonants (/pa/, /ba/, /ma/, /na/, /ʃa/, and /ʒa/) was highly affected by the low-frequency information available in the LF ear regardless of the signal processing condition.

TABLE 6 Pairwise multiple comparisons among consonants for LPF cutoff frequency of 1,000Hz.

| | ka | ga | ma | fa | va |
|---|---|---|---|---|---|
| pa | | | | | *** |
| ba | | | | | *** |
| ta | | | | | *** |
| da | | | | * | *** |
| ka | | | | | *** |
| ga | | | | | ** |
| ma | | | | *** | * |
| sa | | | | * | *** |
| za | | | | *** | * |
| ʃa | | | | | ** |
| ʒa | | | | | *** |

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$.
Consonant pairs with significant difference are only listed.

Our finding that consonant recognition is possible without the full range of spectral information is consistent with existing literature. Lippmann (1996) measured consonant-vowel-consonant syllable recognition in quiet with NH listeners when low-frequency information below 800 Hz was combined with high-frequency information above 4,000 Hz in the same ear. The results showed no significant change in consonant recognition when removing midfrequency consonant information (800–4,000 Hz).

It is not surprising that DSIRs were highly consonant specific, regardless of which signal processing condition was used. Four consonants (/ka/, /ga/, /fa/, and /va/), required the widest amount of spectral information regardless of signal processing and the low-frequency information available in the LF ear. It is known that perception of /fa/ and /va/ requires multiple target frequency regions over wide range of spectrum (Allen, 2005). For a pair /ka/ and /ga/, considerable confusions occurred due to same manner and place of articulation (Miller and Nicely, 1955; Allen, 2005), which results in integration with little salient spectral information (Stevens and Klatt, 1974; Stevens and Blumstein, 1978; Stevens et al., 1992). In contrast, four consonants (/ta/, /da/, /sa/, and /za/) required the least amount of spectral information. Perception of these consonants was easier because major spectral cues for their perception were available at 7,000 Hz and beyond (Li et al., 2010, 2012; Li and Allen, 2011; Yoon, 2021). In this study, Sennheiser HAD-200 circumaural headphones were used, which are optimally calibrated with tones but less optimal with speech stimuli. They show a frequency drop-off of about 10 dB for high frequencies compared to low frequencies and hence need to be (free-field or diffuse-field) equalized (ISO389-8, 2004), which was not done in this study. If done appropriately, SDIRs for these four consonants may be further narrowed because their target frequency regions are extended to around 8 kHz.

Our results are similar to the results reported in Lippmann (1996). In that study, six consonants (/p/, /b/, /t/, /k/, /s/, and /z/) were well perceived when combined frequency information lower

than 800 Hz and higher than 6,300 Hz was presented simultaneously to one ear (Lippmann, 1996). In contrast, four consonants (/d/, /g/, /f/, and /v/) required combined frequency information lower than 800 Hz and higher than 3,150 Hz. Comparing to our results, recognition of /ka/ required less spectral information. Recognition of /da/ required more spectral information. These differences may stem from different testing paradigms: monotic in the Lippmann study vs. dichotic spectral integration in the current study. Ronan et al. (2004) showed that consonant recognition performance was significantly higher in monotic spectral integration than in dichotic spectral integration in listeners with normal hearing. Spehar et al. (2008) also showed that word recognition was approximately 10 percentage points higher (statistically significant) in monotic spectral integration than dichotic spectral integration for young and elderly listeners with normal hearing. Another reason for different DSIRs, for /da/ and /ka/, between two studies would be the use of different contexts of stimuli: consonant-vowel-consonant vs. consonant-vowel syllables. It is well documented that frequency-time regions that support the robust perception that a consonant is changed if different vowels with different positions of consonants (initial, medial, or final) are used as stimuli (Baum and Blumstein, 1987; Hazan and Rosen, 1991; Reidy et al., 2017).

It should be noted that DSIRs for /fa/ and /va/ were negatively affected by the two signal processing conditions. For /fa/, the widest DSIR was required in the target-conflicting condition and then in the unprocessed and target conditions. Our subject response pattern analysis showed that /ma/ was mostly selected in the target-conflicting condition. This result indicates that removing a conflicting frequency range (3–7.8 kHz) for /fa/ causes confusion with /ma/, requiring the widest DSIR. For /va/, the widest DSIR was required in the target condition and then in the unprocessed and target-conflicting conditions. The subject response patterns showed that /fa/ was mostly selected in the target condition. This result indicates that intensifying a target frequency range (0.6–1.4 kHz) for /va/ causes more confusion with /fa/, requiring the widest DSIR even though target time ranges differ.

Another major finding from the current study was that there was no significant effect of both the AI-Gram processed target and target-conflicting regions on DSIR measures except for the case of the 250 Hz cutoff frequency. However, these processed conditions made spectral cues more prominent and DSIRs were numerically narrower (again not statistically significant) for consonant recognition compared to the unprocessed condition. For example, our analyses (Figure 3) revealed that five consonants (/ta/, /da/, /ka/, /va/, and /za/) for the 250 Hz and another five consonants (/pa/, /ta/, /ka/, /na/, and /va/) for the 500 Hz had narrower DSIRs with two signal-processing conditions than those with the unprocessed condition. This trend was also observed for /pa/, /da/, /fa/, /va/, /za/, and /ʒa/ with the 750 Hz and /pa/, /fa/, and /va/ for the 1,000 Hz.

Our correlation analyses (Figure 5) showed that the DSIRs between 250 Hz and 500 Hz were significantly correlated in all three signal-processing conditions. The correlation was strengthened with the two AI-Gram processed conditions except for the target-conflicting condition between 250 and 1,000 Hz. Similar studies for nonsense phoneme perception were not available, but Hall and colleagues compared sentence perception in NH listeners and reported indirect evidence of this relationship (Hall et al., 2008). They first determined the necessary bandwidth for approximately 15–25% correct scores on sentence perception per listener in both quiet and noise listening environments (called criterion speech bandwidths) by adaptively varying the bandwidth of filtered sentences centered either on 500 Hz or 2,500 Hz. This criterion speech bandwidth measure was conducted monaurally. They found no obvious relation between the criterion bandwidths at each center frequency in both quiet and noise: listeners requiring a relatively wide criterion bandwidth at 500 Hz did not necessarily require a wide bandwidth at 2,500 Hz. This result is not surprising as speech information is widely spread out over a wide range of spectral bands, and the importance of each of these spectral bands for speech perception varies. As Hall et al.'s study (Hall et al., 2008) testing settings were different from ours (e.g., monotic and dichotic), any direct comparisons cannot be made. Our results confirm that the normal auditory system integrates lower spectral information, processed by one ear, with different spectral information processed by the opposite ear.

## Clinical implication

The dichotic test setting of the present study with different low-frequency information in the LF ear could be translated into the four different degrees of high-frequency hearing loss in one ear. The approach may be applied to bimodal users who have residual hearing in low-frequency regions (typically below 1,000 Hz) in the hearing aid ear and can have access to wider frequency information through a cochlear implant in the opposite ear (Gifford et al., 2007; Yoon et al., 2012). So, dichotic spectral integration may play an important role. It is expected that some bimodal listeners with limited access to low-frequency information *via* the hearing aid ear require a broader range of spectral information in the cochlear implant ear. The opposite can occur as well. As shown in Figures 3, 4, DSIRs for six consonants (/pa/, /ba/, /ma/, /na/, /ʃa/, and /ʒa/) were highly sensitive to low-frequency information available in the opposite ear. However, perception of four consonants (/ta/, /da/, /sa/, and /za/) required the narrowest DSIRs, while another four consonants (/ka/, /ga/, /fa/, and /va/) were needed the widest DSIRs, regardless of the signal processing. These results suggest that low-frequency sensitive consonants are most affected by interactions of acoustic and electric stimulations. In bimodal hearing, determining the minimum spectral information needed in a cochlear implant ear for consonant-by-consonant perception on an individual, subject-by-subject basis is critical because interactions across ears are highly listener specific (Cullington and Zeng, 2010; Gifford and Dorman, 2019; Shpak

et al., 2020). Fowler et al. (2016) measured speech perception with bimodal listeners as a function of high-pass cutoff frequency for the cochlear implant ear. Speech perception with the cochlear implant ear alone deteriorated as the high-pass cutoff frequency was raised. In contrast, bimodal performance in quiet was improved as the high-pass cutoff frequency was raised for listeners with better residual hearing in a hearing aid ear (< 60 dB HL at 250 and 500 Hz). This result suggests that determining minimum spectral information needed in a cochlear implant ear can reduce spectral interference in bimodal hearing (Fowler et al., 2016). Consonant-specific and listener-specific datasets are also necessary to train a neural network-based deep machine learning algorithm which is currently in progress in our laboratory. Training the deep machine learning algorithm will be effective with our consonant-by-consonant datasets for maximizing algorithm accuracy and minimizing errors (Vaerenberg et al., 2011; Wang, 2017; Wathour et al., 2020). Hence, the present study findings will aid in designing custom bimodal frequency maps for greater consonant intelligibility based on residual hearing available in the hearing aid ear. One caution of direct application into bimodal hearing is that simulating a hearing aid ear requires careful incorporating gains with specific input levels for each band on a patient-by-patient basis using clinical prescription procedures (Zhang et al., 2010; Sheffield et al., 2016), which were not done in the current study.

Currently, our laboratory has conducted a series of bimodal simulation studies to derive the frequency importance function of cochlear implant ear and combined cochlear implant and hearing aid ears. In addition, a spectral integration and interference study is ongoing for vowel and consonant recognition with manipulation of first and second formant frequencies. The present datasets will serve as a control for some ongoing studies. Our long-term goal of the AI-Gram based speech recognition studies is to develop efficient bimodal fitting schemes based on deep machine learning. It is expected that the target and conflicting frequency and time regions, reported in Yoon (2021), in conjunction with the expected results of the bimodal study, the minimum spectral information required for consonant recognition in cochlear implant ears would be effective in training algorithms.

## Limitations

The present study has several limitations. First, using a single female talker creates a clear limitation of talker variability in real listening situations. The target and conflicting regions might differ depending on different talkers (Mullennix et al., 1989; Goldinger et al., 1991; Magnuson and Nusbaum, 2007). Thus, DSIR may also vary widely across talkers, particularly for listeners with hearing loss and hearing devices. However, based on comparable data in the target and conflicting regions between the current study and Li et al. (2010, 2012), different talkers may affect these regions less substantially. Second, the baseline performance for each ear alone was not measured. Our data was likely a result of the dichotic spectral integration. However, it is possible that consonant recognition could be achieved with higher frequency spectral information only, particularly for some consonants such as /sa/ and /ʃa/. Third, the single phonetic environment (consonant+/a/ vowel) was used. Critical spectral-temporal regions that facilitate or limit our ability to integrate auditory information might change if different consonant-vowel combinations are used at different positions (initial, medial, or final) as stimuli (Hayden, 1950; Harris, 1958; Soli, 1981; Viswanathan et al., 2010). Finally, one technical concern is the possibility that optimal spectral integration may occur with different suppression levels to completely remove conflicting frequency and time regions used in the current study. In our pilot study with five NH listeners, a wide range of suppression from −2 to −20 dB in every 2 dB decrement were tested. No additional consonant enhancement was seen with higher than −6 dB for fricative consonants and less than 2% consonant enhancement for stop consonants. With the complete removal of the conflicting regions, speech perception was significantly enhanced for all consonants except /sa/ and /ʃa/, whose perception suffered by 15% compared to the unprocessed condition. Hence, though not studied in the present study, the removal of conflicting frequency and time regions alone as a condition may be studied vastly in future works.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Baylor University (IRB ID: #125371). The patients/participants provided their written informed consent to participate in this study.

## Author contributions

Y-SY conceived and designed the study, analyzed the data, and wrote the draft of the manuscript. DM collected the data. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Institutes of Health under R15DC019240. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

## Acknowledgments

The authors thank the participants for their patience and continuous support. The authors also want to thank Priyanka Jaisinghani, Amanda Haynes, Naomi White, and Reagan Whitaker for their editorial help.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The handling editor declared a past collaboration with one of the authors, Y-SY.

## Publisher's note



## References

Allen, J. B. (2005). Consonant recognition and the articulation index. *J. Acoust. Soc. Am.* 117, 2212–2223. doi: 10.1121/1.1856231

Baum, S. R., and Blumstein, S. E. (1987). Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *J. Acoust. Soc. Am.* 82, 1073–1077. doi: 10.1121/1.395382

Bianchi, F., Fereczkowski, M., Zaar, J., Santurette, S., and Dau, T. (2016). Complex-tone pitch discrimination in listeners with sensorineural hearing loss. *Trends Hear* 20:233121651665579. doi: 10.1177/2331216516655793

Blumstein, S. E., and Stevens, K. N. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am.* 66, 1001–1017. doi: 10.1121/1.383319

Blumstein, S. E., and Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *J. Acoust. Soc. Am.* 67, 648–662. doi: 10.1121/1.383890

Cox, R. M. (2005). Evidence-based practice in provision of amplification. *J. Am. Acad. Audiol.* 16, 419–438. doi: 10.3766/jaaa.16.7.3

Cullington, H. E., and Zeng, F. G. (2010). Bimodal hearing benefit for speech recognition with competing voice in cochlear implant subject with normal hearing in contralateral ear. *Ear Hear.* 31, 70–73. doi: 10.1097/AUD.0b013e3181bc7722

Fowler, J. R., Eggleston, J. L., Reavis, K. M., McMillan, G. P., and Reiss, L. A. (2016). Effects of removing low-frequency electric information on speech perception with bimodal hearing. *J. Speech Lang. Hear. Res.* 59, 99–109. doi: 10.1044/2015_jslhr-h-15-0247

Fox, R. A., Jacewicz, E., and Chang, C. Y. (2011). Auditory spectral integration in the perception of static vowels. *J. Speech Lang. Hear. Res.* 54, 1667–1681. doi: 10.1044/1092-4388(2011/09-0279)

Gifford, R. H., and Dorman, M. F. (2019). Bimodal hearing or bilateral Cochlear implants? *Ask the Patient. Ear Hear* 40, 501–516. doi: 10.1097/aud.0000000000000657

Gifford, R. H., Dorman, M. F., McKarns, S. A., and Spahr, A. J. (2007). Combined electric and contralateral acoustic hearing: word and sentence recognition with bimodal hearing. *J. Speech Lang. Hear. Res.* 50, 835–843. doi: 10.1044/1092-4388(2007/058)

Goldinger, S. D., Pisoni, D. B., and Logan, J. S. (1991). On the nature of talker variability effects on recall of spoken word lists. *J. Exp. Psychol. Learn. Mem. Cogn.* 17, 152–162. doi: 10.1037//0278-7393.17.1.152

Grose, J. H., Porter, H. L., and Buss, E. (2016). Aging and Spectro-temporal integration of speech. *Trends Hear* 20:233121651667038. doi: 10.1177/2331216516670388

Hall, J. W., Buss, E., and Grose, J. H. (2008). Spectral integration of speech bands in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 124, 1105–1115. doi: 10.1121/1.2940582

Happel, M. F., Jeschke, M., and Ohl, F. W. (2010). Spectral integration in primary auditory cortex attributable to temporally precise convergence of thalamocortical and intracortical input. *J. Neurosci.* 30, 11114–11127. doi: 10.1523/jneurosci.0689-10.2010

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Lang. Speech* 1, 1–7. doi: 10.1177/002383095800100101

Hayden, R. E. (1950). The relative frequency of phonemes in general-American English. *Word* 6, 217–223. doi: 10.1080/00437956.1950.11659381

Hazan, V., and Rosen, S. (1991). Individual variability in the perception of cues to place contrasts in initial stops. *Percept. Psychophys.* 49, 187–200. doi: 10.3758/bf03205038

ISO389-8 (2004). ISO 389-8:2004 acoustics–reference zero for the calibration of audiometric equipment–part 8: reference equivalent threshold sound pressure levels for pure tones and circumaural earphones. Available at: https://www.iso.org/standard/32440.html#:~:text=ISO%20389%2D8%3A2004%20specifies,earphones%20(SENNHEISER%20HDA%20200) (Accessed September 22, 2022).

Jürgens, T., Kollmeier, B., Brand, T., and Ewert, S. D. (2011). Assessment of auditory nonlinearity for listeners with different hearing losses using temporal masking and categorical loudness scaling. *Hear. Res.* 280, 177–191. doi: 10.1016/j.heares.2011.05.016

Kong, Y. Y., and Braida, L. D. (2011). Cross-frequency integration for consonant and vowel identification in bimodal hearing. *J. Speech Lang. Hear. Res.* 54, 959–980. doi: 10.1044/1092-4388(2010/10-0197)

Li, F., and Allen, J. B. (2011). Manipulation of consonants in natural speech. *IEEE Trans. Audio Speech Lang. Process.* 19, 496–504. doi: 10.1109/TASL.2010.2050731

Li, F., Menon, A., and Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *J. Acoust. Soc. Am.* 127, 2599–2610. doi: 10.1121/1.3295689

Li, F., Trevino, A., Menon, A., and Allen, J. B. (2012). A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise. *J. Acoust. Soc. Am.* 132, 2663–2675. doi: 10.1121/1.4747008

Lippmann, R. (1996). Accurate consonant perception without mid-frequency speech energy. *IEEE Trans. Speech Audio Process.* 4, 66–69. doi: 10.1109/TSA.1996.481454

Magnuson, J. S., and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 391–409. doi: 10.1037/0096-1523.33.2.391

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338–352. doi: 10.1121/1.1907526

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *J. Acoust. Soc. Am.* 85, 365–378. doi: 10.1121/1.397688

Obuchi, C., Shiroma, M., Ogane, S., and Kaga, K. (2015). Binaural integration abilities in bilateral cochlear implant user. *J. Otol.* 10, 150–153. doi: 10.1016/j.joto.2016.02.001

Patel, R., and McKinnon, B. J. (2018). Hearing loss in the elderly. *Clin. Geriatr. Med.* 34, 163–174. doi: 10.1016/j.cger.2018.01.001

Räsänen, O., and Laine, U. K. (2013). Time-frequency integration characteristics of hearing are optimized for perception of speech-like acoustic patterns. *J. Acoust. Soc. Am.* 134, 407–419. doi: 10.1121/1.4807499

Reidy, P. F., Kristensen, K., Winn, M. B., Litovsky, R. Y., and Edwards, J. R. (2017). The acoustics of word-initial fricatives and their effect on word-level intelligibility in children with bilateral Cochlear implants. *Ear Hear.* 38, 42–56. doi: 10.1097/aud.0000000000000349

Reiss, L. A., Eggleston, J. L., Walker, E. P., and Oh, Y. (2016). Two ears are not always better than one: mandatory vowel fusion across spectrally mismatched ears

in hearing-impaired listeners. *J. Assoc. Res. Otolaryngol.* 17, 341–356. doi: 10.1007/s10162-016-0570-z

Reiss, L. A., Ito, R. A., Eggleston, J. L., and Wozny, D. R. (2014). Abnormal binaural spectral integration in cochlear implant users. *J. Assoc. Res. Otolaryngol.* 15, 235–248. doi: 10.1007/s10162-013-0434-8

Ronan, D., Dix, A. K., Shah, P., and Braida, L. D. (2004). Integration across frequency bands for consonant identification. *J. Acoust. Soc. Am.* 116, 1749–1762. doi: 10.1121/1.1777858

Sheffield, S. W., Simha, M., Jahn, K. N., and Gifford, R. H. (2016). The effects of acoustic bandwidth on simulated bimodal benefit in children and adults with Normal hearing. *Ear Hear.* 37, 282–288. doi: 10.1097/aud.0000000000000281

Shpak, T., Most, T., and Luntz, M. (2020). Phoneme recognition in bimodal hearing. *Acta Otolaryngol.* 140, 854–860. doi: 10.1080/00016489.2020.1780311

Smith-Olinde, L., Besing, J., and Koehnke, J. (2004). Interference and enhancement effects on interaural time discrimination and level discrimination in listeners with normal hearing and those with hearing loss. *Am. J. Audiol.* 13, 80–95. doi: 10.1044/1059-0889(2004/011)

Soli, S. (1981). Second formants in fricatives: acoustic consequences of fricative-vowel coarticulation. *J. Acoust. Soc. Am.* 70, 976–984. doi: 10.1121/1.387032

Spehar, B. P., Tye-Murray, N., and Sommers, M. S. (2008). Intra- versus intermodal integration in young and older adults. *J. Acoust. Soc. Am.* 123, 2858–2866. doi: 10.1121/1.2890748

Stevens, K. N., and Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* 64, 1358–1368. doi: 10.1121/1.382102

Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *J. Acoust. Soc. Am.* 91, 2979–3000. doi: 10.1121/1.402933

Stevens, K. N., and Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *J. Acoust. Soc. Am.* 55, 653–659. doi: 10.1121/1.1914578

SYSTAT (2021). SigmaPlot for graphing and data visualization. *Version 14*.

The MathWorks (2017). MATLAB version 9.3.0.713579 (R2017b): MathWorks.

Tononi, G. (2010). Information integration: its relevance to brain function and consciousness. *Arch. Ital. Biol.* 148, 299–322. doi: 10.4449/aib.v148i3.1224

Vaerenberg, B., Govaerts, P. J., de Ceulaer, G., Daemers, K., and Schauwers, K. (2011). Experiences of the use of FOX, an intelligent agent, for programming cochlear implant sound processors in new users. *Int. J. Audiol.* 50, 50–58. doi: 10.3109/14992027.2010.531294

Varnet, L., Langlet, C., Lorenzi, C., Lazard, D. S., and Micheyl, C. (2019). High-frequency sensorineural hearing loss alters Cue-weighting strategies for discriminating stop consonants in noise. *Trends Hear* 23:2331216519886707. doi: 10.1177/2331216519886707

Viswanathan, N., Magnuson, J. S., and Fowler, C. A. (2010). Compensation for coarticulation: disentangling auditory and gestural theories of perception of coarticulatory effects in speech. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 1005–1015. doi: 10.1037/a0018391

Wang, D. (2017). Deep learning reinvents the hearing aid: finally, wearers of hearing aids can pick out a voice in a crowded room. *IEEE Spectr.* 54, 32–37. doi: 10.1109/mspec.2017.7864754

Wathour, J., Govaerts, P. J., and Deggouj, N. (2020). From manual to artificial intelligence fitting: two cochlear implant case studies. *Cochlear Implants Int.* 21, 299–305. doi: 10.1080/14670100.2019.1667574

Yang, H. I., and Zeng, F. G. (2013). Reduced acoustic and electric integration in concurrent-vowel recognition. *Sci. Rep.* 3:1419. doi: 10.1038/srep01419

Yoho, S. E., and Bosen, A. K. (2019). Individualized frequency importance functions for listeners with sensorineural hearing loss. *J. Acoust. Soc. Am.* 145, 822–830. doi: 10.1121/1.5090495

Yoon, Y. S. (2021). Effect of the target and conflicting frequency and time ranges on consonant enhancement in Normal-hearing listeners. *Front. Psychol.* 12:733100. doi: 10.3389/fpsyg.2021.733100

Yoon, Y. S., Shin, Y. R., and Fu, Q. J. (2012). Clinical selection criteria for a second cochlear implant for bimodal listeners. *Otol. Neurotol.* 33, 1161–1168. doi: 10.1097/MAO.0b013e318259b8c0

Zhang, T., Dorman, M. F., and Spahr, A. J. (2010). Information from the voice fundamental frequency (F0) region accounts for the majority of the benefit when acoustic stimulation is added to electric stimulation. *Ear Hear.* 31, 63–69. doi: 10.1097/aud.0b013e3181b7190c

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 08 November 2022
DOI 10.3389/fnins.2022.1022308



OPEN ACCESS

EDITED BY
Huiming Zhang,
University of Windsor, Canada

REVIEWED BY
John Culling,
Cardiff University, United Kingdom
Herman Myburgh,
University of Pretoria, South Africa

*CORRESPONDENCE
Jörg Encke
joerg.encke@uni-oldenburg.de

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 18 August 2022
ACCEPTED 19 October 2022
PUBLISHED 08 November 2022

CITATION
Encke J and Dietz M (2022) Statistics of the instantaneous interaural parameters for dichotic tones in diotic noise ($N_0S_\psi$). *Front. Neurosci.* 16:1022308. doi: 10.3389/fnins.2022.1022308



# Statistics of the instantaneous interaural parameters for dichotic tones in diotic noise ($N_0S_\psi$)

Jörg Encke[1,2]* and Mathias Dietz[1,2]

[1]Physiology and Modeling of Auditory Perception, Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany, [2]Cluster of Excellence "Hearing4all", University of Oldenburg, Oldenburg, Germany

Stimuli consisting of an interaurally phase-shifted tone in diotic noise—often referred to as $N_0S_\psi$—are commonly used to study binaural hearing. As a consequence of mixing diotic noise with a dichotic tone, this type of stimulus contains random fluctuations in both interaural phase- and level-difference. We report the joint probability density functions of the two interaural differences as a function of amplitude and interaural phase of the tone. Furthermore, a second joint probability density function for interaural phase differences and the instantaneous cross-power is derived. The closed-form expression can be used in future studies of binaural unmasking first to obtain the interaural statistics and then study more directly the relation between those statistics and binaural tone detection.

KEYWORDS

**sound localization, probability density function, interaural level difference, interaural phase difference, tone in noise detection, binaural unmasking**

## 1. Introduction

Tone in noise detection thresholds improve when the interaural configuration of tone and noise differ compared to the diotic case. A rich literature reports on the influence of virtually every parameter of acoustic stimuli on this binaural unmasking (see, e.g., Culling and Lavandier, 2021, for a review). Amongst these parameters, the phase difference $\psi$ introduced between the target tones of the two ear signals is fundamental and was explored already in the first study of dichotic tone in noise detection by Hirsh (1948). Such a signal is commonly referred to as $N_0S_\psi$ where the subscripts indicate the interaural phase difference (IPD) of the noise (N) or signal (S). The difference between the detection threshold for the purely diotic $N_0S_0$ and the $N_0S_\psi$ signal is referred to as the binaural masking level difference (BMLD) and is largest for the case where $\psi = \pi$ (Hirsh, 1948).

Adding a dichotic $S_\psi$ tone to diotic $N_0$ noise reduces the correlation between the left and right signals but also introduces random fluctuations of the interaural phase and level differences (IPD, ILD) (visualized in Figure 1A). The interaural correlation decreases with the tone level, so binaural unmasking and incoherence detection are often treated synonymously (Durlach et al., 1986). However, especially for narrowband noise,

the value of interaural correlation itself was found to be an insufficient predictor for decorrelation detection performance. Instead, detection performance correlated with the amount of IPD and ILD fluctuations as measured by the standard deviation (Goupell and Hartmann, 2006). Similarly, other studies reported the performance in detecting the tone within an $N_0S_\psi$ stimulus to vary considerably depending on the individual noise token. This token to token variability was best accounted for by models that did consider the amount of instantaneous fluctuations in IPD and ILD (Davidson et al., 2009).

Therefore, accounting for binaural tone-in-noise sensitivity can be subdivided into two components: First, the signal-based analysis of how stimulus design parameters such as $\psi$ or the SNR influence the interaural cue statistics. In the second step, binaural sensitivity can then be studied more directly by relating it to the interaural cues. Only relatively few studies, however, have previously treated these statistics. The probability density function (PDF) underlying the statistical distribution of IPDs in (partly) decorrelated noise has been derived in the frame of optical interferometry (Just and Bamler, 1994). Henning (1973) derived the PDF for IPDs in the special case of $N_0S_\pi$ and using a very similar approach for the same stimulus condition, Zurek (1991) additionally derived marginal PDFs for ILDs. Other studies also seemed to have worked on stimuli where the tone IPD did not equal $\pi$, but this work seemed to have remained unpublished (Levitt and Lundry, 1966). The present study closes this gap by deriving a closed form expression for the joint PDF of IPDs and ILDs in the general case of a $N_0S_\psi$ stimulus. From this distribution, the marginal PDFs can also be calculated using numerical integration. These PDFs are especially useful when considering narrowband noises that remain relatively unaffected by the bandpass properties of the auditory periphery. Statistics at the stimulus level should thus well describe statistics of the binaural parameters at the level of binaural integration.

Suppose fluctuations of the IPD are indeed a cue used to detect the tone in an $N_0S_\psi$ stimulus. In that case, the stimulus energy at which these fluctuations occurred might also affect performance. A larger IPD occurring during low-energy stimulus sections can be expected to have less impact than the same IPD occurring at high stimulus energy. Information about the stimulus energy in both ears is captured by the product of the left and right ear stimulus envelope, also called the instantaneous cross-power $P'(t)$. Furthermore, the cross-power plays an essential role in defining the interaural coherence of a stimulus (Encke and Dietz, 2022). Consequently, this study derives the joint PDF for $p'(t)$ and IPD.

## 2. Deriving the probability density functions

The following section will derive the two joint PDF. A computational-notebook that can be used to reproduce these derivations in the computer algebra system sympy (Meurer et al., 2017) can be found as Supplementary material.

If $N(t)$ is a Gaussian noise process with a mean value of zero, the process can be represented using its in-phase and quadrature components $X(t)$ and $Y(t)$:

$$N(t) = X(t)\cos(\omega_0 t) - Y(t)sin(\omega_0 t), \tag{1}$$

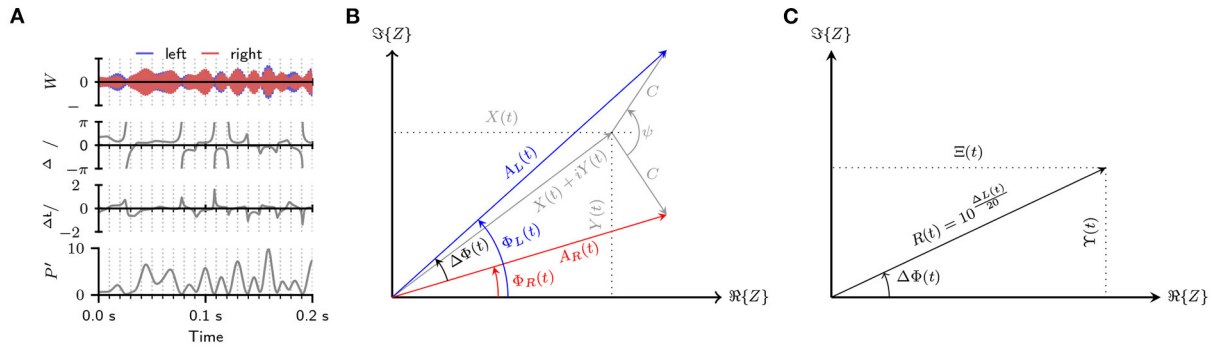



FIGURE 1
**(A)** Visualization of the random fluctuations in IPD $\Delta\Phi(t)$ and ILD $\Delta L(t)$ and $P'(t)$ due to mixing an antiphasic 500 Hz tone with a 500 Hz wide band of diotic noise (SNR = −10 dB). **(B)** Signal model used to derive the PDFs for an $N_0S_\psi$ stimulus. The graphic shows the Complex-plane representation of the basebands of the left and right ear signal: $Z_L(t) = A_L(t)e^{i\Phi_L(t)}$ (blue), and $Z_R(t) = A_R(t)e^{i\Phi_R(t)}$ (red). The left-ear-baseband is constructed by adding a "tone"-vector with length $C$ and angle $+\psi/2$ to the noise baseband $X(t) + iY(t)$. The right-ear-signal is constructed by adding a "tone"-vector with an angle of $-\psi/2$ to the same baseband. The instantaneous IPD $\Delta\Phi(t)$ of the $N_0S_\psi$ signal equal the difference between $\Phi_R$ and $\Phi_L$. **(C)** Complex-plane representation of the interaural-baseband $Z_1(t) = \Xi(t) + i\Upsilon(t)$ which is gained by dividing the left-ears-baseband by the right-ears-baseband. The absolute value of the baseband equals the interaural amplitude ratio $R$ while the phase equals the interaural phase difference $\Delta\Phi$.

where $X(t)$ and $Y(t)$ are orthogonal noise processes with the same variance and mean as $N(t)$. The reference frequency $\omega_0$ is not of relevance for the derivation and can thus be chosen freely. For computational convenience, $\omega_0$ is set to equal the frequency of the tone $S(t)$ which is added with the amplitude $C$ and phase $\psi$:

$$S(t) = C\sin(\omega_0 t + \psi) \tag{2}$$

The resulting signal $W(t) = N(t) + S(t)$ then equals:

$$\begin{aligned} W(t) = &\left[X(t) + C\cos(\psi)\right]\cos(\omega_0 t) \\ &- \left[Y(t) + C\sin(\psi)\right]\sin(\omega_0 t). \end{aligned} \tag{3}$$

When dealing with instantaneous phase and amplitude values, it is beneficial to instead work with the analytic representation $W_a(t)$ of the signal:

$$W_a(t) = \left\{\left[X(t) + C\cos(\psi)\right] + i\left[Y(t) + C\sin(\psi)\right]\right\} e^{i\omega_0 t}, \tag{4}$$

where $i = \sqrt{-1}$ is the imaginary unit. The first term of this expression (enclosed in curly brackets) can be interpreted as an amplitude and phase modulator of the harmonic oscillation $e^{i\omega_0 t}$. This combined modulator will be referred to as the signals complex baseband $Z(t)$

$$Z(t) = \left[X(t) + C\cos(\psi)\right] + i\left[Y(t) + C\sin(\psi)\right] = A(t)e^{i\Phi(t)}, \tag{5}$$

where $A(t)$, $\Phi(t)$ are the instantaneous amplitude and phase of the baseband. In the case of the $N_0S_\psi$ stimulus, a tone with phase $\psi/2$ is added to the noise in the left-ear signal while the phase of the tone in the right-ear signal is $-\psi/2$ resulting in the two basebands:

$$Z_L(t) = \left[X(t) + C\cos\psi/2\right] + i\left[Y(t) + C\sin\psi/2\right] = A_L(t)e^{i\Phi_L(t)} \tag{6}$$

$$\begin{aligned} Z_R(t) &= \left[X(t) + C\cos -\psi/2\right] + i\left[Y(t) + C\sin -\psi/2\right] \\ &= A_R(t)e^{i\Phi_R(t)}. \end{aligned} \tag{7}$$

A vector model of the basebands $Z_R$ and $Z_L$ is shown in Figure 1B where the individual components are visualized as vectors in the complex plane.

Based on these two basebands, PDFs for the interaural parameters will be derived using two separate approaches. In the first approach, the baseband of the left-ear signal $Z_L(t)$ is divided by the baseband of the right-ear signal $Z_R(t)$ resulting in the interaural baseband $Z_1(t)$:

$$\begin{aligned} Z_1(t) &= \frac{Z_R(t)}{Z_L(t)} = \frac{A_R(t)}{A_L(t)}e^{i[\Phi_R(t) - \Phi_L(t)]} \\ &= R(t)e^{i\Delta\Phi(t)}, \end{aligned} \tag{8}$$

where $\Delta\Phi(t)$ and $R(t)$ are the instantaneous IPDs and the interaural amplitudes ratios (IARs), respectively. Instantaneous ILDs can then be calculated as: $\Delta L(t) = 20\log_{10} R(t)$. In the second approach, the PDF for IPDs and the product of the left and right-ear envelope (cross power) $p'$ are derived by multiplying $Z_L(t)$ with the complex conjugate of $Z_R(t)$ resulting in

$$\begin{aligned} Z_2(t) &= Z_R(t)Z_L^*(t) = A_R(t)A_L(t)e^{i[\Phi_R(t) - \Phi_L(t)]} \\ &= P'(t)e^{i\Delta\Phi(t)}. \end{aligned} \tag{9}$$

The process of deriving the PDFs from Equation (8) and Equation (9) follows the exact same rationale so that the process will only be detailed for Equation (8). Results for the second approach will then be stated without further detail.

For the interaural baseband, $Z_L$ and $Z_R$ as resulting from Equations (6) and (7) are inserted into Equation (8) resulting in:

$$\begin{aligned} Z_1(t) &= \frac{[X(t) + C\cos(\psi/2)] + i[C\sin(\psi/2) + Y(t)]}{[X(t) + C\cos(-\psi/2)] + i[C\sin(-\psi/2) + Y(t)]} \\ &= \Xi(t) + i\Upsilon(t) \end{aligned} \tag{10}$$

where $\Xi(t)$ and $\Upsilon(t)$ are the in-phase and quadrature components of the baseband $Z_1(t)$. They can be derived from Equation (10) as:

$$\Xi(t) = \frac{Y^2(t) + [C\cos(\psi/2) + X(t)]^2 - C^2\sin^2(\psi/2)}{[C\sin(\psi/2) - Y(t)]^2 + [C\cos(\psi/2) + X(t)]^2} \tag{11}$$

$$\Upsilon(t) = \frac{2C[C\cos(\psi/2) + X(t)]\sin(\psi/2)}{[C\sin(\psi/2) - Y(t)]^2 + [C\cos(\psi/2) + X(t)]^2}. \tag{12}$$

Figure 1B visualizes the resulting baseband in the complex plane. From this visualization, it can be seen that the instantaneous IPDs and IARs can be calculated as the argument: $\Delta\Phi(t) = \arg\{Z_1(t)\} = \text{arctan2}(\Upsilon(t), \Xi(t))$ and modulus $R(t) = |Z_1(t)| = \sqrt{\Upsilon(t)^2 + \Xi(t)^2}$ of the baseband. Here, arctan2 is the two-argument arctangent that returns the angle in the Euclidean plane.

Both Random Processes $R(t)$ and $\Delta\Phi(t)$ are functions of $X(t)$ and $Y(t)$ which are uncorrelated Gaussian noise processes with the variance $\sigma^2$. The joint PDF $f_{X,Y}(x,y)$ of $X(t)$ and $Y(t)$ is thus that of a bivariate Gaussian distribution:

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma^2}e^{-\frac{1}{2\sigma^2}(x^2+y^2)}, \tag{13}$$

where

$$\int_{-\infty}^{\infty}\int f_{X,Y}(x,y)\,dx dy = 1. \tag{14}$$

Here and in all future equations, lower-case variables will be used to refer to the individual instances generated by a given noise process. $x$ and $y$ are thus two instances generated by the noise processes $X(t)$ and $Y(t)$ and $\xi$, $\upsilon$ are generated by $\Xi(t)$ and $\Upsilon(t)$.

Probability density functions for $\Xi(t)$ and $\Upsilon(t)$ can be gained by applying a coordinate transformation to Equation (13). For this, Equations (11) and (12) are rearranged to calculate $x$ and $y$ given the values of $\xi$ and $\upsilon$:

$$x(\xi,\upsilon) = C\left[\frac{2\upsilon\sin(\psi/2)}{\upsilon^2+(\xi-1)^2} - \cos(\psi/2)\right],$$
$$y(\xi,\upsilon) = \frac{C\left(\upsilon^2+\xi^2-1\right)\sin(\psi/2)}{\upsilon^2+\xi^2-2\xi+1}. \quad (15)$$

These expressions allow us to derive the Jacobian determinant $\left|J(x,y)\right|$. The Jacobian is then used to apply a coordinate transformation from $\mathrm{d}x$ and $\mathrm{d}y$ to $\mathrm{d}\xi$ and $\mathrm{d}\upsilon$:

$$\mathrm{d}x\,\mathrm{d}y = \left|J(x,y)\right|\mathrm{d}\xi\,\mathrm{d}\upsilon = \frac{4C^2\sin^2(\psi/2)}{\left[\upsilon^2+(\xi-1)^2\right]^2}\mathrm{d}\xi\,\mathrm{d}\upsilon. \quad (16)$$

Applying the transformations in Equations (15) and (16) to change the variables of Equation (13) results in:

$$f_{\Xi,\Upsilon}(\xi,\upsilon) = \frac{2C^2\sin^2(\psi/2)}{\pi\sigma^2\left[\upsilon^2+(\xi-1)^2\right]^2}\mathrm{e}^{-\frac{C^2\left[\upsilon^2-2\upsilon\sin(\psi)+\xi^2-2\xi\cos(\psi)+1\right]}{2\sigma^2\left[\upsilon^2+(\xi-1)^2\right]}}. \quad (17)$$

Which is the joint PDF for the two random processes $\Xi(t)$ and $\Upsilon(t)$. To gain the joint PDF $f_{R,\Delta\Phi}(r,\Delta\varphi)$, Equation (17) is transformed from rectangular to polar coordinates (see Figure 1C). This is achieved by using the transforms: $\xi = r\cos\Delta\varphi$, $\upsilon = r\sin\Delta\varphi$, $\mathrm{d}\xi\,\mathrm{d}\upsilon = r\,\mathrm{d}r\,\mathrm{d}\Delta\varphi$ resulting in:

$$f_{R,\Delta\Phi}(r,\Delta\varphi) = \frac{C^2 2r\sin^2(\psi/2)}{\sigma^2\pi h(0)^2}\mathrm{e}^{-\frac{C^2h(\psi)}{\sigma^2 2h(0)}}, \quad (18)$$

where $h(\psi) = r^2 - 2r\cos(\Delta\varphi-\psi)+1$ and $r\in[0,\infty]$, $\Delta\varphi\in[-\pi,\pi]$.

This equation can be interpreted as the distribution of all possible values of the interaural baseband $z_1 = re^{i\Delta\varphi}$ and thus the distribution of all possible combinations of IPDs $\Delta\varphi$ and IARs $r$. It is also apparent from Equation (18) that equal ratios of $C^2/\sigma^2$ result in the same PDF so that PDFs will be referenced using the signal to noise ratio $\mathrm{SNR} = C^2/2\sigma^2$ instead of $\sigma^2$ and $C$. Some examples of these functions are shown in Figures 2A–D. Deriving the joint PDF of $\Delta\varphi$ and ILD $\Delta l$ instead of IAR $r$ is easily done by using transforms $r = 10^{\Delta l/20}$ and $dr = r/20\ln(10)d\Delta l$.

$$f_{\Delta L,\Delta\Phi}(\Delta l,\Delta\varphi) = \frac{C^2 10^{\Delta l/20}\ln(10)\sin^2(\psi/2)}{\sigma^2\pi h(0)^2}\mathrm{e}^{-\frac{C^2h(\psi)}{\sigma^2 2h(0)}}. \quad (19)$$

To derive the joint PDF of $\Delta\Phi(t)$ and $P'(t)$, the process detailed above is repeated based on the interaural baseband $Z_2(t)$ as defined in Equation (9) resulting in the PDF:

$$f_{P',\Delta\Phi}(p',\Delta\varphi) = \frac{e^{-\frac{C^2}{2\sigma^2}-\frac{p'[\cos(\Delta\varphi)-\cos(\Delta\varphi-\psi)]}{2\sigma^2[\cos(\psi)-1]}}p'}{2\pi\sigma^2\sqrt{g}} \quad (20)$$

where $g$ is given by:

$$g = 2C^2\sin^2(\psi/2)\left[2p'\cos(\Delta\varphi) - C^2(\cos(\psi)-1)\right] - p'^2\sin^2(\Delta\varphi). \quad (21)$$

and the range of values is defined by:

$$p'\in[0,\hat{p}'(\Delta\varphi)],\ \Delta\varphi\in[-\hat{\Delta\varphi}(p'),+\hat{\Delta\varphi}(p')], \quad (22)$$

where

$$\hat{p}'(\Delta\varphi) = C^2\frac{\cos(\psi)-1}{\cos(\Delta\varphi)-1}. \quad (23)$$

The function $\hat{\Delta\varphi}(p')$ can be gained by solving Equation (23) for $\Delta\varphi$.

Similar to Equation (19) which defined the distribution of all possible values of $\Delta\varphi$ and $r$, this function can be interpreted as the distribution of all possible combinations of $\Delta\varphi$ and $p'$. However, the range of these combinations is limited by Equation (23) so that large areas of the exemplary PDFs shown Figures 2E–H are undefined. This limitation will be treated further in the discussion.

The marginal PDFs of the IAR $R$, the IPD $\Delta\Phi$ and the cross-power $P'$ can be calculated from the two joint PDFs defined in Equations (19) and (20) by integrating over the other variable.

$$f_{\Delta\Phi}(\Delta\varphi) = \int_0^\infty f_{R,\Delta\Phi}(r,\Delta\varphi)\mathrm{d}r = \int_0^{\hat{p}'(\Delta\varphi)} f_{P',\Delta\Phi}(p',\Delta\varphi)\mathrm{d}p' \quad (24)$$
$$f_R(r) = \int_{-\pi}^{\pi} f_{R,\Delta\Phi}(r,\Delta\varphi)\mathrm{d}\Delta\varphi \quad (25)$$
$$f_{P'}(p') = \int_{-\hat{\Delta\varphi}(p')}^{\hat{\Delta\varphi}(p')} f_{P',\Delta\Phi}(p',\Delta\varphi)\mathrm{d}\Delta\varphi \quad (26)$$
$$f_{\Delta L}(\Delta l) = \int_{-\pi}^{\pi} f_{\Delta L,\Delta\Phi}(\Delta L,\Delta\varphi)\mathrm{d}\Delta\varphi. \quad (27)$$

As previously discussed, the PDFs of $\Delta\varphi$ and $\Delta l$ (and thus $r$) only depend on the SNR and not on the absolute stimulus power. The cross-power $P'$, however, is the product of the left and right stimulus envelope and must thus also depend on stimulus power. For this reason, PDFs for $P'$ will always be shown normalized by $C^2$ so that PDFs only depend on the SNR and are independent of overall stimulus power.

No closed-form solution for Equations (24)–(27) could be found so that numeric integration was used to evaluate them (QUADPACK algorithms QAGS/QAGI; Piessens et al., 1983). Figures 2I–K show some examples of the PDF of $\Delta\Phi$, $\Delta L$, $P'$ and verifies the results by comparing Equations (24)–(25) to PDFs that were numerically estimated from signal waveforms.

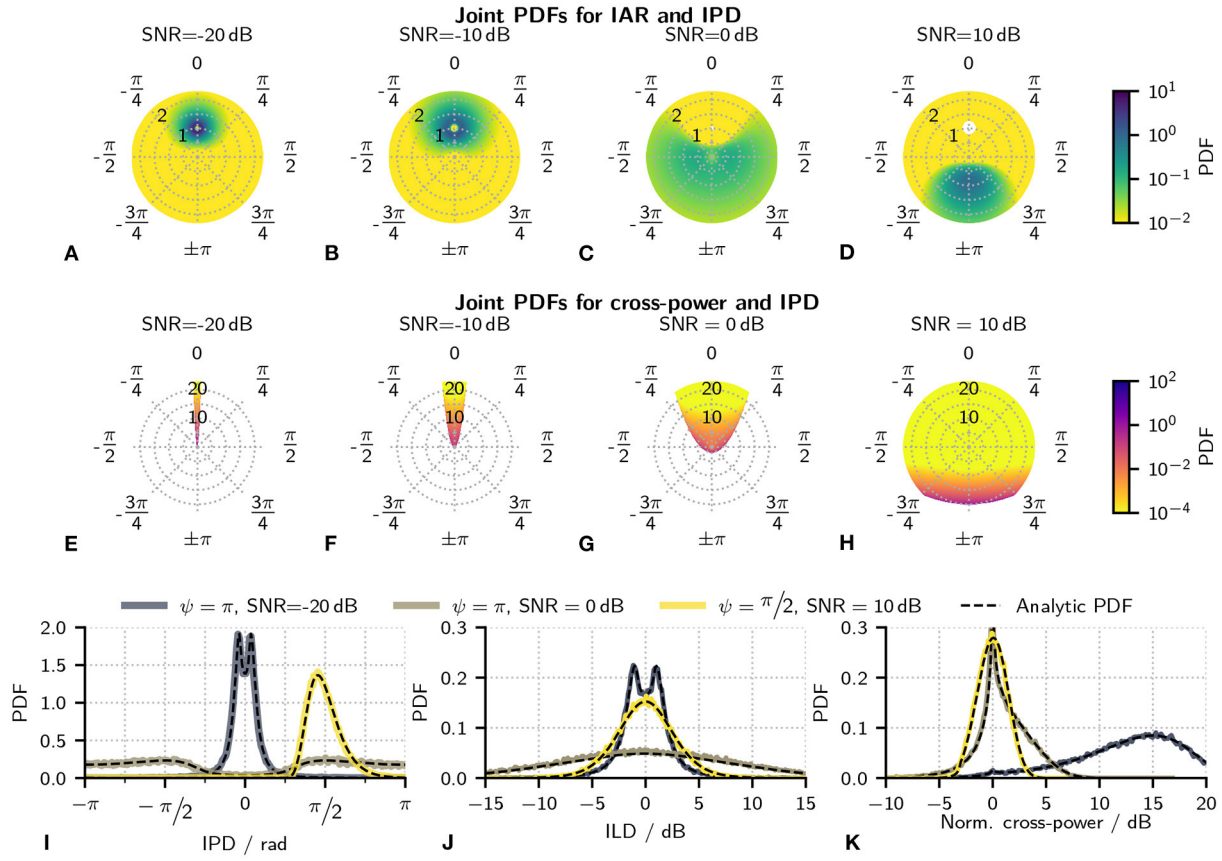



FIGURE 2
**(A–D)** Some examples of the joint PDF of IAR and IPD given in Equation (18). All plots show results for a tone-IPD of $\pi$ with the SNR increasing from left to right. Angles in the polar plot are the IPDs, while the radial variable is the IAR. Colors indicate the probability density. A logarithmically-scaled colormap was used due to the large dynamic range of the PDF. White areas located at an IAR = 1 and IPD = 0 for 0 and 10 dB indicate a probability density of 0. **(E–H)** Joint PDF for cross-power and IPDs given in Equations (20). Results are shown for the same parameters as in **(A–D)**. As in the first row of plots, angles indicate the IPD and color the probability density. The radial variable, however, is the cross-power. These PDFs were calculated for a noise variance of $\sigma^2 = 1$. A logarithmically-scaled colormap was used due to the large dynamic range of the PDF. White areas indicate undefined combinations of cross-power and IPD as defined by Equation (23). **(I–K)** Evaluation of the analytical results by comparing the derived marginal PDFs with numerically estimated PDFs. In all cases, black, dashed lines indicate analytical results gained from Equations (24)–(27). Colored lines indicate results that were instead numerically estimated from waveforms. Panel **(I)** shows marginal PDFS for IPDs $\Delta\Phi$, **(J)** for ILDs $\Delta L$ and k) for the cross-power $P'$.

## 3. Discussion

Figures 2A–D show joint PDFs for IAR and IPD calculated at a tone-IPD of $\psi = \pi$ and different SNRs. Without any tone, this distribution would equal a delta distribution with infinite probability density at an IPD of zero and an IAR of 1. At low SNRs (Figures 2A,B), the antiphasic tone has only a small influence on the noise resulting in probability densities that are still tightly clustered around the IPD of 0 and an IAR of 1. With increasing amplitude of the tone and thus increasing SNR, this clustering becomes less pronounced (Figures 2B,C). When the tone starts to dominate the stimulus, the probability density becomes highest around the tone-IPD of $\pi$ (Figures 2C,D). At large SNRs, the PDF would converge toward a delta distribution at the tone-IPD of $\pi$ and an IAR of 1. Figures 2E–H shows joint PDFs for cross-power and IPD at the same conditions as used in Figures 2A–D. Without the antiphasic tone, the stimulus density would be concentrated on a single line at zero IPD. Also, the signal is diotic so that the cross-power equals the stimulus power so that the cross-power distribution would equal that of the squared envelope. At low SNRs (Figures 2E,F), the addition of the tone starts to introduce IPD fluctuations thus widening the joint PDF. A large area of these joint PDFs are, however, undefined. These undefined areas are determined by Equation (23) and become intuitive when studying the signal model shown in Figure 1B. At low tone amplitudes $C$, it is only possible to gain large IPDs at moments where the envelope of the noise and thus $x + iy$ are small. This also result in a small cross-power $p' = a_L \times a_R$. With increasing $C$, large IPDs can then also appear at increasingly

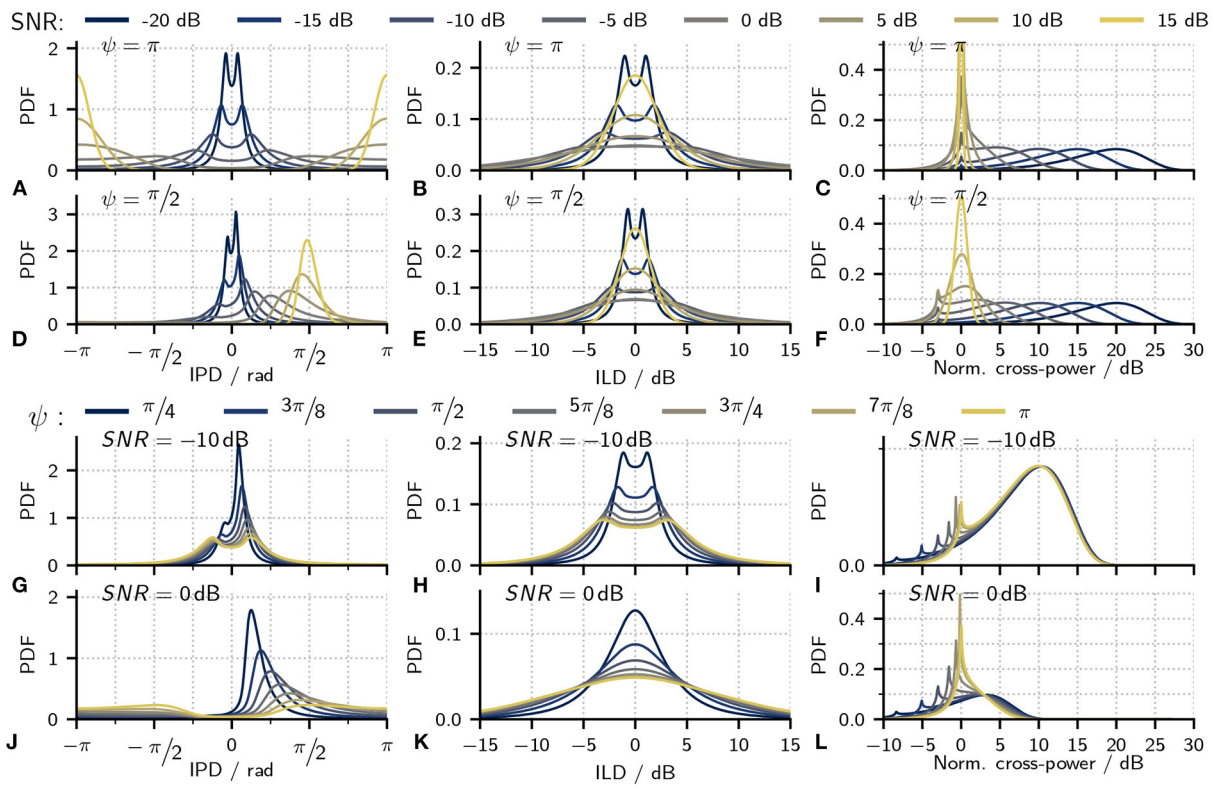



FIGURE 3
Exemplary marginal PDFs for IPDs (first column), ILDs (second column), and the cross-power (third column). For better visualization, the cross-power values were normalized with the squared tone amplitude so that the x-axis shows $10\log_{10}(p'/C^2)$. **(A–F)** PDFs calculated for two fixed signal phases $\psi = \pi$ (top-row) and $\psi = \pi/2$ (bottom row). Different colors indicate results at different SNRs. **(G–L)** PDFs calculated for two fixed SNRS: −10 dB (top-row) and 0 dB (bottom row). Different colors indicate results at different signal phases $\psi$.

large values of $p'$. This is seen in Figure 2G and especially Figure 2H.

While joint PDFs are the main contribution of this study, they are hard to visualize and, consequently, difficult to discuss in detail. Instead, the following section discusses marginal PDFs for IPDs, cross-power, and ILDs as a function of different stimulus properties. These PDFs lack information about the interaction between the individual metrics, such as IPD and cross-power or ILD. However, they do convey the impact of different metrics more intuitively. Figures 3A,D show examples of the marginal IPD PDFs for $\psi = \pi$ and $\psi = \pi/2$ while varying the SNR. The instantaneous IPD $\Delta\varphi$ can be interpreted as a result of the mixture of zero IPD due to the diotic noise and the IPD $\psi$ of the tone. The weighting of the two IPDs is determined by the noise's instantaneous power relative to the tone's power. Thus, at large negative SNRs where the stimulus is dominated by noise, IPD PDFs show a mean value close to zero and only little variance. With increasing SNR, the IPDs are increasingly influenced by the tone-IPD so that the distributions mean moves toward $\psi$ and variance increases. At larger positive SNRs, where the noise power is small compared to the tone, the IPDs are dominated by the tone-IPD $\psi$ so that the variance decreases again. In the two extreme cases where the SNR would either be $-\infty$ or $+\infty$, the signal consists of only the noise or the tone so that neither IPD nor ILD fluctuates—both PDFs are then $\delta$-distributions. For the IPD, this distribution is either be located at zero (SNR=$-\infty$) or at $\psi$ (SNR=$+\infty$) while the ILD distribution is always centered at 0 dB. Figures 3B,E show ILD PDFs for the same parameters as used for the IPD PDFs in Figures 3A,D. Instantaneous ILDs $\Delta l$, are a direct result of the relative energy of the instantaneous noise and the tone. As a result, ILD PDFs exhibit the same change of variance as discussed for the IPDs, low variance at both high or low SNRs where the stimulus is either dominated by the tone or noise and an increase of variance at intermediate SNRs. Figures 3C,F show distributions for the remaining parameter $P'$ plotted in decibels relative to the squared amplitude of the tone. For large SNRs, the signal is dominated by the tone, $p'/C^2$ is thus narrowly distributed around 0 dB. With decreasing SNR, the noise power increases relative to $C^2$ so that the peak of the distribution shifts toward larger values of $p'/C^2$ with the overall shape of the distribution remaining largely unchanged.

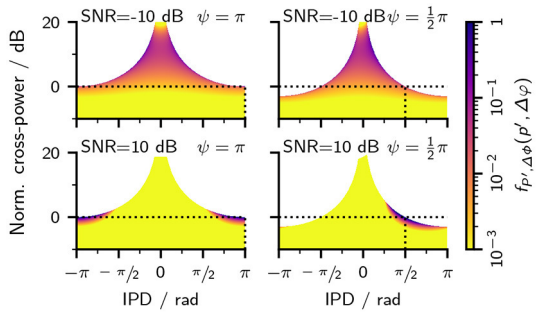



FIGURE 4
Joint probability functions of the cross-power $P'$ and IPD as defined in Equation (20). For better comparison, the y-axis was normalized with the squared tone amplitude so that the y-axis shows $10\log_{10}(p'/C^2)$. The top row shows PDFs at an SNR of −10 dB, while the bottom row shows PDFs at an SNR of 10 dB. Columns show Each panel shows a PDF at different SNRs and Tone-IPDs $\psi$. The horizontal dashed black lines indicate the location where $p' = C^2$ so that the normalized cross-power is 0 dB. The vertical black lines indicate where the IPD matches the tone-phase $\Delta\varphi = \psi$. Note that the color map is logarithmically-scaled and that changes in the scale were limited to values between 1 and $10^{-3}$.

Figures 3G–L additionally show IPD, ILD, and $P'$ PDFs for cases where the SNR was fixed while varying $\psi$. From the vector summation shown in Figure 1B, it is intuitive that, at the same tone amplitude $C$, a smaller value of $\psi$ also results in smaller IPDs. As a direct consequence, IPD and ILD PDFs also show less variance for smaller values of $\psi$. The PDFs for $P'$, however, are largely uninfluenced by $\psi$—with the notable exception of a sharp peak located at $p'/C^2 = \sin^2(\psi/2)$. This peak is a consequence of Equation (23), which limits the possible combinations of IPDs and $P'$. To better understand the origin of this peak, Figure 4 shows joint PDFs of IPD and $P'$. Notably, the probabilities are heavily clustered close to the limit defined by Equation (23). The low slope of the limiting $\hat{p}'$ function toward $\pm\pi$ in combination with the accumulation of probability density along this limit results in the observed peak in the cross-power PDFs. From Equation (23) follows that $\hat{p}'(\Delta\varphi = \pm\pi) = C^2\sin^2(\psi/2)$ which is the location of the peaks in Figures 3I,L.

All PDFs derived above show discontinuities for $\Delta\varphi \in \{0, \pm\pi\}$ for which the probability densities approach zero. Or, in other words, a $N_0S_\psi$ stimulus will never contain IPDs that are exactly zero or $\pi$. Both discontinuities can be understood when keeping in mind that the IPD is defined by $\Delta\varphi = \arctan2(\upsilon, \xi)$. Which can only result in a value of 0 or $\pm\pi$ if $\upsilon = 0$. This is only the case when $x = -C\cos(\psi/2)$. As the probability of $x$ to take this exact value approaches zero, the joint PDFs will also approach zero. For further discussion of the PDFs, however, this discontinuity was not shown explicitly in the plots above as its implication in practice is limited.

Furthermore, the PDFs derived in this study are independent of noise spectrum and bandwidth. They are thus valid for any Gaussian noise with zero mean. Further, the tone frequency does not need to be located within the noise spectrum. However, with auditory processing, especially peripheral filtering, the spectrum will influence the effective SNR at the level of binaural interaction and, thus, the PDFs of the encoded binaural cues. In these cases, PDFs will be determined by the effective SNR of the stimulus as processed, meaning after considering the bandpass properties of the auditory periphery. While all PDFs were derived for the diotic noise case $N_0S_\psi$, they can easily be generalized to cases where an additional phase delay $\psi_2$ is applied to the whole stimulus. Such a signal could then be referred to as $(N_0S_\psi)_{\psi_2}$ and would result in identical IPD distributions as in the $N_0S_\psi$ case but shifted by $\psi_2$ with ILD and $P'$ distributions remaining unchanged.

## 3.1. Quantifying IPD and ILD variability

Multiple studies have used models making use of the variability of IPDs, ILDs, or a combination of the two, as a detection cue for tone in noise experiments (e.g., Davidson et al., 2009; Dietz et al., 2021; Encke and Dietz, 2022; Eurich et al., 2022) or for decorrelation detection (Goupell and Hartmann, 2007). Based on the derived PDFs, the following section will thus discuss different measures for the amount of IPD and ILD fluctuation for the special case of $N_0S_\pi$.

The amount of ILD fluctuations can be quantified by calculating the variance $V$ of the underlying distribution defined as:

$$V = <\Delta L(t)^2> = \int_{-\pi}^{\pi}\int_{-\infty}^{\infty} \Delta l^2 f_{\Delta L,\Delta\Phi}\, \mathrm{d}\Delta l\, \mathrm{d}\Delta\varphi, \qquad (28)$$

where the angular brackets symbolize the ensemble average. The resulting variance as a function of SNR is shown in Figure 5A. As expected from the plots in Figure 3, ILD variance first increases with SNR until reaching its maximum around an SNR of −0.73 dB from where the variance decreases as the tone starts to dominate the stimulus.

Most previous studies relied on the regular variance (or standard deviation $\sqrt{V}$) as defined in Equation (28) when quantifying IPD variance (Goupell and Hartmann, 2007; Davidson et al., 2009). This approach makes sense at low SNRs where IPDs are narrowly distributed around 0. At higher SNRs, however, the distribution starts to move toward a mean value of $\pi$, and calculating the regular variance is of little significance. An alternative and better-suited metric for quantifying the IPD variability is the circular variance $V_{\mathrm{circ}}$ (Fisher, 1993)

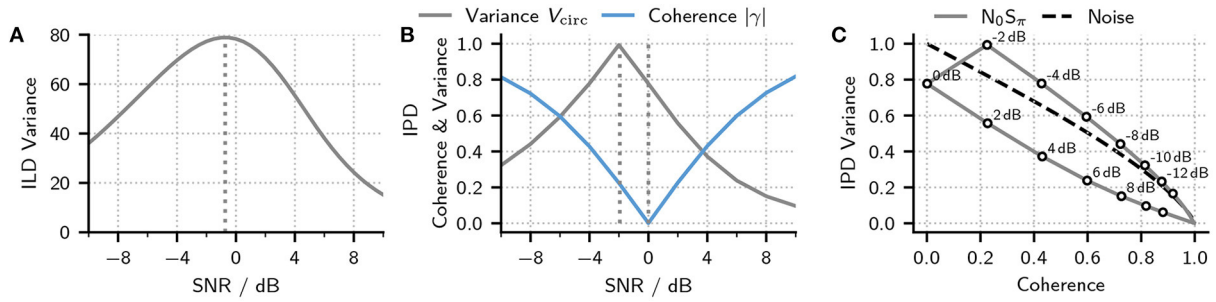



FIGURE 5
**(A)** Variance of ILDs in an $N_0S_\pi$ signal calculated at different SNRs. The dashed line marks the maximum of the function **(B)** Circular IPD variance in an $N_0S_\pi$ signal calculated at different SNRs (blue line) and the matching interaural coherence (gray line). Dotted lines indicate the location of the maximum in variance and minimum in coherence. **(C)** Circular IPD variance as a function of stimulus coherence for an $N_0S_\pi$ stimulus (gray line and symbols) as well as (partly) decorrelated noise (dashed black line) (Just and Bamler, 1994). Symbols and labels indicate SNRs resulting in a given combination of coherence and variance.

defined as:

$$V_{\text{circ}} = 1 - \left|\left\langle e^{i\Delta\Phi(t)}\right\rangle\right| = 1 - \left|\int_{-\pi}^{\pi}\int_{0}^{\hat{p}'(\Delta\varphi)} e^{i\Delta\varphi} f_{P',\Delta\Phi}\, dp'\, d\Delta\varphi\right|, \tag{29}$$

where the angular brackets symbolize the ensemble average, $V_{\text{circ}}$ can take values between 0 and 1 with a value of 0 indicating no IPD fluctuations. In contrast, a value of 1 indicates a wide distribution of IPDs (but not necessarily a uniform distribution). The gray line shows the circular variance as a function of SNR in Figure 5B. Like the ILD variance, IPD variance increases with increasing SNR until reaching its maximum around an SNR of −1.93 dB from where the variance starts to decrease.

A second and alternative metric for quantifying the amount of IPD fluctuations has recently been shown to directly account for the detection performance in a variety of tone in noise tasks: The interaural coherence[1] $|\gamma|$ (Encke and Dietz, 2022; Eurich et al., 2022). The interaural coherence is defined as the modulus of the complex-valued correlation coefficient and can be calculated as:

$$|\gamma| = \frac{|\langle R_a(t)L_a^*(t)\rangle|}{<\sqrt{|R_a(t)|^2><|L_a(t)|^2>}} = \frac{\left|\left\langle P'(t)e^{i\Delta\Phi(t)}\right\rangle\right|}{<\sqrt{|R_a(t)|^2><|L_a(t)|^2>}}, \tag{30}$$

$$= \frac{1}{2\sigma + C^2}\left|\int_{-\pi}^{\pi}\int_{0}^{\hat{p}'(\Delta\varphi)} p' e^{i\Delta\varphi} f_{P',\Delta\Phi}\, dp'\, d\Delta\varphi\right|, \tag{31}$$

where $R_a, L_a$ are the analytical representation of the left and right ear signals, the asterisk symbolizes the complex conjugate, and $\sigma^2$ and $C$ are the variance of the noise and the amplitude of the tone, respectively. Comparing this equation to the definition of $V_{\text{circ}}$ in Equation (29), shows that the two measures are closely related, with the main difference being that $|\gamma|$ weights the IPDs by $p'$ before averaging. This weighting requires a normalization achieved by the term before the integrals. In addition to this, the two metrics show inverse behavior. A stimulus with no IPD fluctuations will result in an interaural coherence of $|\gamma| = 1$ while the circular variance would be $V = 0$.

An interesting property of $|\gamma|$ is that any stimulus with a real-valued cross power density spectrum such as $N_0S_\pi$ also results in a real-valued $\gamma$ which then equals the interaural (Pearson) correlation. Figure 5B shows the interaural coherence (and thus correlation) as a function of SNR (blue line). As expected from the previous discussions, the coherence decreases with increasing SNR until reaching a coherence of zero at an SNR of 0 dB from where it starts to increase. Surprisingly, however, the minimum in coherence does not match the maximum in IPD or ILD variability. Figure 5C thus shows the same data as in panel b but plotting IPD variance as a function of coherence. The same plot also shows the IPD variance of two partly correlated noise tokens as a function of coherence. From this figure, one can appreciate that, depending on the stimulus, the same coherence can result in different amounts of IPD variance. These differences are caused by the $p'$ weighting

1 Note that there are several different definitions of coherence. Our use of coherence as $|\gamma|$ is a typical time-domain definition (Saleh, 2007). In general signal processing, the coherence function is instead often defined in the frequency domain and calculated as the normalized absolute value of the cross-spectral power density (CSPD) (Shin, 2008). The two definitions are closely related, as the time-domain coherence can also be defined by using a Fourier transform of the CSPD. In binaural research, a third definition exists, where interaural coherence is sometimes used to refer to the maximum of the real-valued cross-correlation function (Blauert, 1983).

of IPDs that is included when calculating $|\gamma|$ (see Equation 31). two stimuli that share the same IPD PDF but differing $P'$ PDFs would thus show also differ in their coherence.

## 4. Summary

This study aimed to derive the joint PDF for ILDs (IARs) and IPDs as well as IPDs and $P'$. The two functions are given by the Equations (19) and (20). The two equations are a key component for understanding how the SNR and $\psi$ influence the magnitude of binaural unmasking when considering IPD and ILD variance as the underlying cue. The approach applied to derive PDFs can further be used as a template for other types of binaural signals. In the future, it will hopefully help to get a better understanding of how different stimulus statistics influence binaural unmasking.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author/s.

## Author contributions

JE and MD designed the research and wrote the paper. JE conducted the calculations, analyzed the data, and produced the figures. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme grant agreement No. 716800 (ERC Starting Grant to MD).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1022308/full#supplementary-material

## References

Blauert, J. (1983). *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press.

Culling, J. F., and Lavandier, M. (2021). " Chapter 8: Binaural unmasking and spatial release from masking," in *Binaural Hearing*, eds R. Y. Litovsky, M. J. Goupell, R. R. Fay, and A. N. Popper (Cham: Springer International Publishing), 209–241. doi: 10.1007/978-3-030-57100-9_8

Davidson, S. A., Gilkey, R. H., Colburn, H. S., and Carney, L. H. (2009). An evaluation of models for diotic and dichotic detection in reproducible noises. *J. Acoust. Soc. Am.* 126, 1906. doi: 10.1121/1.3206583

Dietz, M., Encke, J., Bracklo, K. I., and Ewert, S. D. (2021). Tone detection thresholds in interaurally delayed noise of different bandwidths. *Acta Acust.* 5, 60. doi: 10.1051/aacus/2021054

Durlach, N. I., Gabriel, K. J., Colburn, H. S., and Trahiotis, C. (1986). Interaural correlation discrimination: II. Relation to binaural unmasking. *J. Acoust. Soc. Am.* 79, 1548–1557. doi: 10.1121/1.393681

Encke, J., and Dietz, M. (2022). "A hemispheric two-channel code accounts for binaural unmasking in humans," in *Communications Biology*. 5. doi: 10.1038/s42003-022-04098-x

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (2022). Lower interaural coherence in off-signal bands impairs binaural detection. *J. Acoust. Soc. Am.* 151, 3927–3936. doi: 10.1121/10.0011673

Fisher, N. I. (1993). " Chapter 2: Descriptive methods," in *Statistical Analysis of Circular Data* (Cambridge, UK: Cambridge University Press), 15–38. doi: 10.1017/CBO9780511564345.004

Goupell, M. J., and Hartmann, W. M. (2006). Interaural fluctuations and the detection of interaural incoherence: bandwidth effects. *J. Acoust. Soc. Am.* 119, 3971–3986. doi: 10.1121/1.2200147

Goupell, M. J., and Hartmann, W. M. (2007). Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models. *J. Acoust. Soc. Am.* 122, 1029–1045. doi: 10.1121/1.2734489

Henning, G. B. (1973). Effect of interaural phase on frequency and amplitude discrimination. *J. Acoust. Soc. Am.* 54, 1160–1178. doi: 10.1121/1.1914363

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *J. Acoust. Soc. Am.* 20, 536–544. doi: 10.1121/1.1906407

Just, D., and Bamler, R. (1994). Phase statistics of interferograms with applications to synthetic aperture radar. *Appl. Opt.* 33, 4361. doi: 10.1364/AO.33.004361

Levitt, H., and Lundry, E. A. (1966). Binaural vector model: relative interaural time differences. *J. Acoust. Soc. Am*. 40, 1251–1251. doi: 10.1121/1.1943044

Meurer, A., Smith, C. P., Paprocki, M., Čertík, O., Kirpichev, S. B., Rocklin, M., et al. (2017). Sympy: symbolic computing in Python. *PeerJ Comput. Sci*. 3, e103. doi: 10.7717/peerj-cs.103

Piessens, R., de Doncker-Kapenga, E., Überhuber, C. W., and Kahaner, D. K. (1983). *Quadpack*. Berlin; Heidelberg: Springer. doi: 10.1007/978-3-642-61786-7

Saleh, B. (2007). *Fundamentals of Photonics*. Hoboken, NJ: Wiley-Interscience.

Shin, K. (2008). *Fundamentals of Signal Processing for Sound and Vibration Engineers*. Chichester, Hoboken, NJ: John Wiley & Sons.

Zurek, P. M. (1991). Probability distributions of interaural phase and level differences in binaural detection stimuli. *J. Acoust. Soc. Am*. 90, 1927–1932. doi: 10.1121/1.401672

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 23 November 2022
DOI 10.3389/fnins.2022.1059639



OPEN ACCESS

EDITED BY
Jyrki Ahveninen,
Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, United States

REVIEWED BY
Elin Roverud,
Boston University, United States
Mary M. Flaherty,
University of Illinois at Urbana-Champaign, United States

*CORRESPONDENCE
Yonghee Oh
yonghee.oh@louisville.edu

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 01 October 2022
ACCEPTED 07 November 2022
PUBLISHED 23 November 2022

CITATION
Oh Y, Hartling CL, Srinivasan NK, Diedesch AC, Gallun FJ and Reiss LAJ (2022) Factors underlying masking release by voice-gender differences and spatial separation cues in multi-talker listening environments in listeners with and without hearing loss.
*Front. Neurosci.* 16:1059639.
doi: 10.3389/fnins.2022.1059639



# Factors underlying masking release by voice-gender differences and spatial separation cues in multi-talker listening environments in listeners with and without hearing loss

Yonghee Oh[1,2]*, Curtis L. Hartling[3], Nirmal Kumar Srinivasan[4], Anna C. Diedesch[5], Frederick J. Gallun[2,3] and Lina A. J. Reiss[2,3]

[1]Department of Otolaryngology and Communicative Disorders, University of Louisville, Louisville, KY, United States, [2]National Center for Rehabilitative Auditory Research, VA Portland Health Care System, Portland, OR, United States, [3]Department of Otolaryngology, Oregon Health & Science University, Portland, OR, United States, [4]Department of Speech-Language Pathology & Audiology, Towson University, Towson, MD, United States, [5]Department of Communication Sciences and Disorders, Western Washington University, Bellingham, WA, United States

Voice-gender differences and spatial separation are important cues for auditory object segregation. The goal of this study was to investigate the relationship of voice-gender difference benefit to the breadth of binaural pitch fusion, the perceptual integration of dichotic stimuli that evoke different pitches across ears, and the relationship of spatial separation benefit to localization acuity, the ability to identify the direction of a sound source. Twelve bilateral hearing aid (HA) users (age from 30 to 75 years) and eleven normal hearing (NH) listeners (age from 36 to 67 years) were tested in the following three experiments. First, speech-on-speech masking performance was measured as the threshold target-to-masker ratio (TMR) needed to understand a target talker in the presence of either same- or different-gender masker talkers. These target-masker gender combinations were tested with two spatial configurations (maskers co-located or 60° symmetrically spatially separated from the target) in both monaural and binaural listening conditions. Second, binaural pitch fusion range measurements were conducted using harmonic tone complexes around a 200-Hz fundamental frequency. Third, absolute localization acuity was measured using broadband (125–8000 Hz) noise and one-third octave noise bands centered at 500 and 3000 Hz. Voice-gender differences between target and maskers improved TMR thresholds for both listener groups in the binaural condition as well as both monaural (left ear and right ear) conditions, with greater benefit in co-located than spatially separated conditions. Voice-gender difference benefit was correlated

with the breadth of binaural pitch fusion in the binaural condition, but not the monaural conditions, ruling out a role of monaural abilities in the relationship between binaural fusion and voice-gender difference benefits. Spatial separation benefit was not significantly correlated with absolute localization acuity. In addition, greater spatial separation benefit was observed in NH listeners than in bilateral HA users, indicating a decreased ability of HA users to benefit from spatial release from masking (SRM). These findings suggest that sharp binaural pitch fusion may be important for maximal speech perception in multi-talker environments for both NH listeners and bilateral HA users.

KEYWORDS

**voice-gender release from masking, spatial release from masking, binaural pitch fusion, localization acuity, hearing loss, hearing aid (HA)**

## Introduction

Multi-talker listening environments occur when multiple talkers with various voice characteristics and spatial locations interact with each other. Those multi-talker listening situations present a challenging auditory environment which can make the task of target speech perception remarkably difficult for listeners due to masking effects created by the abundance of interfering background talkers (maskers). This situation is often referred to as the "cocktail party" phenomenon (Cherry, 1953).

Many previous studies have reported that there are two major acoustic cues that can improve speech segregation performance of a target message in listening environments like the "cocktail party" (Brungart, 2001; Albogast et al., 2002; Darwin et al., 2003; Ericson et al., 2004; Allen et al., 2008; Brungart et al., 2009; Best et al., 2011; Litovsky, 2012; Gallun et al., 2013; Srinivasan et al., 2016; Gaudrain and Başkent, 2018; Oh et al., 2021). One of these acoustic cues is vocal-characteristic differences between target and maskers that are a result of differences in talker gender (e.g., fundamental frequency differences, vocal-tract length differences, etc.) and the other is spatial separation between target and maskers (e.g., co-located vs. spatially separated talkers). Here, the improvement they can provide for speech segregation is referred to as "release from masking." Specifically, the release from masking by the cues from vocal-characteristic differences is termed "voice-gender release from masking" (VGRM), and the masking release by spatial separation cues is termed "spatial release from masking" (SRM). It should be noted that the term VGRM was originally proposed in the study by Oh and Reiss, 2017a,b and used in their other studies (Oh et al., 2021, 2022). Here, "gender" denotes the classical categorization of a talker's voice with their assigned sex at birth. Different terms have been used in previous speech-on-speech masking studies (e.g., "sex-mismatch benefits" Richter et al., 2021).

Previous studies have explored VGRM in isolation and have found that differences in voice characteristics between talkers of different genders lead to greater masking release than the differences in voice characteristics between talkers of the same gender for normal hearing (NH) listeners (Brungart, 2001; Ericson et al., 2004; Brungart et al., 2009). Studies have also explored SRM in isolation and have established that NH listeners benefit significantly from spatial separation cues between the target and competing maskers, beginning at separations as small as 2°, and that SRM benefit generally improves with increasing degrees of separation (Allen et al., 2008; Best et al., 2011; Litovsky, 2012; Gallun et al., 2013; Srinivasan et al., 2016; Yost, 2017).

While these findings are important, few studies have explored the interaction between these two cues together and their influences on SRM and VGRM. One recent study by Oh et al. (2021) found that there is an unequal perceptual weighting between the VGRM and SRM that NH listeners achieve across a spatial field. That is, at smaller spatial separations (up to 15–30°) between target and maskers, VGRM is more dominant than SRM, and at larger separations, (greater than 30 up to 60°) the perceptual weighting is reversed and SRM is more dominant than VGRM. Additionally, there was a clear point of intersection between this reversal of VGRM and SRM dominance where the magnitude of masking release for SRM and VGRM was equal.

In hearing-impaired (HI) listeners, bilateral device use including hearing aid (HA) and/or cochlear implant (CI) can be a major factor for binaural listening advantages in both voice-gender difference and spatial separation cues (Litovsky et al., 2006; Marrone et al., 2008; Visram et al., 2012; Bernstein et al., 2016). However, benefits from bilateral devices are highly variable, and often provide little speech perception benefit or even interfere with speech perception, compared to monaural device use (Litovsky et al., 2006; Ching et al., 2007; Reiss et al., 2016; Reiss and Molis, 2021).

Reduced benefits of voice-gender differences in HI listeners could be attributed to poorer monaural frequency resolution for representation of pitch or even vocal tract length cues for voice pitch discrimination. Alternatively, recent findings suggest that reduced benefits from voice-gender difference could be explained by an increased likelihood to integrate dichotic stimuli that evoke different pitches between two ears into a single fused sound, which is termed binaural pitch fusion (Reiss and Molis, 2021; Oh et al., 2022). Generally, binaural pitch fusion is narrow in NH listeners because the two ears provide essentially matched spectral information for a given signal. In contrast, HI listeners can exhibit abnormally broad binaural pitch fusion, i.e., can fuse stimuli with pitches differing by up to 3–4 octaves across ears into a single percept (Reiss et al., 2014, 2017, 2018a,b; Oh and Reiss, 2017b, 2020). Thus, broad binaural pitch fusion appears to be detrimental, and could negatively impact the ability to segregate out multiple voices of different pitches in complex environments. In the current study, as the first goal, we investigated whether variability in binaural pitch fusion may explain some of the variability in voice-gender difference benefits in a common speech-on-speech masking task similar to those used in the previous studies.

Similarly, reduced benefits of spatial separation have previously been attributed to aging, hearing loss, poor sound source localization abilities, and a combination of those factors (Gallun et al., 2005, 2013; Best et al., 2011; Gifford et al., 2014; Füllgrabe et al., 2015; Srinivasan et al., 2016, 2021; Swaminathan et al., 2016; Ellinger et al., 2017; Baltzell et al., 2020). Their studies found aging and hearing loss could contribute to the reduction in SRM interdependently or independently (Gallun et al., 2013; Srinivasan et al., 2016). In addition, reduced temporal and spectral processing caused by either aging or hearing loss could reduce the ability to use spatial cues to segregate different auditory streams (Best et al., 2011; Füllgrabe et al., 2015; Srinivasan et al., 2016). There has also been some evidence showing that absolute sound localization ability from the processing of interaural time differences (ITDs) and interaural level differences (ILDs) could contribute to SRM (Gallun et al., 2005; Gifford et al., 2014; Srinivasan et al., 2016; Swaminathan et al., 2016; Ellinger et al., 2017; Baltzell et al., 2020). Most of their studies argued that the limited access to those localization cues could be explained by the interaction between aging and hearing loss. In the current study, as the second goal, we investigated whether variability in listener's absolute sound localization ability may explain some of the variability in SRM in speech-on-speech masking.

The overall goal of this study was to measure two different types of masking releases due to (1) the voice-gender differences between talkers (i.e., VGRM); and (2) the spatial separation between talkers (i.e., SRM), and investigate how these differ in bilateral HA users from age-matched NH listeners. Further, measurements of binaural pitch fusion and absolute localization acuity were conducted on the same subject groups that participated in the speech-on-speech masking experiment. We explored whether variability in pitch fusion and localization acuity could explain the variability in VGRM and SRM, respectively. In order to check that these correlations are truly due to binaural processing, speech-on-speech masking experiments were repeated in two monaural (left ear and right ear) listening conditions, and their results were compared with those in the bilateral listening conditions. Our primary hypothesis was that broad binaural pitch fusion would be associated with reduced benefit from the voice-gender difference cue, and conversely that narrow binaural pitch fusion would be associated with a greater advantage in the use of this cue. In other words, the benefit from the voice gender difference cue (VGRM) would be negatively correlated with the binaural pitch fusion ranges. We also hypothesized a negative correlation between sound localization acuity and masking release by spatial separation (SRM). That is, poor localization acuity would be associated with reduced SRM, and conversely that acute localization acuity would be associated with a greater advantage in SRM. Finally, we expected that no correlations would be observed with the monaural listening conditions.

# Materials and methods

## Participants

All measurements were conducted according to the guidelines for the protection of human subjects as set forth by the Institutional Review Boards (IRBs) of both Oregon Health and Sciences University and the Portland VA Medical Center, and the methods employed were approved by these IRBs. Twenty-three adult subjects, consisting of eleven NH listeners ranging in age from 36 to 67 years (mean and standard deviation (std) = 50.0 ± 9.9 years; 7 females), twelve bilateral HA users ranging in age from 30 to 75 years (mean and std = 53.8 ± 16.7 years; 10 females; **Table 1**), participated in this study. A Kruskal–Wallis $H$-test showed that there were no significant age differences between these two listener groups [$H(1) = 1.817$, $p = 0.611$]. All subjects were native English speakers and screened for normal cognitive function using the 10-min Mini Mental Status Examination (MMSE) with a minimum score of 27 out of 30 required to qualify (Folstein et al., 1975; Souza et al., 2007), ruling out cognitive impairment that would potentially influence performance.

Normal hearing was defined as air conduction thresholds ≤25 dB hearing level (HL) from 125 to 4000 Hz. Mean pure-tone averages at octave interval frequencies between 125 and 4000 Hz for NH subjects were 12.6 ± 2.2 dB HL for the left ear and 11.5 ± 1.4 dB HL for the right ear. Bilateral HA users had moderate to severe hearing losses in both ears and relatively symmetric losses between ears, with the exception of subject H1. Mean pure-tone averages were 56.5 ± 10.8 and 57.7 ± 10.5 dB

TABLE 1 Demographic information for hearing-aid (HA) users: age, sex, etiology of hearing loss, and reference ear.

| Subject ID | Age (years) | Sex | Etiology of hearing loss | Reference ear |
|---|---|---|---|---|
| H1 | 75 | Male | Unknown | Right |
| H2 | 30 | Female | Genetic | Right |
| H3 | 39 | Female | Genetic | Right |
| H4 | 67 | Female | Genetic | Right |
| H5 | 34 | Female | Unknown | Right |
| H6 | 39 | Male | Genetic | Right |
| H7 | 71 | Female | Unknown | Left |
| H8 | 47 | Female | Noise | Left |
| H9 | 67 | Female | Unknown | Right |
| H10 | 73 | Female | Unknown | Left |
| H11 | 43 | Female | Genetic | Left |
| H12 | 60 | Female | Unknown | Left |
| M | 53.8 | | | |
| SD | 16.7 | | | |

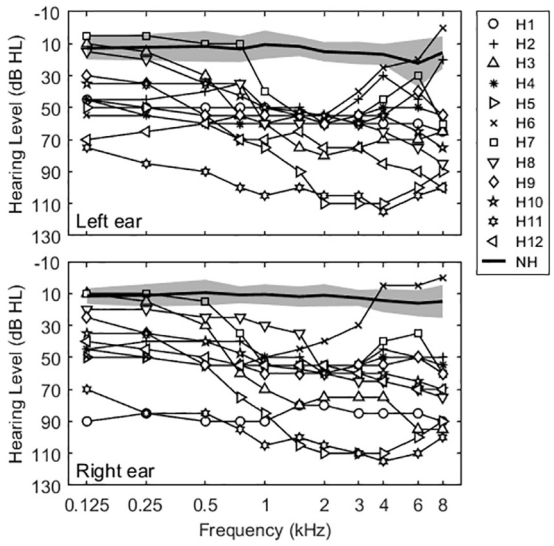



FIGURE 1
Unaided audiograms for the NH and bilateral HA subjects in this study. Solid thin lines show individual thresholds for bilateral HA users. Solid thick lines and shaded areas represent averaged thresholds and standard deviations for NH subjects.

HL for left and right ears, respectively. Figure 1 shows group-averaged audiograms for NH subjects (thick solid lines) and individual audiograms for bilateral HA subjects (lines with open symbols) for left and right ears.

All bilateral HA users were required to have at least 1 year of experience with bilateral HA use and have monaural word intelligibility scores of 65% or higher on the Consonant Nucleus Consonant (CNC) word test with both devices. For the speech-on-speech masking experiment and the sound localization acuity experiment, all HA users used lab loaner HA devices (Phonak Ambra). All extra processing features for hearing devices were disabled, including adaptive/automatic gain control, frequency lowering, directional microphones, and noise reduction. HAs were verified to meet NAL-NL2 (National Acoustics Laboratories-Non-Linear2, Australia) targets (speech stimuli at 50, 65, and 75 dB SPL) using real-ear measurements in order to provide suitable amplification for a subject's hearing loss, and all subjects met the target criteria. In both subject groups, tympanometry was also conducted to verify normal middle ear function. Additional details of etiology of hearing loss of the HA users are shown in Table 1. All subjects were paid an hourly wage and completed all experiments in between four to seven sessions of 2–3 h each. No prior experience with psychophysical research was required for participation; however, practice tutorials (20–30 min) were provided to all subjects in order to assure familiarity with the procedures.

## Stimuli and procedures

Three main experiments were conducted in this study: speech recognition threshold measurement in competing speech, binaural pitch fusion range measurement, and localization acuity measurement. The measurements of both speech recognition threshold and localization acuity were conducted in the anechoic chamber located at the National Center for Rehabilitative Auditory Research (NCRAR). The measurement of binaural pitch fusion range was conducted in a double–walled, sound attenuated booth at the Oregon Hearing Research Center (OHRC). All statistical analyses were conducted in SPSS (version 25, IBM).

### Speech-on-speech masking measurement

All speech stimuli were digitally processed in MATLAB to have a sampling rate of 44.1 kHz. Stimuli were presented through a bank of three eight-channel amplifiers (Ashlys/ne4250) and 24 frequency-equalized loudspeakers calibrated by a Brüel and Kjaer sound level meter. The loudspeakers were arranged in a circle in the horizontal plane with 15° increments surrounding the listener and equidistant at 2 m from the listener's head.

All speech stimuli were drawn from the Coordinate Response Measure (CRM; Bolia et al., 2000) speech corpus, which consists of sentences in the form "*Ready* [*call sign*] *go to* [*color*] [*number*] *now*." In this study, speech stimuli were presented with a 20% slower speaking rate than the original CRM corpus stimuli because some HA users had difficulties in understanding target-only stimuli at the original speaking rate. A custom MATLAB implementation of a modified pitch synchronous overlap add (PSOLA) technique (Moulines and Laroche, 1995) was used to time-stretch CRM sentences by 20%. There are eight possible *call signs* (Arrow, Baron, Charlie,

Eagle, Hopper, Laker, Ringo, and Tiger), and 12 keywords: four *colors* (red, green, white, and blue) and the *numbers* (1–8). All possible combinations of the *call signs*, *colors*, and *numbers* were spoken by four male ($F_0$ = 100 ± 7 Hz) and four female talkers ($F_0$ = 204 ± 12 Hz). Note that fundamental frequency ($F_0$), which represents the voice pitch, was estimated using the cepstrum algorithm in MATLAB where the output is the Fourier transform of the log of the magnitude spectrum of the input waveform (Flanagan, 1965). $F_0$ for each talker was averaged across all of that talker's CRM speech stimuli.

Each subject was presented with three simultaneous sentences from the CRM corpus (1 target and 2 simultaneous maskers). Subjects identified keywords associated with one target sentence while attempting to ignore two masker sentences. Target speech stimuli were presented from directly in front of the listener with a fixed sound presentation level of 60 dB SPL. Masker speech stimuli were presented in one of two spatial configurations: co-located (target at 0°, maskers at 0°) or 60° symmetrical separations (target at 0°, maskers at ± 60°). Only symmetrical target-masker separation conditions were considered in order to minimize availability of the better ear cue (monaural head shadow effect; Shaw, 1974; Kidd et al., 1998) and maximize reliance on spatial cues or voice-gender cues for source segregation.

These two spatial conditions were tested with four different gender combinations of target and maskers: MM (male target, male maskers), MF (male target, female maskers), FF (female target, female maskers), and FM (female target, male maskers), for a total of 2 × 4 = 8 conditions. In each trial, the subject was instructed to face the front speaker and attend to the target sentence, always identified here by the *call sign* "Charlie," and indicate the target *color* and *number* keywords from the 32 possible *color*/*number* combinations. The masker sentences had exactly the same form as the target but a different *call sign*, *color*, and *number*, randomly selected on each trial. The one target and two masker sentences were randomized from eight talkers (four males and four females) for each target-masker gender combination at each trial, and they were temporally aligned at the beginning and were roughly the same total duration.

Responses were obtained using a touch screen monitor located on a stand within arm's reach of the listener seated in the middle of the anechoic chamber. The monitor was directly in front of the listener but below the plane of the loudspeakers. Subjects were asked to look straight ahead and to hold their heads stead during a stimulus presentation. Feedback was given after each presentation in the form of "Correct" or "Incorrect." Approximately one second of silence followed the response being registered, prior to the next stimulus presentation.

The masker sound presentation level was adaptively varied at each trial to find the target-to-masker ratio (TMR), or the masker level yielding 50% correct recognition of both target *color* and *number* (i.e., 1/32 chance), using a one-up/one-down procedure (Levitt, 1971). The initial level for the masker sentence was set at 30 dB SPL and increased in level by 5 dB for each correct response until an incorrect response occurred, then decreased in level for each incorrect response until a correct response, and so on. This was repeated until three reversals in direction were obtained, at which point the step size was changed to 1 dB and six more reversals were measured. The TMR was estimated as the average of the last six reversals. Note that TMR indicates the difference in level between the target and each masker in the symmetrical target-masker separation conditions, while signal-to-noise ratio (SNR) refers to difference between the target and the combined masker level. For example, if the target level is 60 dB SPL and each masker is also 60 dB SPL, the TMR would be 0 dB, and the overall SNR would be approximately −3 dB. All subjects were tested in binaural listening conditions and in both monaural listening conditions with the non-test ear plugged and muffed. Thresholds were averaged over three separate runs for each condition.

## Binaural pitch fusion measurement

All stimuli were digitally generated at a sampling rate of 44.1 kHz with MATLAB, delivered using an ESI Juli sound card, TDT PA5 digital attenuator and HB7 headphone buffer, and presented over Sennheiser HD-25 headphones. Headphone frequency responses were equalized using calibration measurements obtained with a Brüel and Kjaer sound level meter with a 1-inch microphone in an artificial ear.

Prior to the binaural fusion range measurements, loudness balancing was conducted sequentially across frequencies and across ears using a method of adjustment. For both listener groups, 300-ms tones at 0.125, 0.25, 0.375, 0.5, 0.625, 0.75, 0.875, 1, 1.25, 1.5, 2, 3, and 4 kHz in the reference ear were initialized to "medium loud and comfortable" levels corresponding to a 6 or "most comfortable" on a visual loudness scale from 0 (no sound) to 10 (too loud). Loudness for the comparison ear was then adjusted for each frequency to be equally loud to a tone in the reference ear during sequential presentation across the ears, based on subject feedback. Here, all loudness balancing adjustments were repeated with a fine attenuation resolution (0.1 dB steps for bilateral HA and 0.5 dB steps for NH listeners) until equal loudness was achieved with all comparison sequences within and across ears, with a reference to a 500-Hz tone in the reference ear. The averaged comfortable sound levels were 65 ± 4/65 ± 4.1 dB sound pressure level, SPL (left/right ear) for NH listeners and 90 ± 1.4/91 ± 1.7 dB SPL (left/right ear) for bilateral HA users. The frequencies and order of presentation were randomized to minimize the effect of biases such as time-order error and underestimation or overestimation of the loudness (Florentine et al., 2011). This loudness balancing procedure was performed to minimize use of level-difference cues and maximize focus on pitch differences as the decision criteria. Using the same program, each ear was then checked for poor within-ear pitch ranking ability by asking subjects to rank which tone was higher in pitch for all frequency combinations.

Binaural pitch fusion range measurements were then performed to measure the fusion ranges over which dichotic pitches were fused with dichotic 1500-ms harmonic tone complexes. The method of constant stimuli procedure was used: the reference stimulus was fixed in the designated "reference ear," and the contralateral, comparison stimulus was varied across trials. For NH listeners, the reference ear was randomized. For bilateral HA users, if one ear had poor within-ear frequency discrimination as assessed during the loudness balancing procedure, that ear was assigned to be the reference ear so that the resolution of comparison stimulus testing would be maximized using the contralateral better ear, instead of limited by the worse ear. The reference fundamental frequency ($F_{0ref}$) was fixed at 200 Hz, and the comparison stimuli consisted of other harmonic complexes with fundamental frequencies ($F_{0comp}$) sampled around the reference with 1/64 to 1/16 octave steps and varied pseudo-randomly across trials. The number of harmonic components was fixed at four.

At each trial, subjects were asked to indicate whether they heard a single fused sound or two different sounds through a touch screen monitor. If a single sound was heard, subjects were instructed to indicate whether they heard that sound as a single fused sound ("Same"). If two different sounds were heard, subjects were instructed to indicate which ear had the higher pitch ("Left higher" or "Right higher") as a check of whether two sounds were really heard. A "Repeat" button was also provided to allow subjects to listen to the stimuli again. No feedback was given during the run. Binaural pitch fusion ranges were averaged over three separate runs.

### Localization acuity measurement

Three Gaussian noise-band stimuli with 500-ms duration were generated with sixth-order Butterworth filter and processed in MATLAB to have a sampling rate of 44.1 kHz. The broadband stimulus was band-pass noise filtered between 125 and 8000 Hz, and two narrowband stimuli were band-pass noises centered at 500 and 3000 Hz with 1/3-octave-wide bands. All stimuli were presented through the same 24-loudspeaker array system and equipment configuration as used in the speech-on-speech masking experiment.

Prior to the localization acuity measurements, threshold estimates of "quiet detection threshold" were performed to ensure the audibility of each noise stimulus. A one-up/two-down adaptive procedure tracking the 70.7% correct point (Levitt, 1971) was used with a four-interval (two-cue, two-alternative). On each trial, the target sound was assigned to the second or third interval with equal probability, and no signal was presented in the first and the fourth intervals. The initial level was set at 50 dB SPL and decreased in level for two consecutive correct responses until an incorrect response occurred, then increased in level for each incorrect response until a correct response, and so on. This was repeated until three reversals in direction were obtained, at which point the step size was decreased by half for each reversal. The average of the last six reversals with a 1-dB step size was used to estimate thresholds. The averaged quiet threshold levels were 21 ± 3.2/24 ± 5.1/25 ± 5.3 dB SPL (broadband/500-Hz band-pass noise/3000-Hz band-pass noise) for NH listeners and 32 ± 9.4/40 ± 8.7/43 ± 7.2 dB SPL (broadband/500-Hz band-pass noise/3000-Hz band-pass noise) for bilateral HA users.

Localization acuity measurements were then performed with the method of constant stimuli procedure for each stimulus condition: three presentations of the 24 speakers in random order (i.e., 72 trials for each stimulus condition). The stimulus level was fixed at 30 dB sensational level (SL). Subjects were asked to look straight ahead and to hold their heads steady during a stimulus presentation and asked to identify the location of the sound through the touchscreen (a circle with a radius of 5 cm without a visual representation of all speakers) after stimulus presentation. No feedback was given during the run. Localization acuity was averaged over three separate runs for each stimulus condition.

## Results

### Effects of voice-gender differences and spatial separation on speech recognition thresholds in noise

**Figures 2**, **3** show the results of the speech-on-speech masking experiment for NH and HA user groups, respectively. Note that the TMR thresholds of the two same-gender conditions (MM and FF) were similar at each spatial configuration in both groups, as were those of the two different-gender conditions (MF and FM), and these TMR threshold similarities between talker-masker gender combinations were also reported in the previous studies (Gallun et al., 2013; Oh et al., 2021) that used the same experimental setup as the current study. Thus, the TMR thresholds averaged in the same-gender vs. the different-gender conditions were used for all plots and statistical analyses in this study.

The top row of **Figure 2** shows individual and mean TMR thresholds as a function of target-maskers spatial separation (0 and ± 60°) for three listening conditions (binaural, left only, and right only) in NH listeners. Generally, smaller or more negative TMR thresholds indicate better (or improved) speech recognition ability in noise. In the binaural listening condition, the results show that the same-gender condition (3.16 ± 0.56 dB) exhibits larger (poorer) TMR thresholds than the different-gender condition (−5.18 ± 2.19 dB) in the co-located target-maskers spatial configuration. A similar trend was observed in the spatially (± 60°) separated configuration (the same-gender condition: −8.31 ± 3.14 dB; the different-gender condition: −11.09 ± 2.65 dB). In both spatial configurations, the lower TMR values for the different-gender conditions relative

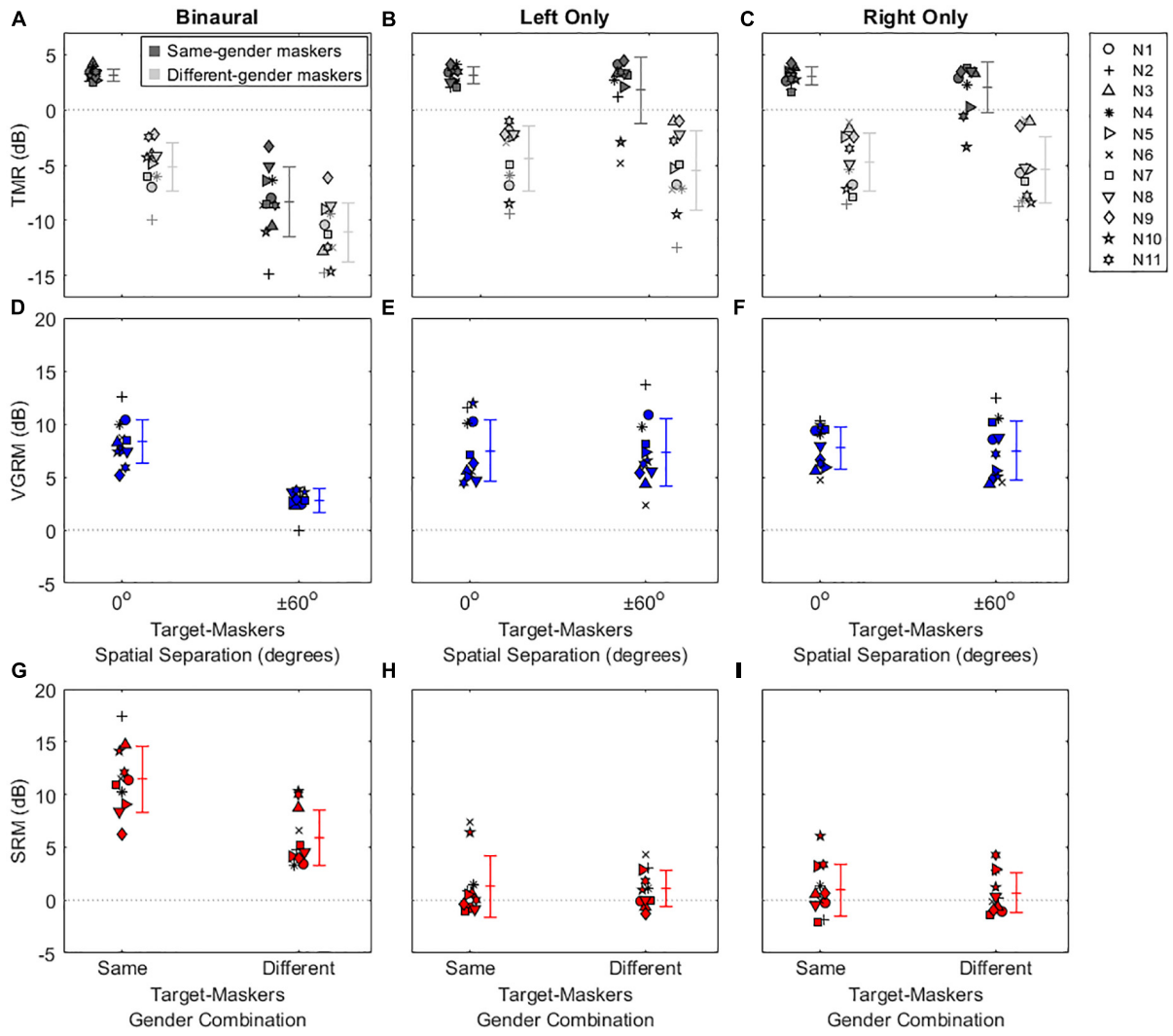



FIGURE 2
Individual and average target-to-masker ratio (TMR) thresholds and voice gender release from masking (VGRM) and spatial release from masking (SRM) for NH listeners. The left, middle, and right columns refer to the binaural, left only, and right only listening conditions, respectively. The upper panels show the TMR thresholds **(A–C)** as a function of target-masker spatial separation (0 and ± 60°). Dark-shaded and light-shaded symbols indicate TMR thresholds for the same-gender masker and the different-gender masker conditions, respectively. The middle panels show VGRMs **(D–F)** as a function of target-masker spatial separation (0 and ± 60°). The lower panels show SRMs **(G–I)** as a function of target-masker gender combination (same-gender and different-gender). Error bars represent standard deviation around the mean. Horizontal dotted lines represent reference zero values.

to the same-gender conditions are indicative of the amount of VGRM, which shows how much speech recognition thresholds in noise are improved by differences in gender between the target and maskers. The amount of VGRM (Figure 2D) was calculated by the difference in TMR thresholds between same-gender (dark-gray symbols) and different-gender (light-gray symbols) conditions at each spatial configuration. The VGRM for NH listeners (Figure 2D) ranged between −0.11 and 12.52 dB, and the mean VGRM was greater in the co-located spatial configuration (8.34 ± 2.07 dB) than in the spatially separated configuration (2.79 ± 1.10 dB).

Another interesting finding in NH listeners is that spatial separation of the maskers to ± 60° relative to the target at 0° led to smaller (better) TMR thresholds for all target-masker gender combinations. This reduction is indicative of the amount of SRM, which shows how much speech recognition thresholds are improved by spatial separation of the talker from the maskers. The amount of SRM (Figure 2G) is defined as the spatial separation benefits at each target-masker gender combination [i.e., differences between dark-gray (or light gray) symbols at 0° and at ± 60° in Figure 2A]. The SRM for NH listeners ranged between 3.42 and 17.29 dB, and the mean SRM was greater in

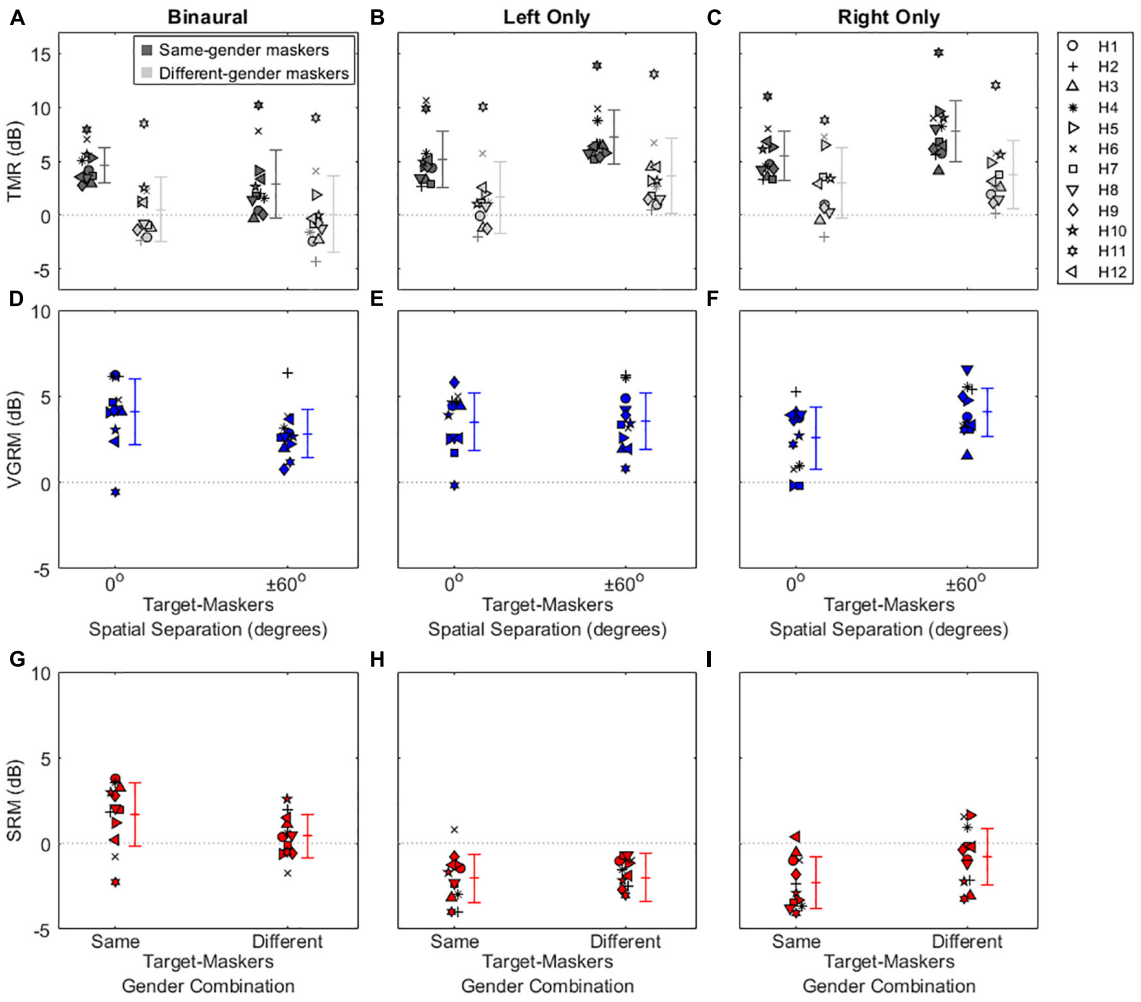



FIGURE 3
Individual and average target-to-masker ratio (TMR) thresholds and voice gender release from masking (VGRM) and spatial release from masking (SRM) for bilateral HA users. Plotted as in **Figure 2**, with different scales on the *y*-axis.

the same-gender target-maskers combination (11.47 ± 3.12 dB) than in the different-gender combination (5.91 ± 2.61 dB).

Compared to the binaural listening condition, the two monaural listening conditions elicited TMR threshold changes, especially in the spatially separated target-maskers configuration, and thus different results in VGRM and SRM. First, the TMR thresholds in the left-only (**Figure 2B**) and right-only (**Figure 2C**) listening conditions were similar to those in the binaural listening condition (**Figure 2A**) at the co-located target-maskers configuration (left only: 3.11 ± 0.78 dB same-gender/−4.38 ± 2.92 dB different-gender; right only: 3.02 ± 0.82 dB same-gender/−4.72 ± 2.64 dB different-gender). However, the monaural TMR thresholds were essentially unchanged compared to the co-located condition when the target and maskers were spatially separated (left only: 1.81 ± 2.99 dB same-gender/−5.48 ± 3.60 dB different-gender; right only: 2.06 ± 2.32 dB same-gender/−5.40 ± 2.99 dB different-gender). The masking release results in the two monaural listening conditions show that the VGRM remained steady at around 8 dB regardless of spatial separation between target and maskers (**Figures 2E,F**), while SRM was decreased to near zero regardless of target-maskers gender differences (**Figures 2H,I**).

The results in the top row of **Figure 3** show that bilateral HA users exhibited overall poorer speech recognition performance (i.e., more positive TMR thresholds with a range between −4.38 and 15.09 dB) throughout all listening conditions compared to NH listeners (TMR thresholds with a range between −14.85 and 4.43 dB). Interestingly, spatial separation between target and maskers didn't improve TMR thresholds for HA users even in the binaural listening condition (differences between 0 and ± 60° in the same-colored symbols). The mean SRMs for bilateral HA users (**Figure 3G**) were 1.70 ± 1.84 dB and 0.41 ± 1.24 dB for the same-gender and different-gender

talker combinations, respectively. In contrast, benefits from voice-gender differences existed in both spatial separation configurations, thus positive mean VGRMs (Figure 3D) were observed (4.11 ± 1.89 dB for the 0° and 2.83 ± 1.42 dB for ± 60° spatial separations). In the two monaural listening conditions, the SRM performance was more degraded (−2 dB shown in Figures 3H,I) than in the binaural listening conditions; however, the VGRM performance was remained steady at around 4 dB (Figures 3E,F).

Since the primary goal of this study was to investigate masking release by voice-gender differences (VGRM) and spatial separations (SRM), only the masking release data were analyzed in each masking release type using linear mixed model (LMM) analyses with the amount of masking release (VGRM or SRM) as a dependent variable, the subject group (NH vs. bilateral HA), the listening conditions (binaural vs. left only vs. right only), and the target-maskers conditions (spatial separation for VGRM: 0° vs. ± 60°; gender difference for SRM: same-gender vs. different gender) as fixed effects, and the subject as a random effect. The results for both VGRM and SRM showed significant main effects of all fixed factors ($p < 0.006$ for all cases) and significant interactions between any two combinations of the fixed factors ($p < 0.006$ for all cases). *Post-hoc* pairwise comparisons using Bonferroni correction were computed to better understand the interaction between those fixed factors. The results demonstrated that the VGRM at the ± 60° in the binaural listening condition was significantly lower than all other VGRMs in the NH listeners ($p < 0.001$ for all cases), but no VGRMs were significantly different in bilateral HA users ($p = 1.000$ for all cases). In addition, the results demonstrated that the SRM in NH listeners was significantly higher in the binaural listening condition than in two monaural listening conditions ($p < 0.001$ for all cases), and the same-gender target-maskers combination elicited a significantly higher masking release than the different-gender combination in the binaural listening condition ($p < 0.001$). A similar binaural listening benefit in SRM was also observed in the bilateral HA user group ($p < 0.05$ for all cases), but the SRMs were not significantly different between the two target-maskers gender combinations in the binaural listening condition ($p = 1.000$). Please see the Supplementary material for the detailed LMM specifications and results.

## Binaural pitch fusion and its relationship with voice gender release from masking

Figure 4 shows individual harmonic tone fusion range results for NH listeners (Figure 4A) and bilateral HA users (Figure 4B). As shown in the example fusion functions in the insets of Figure 4, fusion functions were computed as the averages of the subject responses to the multiple (six to seven) presentations of each reference and comparison stimulus pair, expressed as a function of comparison tone fundamental frequency. Values near 0 indicate comparison stimuli that did not often fuse with the reference stimulus (were heard as two sounds), while values near 1 indicate comparison stimuli that were often fused with the reference stimulus (were heard as one sound). Vertical dotted lines indicate 50% points on the fusion function, and the fusion range was defined as the range between these two lines (horizontal arrows), i.e., frequencies were fused more than 50% of the time. Fusion range is thus a measure of the breadth of fusion. The NH subjects (Figure 4A) exhibited narrow harmonic tone fusion ranges (0.14 ± 0.12 octaves), while bilateral HA users (Figure 4B) showed significantly broader harmonic tone fusion ranges [0.53 ± 0.57 octaves; $t(21) = -2.25$, $p = 0.036$].

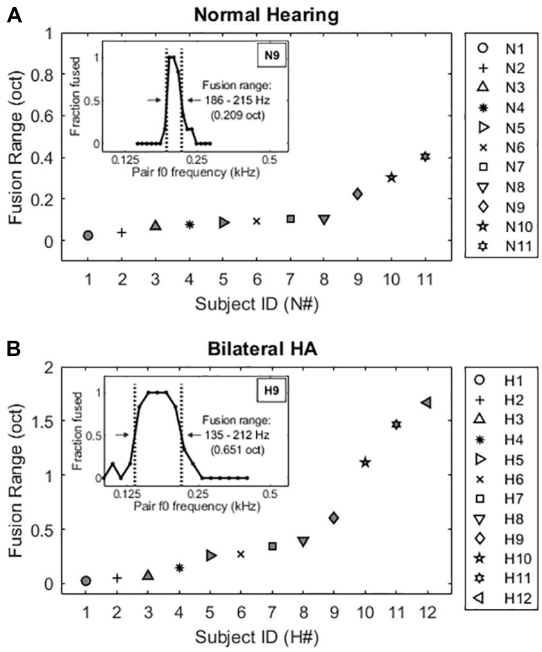



FIGURE 4
Individual harmonic tone fusion range results in an octave scale for NH listeners **(A)** and bilateral HA users **(B)**. A sample fusion function inset within each panel illustrates the fusion ranges of the 50% points (vertical dotted lines) on the fusion function.

The next step was to determine whether VGRM, the release from masking due to voice-gender differences between target and maskers, is related to the width of binaural pitch fusion. Multiple regression analyses were conducted to measure a linear relationship between two variables. Figure 5 shows individual VGRMs plotted as a function of fusion ranges in the co-located target-maskers configuration for NH listeners (left column) and bilateral HA users (right column). In the binaural listening condition, VGRM was significantly correlated with the fusion range in both subject groups (NH listeners: $r = -0.710$, $p = 0.014$ in Figure 5A; bilateral HA users: $r = -0.850$, $p < 0.001$ in

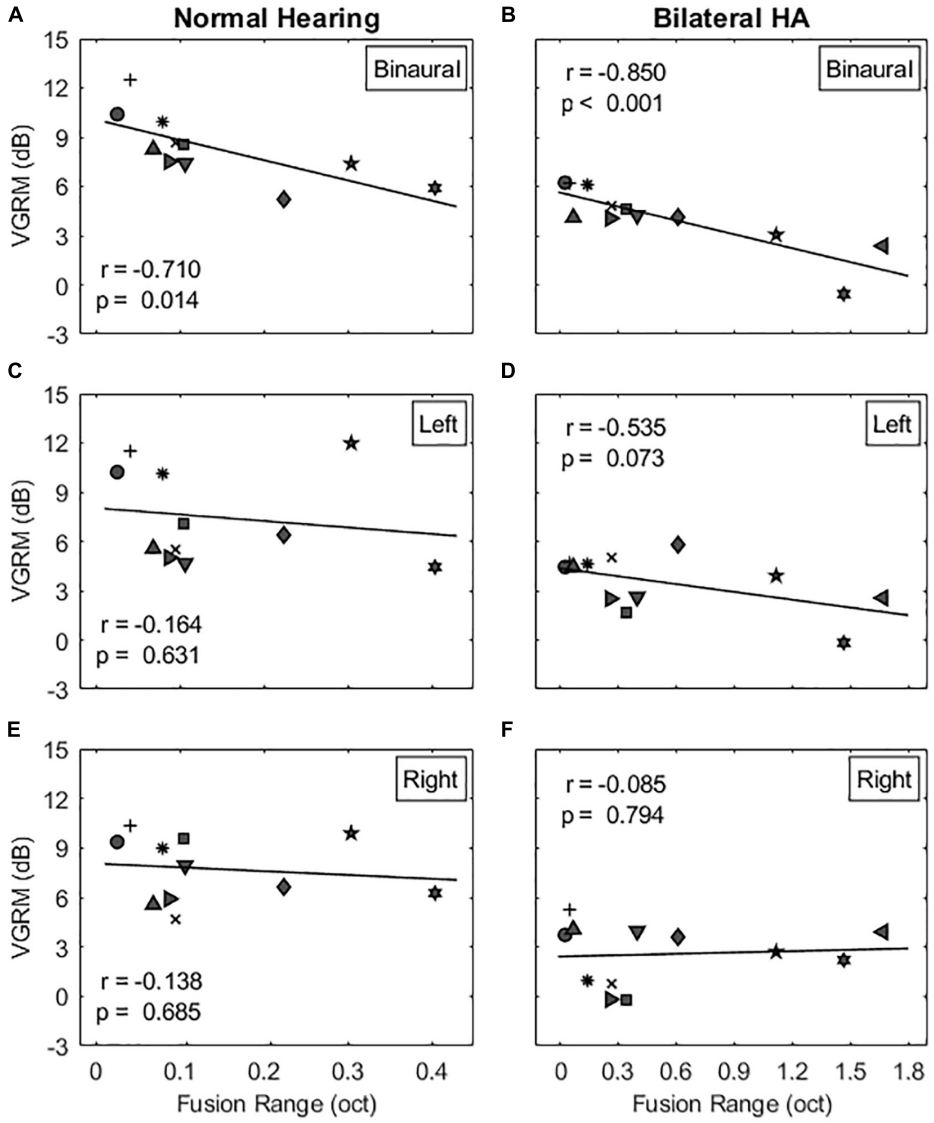



FIGURE 5
Correlations between voice gender release from masking (VGRM) and binaural pitch fusion range for the co-located target-maskers configuration. The left and right columns show the correlation results for NH and bilateral HA user groups, respectively. The panels **(A–F)** show the correlation results for the binaural, left, and right listening conditions, respectively. Table 2 shows the correlation results for the spatially separated target-maskers configuration.

Figure 5B). In other words, listeners with narrow binaural pitch fusion ranges had larger VGRM (larger differences in TMR thresholds between same-gender and different-gender maskers) than did listeners with broad fusion. However, this negative correlation between VGRM and fusion range was eliminated in the two monaural listening conditions in both listener groups (see Figures 5C–F: $p > 0.073$ for all cases). Note also that some listeners with broad fusion had greater VGRM in one or both monaural conditions compared to the binaural condition (e.g., N10 and H9, indicated by star and diamond symbols in Figures 5A–E, respectively). As provided in Table 2, no significant correlation was observed in the spatially separated target-maskers configuration as well ($p > 0.163$ for all cases).

## Localization acuity and its relationship with spatial release from masking

Figure 6 shows individual minimum audible angle results for NH listeners (Figure 6A) and bilateral HA users (Figure 6B). Example localization scatter plots were shown in the insets of Figure 6. The subject's response angles were plotted as a function of the source angles, and ideal performance would be

TABLE 2 Regression coefficients between voice gender release from masking (VGRM) and binaural pitch fusion range widths for NH and bilateral HA user groups in each spatial separation and listening condition.

| Target-maskers spatial separation | Listening condition | Correlation *r*-values (significance) | |
|---|---|---|---|
| | | NH | Bilateral HA |
| 0 degree | Binaural | **−0.710 (0.014)*** | **−0.850 (<0.001)***** |
| | Left only | −0.164 (0.631) | −0.535 (0.073) |
| | Right only | −0.138 (0.685) | −0.085 (0.794) |
| ± 60 degree | Binaural | −0.428 (0.195) | −0.586 (0.063) |
| | Left only | −0.338 (0.310) | −0.260 (0.414) |
| | Right only | −0.350 (0.291) | −0.249 (0.435) |

Correlation values in bold face indicate significant results (***$p < 0.001$; *$p < 0.05$).

represented by all points lying on the diagonal lines. The root-mean-square (RMS) angular errors were calculated to quantify a subject's accuracy in localizing sound sources (Lorenzi et al., 1999). It should be noted that the circle and plus symbols in the insets of Figure 6 indicate the subject's responses to any given source locations in the front and rear source fields, respectively, and that front-back confusions were excluded for estimating the absolute localization ability in this study. The NH subjects (Figure 6A) exhibited fine localization acuity with all stimuli tested in this study (broadband: 5.75 to 13.75°; 500-Hz band-pass noise: 6.2 to 12.35°; 3000-Hz band-pass noise: 7.25 to 11.65), while bilateral HA users (Figure 6B) showed significantly poorer localization acuity [broadband: 10 to 26.4°; 500-Hz band-pass noise: 9.4 to 28.2°; 3000-Hz band-pass noise: 11.2 to 26.35 degree; $t(48.5) < -4.61$, $p < 0.001$ for all stimulus cases]. The localization acuity was not significantly different across the stimulus types for each subject group [NH: $t(42) > -0.760$, $p = 1$; bilateral HA: $t(42) > -0.619$, $p = 1$].

The next step was to determine whether SRM, the release from masking due to spatial separation between target and maskers, is related to the absolute localization ability quantified as the RMS angular error. Multiple regression analyses were conducted to measure a linear relationship between two variables. Figure 7 shows individual SRMs plotted as a function of RMS angular errors in the same-gender target-maskers combination for NH listeners (left column) and bilateral HA users (right column). Results show that the SRM was correlated, but the correlation was not statistically significant, with the RMS angular errors for all stimuli tested in this study ($p > 0.077$). In other words, there was a tendency for listeners with sharp

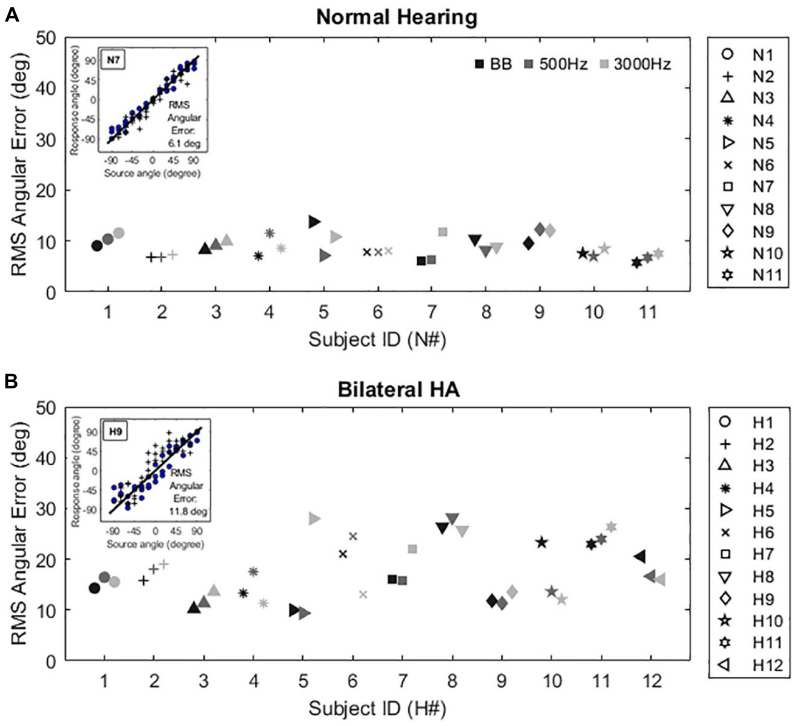



FIGURE 6
Individual localization acuity results with three different stimuli (BB: broadband, 500 Hz: band-passed noises centered at 500 Hz, 3000 Hz: band-passed noises centered at 3000 Hz) for NH listeners **(A)** and bilateral HA users **(B)**. A sample subject response inset within each panel illustrates the mean root mean square (RMS) angular error calculated by the difference between the perfect localization (diagonal line) and the listener's response (symbols) angles. The circle and plus symbols indicate the subject's responses to any given source locations in the front and rear source fields, respectively.

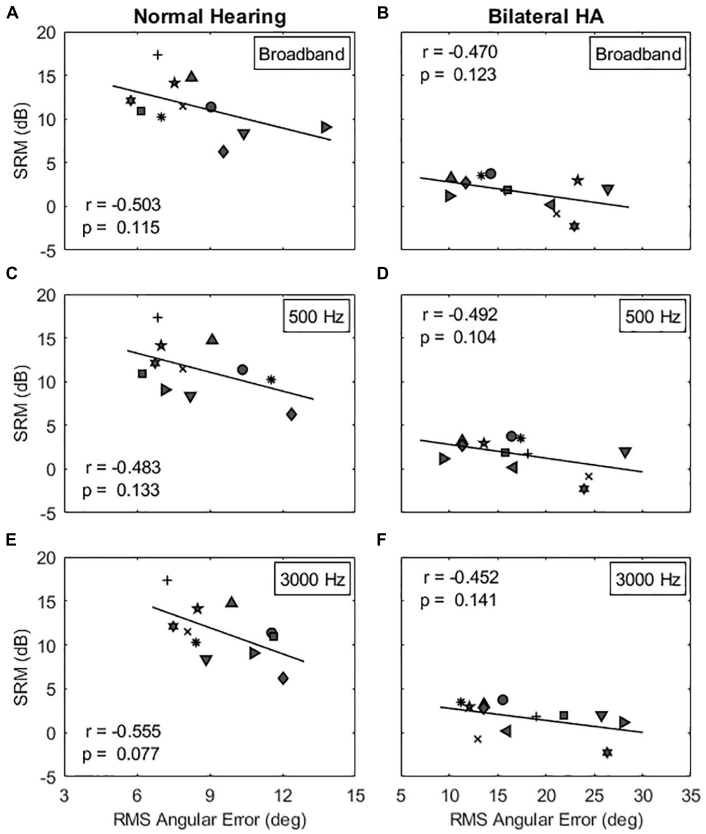



FIGURE 7
Correlations between spatial release from masking (SRM) and localization acuity for the same-gender target-maskers condition. The left and right columns show the correlation results for NH and bilateral HA user groups, respectively. The panels **(A–F)** show the correlation results for the broadband, 500 and 3000 Hz stimulus conditions, respectively. Table 3 shows the correlation results for the different-gender target-maskers condition.

localization acuity to have larger SRM (larger differences in TMR thresholds between co-located and spatially separated maskers) compared to listeners with poor localization acuity. In addition, this correlation was reduced in the different-gender target-maskers combination (not shown). The model summary of the regression analysis is provided in Table 3.

TABLE 3 Regression coefficients between spatial release from masking (SRM) and localization acuity for NH and bilateral HA user groups in each stimulus and gender-combination conditions.

| Target-maskers gender combination | Stimulus type | Correlation *r*-values (significance) | |
|---|---|---|---|
| | | NH | Bilateral HA |
| Same-gender | Broadband | −0.503 (0.115) | −0.470 (0.123) |
| | 500 Hz | −0.483 (0.133) | −0.492 (0.104) |
| | 3000 Hz | −0.555 (0.077) | −0.452 (0.141) |
| Different-gender | Broadband | −0.210 (0.491) | −0.128 (0.692) |
| | 500 Hz | −0.379 (0.536) | −0.228 (0.477) |
| | 3000 Hz | −0.236 (0.456) | −0.263 (0.408) |

## Discussion

The ability to segregate a target talker from competing masker talkers is important for speech perception in multi-talker listening environments. The current study measured speech-on-speech masking performance by varying voice-gender differences and spatial separation cues between target and maskers in both NH listeners and bilateral HA users, and examined how this performance relates to binaural pitch fusion range and localization acuity.

The results from NH listeners showed that VGRM, the average masking release *via* voice-gender differences, was maximized at 8.34 dB in the co-located spatial configuration and reduced to 2.79 dB in the separated spatial configuration. Similarly, SRM, the average masking release *via* talker spatial separation, was maximized at 11.47 dB when the target was presented with the same-gender maskers and reduced to 5.91 dB when the different-gender target-maskers were presented.

TABLE 4 Regression coefficients for masking release by voice-gender differences (VGRM) and spatial separation (SRM), binaural pitch fusion range, and absolute localization acuity predicted by age and pure tone average (PTA).

| Measurement | Condition | Predictor variable | Correlation *r*-values (significance) |
|---|---|---|---|
| VGRM | Co-located | Age | −0.262 (0.227) |
| | target-maskers | PTA | **−0.713 (<0.001)***** |
| | Spatially separated | Age | −0.254 (0.243) |
| | target-maskers | PTA | **−0.537 (0.008)**** |
| SRM | Same-gender | Age | −0.126 (0.565) |
| | target-maskers | PTA | **−0.636 (0.001)**** |
| | Different-gender | Age | −0.092 (0.677) |
| | target-maskers | PTA | **−0.423 (0.045)*** |
| Binaural pitch | – | Age | 0.200 (0.371) |
| fusion range | | PTA | **0.534 (0.009)**** |
| Absolute | Broadband | Age | 0.227 (0.297) |
| localization | | PTA | **0.627 (0.001)**** |
| acuity | 500 Hz | Age | 0.087 (0.692) |
| | | PTA | **0.588 (0.003)**** |
| | 3000 Hz | Age | 0.088 (0.690) |
| | | PTA | **0.763 (<0.001)***** |

Correlation values in bold face indicate significant results (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$).

Consistent with previous studies, these findings demonstrate a trading relationship between the perceptual weights applied to voice-gender difference and those to spatial separation cues. This trading relationship of masking release was also partially discussed in previous literature (Misurelli and Litovsky, 2012, 2015; Gallun and Diedesch, 2013; Gallun et al., 2013; Oh et al., 2021). The current study results also indicate that this trading relationship is eliminated in monaural listening conditions. SRM was minimized at around 1 dB regardless of the talkers' gender difference cue, while VGRM was maintained at around 8 dB regardless of the talkers' spatial separation cue. Hence, the trading relationship between SRM and VGRM appears to be related to the presence of binaural cues.

The results from bilateral HA users showed that average VGRM was 4.11 and 2.83 dB for co-located and spatially separated conditions, while average SRM was 1.7 and 0.41 dB for the same-gender and different-gender maskers. As in NH listeners, a trading relationship was observed between the two masking release types, though not as pronounced. In addition, both voice gender difference and spatial separation benefits were reduced in HA users compared to NH listeners.

Previous studies have reported that reduced masking release performance observed in bilateral HA users could be attributed to reduced ability to access monaural spectro-temporal cues and/or binaural cues caused by either aging or hearing loss (Best et al., 2011, 2012; Gallun et al., 2013; Füllgrabe et al., 2015; Srinivasan et al., 2021). In this study, we also conducted multiple regression analyses to find a linear relationship between two different types of masking releases (VGRM and SRM; combined both NH and HA subjects' data) and subject factors (e.g., age and degree of hearing loss). The results showed that the pure-tone average (PTA from 125 and 4000 Hz) accounted for more than 18% ($R^2$ predictor, $p < 0.045$) of the variance in both VGRM and SRM; however, age couldn't explain VGRM and SRM variances ($p > 0.227$). The model summary of the regression analysis is provided in **Table 4**. However, as will be discussed, broad binaural pitch fusion and poor sound localization abilities might be other factors reducing overall SRM and VGRM.

One likely reason for the reduced SRM, though, for bilateral HA users is that they have limited access to binaural cues on the horizontal plane such as ITD and ILD cues. Previous studies have shown that ITD sensitivity is particularly important for localization performance and speech perception in noise (Gallun et al., 2005; Gallun and Diedesch, 2013; Gifford et al., 2013, 2014; Swaminathan et al., 2016; Ellinger et al., 2017). Phase-locking and ITD sensitivity can both be impaired with hearing loss (Henry and Heinz, 2013; Dai et al., 2018). In addition, bilateral HA users have reduced access to ongoing ITD cues, because the hearing devices are not designed to coordinate their timing of stimulation of the auditory nerves across the ears (Brown et al., 2016). Thus, they do not communicate their processing schemes (such as compression ratio) across the devices, especially for old hearing devices, which could alter ILDs (Byrne and Noble, 1998; Wiggins and Seeber, 2013). To minimize any potential interaural cue distortion, the current study used symmetrical target-masker configurations (co-location and ± 60° separation) so that the image of both target and masker signals can appear in front, as opposed to the left or right due to reduced ILD, and all additional processing features for hearing devices were disabled to avoid altered ILD cues. Note that in this study, effects of head shadow were also

minimized due to the symmetrical target-masker configuration. In addition, all HA users used lab loaner HA devices (Phonak Ambra) with all extra processing features disabled. Due to lack of acclimation, overall performance may be reduced with the loaner devices compared with the subjects' own hearing devices. However, for evaluation of VGRM and SRM in this study, it is important to disable these extra processing features, which often include noise reduction and directional microphones.

There was also significant variation in listeners' masking release performance for both NH and HI listeners. The findings of this study show that, as hypothesized, binaural pitch fusion range is a strong predictor for variation in VGRM. In contrast, localization ability does not seem to predict variation in SRM, though a non-significant trend was observed.

Regarding the relationship of binaural fusion to VGRM, a strong negative correlation was observed. Previous studies have found that differences in age or hearing loss (alone or in combination) can explain some of the variance across subjects (Glyde et al., 2013; Besser et al., 2015). The proportion of variance accounted for by either factor was between 24 and 39% ($R^2$ predictor, $p < 0.01$). In this study, stronger negative correlations were observed between binaural fusion range and VGRM for both NH listeners and bilateral HA users, especially in the co-located target-masker configuration. As reported in **Table 2**, the proportion of variance accounted for by binaural pitch fusion for VGRM was 50% ($R^2$ predictor, $p = 0.014$) for NH listeners, and 72% ($R^2$ predictor, $p < 0.001$) for bilateral HA users, which are higher than the amount of variance explained by age ($R^2 = 0.07$, $p = 0.23$ in the current study; $R^2 = 0.02$, $p < 0.52$ in Glyde et al., 2013) or hearing loss ($R^2 = 0.51$, $p < 0.01$ in the current study; $R^2 = 0.39$, $p < 0.001$ in Glyde et al., 2013) alone. Hence, broad binaural fusion could be a stronger predictor for reduced VGRM than age or hearing loss. It should be noted that the significance of this proportion of variance was observed only in the co-located target-maskers spatial configuration. We also confirmed that significance of the correlation was eliminated when binaural cues were not provided (i.e., at two monaural listening conditions; see **Table 2**), indicating that the correlation is not explained by poorer frequency discrimination or other factors that might also lead to broad binaural fusion. In particular, some subjects with broad fusion had larger VGRM under monaural listening compared to binaural listening, consistent with an interpretation of binaural interference arising from broad binaural fusion.

Regarding the relationship of sound localization acuity to SRM, a negative correlation was observed, but was not statistically significant. As reported in **Table 3**, the proportion of variance accounted for by localization acuity for SRM was low at 25% ($R^2$ predictor, $p = 0.115$) for NH listeners and 22% ($R^2$ predictor, $p = 0.123$) for bilateral HA users. A similar finding was also reported in the study by Srinivasan et al. (2021) with 22% of variance ($R^2$ predictor, $p = 0.033$) accounted for NH listeners. The lack of statistical significance in this study is likely due to the small sample size for each listener group, along with the small effect size. There is likely to be an effect of localization acuity, but this effect seems to be small. One reason for the small effect size is that localization acuity with multiple sounds from multiple sound sources may differ from that for a single sound, especially when there is broad binaural fusion. In such cases, fusion of multiple sounds from different spatial locations may occur, leading to an illusion of a single sound source with a diffuse spatial percept, and thus poor localization acuity. Thus, a better predictor of ability to benefit from SRM may be localization ability of more than one sound source presented simultaneously. It should also be noted that the current study estimated the absolute localization acuity without considering front-back confusion in the subject's responses. In this study, three NH and four HA subjects showed some degree of front-back confusion rates in their absolute localization acuity measurements, especially for the two narrowband signal conditions. The application of a more rigorous angular analysis, perhaps one in which front-back errors are considered, should be explored in future studies.

Interestingly, the multiple regression analysis results (**Table 4**) showed that the pure-tone average was a strong predictor for the variations of all outcomes measured in this study: (1) the masking release (>18% as $R^2$ predictor, $p < 0.045$); (2) the binaural pitch fusion range (29% as $R^2$ predictor, $p = 0.009$); and (3) the absolute localization acuity at three different stimuli (>35% as $R^2$ predictor, $p < 0.003$). However, age couldn't predict those variations ($p > 0.227$). These results indicate that the degree of hearing loss itself could be a common factor to explain degraded binaural sensitivity involved in speech-on-speech masking performance and related to pitch and spatial perception. In addition, although the correlation between age and degree of hearing loss was not found in the current study ($r = 0.078$, $p = 0.724$), it is well known that the age of the listeners is often allowed to covary with hearing loss. Furthermore, as mentioned in the introduction, the reduce binaural sensitivity could be caused by a reduction in higher-order processing such as cognitive and linguistic abilities (Besser et al., 2015). Therefore, future work will need to involve listeners who vary widely in age regardless of hearing status to separately examine the effects of age and hearing loss as factors.

In conclusion, this is the first study to demonstrate an important role of abnormally broad binaural pitch fusion in reduced binaural benefits for speech perception in multi-talker listening environments for both NH and HI listeners. The findings demonstrate that masking release from both voice gender and spatial cues is much smaller for HA users than NH listeners, and that the reduced benefit from voice gender cues is explained by abnormally broad binaural pitch fusion. Thus, for HI listeners, it will be critically important to help restore sharply tuned pitch fusion across ears for optimal binaural

benefit in noise environments, especially when benefit from spatial cues is limited. Increased understanding of factors that affect binaural benefits for speech perception for HI listeners is clinically essential for the future design of training- and device-based rehabilitative strategies to improve speech perception in quiet and noise.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by the Institutional Review Boards (IRBs) of both Oregon Health and Sciences University and the Portland VA Medical Center. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

YO, FG, and LR designed the experiments. YO, CH, NS, and AD performed the experiments. YO analyzed the data. All authors contributed to the article, discussed the results at all states, and approved the submitted version.

## Funding

This research was supported by grants R01 DC013307, P30 DC005983, and F32 DC016193 from the National Institutes of Deafness and Communication Disorders, National Institutes of Health and the 2020 New Century Scholars Research Grant by an American Speech-Language-Hearing Foundation.

## Acknowledgments

We would like to thank all of the participants who volunteered their time to be involved in this experiment.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1059639/full#supplementary-material

## References

Albogast, T. L., Mason, C. R., and Kidd, G. Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. Soc. Am.* 112, 2086–2098. doi: 10.1121/1.1510141

Allen, K., Charlile, S., and Alais, D. (2008). Contributions of talker characteristics and spatial location to auditory streaming. *J. Acoust. Soc. Am.* 123, 1562–1570. doi: 10.1121/1.2831774

Baltzell, L. S., Swaminathan, J., Cho, A. Y., Lavandier, M., and Best, V. (2020). Binaural sensitivity and release from speech-on-speech masking in listeners with and without hearing loss. *J. Acoust. Soc. Am.* 147, 1546–1561. doi: 10.1121/10.0000812

Bernstein, J. G. W., Goupell, M. J., Schuchman, G. I., Rivera, A. L., and Brungart, D. S. (2016). Having two ears facilitates the perceptual separation of concurrent talkers for bilateral and single-side deaf cochlear implantees. *Ear Hear.* 37, 289–302. doi: 10.1097/AUD.0000000000000284

Besser, J., Festen, J. M., Goverts, T., Kramer, S. E., and Pichora-Fuller, M. K. (2015). Speech-in-speech listening on the LiSN-S test by older adults with good audiograms depends on cognition and hearing acuity at high frequencies. *Ear Hear.* 36, 24–41. doi: 10.1097/AUD.0000000000000096

Best, V., Marrone, N., Mason, C., and Kidd, G. Jr. (2012). The influence of non-spatial factors on measures of spatial release from masking. *J. Acoust. Soc. Am.* 131, 3103–3110. doi: 10.1121/1.3693656

Best, V., Mason, C. R., and Kidd, G. Jr. (2011). Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing Talkers. *J. Acoust. Soc. Am.* 129, 1616–1625. doi: 10.1121/1.3533733

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (2000). A speech corpus for multitalker communications research. *J. Acoust. Soc. Am.* 107, 1065–1066. doi: 10.1121/1.428288

Brown, A. D., Rodriquez, F. A., Portnuff, C. D. F., Goupell, M. J., and Tollin, D. J. (2016). Time-varying distortions of binaural information by bilateral hearing aids: Effects of nonlinear frequency compression. *Trends Hear.* 20, 1–15. doi: 10.1177/2331216516668303

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2009). Multitalker speech perception with ideal time-frequency segregation: Effects of voice characteristics and number of talkers. *J. Acoust. Soc. Am.* 125, 4006–4022. doi: 10.1121/1.3117686

Byrne, D., and Noble, W. (1998). Optimizing sound localization with hearing aids. *Trends Amplif.* 3, 51–73. doi: 10.1177/108471389800300202

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Ching, T. Y., van Wanrooy, E., and Dillong, H. (2007). Binaural-bimodal fitting or bilateral implantation for managing severe to profound deafness: A review. *Trends Amplif.* 11, 161–192. doi: 10.1177/1084713807304357

Dai, L., Best, V., and Shinn-Cunningham, B. G. (2018). Sensorineural hearing loss degrades behavioral and physiological measures of human spatial selective auditory attention. *Proc. Natl. Acad. Sci. U.S.A.* 115, E3286–E3295. doi: 10.1073/pnas.1721226115

Darwin, C. J., Brungart, D. S., and Simpson, B. D. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *J. Acoust. Soc. Am.* 114, 2913–2922. doi: 10.1121/1.1616924

Ellinger, R. L., Jakien, K. M., and Gallun, F. J. (2017). The role of interaural differences on speech intelligibility in complex multi-talker environments. *J. Acoust. Soc. Am.* 141, EL170–EL176. doi: 10.1121/1.4976113

Ericson, M. A., Brungart, D. S., and Simpson, B. D. (2004). Factors that influence intelligibility in multitalker speech displays. *Int. J. Aviat. Psychol.* 14, 313–334. doi: 10.1207/s15327108ijap1403_6

Flanagan, J. L. (1965). *Speech analysis, synthesis and perception*. New York, NY: Springer-Verlag, 176–184. doi: 10.1007/978-3-662-00849-2

Florentine, M., Popper, A. N., and Fay, R. R. (2011). "Chapter 2," in *Loudness*, eds L. E. Marks, and M. Florentine (New York, NY: Springer), 17–56.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. (1975). "Mini-mental state": A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* 12, 189–198. doi: 10.1016/0022-3956(75)90026-6

Füllgrabe, C., Moore, B. C., and Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6:347. doi: 10.3389/fnagi.2014.00347

Gallun, F. J., and Diedesch, A. C. (2013). Exploring the factors predictive of informational masking in a speech recognition task. *Proc. Meet. Acoust.* 19:060145. doi: 10.1121/1.4799107

Gallun, F. J., Kampel, S. D., Diedesch, A. C., and Jakien, K. M. (2013). Independent impacts of age and hearing loss on spatial release in a complex auditory environment. *Front. Neurosci.* 7:252. doi: 10.3389/fnins.2013.00252

Gallun, F. J., Mason, C. R., and Kidd, G. (2005). Binaural release from informational masking in a speech identification task. *J. Acoust. Soc. Am.* 118, 1614–1625. doi: 10.1121/1.1984876

Gaudrain, E., and Başkent, D. (2018). Discrimination of voice pitch and vocal-track length in cochlear implant users. *Ear Hear.* 39, 226–237. doi: 10.1097/AUD.0000000000000480

Gifford, R. H., Dorman, M. F., Skarzynski, H., Lorens, A., Polak, M., Driscoll, C. L. W., et al. (2013). Cochlear implantation with hearing preservation yields significant benefit for speech recognition in complex listening environments. *Ear Hear.* 34, 413–425. doi: 10.1097/AUD.0b013e31827e8163

Gifford, R. H., Grantham, D. W., Sheffield, S. W., Davis, T. J., Dwyer, R., and Dorman, M. F. (2014). Localization and interaural time difference (ITD) thresholds for cochlear implant recipients with preserved acoustic hearing in the implanted ear. *Hear. Res.* 312, 28–37. doi: 10.1016/j.heares.2014.02.007

Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2013). The effects of hearing impairment and aging on spatial processing. *Ear Hear.* 34, 15–28. doi: 10.1097/AUD.0b013e3182617f94

Henry, K. S., and Heinz, M. G. (2013). Effects of sensorineural hearing loss on temporal coding of narrowband and broadband signals in the auditory periphery. *Hear Res.* 303, 39–47. doi: 10.1016/j.heares.2013.01.014

Kidd, G., Mason, C. R., Rohtla, T. L., and Deliwala, P. S. (1998). Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J. Acoust. Soc. Am.* 104, 422–431. doi: 10.1121/1.423246

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Litovsky, R. Y. (2012). Spatial release from masking. *Acoust. Today* 8, 18–25. doi: 10.1121/1.4729575

Litovsky, R. Y., Johnstone, P. M., Godar, S., Agrawal, S., Parkinson, A., Peters, R., et al. (2006). Bilateral cochlear implants in children: Localization acuity measured with minimum audible angle. *Ear Hear.* 27, 43–59. doi: 10.1097/01.aud.0000194515.28023.4b

Lorenzi, C., Gatehouse, S., and Lever, C. (1999). Sound localization in noise in normal-hearing listeners. *J. Acoust. Soc. Am.* 105, 1810–1820. doi: 10.1121/1.426719

Marrone, N., Mason, C. R., and Kidd, G. (2008). Evaluating the benefit of hearing aids in solving the cocktail party problem. *Trends Amplif.* 12, 300–315. doi: 10.1177/1084713808325880

Misurelli, S. M., and Litovsky, R. Y. (2012). Spatial release from masking in children with normal hearing and with bilateral cochlear implants: Effect of interferer asymmetry. *J. Acoust. Soc. Am.* 132, 380–391. doi: 10.1121/1.4725760

Misurelli, S. M., and Litovsky, R. Y. (2015). Spatial release from masking in children with bilateral cochlear implants and with normal hearing: Effect of target-interferer similarity. *J. Acoust. Soc. Am.* 138, 319–331. doi: 10.1121/1.4922777

Moulines, E., and Laroche, J. (1995). Non-parametric techniques for pitch-scale modification of speech. *Speech Commun.* 16, 175–205. doi: 10.1016/0167-6393(94)00054-E

Oh, Y., and Reiss, L. (2017a). Voice gender release from masking in cochlear implant users is correlated with binaural pitch fusion. *J. Acoust. Soc. Am.* 141:3816. doi: 10.1121/1.4988444

Oh, Y., and Reiss, L. A. (2017b). Binaural pitch fusion: Pitch averaging and dominance in hearing-impaired listeners with broad fusion. *J. Acoust. Soc. Am.* 142, 780–791. doi: 10.1121/1.4997190

Oh, Y., and Reiss, L. A. (2020). Binaural pitch fusion: Binaural pitch averaging in cochlear implant users with broad binaural fusion. *Ear Hear.* 41, 1450–1460. doi: 10.1097/AUD.0000000000000866

Oh, Y., Bridges, S. E., Schoenfeld, H., Layne, A. O., and Eddins, D. (2021). Interaction between voice-gender difference and spatial separation in release from masking in multi-talker listening environments". *JASA Express Lett.* 1:084404. doi: 10.1121/10.0005831

Oh, Y., Srinivasan, N. K., Hartling, C. L., Gallun, F. J., and Reiss, L. A. J. (2022). Differential effects of binaural pitch fusion range on the benefits of voice gender differences in a "cocktail party" environment for bimodal and bilateral cochlear implant users. *Ear Hear.* Available online at: https://journals.lww.com/ear-hearing/Abstract/9900/Differential_Effects_of_Binaural_Pitch_Fusion.70.aspx

Reiss, L. A., Eggleston, J. L., Walker, E. P., and Oh, Y. (2016). Two ears are not always better than one: Mandatory vowel fusion across spectrally mismatched ears in hearing-impaired listeners. *J. Assoc. Res. Otolaryngol.* 17, 341–356. doi: 10.1007/s10162-016-0570-z

Reiss, L. A., Fowler, J. R., Hartling, C. L., and Oh, Y. (2018a). Binaural pitch fusion in bilateral cochlear implant users. *Ear Hear.* 39, 390–397. doi: 10.1097/AUD.0000000000000497

Reiss, L. A., Ito, R. A., Eggleston, J. L., and Wozny, D. R. (2014). Abnormal binaural spectral integration in cochlear implant users. *J. Assoc. Res. Otolaryngol.* 15, 235–248. doi: 10.1007/s10162-013-0434-8

Reiss, L. A., and Molis, M. R. (2021). An alternative explanation for difficulties with speech in background talkers: Abnormal fusion of vowels across fundamental frequency and ears. *J. Assoc. Res. Otolaryngol.* 22, 443–461. doi: 10.1007/s10162-021-00790-7

Reiss, L. A., Molis, M., Simmons, S., and Katrina, L. (2018b). Effects of broad binaural fusion and hearing loss on dichotic concurrent vowel identification. *J. Acoust. Soc. Am.* 143:1942. doi: 10.1121/1.5036358

Reiss, L. A., Shayman, C. S., Walker, E. P., Bennett, K. O., Fowler, J. R., Hartling, C. L., et al. (2017). Binaural pitch fusion: Comparison of normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 143, 1909–1920. doi: 10.1121/1.4978009

Richter, M. E., Dillon, M. T., Buss, E., and Leibold, L. J. (2021). Sex-mismatch benefit for speech-in-speech recognition by pediatric and adult cochlear implant users. *JASA Express Lett.* 1:084403. doi: 10.1121/10.0005806

Shaw, E. A. G. (1974). Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *J. Acoust. Soc. Am.* 56, 1848–1861. doi: 10.1121/1.1903522

Souza, P. E., Boike, K. T., Witherall, K., and Tremblay, K. (2007). Prediction of speech recognition from audibility in older listeners with hearing loss: Effects of age, amplification, and background noise. *J. Am. Acad. Audiol.* 18, 54–65. doi: 10.3766/jaaa.18.1.5

Srinivasan, N. K., Jakien, K. M., and Gallun, F. J. (2016). Release from masking for small separations: Effects of age and hearing loss. *J. Acoust. Soc. Am.* 140, EL73–EL78. doi: 10.1121/1.4954386

Srinivasan, N. K., Staudenmeier, A., and Clark, K. (2021). Effect of gap detection threshold and localisation acuity on spatial release from masking in older adults. *Int. J. Audiol.* 18, 1–8.

Swaminathan, J., Mason, C. R., Streeter, T., Best, V., Roverud, E., and Kidd, G. (2016). Role of binaural temporal fine structure and envelope cues in cocktail-party listening. *J. Neurosci.* 36, 8250–8257. doi: 10.1523/JNEUROSCI.4421-15.2016

Visram, A. S., Kulk, K., and McKay, C. M. (2012). Voice gender differences and separation of simultaneous talkers in cochlear implant users with residual hearing. *J. Acoust. Soc. Am.* 132, EL135–EL141. doi: 10.1121/1.4737137

Wiggins, I. M., and Seeber, B. U. (2013). Linking dynamic-range compression across the ears can improve speech intelligibility in spatially separated noise. *J. Acoust. Soc. Am.* 133, 1004–1016. doi: 10.1121/1.4773862

Yost, W. A. (2017). Spatial release from masking based on binaural processing for up to six maskers. *J. Acoust. Soc. Am.* 141, 2093–2106. doi: 10.1121/1.4978614

frontiers | Frontiers in Psychology

TYPE Original Research
PUBLISHED 24 November 2022
DOI 10.3389/fpsyg.2022.994047



OPEN ACCESS

EDITED BY
Yi Zhou,
Arizona State University,
United States

REVIEWED BY
Ming Yang,
HEAD Acoustics,
Germany
Michael Torben Pastore,
Arizona State University,
United States

*CORRESPONDENCE
Thomas Biberger
thomas.biberger@uol.de

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

RECEIVED 14 July 2022
ACCEPTED 26 September 2022
PUBLISHED 24 November 2022

CITATION
Biberger T and Ewert SD (2022) Binaural detection thresholds and audio quality of speech and music signals in complex acoustic environments.
*Front. Psychol.* 13:994047.
doi: 10.3389/fpsyg.2022.994047



# Binaural detection thresholds and audio quality of speech and music signals in complex acoustic environments

Thomas Biberger* and Stephan D. Ewert

Department of Medical Physics and Acoustics and Cluster of Excellence Hearing4all, University of Oldenburg, Oldenburg, Germany

Every-day acoustical environments are often complex, typically comprising one attended target sound in the presence of interfering sounds (e.g., disturbing conversations) and reverberation. Here we assessed binaural detection thresholds and (supra-threshold) binaural audio quality ratings of four distortions types: spectral ripples, non-linear saturation, intensity and spatial modifications applied to speech, guitar, and noise targets in such complex acoustic environments (CAEs). The target and (up to) two masker sounds were either co-located as if contained in a common audio stream, or were spatially separated as if originating from different sound sources. The amount of reverberation was systematically varied. Masker and reverberation had a significant effect on the distortion-detection thresholds of speech signals. Quality ratings were affected by reverberation, whereas the effect of maskers depended on the distortion. The results suggest that detection thresholds and quality ratings for distorted speech in anechoic conditions are also valid for rooms with mild reverberation, but not for moderate reverberation. Furthermore, for spectral ripples, a significant relationship between the listeners' individual detection thresholds and quality ratings was found. The current results provide baseline data for detection thresholds and audio quality ratings of different distortions of a target sound in CAEs, supporting the future development of binaural auditory models.

KEYWORDS

**audio quality, detection thresholds, complex acoustic environments, auditory modeling, reverberation**

## Introduction

In daily life, a sound attended to (target) is often interfered with other (masking) sounds as well as by sound reflections and reverberation in enclosed spaces (referred to as a complex acoustic environment, CAE). However, in psychoacoustics, masking is typically assessed under optimal (anechoic) conditions, using abstracted and simplified stimuli (see, e.g., Ewert, 2020), such as pure tones and stationary noise. Such stimuli are suited for investigating basic sensory abilities and limitations of the auditory system, while minimizing

cognitive aspects. Here, additional energetic masking (EM), caused by spectral and temporal overlap of the target and the masker in the auditory periphery, plays an important role, degrading the internal representation of the target. In reverberation, additional self-masking and overlap-masking elicited by early and late room reflections (e.g., Bolt and MacDonald, 1949) occur.

Psychoacoustic (e.g., Kopčo and Shinn-Cunningham, 2003) and speech intelligibility (e.g., Cherry, 1953; Best et al., 2015; Ewert et al., 2017) studies showed that in comparison to a co-located condition, listeners benefit from spatially separated target and maskers, referred to as spatial release from masking (SRM). SRM was reduced in echoic environments (e.g., Plomp, 1976; Lavandier and Culling, 2010; Biberger and Ewert, 2019). Reverberation degrades binaural cues (e.g., Rakerd and Hartmann, 1985; Rakerd and Hartmann, 1986) such as interaural level and time differences (ILDs and ITDs). Moreover, amplitude modulations are reduced in the presence of reverberation, lowering the chance listening into the dips of fluctuating masker signals (Houtgast et al., 1980), where EM of the target is lowest.

In many situations, target sounds are transmitted by electroacoustic systems, e.g., a TV set, conference system or earphones, typically involving audio-signal processing. In this case, linear and non-linear distortions introduced by the signal processing and the transmission chain might be perceptible, affecting the perceived audio quality. Accordingly, detectability, as well as the supra-threshold salience of such distortions, are of interest. Comparable to fundamental psychoacoustic research, the consequences of different distortions on audio quality have often been examined under optimal conditions, without maskers and reverberation, including for the development and evaluation of instrumental quality measures (e.g., van Buuren et al., 1999; Moore and Tan, 2003; Tan et al., 2003; Fleßner et al., 2017, 2019). Only a few studies (e.g., Toole and Olive, 1988; Schobben and van de Par, 2004; Schepker et al., 2019) examined the influence of reverberation on the detectability of signal distortions. Toole and Olive (1988) observed a better detectability of signal resonances in reverberant rooms compared to anechoic conditions. Schobben and van de Par (2004) examined the effect of reverberation and loudspeaker cross-talk on the subjective quality of low-bitrate audio coding. They found reduced audibility of coding artifacts in reverberation. Schepker et al. (2019) evaluated the audio quality of a hearing device prototype, aiming at acoustical transparency (i.e., without any perceptible distortion) in rooms with different reverberation times. No large effect of reverberation time was found, suggesting that the use of only a single or few reverberation times might be sufficient for the audio quality assessment of such devices. Only a few approaches, e.g., Cauchi et al. (2019), and Biberger et al. (2021) considered aspects of reverberation affecting quality predictions. Biberger et al. (2021) found monaural spectral cues, capturing spectral coloration distortions of hearing devices aiming at acoustically transparency, to be more reliable for quality predictions in reverberation than cues based on the temporal fine structure or cepstrum correlation.

One other important aspect in CAEs is the number of spatially distributed sound sources (e.g., Weisser et al., 2019; Fichna et al., 2021) interfering with a target sound. However, neither the effect of interfering sounds on the perceived audio quality of a target sound, nor the applicability of existing instrumental audio quality measures to CAEs have yet been systematically examined. Instrumental quality measures have mainly been applied under anechoic conditions without maskers (Beerends et al., 2002; Moore and Tan, 2004; Huber and Kollmeier, 2006; Kates and Arehart, 2010; Harlander et al., 2014). Some auditory perception models have been applied to isolated aspects of CAEs. One example is the (monaural) Generalized Power Spectrum Model (GPSM), which has been applied to psychoacoustic masking with simplified psychoacoustic stimuli (Biberger and Ewert, 2016, 2017) as well as to audio quality for various distortions in anechoic and echoic conditions without maskers (Biberger et al., 2018, 2021).

Overall, relatively little is known about the detectability of distortions and (supra-threshold) audio quality perception in CAEs. It is unclear whether the results of "classical" quality measurements in anechoic conditions can be transferred to acoustic environments of different complexity, and whether existing audio quality models can be straightforwardly applied.

This study investigates the detectability and supra-threshold perception of a variety of prototypical audio signal distortions in CAEs of different complexity: The effect of room reverberation was assessed by using an anechoic (reference) and two echoic rooms with mild and moderate reverberation times (T60) of 0.35 s (resembling a typical living room) and 1.5 s (resembling a larger auditorium, parking lot, or church). The effect of maskers was assessed by configurations with no (reference), one, and two maskers that were either spatially co-located with the frontal target, or spatially separated to both sides of the target. Four types of distortions were applied to the target signal: i) spectral ripples (linear distortion), ii) a saturating, instantaneous non-linearity (non-linear distortion), iii) differences in the target sound-source intensity, and iv) a variation of the spatial position of the target (azimuthal direction of 0°, 4°, and 30° relative to the listener's viewing direction). The target was either speech, an acoustic guitar (representing a musical instrument), and a pink noise (representing environmental background noise). These targets differ in their spectro-temporal characteristics and might be differently affected by the distortions. While the acoustic guitar shows strong transients, the pink noise is stationary and produces a broadband excitation of auditory filters more equally than speech and the guitar. Speech was considered as the most relevant target in daily life and thus applied to all experiments of this study, while guitar music and noise were only applied to a subset of experiments. The International Speech Test Signal (ISTS; Holube et al., 2010) and a pop music excerpt were used as maskers, reflecting typical (disturbing) sounds in CAEs.

In the first experiment, detection thresholds for distorted signals were measured for a subset of the conditions, while in the

second experiment, supra-threshold audio quality ratings were obtained for two different degrees of distortion. Based on the systematic data set obtained, it was investigated (a) whether room reverberation and masker configuration affects detection thresholds and quality ratings for distorted signals; (b) whether distortion-detection thresholds and quality ratings are related, allowing adjustments of signal processing, as well as individualized perception models, based only on distortion-detection thresholds; (c) whether the individual listeners' overall performance to detect or rate the target distortions is correlated across conditions having different amount of reverberation and maskers; (d) the extent to which existing auditory models are applicable to distortion detection and audio quality ratings in such CAEs.

# Materials and methods

## Listeners

Sixteen self-reported normal-hearing listeners (7 female, 9 male) with a mean age of 28.7 years (all native German speakers) participated in the experiments. Ten of the sixteen participants received an hourly compensation. The other participants were employed by the Department of Medical Physics and Acoustics at the University of Oldenburg. All listeners had prior experience in psychoacoustic measurements.

## Stimuli

### Target and masker signals

German speech (spoken language), acoustic guitar, and pink-noise stimuli from the study of Fleßner et al. (2019), having different spectro-temporal properties, were used as target. The speech stimulus "ein Haus, keine Brücke" ("a house, no bridge") was spoken by a female speaker. The speech stimulus shows slow amplitude modulations (5-Hz range) and a relatively narrowband spectrum. The excerpt of a guitar piece comprised many transients and a wider bandwidth. The pink noise was a stationary stimulus with a broadband spectrum, covering the entire audible frequency range. All target signals had a duration of 2 s.

A male-transformed version of the ISTS speech signal (Holube et al., 2010) as applied in Schubotz et al. (2016) and Ewert et al. (2017) and a pop-music excerpt taken from Fleßner et al. (2019) were used as maskers. ISTS is nonsense speech generated from six different speakers in different languages (American-English, Arabic, Mandarin, French, German, and Spanish). The music signal includes multiple instruments and vocals, with a rather broadband spectrum. The maskers had a duration of 2.5 s and started 0.5 before the target onset. Raised-cosine ramps of 10 ms were applied to the masker and target stimuli. All signals were convolved with binaural room impulse responses (BRIRs) to define their spatial position and to simulate room reverberation (see Section "Rooms and masker configurations").

### Target stimulus distortions

The target stimuli were subjected to four different types of distortions; spectral ripples, non-linear saturation, intensity-based, and spatial:

Spectral ripples (linear distortions) were introduced as described in Fleßner et al. (2019), using sinusoidal modulation of the spectral envelope. Ten periods of the spectral sinusoidal modulation were applied between 50 Hz and 16 kHz, with equidistant spacing on a logarithmic frequency axis, corresponding to about 1.2 spectral ripples per octave. The spectral modulation depth (peak-to-valley ratio in dB) was adjusted to change the amount of distortion.

Non-linear distortions caused by a simple instantaneous symmetric saturating input–output (I/O) characteristic (referred to as non-linear saturation) simulated signal distortions caused by, e.g., large displacements of the loudspeaker diaphragm at high signal levels. The I/O characteristic was implemented as $y(t) = x(t) - \alpha \times (x(t))^3$, where x(t) and y(t) are input and output signals, respectively. The factor α weights the cubic term relative to x(t), and thus controls the nonlinearity of the I/O characteristic. Input values were limited to the range $\pm\sqrt{\frac{1}{3 \cdot \alpha}}$ where the non-linear I/O characteristic completely saturates (soft clipping). This saturating I/O function resulted in pronounced harmonic distortions at higher signal levels, typically occurring at signal onsets and transients. These additionally introduced frequency components likely provided spectral or amplitude modulation cues to the listeners.

Intensity-based distortions were introduced by adjusting the overall sound level in dB relative to the level of the reference signal. In contrast to spectral ripples and non-linear saturation, no spectral amplitude modulation cues were introduced.

Spatial (binaural) distortions were introduced by changing the azimuth location of the target using the appropriate BRIRs. The reference target was always presented in front (0° azimuth) of the listeners, while the spatially distorted target was shifted to the right side (relative to the viewing direction of the listener).

Anchor signals were generated by applying a 3.5 kHz low-pass filter, non-linear saturation and spatial distortion to the reference signals. The non-linear saturation ($\alpha_{speech} = 0.25$, $\alpha_{music} = 0.34$, $\alpha_{noise} = 0.4$) and spatial distortion (position at 40° azimuth) in the anchor were more pronounced than the distortions applied in the other stimuli of this study.

For the detection experiment, the strength of distortion was adjusted during the experiment according to the listener's response (see Section "Apparatus, procedure, and statistical analysis"), while for the quality rating experiments distortions were applied in two different "effect strengths," denoted as mild and moderate distortions, using the parameters provided in Table 1. For non-linear saturation, Table 1 provides values for the dimensionless parameter α and the maximum total harmonic distortion (THD) for the peak value of the reference signals in percent.

TABLE 1 Experimental parameters (and units) controlling the amount of distortions (columns) in the detection and discrimination experiments and for the quality rating experiments (mild/moderate).

| | Spectral ripples (peak-to-valley ratio in dB) | Non-linear saturation (dimensionless parameter α) | Intensity (ΔdB re reference) | Spatial (Δ azimuth ° re reference) |
|---|---|---|---|---|
| **Detection and discrimination experiments** | | | | |
| Starting value | 18 | 0.62(33.8) | 6.5 | 18 |
| Initial step size | 5 | 0.2 | 2 | 4 |
| Minimum step size | 1.5 | 0.035 | 0.2 | 0.3 |
| **Supra-threshold quality ratings** | | | | |
| Speech | 12/18 | 0.11(15.3)/0.17(21.4) | 1.5/4 | 4°/30° |
| Guitar | 2.5/5 | 0.18(22.1)/0.28(27.1) | 1/4 | 4°/30° |
| Noise | 8/14 | 0.18(22.1)/0.37(29.7) | 1.5/4.5 | 4°/30° |

For a better representation of the amount of non-linear distortion (second column), the THD@peak-values given in percent are provided in parentheses in addition to the dimensionless parameter α.

## Rooms and masker configurations

Three room conditions were realized using headphone auralization and BRIRs generated by the room acoustics simulator (RAZR; Wendt et al., 2014). RAZR calculates early reflections up to the third order using the image source model (Allen and Berkley, 1979), while later reflections were calculated by a feedback delay network (Jot and Chaigne, 1991). An assessment of various common room acoustical parameters and subjective ratings of perceived room acoustical attributes showed a good correspondence between simulated and real rooms (see Wendt et al., 2014; Brinkmann et al., 2019).

An anechoic room served as the reference, only providing the direct sound. A small room with dimensions of 5.28 × 3.5 × 2.5 $m^3$ (length x width x height) and a room volume of 46 $m^3$, was realized with an average reverberation time of T60 of 0.35 s (0.4, 0.37, 0.35, 0.32, and 0.29 s were observed for 0.25, 0.5, 1, 2, and 4 kHz). These parameters were motivated by the average values of reverberation time measurements in furnished living rooms (Díaz and Pedrero, 2005). A large room with dimensions of 7.5 × 4.52 × 3 $m^3$ (~100 $m^3$) was used with an average T60 of 1.5 s (1.53, 1.53, 1.56, 1.44, and 1.45 s at 0.25, 0.5, 1, 2, and 4 kHz). The volume of the large room is similar to the largest furnished living rooms measured by Díaz and Pedrero (2005) which had on average a room volume of about 95 $m^3$ and a T60 of about 0.6 s. The longer T60 of 1.5 s was chosen to better represent environments with pronounced reverberation.

The target and masker sources were convolved with the BRIRs such that they were placed on each of the positions as indicated in Figure 1 for the target and maskers. In each of the three rooms, the receiver and target had identical positions. Different masker configurations were only examined in the small room. Figure 1 illustrates the condition $2M_{sep}$, with two spatially separated maskers at ±45° azimuth from the target position in the small room. In the $1M_{sep}$ condition, only the left spatially separated masker was presented. In the co-located masker configuration, $2M_{co}$, the two maskers were spatially co-located with the target (that always remained in the same position). In the separated conditions, the masker to the left was always the ISTS speech signal, while the masker to the right was always the pop music excerpt. The direct-to-reverberant ratios between target and receiver ($DRR_T$), between left masker and receiver ($DRR_{ML}$), and between right masker and receiver ($DRR_{MR}$) are given in Table 2 for all three rooms.

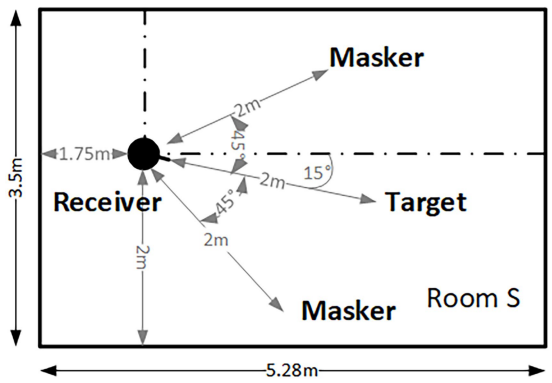



FIGURE 1
Illustration of the target, receiver, and masker positions in the small room. The same target-receiver-masker positions relative to the lower-left corner as the room origin were used in the anechoic and large room. Accordingly, the distance between the left masker and the upper wall increased from 0.5m to 1.5m.

The receiver-target-masker positions were asymmetrically arranged in the room, with a distance of 2 m between the target/maskers and the receiver. All sources and the receiver were positioned at a height of 1.7 m above the floor. Such an asymmetric arrangement in the room is more likely to occur in daily life than an unnatural, completely symmetrical arrangement. The asymmetric arrangement in the room results in small long-term level differences between the ears caused by early reflections, while no such differences are present for the direct sound. The fixed distance of 2 m, independent of the room, was chosen to represent a typical distance between the receiver and the sound-emitting device, e.g., a TV.

## Apparatus, procedure, and statistical analysis

Listeners performed the experiments with dichotically presented stimuli *via* Sennheiser HD 650 headphones, while

TABLE 2 Room acoustical properties of the three different rooms.

| Room | Volume ($m^3$) | T60 (s) | $DRR_{T,0°}$ (dB) | $DRR_{T,4°}$ (dB) | $DRR_{T,30°}$ in (dB) | $DRR_{ML}$ in (dB) | $DRR_{MR}$ in (dB) |
|---|---|---|---|---|---|---|---|
| Anechoic | – | 0 | ∞ | ∞ | ∞ | ∞ | ∞ |
| Small | 46 | 0.35 | −4.7/−4.3 | −5.8/−3.5 | −11/−1.8 | −2.5/−12.8 | −13.9/−1.6 |
| Large | 100 | 1.5 | −6.7/−6.3 | −7.8/−5.5 | −13.5/−3.7 | −3.1/−15.6 | −16.6/−2.8 |

In each room the receiver-target/masker-source distance was 2 m. The third column is the reverberation time T60 in s. All DRRs are provided for the left/right ear. $DRR_{T,0°}$ refers to the target and the co-located maskers placed at 0° azimuth, while $DRR_{T,4°}$ and $DRR_{T,30°}$ refer to the target source positions at 4° and 30° azimuth (spatial distortions). $DRR_{ML}$ and $DRR_{MR}$ refer to the spatially separated masker on the left (ISTS) and right (pop music), respectively. The values of $DRR_{T,0°}$, $DRR_{T,4°}$, and $DRR_{T,30°}$ were calculated from the receiver-target-source BRIRs. $DRR_{ML}$ and $DRR_{MR}$ were calculated from the receiver-left-masker and receiver-right-masker BRIRs.

seated in a double-walled, sound-attenuated booth. The transfer function of the headphones was digitally equalized to obtain a flat frequency response in the artificial ear (B&K Type 4153). The level of the reference and masker signals at 0° in the anechoic condition was 61 dB sound pressure level (SPL). Depending on the reverberation time of the simulated room and the number of maskers, the overall level could reach up to about 78 dB SPL. Subjects responded *via* a touchscreen. All audio files had a sampling rate of 44.1 kHz.

All listeners started with the detection experiment, where only speech signals were used. A three-alternative, forced-choice (3-AFC) procedure was used to determine distortion-detection thresholds. Three intervals were presented, and listeners had to identify the randomly chosen interval containing the distorted speech signal (target). The strength of the distortion was varied according to a 1-up, 2-down procedure for estimating the 70.7% correct point of the psychometric function (Levitt, 1971). To reduce the measurement time, the 2-s speech target was separated into two 1-s-segments that were randomly selected per trial. Stimuli in each trial were separated by 300-ms silent intervals. The initial and minimum step sizes used in the experiments are provided in Table 1. After the minimum step size was reached, six reversals were measured, from which the mean threshold was calculated. The final threshold was the mean of the estimates from two measurement runs. All measurements were performed using the AFC-framework (Ewert, 2013). The detection experiment was divided into two 45-min sessions. The order of presentation of distortions was Latin-Square balanced, while the order of the room conditions Anechoic, Small, Small,$2M_{sep}$, and Large, was randomized. Prior to the actual measurement, a randomly selected room condition was used as training run for each type of distortion.

For the (supra-threshold) audio-quality ratings, distorted speech, guitar music, and noise were used as the target. A measurement procedure applied in previous studies of Fleßner et al. (2017, 2019) was used, similar to the Multiple Stimulus Test with Hidden Reference and Anchor (MUSHRA, ITU-R, 2014). Listeners had to rate quality differences between several distorted targets, also denoted as test signals, and a given (unprocessed) reference target, by using a numerical rating scale ranging from 0 ("very strong difference") to 100 ("no difference"). To ensure that listeners used the full range of the rating scale and to test the reliability of the listeners' ratings, a hidden reference (without any distortions) and a strongly distorted anchor signal were included. The audio signals were played in a loop and the listeners could listen as long as they wished. Listeners could also switch between the different test signals at any time, in which case the audio restarted at the beginning. The quality rating experiment was divided into three sessions: In the first (test) and third (retest) session the *Effect of room* was assessed, and in the second session the *Effect of masker configuration* (test–retest) was assessed. In the *Effect of room* sessions, participants rated audio quality for distorted speech, guitar, and noise targets randomly presented in the Anechoic, Small and the Large room. In the *Effect of masker configuration* session, participants rated distorted speech targets for different configurations of interfering maskers in the Small room. Prior to the actual measurement phase in the first and second session a training run to familiarize the participants with the procedure was performed.

The results of the initial detection experiment were used as the criterion for participation in this study. The mean values of the listener's detection and discrimination thresholds had to be below the values given in Table 1 for the speech target with mild distortions. Five listeners had intensity JNDs slightly above the intended limit of 1.5 dB, but were included given that they clearly fulfilled the entrance criterion for the other three distortions. In total, nine of 25 initially screened listeners did not pass the criterion, resulting in the 16 participants of this study.

For statistical analysis, repeated-measures analysis of variance (ANOVA) was applied using IBM SPSS. Greenhouse–Geisser correction was applied if sphericity was violated. Bonferroni correction was applied in the post-hoc pairwise comparisons. The effect size of contrasts was calculated as $ES = \sqrt{\left[F\left(1,df\right)\right] / \left[F\left(1,df\right)+df\right]}$, where $F$ and $df$ refer to the F-ratio and the residual degrees of freedom, respectively.

# Results

## Detection and discrimination thresholds

In the following, the mean distortion detection thresholds and discrimination JNDs for the speech target based on the average across sixteen listeners are reported.

Figure 2 shows detection thresholds for the four types of distortions as black filled symbols in the different panels. The abscissa represents the four room configurations: Anechoic, Small (mild reverberation), Small,$2M_{sep}$ (mild reverberation plus two spatially separated maskers), and Large (moderate reverberation).

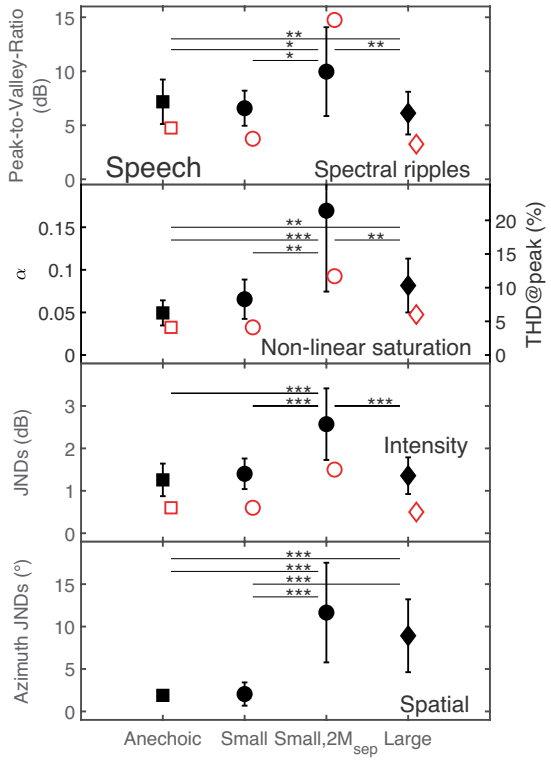



FIGURE 2
The panels show detection thresholds (black closed symbols) for the four types of distortions in the speech target presented without reverberation (Anechoic room), with mild reverberation (Small room), with mild reverberation and two maskers (Small,2M$_{sep}$), and with moderate reverberation (Large room). The rooms are represented by squares, circles, and diamonds, respectively. Here and in the following figures, statistically significant pair-wise comparisons based on levels of 0.05, 0.01, and 0.001 are reported by *, ** and ***, respectively. Red open symbols refer to predicted data. For non-linear saturation, in addition to the dimensionless parameter α, the right y-axis provides the THD@peak in percent.

Detection thresholds for spectral ripples are given as peak-to-valley ratio in dB, for non-linear saturation as the value of the dimensionless non-linearity parameter α (left y-axis), and the THD for the peak value of the reference signal (THD@peak) in percent (right y-axis). Discrimination thresholds for intensity-based distortions are reported as intensity JNDs in dB SPL, and spatial distortions are given as azimuth JNDs in degrees.

As shown in the upper panel of Figure 2, speech signals distorted by spectral ripples had significantly higher detection thresholds in the Anechoic (peak-to-valley ratio of 7.2 dB) than in the Large room (peak-to-valley ratio of 6.1 dB). Conversely, for non-linear saturation, listeners had significantly lower detection thresholds in the Anechoic than in the Large room with moderate reverberation. Intensity JNDs for the room configurations Anechoic, Small and Large ranged between 1.3 and 1.4 dB, while a JND of 2.6 dB was observed for Small,2M$_{sep}$. Post-hoc comparisons showed no significant intensity JND differences between the room configurations Anechoic, Small, and Large. Similar azimuth JNDs of 1.9° and 2° were observed for the Anechoic and the Small room, while significant higher JNDs of about 11.7° and 8.9° were found for the Small,2M$_{sep}$ and the Large room.

A 2-way repeated-measures ANOVA [*distortion* (spectral ripples, non-linear saturation, intensity, spatial), *room* (Anechoic, Small, Small,2M$_{sep}$, Large)] showed a significant main effect of the factors *distortion*, $F(2, 29.8) = 87$, $p < 0.001$, and *room*, $F(1.4, 20.5) = 39$, $p < 0.001$. Moreover a significant two-way interaction between the factors *distortion* and *room*, $F(2.4, 35.3) = 33$, $p < 0.001$ was found. Statistically significant differences (post-hoc test) based on levels of 0.05, 0.01, and 0.001 are indicated in Figure 2 by *, **, and ***, respectively.

In summary, it can be concluded that the presence of maskers had a strong effect ($ES = 0.86$), while mild reverberation alone (Small room) had only a small effect ($ES = 0.13$), suggesting that results in anechoic conditions are transferable to conditions with mild reverberation.

TABLE 3 Individual test–retest PCC for each of the 16 listeners.

| Listener | "Rooms" | "Maskers" | Overall |
|---|---|---|---|
| #1 | 0.92 | 0.99 | 0.95 |
| #2 | 0.94 | 0.93 | 0.93 |
| #3 | 0.88 | 0.97 | 0.91 |
| #4 | 0.9 | 0.95 | 0.92 |
| #5 | 0.99 | 0.99 | 0.99 |
| #6 | 0.91 | 0.93 | 0.92 |
| #7 | 0.88 | 0.93 | 0.9 |
| #8 | 0.92 | 0.93 | 0.92 |
| #9 | 0.89 | 0.95 | 0.9 |
| #10 | 0.89 | 0.97 | 0.91 |
| #11 | 0.9 | 0.93 | 0.91 |
| #12 | 0.93 | 0.91 | 0.93 |
| #13 | 0.92 | 0.97 | 0.93 |
| #14 | 0.96 | 0.96 | 0.96 |
| #15 | 0.89 | 0.93 | 0.9 |
| #16 | 0.9 | 0.9 | 0.9 |

The first column refers to the listener, while the second and third columns refer to PCC scores based on audio quality ratings for the experiments *Effect of room* and *Effect of masker configuration*, respectively. The last column shows the overall PCC for both experiments.

## Supra-threshold quality ratings

Listener's individual scores were averaged across test and retest. The test–retest Pearson-Correlation-Coefficient (PCC) of the data was 0.91 and 0.95 for *Effect of room* and *Effect of masker configuration*, respectively. For more details, test–retest PCCs for each of the 16 listeners are provided in Table 3.

### Effect of room

In Figure 3, the subjective quality scores (averaged across all 16 listeners; error bars indicate one inter-individual standard deviation) for speech (upper panel), guitar (middle panel) and

noise (lower panel) signals impaired by spatial, non-linear, spectral, and intensity distortions are shown for the Anechoic, Small, and Large rooms, indicated by black-filled squares, circles, and diamonds, respectively. The ordinate shows the quality scores, ranging from 0 ("very strong difference") to 100 ("no difference"). The abscissa indicates the hidden reference, anchor, and each of the four distortions having mild and moderate amounts.

A clear difference of about 21 points on the MUSHRA scale between listeners' ratings for mildly and moderately distorted signals can be observed for each of the four distortions. The hidden reference always received the highest rating, while the anchor signal always received the lowest rating, as intended by the experimental design. For the speech target (upper panel), only slight differences in the quality scores between the three rooms were observed. A stronger effect of reverberation was observed for

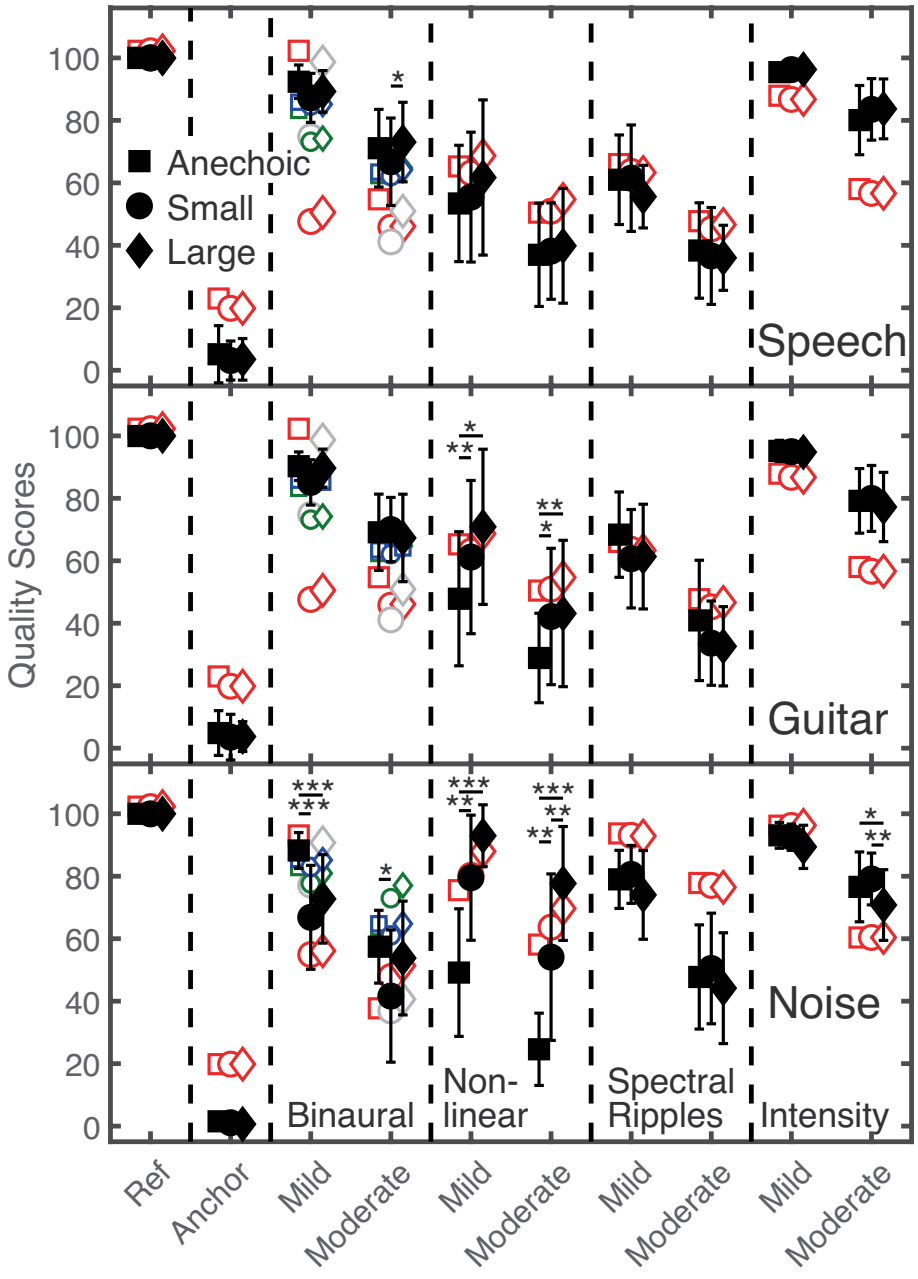



FIGURE 3
The upper, middle, and lower panels show supra-threshold audio quality ratings for speech, guitar music, and noise (black filled symbols). The ordinate represents quality scores ranging from 0 ("very strong difference") to 100 ("no difference"). The abscissa represents the hidden reference, anchor, and type of distortion (mild and moderate amount). The Anechoic, Small, and Large room are represented by squares, circles, and diamonds, respectively. Statistically significant pairwise comparisons between rooms are indicated by the asterisks. The red and small green open symbols refer to GPSM$^q$ and BAM-Q predictions, while gray and small blue symbols refer to GPSM$^q$ and BAM-Q predictions for which only the direct sound and early reflections of the BRIRs were considered.

guitar music and noise (middle and lower panels). Here, reverberation showed a particularly strong impact on spatial and non-linear distortions. For non-linear saturation, quality ratings increased with increasing T60 and decreasing DRR. For spatial distortions, quality ratings were lower for noise presented in the Small room than in the other two rooms. Although counterintuitive, such a behavior was – to some extent - also observed for speech and guitar signals.

A 4-way, repeated-measures ANOVA [*distortions* (spatial, non-linear, spectral, intensity), *room* (Anechoic, Small, Large), *stimuli* (speech, guitar, noise), *effect strength* (mild, moderate)] showed a significant main effect of the factors *distortion*, $F(1.5, 22) = 72$, $p < 0.001$, *room*, $F(2, 30) = 7.5$, $p < 0.01$, and *strength*, $F(1, 15) = 185$, $p < 0.001$, while no significant effect was found for *stimuli*, $F(2,30) = 2.7$, $p = 0.84$.

Focusing on the effect of room in the data, only significant interactions including the factor *room* are reported: There were significant two-way interactions between the factors *room* and *stimuli*, $F(4, 60) = 5$, $p < 0.01$, and between the factors *distortions* and *room*, $F(3.2, 48) = 39$, $p < 0.001$. Moreover, three-way interactions between the factors *stimuli*, *room*, and *distortion*, $F(4.7, 71) = 15.3$, $p < 0.001$ and between the factors *room*, *stimuli*, and *effect strength*, $F(4, 60) = 4.3$, $p < 0.01$ were found.

Taken together, the room had a significant effect on quality ratings, depending on the type of distortion and the stimulus: For speech, only slight differences across the anechoic and the two echoic rooms were observed, in contrast to guitar music and noise. Thus, for the assessment of speech quality, room reverberation only appears to have a small effect. Regarding the type of distortion, quality ratings for non-linear saturation depended most strongly on the amount of reverberation.

## Effect of masker configuration

Figure 4 shows average subjective quality scores and inter-individual standard deviations (black filled circles) for the speech target with spectral ripples (upper-left panel), non-linear saturation (lower-left panel), intensity (upper-right panel), and spatial (lower-right panel) distortions in the Small room as a function of masker configuration.

The hidden reference without maskers (Ref) always obtained the highest ratings, while the hidden reference with two spatially separated maskers (Ref $2M_{sep}$) received about 9 point lower scores. The anchor always received the lowest ratings, and no substantial

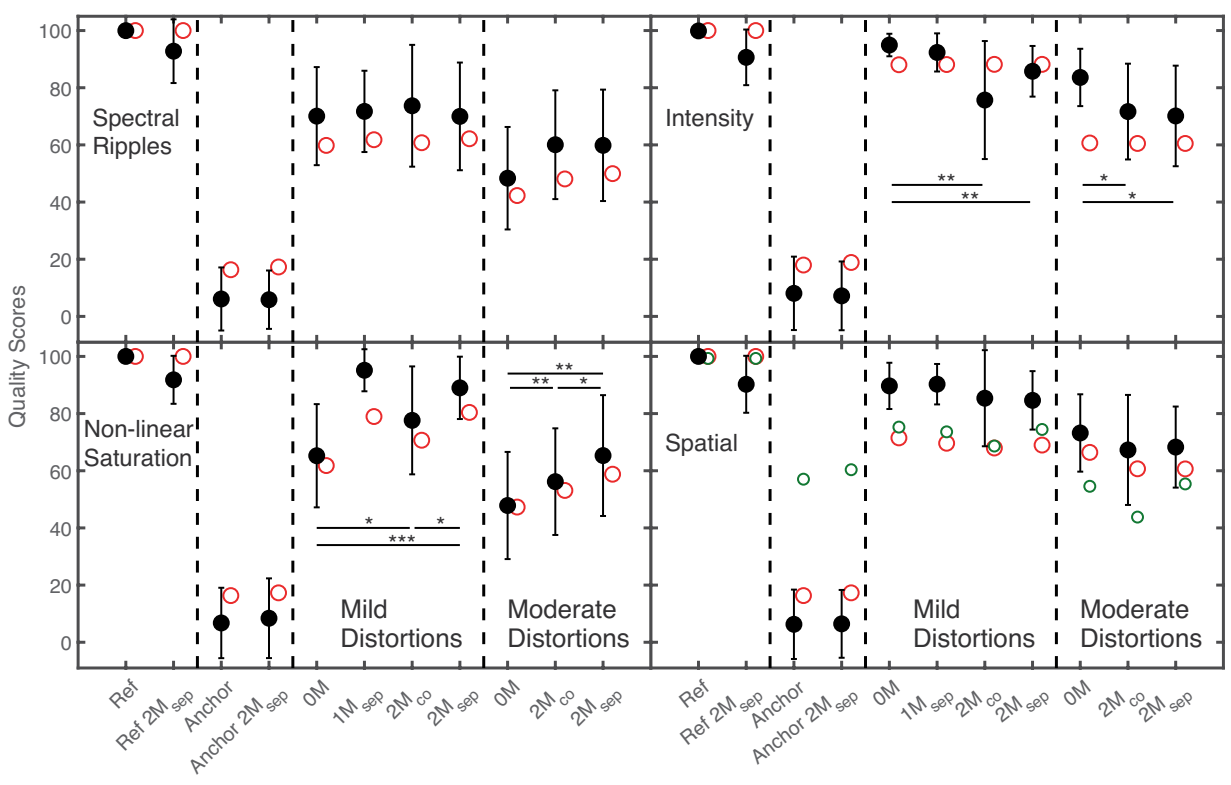



FIGURE 4
Quality ratings (black filled symbols) for the speech target with spectral ripples, intensity distortion, non-linear saturation, and spatial distortions in the Small room as a function of the masker configuration. Ref and Anchor refers to the hidden reference and the anchor signals. 0M indicates no masker. In the configuration $1M_{sep}$, the ISTS masker was presented at −45° azimuth relative to the viewing direction of the listener. In configurations $2M_{co}$, and $2M_{sep}$ the ISTS and music maskers were presented at 0°, and +−45° azimuth. Statistically significant pairwise comparisons involving masker configurations 0M, $2M_{co}$, $2M_{sep}$ are indicated by the asterisks. Red and small green open circles refer to GPSM[q] and BAM-Q predictions.

differences between the anchors with and without maskers were observed.

The speech signal with mild spectral ripples was hardly affected by the presence of maskers, indicated by similar ratings of about 71 points. For moderate spectral ripples, a higher rating was observed for spatially co-located ($2M_{co}$) or spatially separated ($2M_{sep}$) maskers compared to the condition without (0 M) masker (60 vs. 48 points, respectively). However, a pairwise comparison showed no significant difference.

Lower quality scores were found for non-linear saturation without maskers than with maskers. For both mild and moderate distortions, listeners provided lower scores of about 10 points for $2M_{co}$ than for $2M_{sep}$.

For intensity distortions, higher quality scores were obtained without maskers than with maskers: Quality scores for mild intensity distortions were slightly lower (about 10 points) for co-located ($2M_{co}$) than for spatially separated ($2M_{sep}$) maskers, while similar quality scores (about 71 points) were obtained for moderate distortions under these two masker conditions.

The presence of maskers had only a slight effect on the perception of spatial distortions. For both mild and moderate distortions, only small differences of about 5 points between the conditions 0 M and $2M_{co}$, and between the conditions 0 M and $2M_{sep}$ were observed.

A 3-way, repeated-measures ANOVA (*distortion, effect strength, masker*) showed significant main effects of *distortion*, $F(2.2, 32.3) = 14$, $p < 0.001$ and *effect strength*, $F(1, 15) = 74$, $p < 0.001$, while no significant effect was found for *masker*.

Significant two-way interactions between the factors *distortion* and *effect strength*, $F(1.9, 28.4) = 5$, $p < 0.01$, and between the factors *distortion* and *masker*, $F(2.9, 44.1) = 13.9$, $p < 0.001$, were found, together with a three-way interaction between factors *distortion*, *effect strength* and *masker*, $F(6, 90) = 3.5$, $p < 0.01$. Pairwise comparisons (indicated by the asterisks in Figure 4) showed some significant effects of masker for non-linear saturation and intensity distortions. Thus, although no main effect of masker was found, quality ratings for non-linear saturation are more affected by maskers than the other distortions, as also observed for the effect of room.

## Comparison of individual results across conditions and outcome measures

To assess the relation of listener's individual distortion detection and discrimination thresholds, a one-tailed correlation analysis was performed. The upper right side of Table 4 shows significant correlations (indicated as asterisks) between the listeners' thresholds in the room configurations Anechoic (A), Small (S), Small,$2M_{sep}$ (S,$2M_{sep}$), and Large (L) for spectral ripples, non-linear saturation, and intensity distortions. Such a relationship was not observed for spatial distortions. Thus, for (monaural) spectral ripples, non-linear saturation, and intensity distortions, the listeners' performance in anechoic, "classical psychoacoustic test" conditions might be a good indicator for their performance in echoic rooms with mild to moderate reverberation, and for acoustic environments with maskers.

Table 4 further indicates a significant correlation between spectral ripples and non-linear saturation. However, no relationship was found between spectral ripples and intensity distortions nor between spectral ripples and spatial distortions.

For the most complex scene in the detection experiment involving two maskers, Small,$2M_{sep}$, significant correlations are shown in Table 4 for most of the distortions. Here, the presence of the maskers (and the corresponding masking of the target) likely dominates effects, resulting in the significant correlations.

The same correlation analysis as applied to distortion detection thresholds was also applied to the quality ratings, and is shown on the lower left side of Table 4. For clarity, only correlations for mildly distorted signals were reported in Table 4, which are comparable to those from the moderate distortions. For spectral ripples, non-linear saturation, and intensity distortion in different rooms (see Figure 3), a similar correlation pattern as observed for detection thresholds was found for speech and guitar music, but not for noise signals.

Quality ratings (see Figure 4) under conditions without maskers (0 M), with two co-located ($2M_{co}$) and separated maskers ($2M_{sep}$) more often revealed significant correlations between the four types of distortions and distortion strength (mild, moderate) within a certain masker configuration, than between the different masker configurations (not shown). This indicates that for a certain masker configuration (e.g., $2M_{co}$), listeners provided consistent individual ratings across the different types of distortions and distortion strength, but not across different masker configurations. This observation agrees with the significant correlation found between individual detection and discrimination thresholds for each of the four distortions in the condition Small,$2M_{sep}$, and suggests that the perception of distorted signals in CAEs may depend on the individual ability to separate the distorted speech target from the maskers.

A one-tailed correlation analysis was used to examine a potential relationship between the listeners' performance in the detection/discrimination thresholds and the supra-threshold quality ratings. A significant correlation was only found for spectral ripples, as indicated in Figure 5 that shows the individual quality scores as a function of the detection thresholds for the Anechoic (upper panel) and the Small room (lower panel). Quality scores increased with increasing detection thresholds, indicating that (sensitive) listeners provided lower quality scores than listeners with higher detection thresholds.

## Applicability of auditory models to CAEs

The application of (reference-based) auditory models is a common way to assess the contribution of energetic and

TABLE 4 Statistically significant correlations for the detection and discrimination thresholds are represented as black asterisks in the upper right segment.

| Distortions | Room | Spectral ripples | | | | Non-linear | | | | Intensity | | | | Spatial | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | S | $S,2M_{sep}$ | L | A | S | $S,2M_{sep}$ | L | A | S | $S,2M_{sep}$ | L | A | S | $S,2M_{sep}$ | L |
| Spectral ripples | A | – | ** | ** | ** | ** | * | ** | ** | | | | | | | | |
| | S | ** | – | ** | ** | ** | * | ** | ** | | | * | | | | * | |
| | | * | | | | | | | | | | | | | | | |
| | $S,2M_{sep}$ | | | – | ** | ** | | ** | ** | | | * | | | | * | |
| | L | ** | | | – | ** | * | ** | ** | | | * | | | | | |
| | | * | | | | | | | | | | | | | | | |
| Non-linear | A | ** | | | * | – | | ** | * | | | * | | | | * | |
| | | ** | | | * | | | | | | | | | | | | |
| | S | * | | | * | ** | – | | * | ** | * | ** | ** | | | | |
| | | ** | | | * | ** | | | | | | | | | | | |
| | | | | | * | * | | | | | | | | | | | |
| | $S,2M_{sep}$ | | | | | | | – | ** | | | * | | | | * | |
| | L | * | ** | | * | ** | ** | | – | | | ** | | | | * | |
| | | ** | | | * | ** | ** | | | | | | | | | | |
| Intensity | A | | | | | | | | | – | ** | ** | ** | | | | ** |
| | S | * | * | | | | | | | * | – | * | ** | | | | |
| | | | * | | | | | | | ** | | | | | | | |
| | $S,2M_{sep}$ | | | | | | | | | | | – | ** | | | ** | ** |
| | L | | * | | | | | | | * | ** | | – | * | | | |
| | | | | | | | | | | ** | ** | | | | | | |
| | | | | | | | | | | * | ** | | | | | | |
| Spatial | A | ** | | | | * | | | | | * | | | – | | | |
| | | ** | | | | | | | | | | | | | | | |
| | S | | | | | * | * | | * | | | | | | – | | |
| | | | | | | * | | | | | | | | | | | |
| | $S,2M_{sep}$ | | | | | | | | | | | | | | | – | ** |
| | L | * | ** | | * | ** | * | | | | * | | | ** | ** | | – |
| | | ** | ** | | ** | | | | | | * | | | | | | |

Significant correlations for the quality ratings for mildly distorted speech, guitar, and noise signals are shown as red, blue, and green asterisks in the lower left segment. Significance levels of 0.05, and 0.01 are reported by *, and **, respectively. Distortions and room configurations are given in the headers of the rows and columns. The abbreviations A, S, and L refer to the Anechoic, Small, and Large room, while $S,2M_{sep}$ refers to the Small room with two spatially separated maskers.

amplitude modulation masking. Here, the GPSM (Biberger and Ewert, 2017), which has been shown to account for several psychoacoustic detection and masking experiments with less complex stimuli (e.g., pure tones and noise), was used as monaural auditory model for predicting data from the detection and discrimination experiments. The monaural audio quality model GPSM$^q$ (Biberger et al., 2018), which previously successfully predicted subjective quality ratings for different types of distortions and stimuli (see also, Fleßner et al., 2019; Biberger et al., 2021), was applied to the quality ratings. Both models are based on short-term power and envelope power SNRs. Additionally, the binaural auditory model for audio quality (BAM-Q; Fleßner et al., 2017), based on the binaural psychoacoustic model front end of Dietz et al. (2011), was applied for the spatial distortion. BAM-Q predictions are based on the combination of the sub-measures ILDs, ITDs, and interaural vector strength. The same AFC-framework (Ewert, 2013) and the stimuli as used for the detection and discrimination measurements were also used for GPSM simulations. For audio quality predictions, the same sound files presented to the listeners during the audio quality rating experiments were provided to the quality models, whereas the left and right ear channels were concatenated to a one-dimensional vector when the monaural GPSM$^q$ was applied. For quality predictions with maskers, an additional preprocessing step was introduced that removed signal parts with a signal-to-noise ratio (SNR) below −10 dB. The preprocessing had perfect *a-priori* knowledge about the target and the masker signals, similar to the assumption in models, to form an ideal binary mask (IBM; e.g., Wang, 2005; Brungart et al., 2006) to examine the consequences of energetic masking on speech intelligibility. A linear transformation was applied, to map the predicted quality scores onto the same scale, ranging from 0 ("very strong difference") to 100 ("no difference"), as used for listener ratings. In Figures 2–4, predictions of GPSM and GPSM$^q$

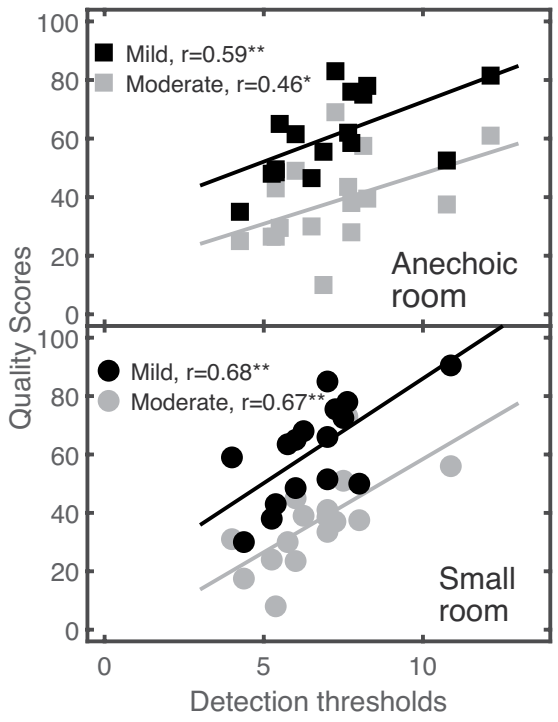



FIGURE 5
Relationship between listeners' quality scores (ordinate) and detection thresholds (abscissa) for linear distortions in the speech target in the Anechoic and Small rooms (upper and lower panel). Mild and moderate distortions are represented by black and gray symbols, respectively. Linear regression models fitted to data are shown as solid lines. Statistically significant effects based on levels of 0.05, and 0.01 are indicated as * and **.

are shown as red open symbols, whereas BAM-Q predictions are shown as small green open symbols.

## Detection and discrimination thresholds

The GPSM captures only monaural cues, and therefore no predictions are shown for the spatial distortions. For spectral ripples, lower thresholds were predicted for the Small and the Large room than for the Anechoic room, similar to measured thresholds. For non-linear saturation, higher thresholds were predicted for the Large room than for the Anechoic and Small room, also in agreement with the measured data. Predicted intensity JNDs showed no systematic differences between the Anechoic, Small and Large rooms. With the exception of the predicted threshold for speech with spectral ripples in small,$2M_{sep}$, model predictions consistently showed lower thresholds and smaller JNDs than measured data. Despite this, generally the higher sensitivity, GPSM-based predictions captured most of the room- and masker related consequences on thresholds and JNDs for those distortions. Accordingly, it is expected that the GPSM is generally applicable for the prediction of distorted speech signals in CAEs, while the higher sensitivity hints in the direction of (higher-level) cognitive effects not covered by the modeled energetic- and amplitude-modulation masking.

## Supra-threshold quality ratings

Regarding the effect of room, GPSM$^q$ scores for speech, guitar and noise signals with mild distortions were always higher than with moderate distortions (see Figure 3). Monaural GPSM$^q$ quality predictions for binaural distortions in the speech and guitar signals largely agree with listener's quality ratings in the anechoic condition, but not with those in the echoic Small and Large rooms. In these monaural predictions only spectral differences were taken into account: The change of the target source position of 4° azimuth relative to the reference position (0° azimuth) resulted in large spectral differences for the echoic conditions in GPSM$^q$, related to differences in the sound-reflection patterns and late reverberation for the two target source positions. The similarity of the listeners' quality scores under those conditions suggests that only differences in the (unaltered) direct sound were considered by the listeners, ignoring the effects of reverberation. This was tested by considering only the direct sound (and early reflections) of the BRIRs in the model, which is conceptually similar to the approaches used by, e.g., Rennies et al. (2014) and Leclere et al. (2015) to simulate the effect of reverberation on (binaural) speech intelligibility by separating the early (useful) from the late (detrimental) room reflections. Gray open symbols in Figure 3 represent GPSM$^q$ predictions using a 5-ms window, starting with the direct sound. This modification clearly improved prediction accuracy for the mild spatial distortions and did not degrade prediction accuracy for moderate spatial distortions. For moderate spatial distortions, it can be expected that monaural spectral differences would have had only a minor effect on listeners' quality ratings. Binaural predictions were also improved when using the same 5-ms window (small blue symbols), less, however, than for the monaural predictions.

For non-linear distortion, there are more pronounced differences in the listeners' quality ratings between room conditions than observed in predicted scores, particularly for guitar and noise signals. GPSM$^q$ mainly predicts higher scores for speech in the Large room than in the other two rooms. Measured and predicted quality scores of speech, guitar and noise signals with spectral ripples showed no substantial effect of room reverberation. Similarly, no room dependence was observed in the measured and predicted quality scores for intensity distortions in speech and guitar signals. A room effect was only observed for noise signals with moderate intensity distortions, where listeners provided significantly higher ratings for the Anechoic and Small rooms, than for the Large room. Such differences were not observed by GPSM$^q$ prediction. Given that the intensity distortions cause loudness differences between the reference and the distorted signal, loudness models (e.g., Chalupper and Fastl, 2002; Pieper et al., 2018) may account for the observed differences. However, loudness predictions (not shown) of the dynamic loudness model

(DLM; Chalupper and Fastl, 2002) provided a similar loudness ratio between reference and test signal of about 1.3 in the three rooms used, suggesting similar perceived loudness differences. Thus, neither the GPSM$^q$ nor the DLM predicted the observed effect of room for moderate intensity distortions with noise. Despite such deviations, GPSM$^q$ achieved an overall good prediction performance for audio quality of distorted signals in rooms with different reverberation times, indicated by a Pearson correlation coefficient of 0.8 and a Spearman rank correlation coefficient of 0.75.

For the different masker configurations (speech target), predicted quality scores of GPSM$^q$ are shown in Figure 4. The preprocessing only kept unmasked and thus reliable time segments of the target. The predicted quality scores for anchor signals and spectral ripples, non-linear saturation, and spatial distortion agree well with the data. The preprocessing assumes that masked segments of the distorted target do not affect the listeners' quality ratings of the entire distorted signal. Given the accurate predictions, this assumption appears to be valid for anchor signals and spectral ripples, non-linear saturation, and spatial distortion, suggesting a certain degree of invariance of the perceptual quality attributes of the target auditory object in the presence of maskers. For intensity distortions, lower quality scores were obtained with maskers than without maskers by the listeners, whereas similar scores were predicted with and without maskers. For intensity distortion, listeners' quality ratings are likely based on a comparison between the target loudness of the reference and the test signal. Hypothetically, the reduced number of spectro-temporal segments (or observations) of the target available to the auditory system in the presence of the masker decreases the perceived target loudness. A comparable effect of maskers on the target loudness was also observed by Fichna et al. (2021), where the loudness of the target speaker decreased with an increasing number of maskers. Consequently, target loudness (as the presumably underlying quality attribute for the intensity distortions) is invariant in the presence of other interfering auditory objects and masking effects have to be taken into account (see upper right panel of Figure 4). Overall, GPSM$^q$ predictions agreed well with subjective quality ratings for distorted signals in the presence of maskers, indicated by a Pearson correlation coefficient of 0.87 and a Spearman rank correlation of 0.89. Binaural BAM-Q predictions for the spatially distortions show a similar pattern as observed in the measurements. Surprisingly, BAM-Q predicted higher quality scores for the anchor signal (target position at 40°) than for the speech target with moderate spatial distortions (target position at 30°). While BAM-Q observed larger ITD differences for the 40° target position than for 30°, lower ILD differences were observed, with no substantial differences for IVS. The final quality measure provided by BAM-Q was obtained by combining ILD, ITD, and IVS differences, with ILDs receiving the strongest weighting (see Section "Applicability of auditory models to CAEs" in Fleßner et al., 2017), thus explaining the surprisingly high quality ratings for the anchor signal.

# Discussion

## Detection and discrimination thresholds

No statistically significant differences between the Anechoic and the Small room occurred for the detection thresholds of the four distortions, suggesting that anechoic thresholds are also representative for rooms with mild reverberation (T60: 0.35 s), as typically encountered in home environments. Conversely, with the exception of intensity distortion, significant threshold differences were found between the Large room with moderate reverberation (T60: 1.5 s) and the other two rooms. The absence of any effect of intensity distortion can be expected, given that neither spectral, nor amplitude modulation, nor spatial changes were introduced. All room acoustic features, such as the pattern of early reflections and the DRR, were invariant to the level changes introduced in the intensity distortion.

For spectral ripples, one reason for lower thresholds in the Large room might be an improved audibility of spectral ripples in certain frequency regions because of the room's modal structure. Similarly, Toole and Olive (1988) observed a better detectability of signal resonances in echoic than in anechoic rooms, which was presumably a result of an improved audibility of such resonances. According to the representation of power-based SNRs in the auditory model GPSM, the most dominant spectral differences between the Anechoic and the Large room occurred below 800 Hz.

Non-linear saturation resulted in additional frequency components at higher signal levels, which likely provided spectral or amplitude modulation cues to the listeners. A comparison of the power- and envelope-power SNR representation (across auditory and modulation filters) showed increased energy between 2 kHz and 3.15 kHz for non-linear saturation under anechoic conditions. Particularly large differences were observed in high modulation filters (above 64 Hz) at signal onsets. Such differences were substantially reduced with moderate reverberation in the Large room.

Substantially increased position JNDs for the target in the Large room suggest a degradation of binaural cues in the signal onsets. For sound localization, e.g., Wallach et al. (1949), Blauert (1971) have shown that the direction of the sound that arrives at the ears first dominates perception compared to later-arriving reflections from other directions. Accordingly, signal onsets are important for sound localization in real rooms (e.g., Stecker and Moore, 2018), as the onsets may be less impaired by overlap masking. To interpret the current results, a binaural auditory model (Dietz et al., 2011) was also applied here (not shown). Only the direct sound and early reflections up to 50 ms after the direct sound were analyzed, reflecting a simplistic simulation of the precedence effect motivated by, e.g., Haas (1972) and Lochner and Burger (1964). Consistently pointing ILDs (> 1,500 Hz) and ITDs (< 1,500 Hz) were found for the Anechoic and Small rooms, but more strongly fluctuating ITDs were found for the Large room. Only slight differences in ILDs (> 1,500 Hz) were observed

between the Small and the Large rooms, suggesting that ITDs served as a main cue under the current room conditions.

Maskers (here the interfering ISTS speech signal and pop music) caused a substantial increase of detection and discrimination thresholds for all four distortions. As supported by the model simulations, this is a direct consequence of the reduced amount of distorted spectro-temporal speech segments available to the listeners, hampering the detection of distortion effects in the target. Thus, particularly for CAEs with mild reverberation, the effect of masking caused by interfering sounds is most relevant.

The correlation analysis for listener's individual thresholds (see Section "Comparison of individual results across conditions and outcome measures") indicated that well-performing listeners, who obtained low detection thresholds for linear, non-linear saturation and intensity distortions in the Anechoic room, mostly remained good performers in the echoic rooms with and without maskers. Conversely, this was not observed for spatially distorted speech. Overall, findings of the correlation analysis suggest that for spectral ripples, non-linear saturation, and intensity distortion, the individual listener's performance under anechoic conditions might be a good indicator for their performance in CAEs with mild- to moderate reverberation and maskers, but not for spatial distortion.

## Supra-threshold quality ratings – Effect of room

Overall, the supra-threshold perception of distortions was affected by reverberation, as supported by the significant main effect of *room* on quality ratings. However, the effect depended on the stimulus and the type of distortion, as indicated by the significant interactions reported in Section "Effect of room":

Quality ratings for spectral ripples were hardly affected by reverberation. Although no significant effects were found, the trend that for all three stimuli, spectral ripples were rated higher in the Anechoic room than in the Small and Large rooms, agrees with the effects found in the detection experiment.

For signals distorted by non-linear saturation, higher quality ratings were obtained in reverberation than in the Anechoic room. Here, as observed for detection thresholds, reverberation is expected to mask distorted parts of the signals. As shown in Figure 3 and indicated by the interaction between factors *stimuli* and *room*, non-linear saturation in guitar music and noise was more effectively masked by reverberation than in the speech signal. This is presumably based on differences in the signal properties of the fluctuating speech, and guitar signals and the stationary noise signal: Non-linear saturation mainly affects signal peaks in fluctuating targets, which provide high SNRs in reverberation, while harmonic distortions in noise mainly result in perceivable spectral coloration changes.

For intensity distortion, room reverberation had no effect on the listeners' quality ratings, except for moderate intensity distortion in noise. Here, the lower quality ratings in the Large room compared to the Anechoic and Small rooms, imply larger perceived differences. The dominating supra-threshold cue associated with intensity is loudness. Accordingly, a loudness model (Chalupper and Fastl, 2002) was applied in Section "Applicability of auditory models to CAEs," but did not explain the lower quality ratings in the Large room for that specific condition (see lower panel in Figure 3), but agreed with the other quality ratings for intensity distortions. Overall, for intensity distortions it can be summarized that reverberation had no, or only a minor, effect on quality ratings as already observed for intensity JNDs (see Figure 2).

For spatially distorted noise signals, lower quality ratings were obtained in the Small room than in the Anechoic and Large rooms. This appears counterintuitive, given that a smaller effect of reverberation would be assumed for the Small than for the Large room. Here, listeners may have rated spectral differences instead of spatial differences: A comparison of the (third-octave-smoothed) frequency spectra of the noise target at 0° and 4° in the Small room shows level differences between frequencies of 850 Hz to 1,440 Hz of up to 3 dB, while only slight level differences were observed in the Large room. Therefore, spectral as well as binaural cues appear relevant for the perception of spatial distortions, depending on the specific echoic environment.

A central question of this study was if listeners who showed lower detection thresholds than other listeners were also more sensitive in the quality ratings than the others. Such relationship would allow making individual adjustments in, e.g., hearing devices, purely based on distortion detection thresholds. The correlation analysis revealed a significant correlation between detection thresholds and quality ratings for speech signals with spectral ripples. Therefore, information about the listener's threshold for spectral ripples might be sufficient for an individualized adjustment of hearing devices concerning spectral ripples when focusing on speech quality.

## Supra-threshold quality ratings – Effect of masker configuration

Based on the models applied in this study and the concept of energetic masking, it is expected that listeners base their quality judgements on reliable (unmasked) spectro-temporal segments of the distorted target in the presence of fluctuating maskers. For equally distributed distortions over time, it thus appears plausible to expect only slight differences between quality ratings with and without maskers, given that the effect of distortion is observable in the unmasked spectro-temporal segments. For non-equally distributed distortions, differences can be expected when, e.g., more-strongly distorted segments are masked, while mildly distorted segments are not masked. Such a behavior was observed for (moderate) spectral ripples and non-linear saturation where listeners rated quality higher for the masked than for the unmasked distorted target (speech) as shown in Figure 4. Here, saturation distortions considered in this study were unequally distributed over time, as they only occurred at higher signal levels.

Although the spectral ripples applied in this study are in principle equally distributed over time, the spectral composition of the target changed over time, and thus provided spectro-temporal segments where the distortions were easier to detect than in other segments. This interpretation agrees well with the quality predictions shown in Figure 4, where audio quality was estimated using only reliable and unmasked segments of the distorted target (with an SNR >= −10 dB).

For intensity or spatial distortions, the presence of maskers lowered the perceived quality. Intensity distortions were introduced by decreasing the overall level of the target. Therefore, in the quality ratings for intensity distortions, listeners likely rated loudness in comparison to the reference. Accordingly, the observed lower scores in the presence of maskers might reflect a lower perceived loudness of the target, as parts were masked and not accessible to the listeners. A similar observation was made in Fichna et al. (2021), where the loudness of a target speaker was decreased as the number of the maskers was increased. A masker-induced loudness reduction was also observed in the data of a "classical" loudness experiment presented in Figures 8–10 in Fastl and Zwicker (2006) where the loudness of a 1-kHz tone was reduced as a stationary pink-noise masker was added to the tone.

As for intensity distortions, a slight tendency for lower quality ratings in the presence of maskers was also observed for spatial distortions. Surprisingly, no difference between quality ratings was observed for co-located and spatially separated maskers. Here, it might be expected that the target at 4° azimuth (mild spatial distortion) was more efficiently masked by co-located maskers (at 0°) than by separated maskers (at ±45°), while the moderate spatial distortion at 30° azimuth was more efficiently masked by the separated maskers than by co-located maskers. However, the diversity of the ISTS and pop-music maskers and the speech target may have facilitated segregation and direction estimation of these perceptually very different sound sources.

Another interesting effect was observed for the quality ratings assigned to the reference with and without maskers. On average, listeners rated the reference with maskers ($2M_{sep}$) 9 points lower than without maskers. Here, the maskers likely introduced an uncertainty about the reference and affected the overall rating. Only one listener ignored the maskers and provided a rating of 100 points. This uncertainty effect is an important finding for reference-based audio quality predictions, as quality differences between the reference with and without maskers cannot simply be predicted by only taking unmasked spectro-temporal segments of the reference signal into account (which would not predict any quality difference). Accordingly, for audio quality models, an uncertainty has to be considered, which may depend on the spatial position of the masker, the number and the type of maskers.

## Implications for auditory models

Detection and discrimination thresholds were more accurately predicted than quality ratings, showing that basic sensory cues are reasonably well represented in the model's auditory preprocessing. As shown in Figure 2, GPSM consistently predicted lower thresholds and JNDs than observed in the data. Such higher sensitivity of the model compared to the listeners could be reduced by introducing additional internal noise as suggested in earlier studies (Dau et al., 1997; Wallaert et al., 2017; Ewert et al., 2020) to represent further cognitive effects, which might be related to segregation of the target from the scene.

While GPSM$^q$ captured most of the effects of reverberation on quality ratings, it strongly overestimated the spectral differences related to differences in the sound reflection patterns between target positions of 0° and 4°. GPSM$^q$ predictions can be improved when only the direct sound and very early reflections of up to 5 ms are analyzed, both considered as "useful," whereas late room reflections are considered as masker ("detrimental"). The same 5-ms temporal window also improved binaural quality predictions of BAM-Q for spatially distorted speech signals in echoic rooms. The underlying cognitive effects of separating and segregating direct sound and (typically correlated) early sound reflections, from typically uncorrelated late reverberation, representing a background "masker," have to be considered for future modeling.

For quality predictions in the presence of maskers, a preprocessing was applied to the waveform of the signals, removing "unreliable" temporal segments with an SNR below −10 dB. In contrast to the data, without such a preprocessing, GPSM$^q$ would predict higher quality scores for conditions with maskers, because the model would observe reduced differences between the test and reference signal for temporal segments dominated by the masker. As shown in Figure 4, quality predictions of GPSM$^q$ with preprocessing for spectral ripples, non-linear saturation, and spatial distortion agreed well with data; they did not, however, capture the effect of maskers for intensity distortions. Here, instrumental measures would have to predict an apparent lower overall target loudness for acoustic environments with maskers than without maskers.

## Summary and conclusion

Detection thresholds and supra-threshold audio quality ratings of spectral ripples, non-linear saturation, intensity, and spatial distortions of a target in complex acoustic environments was investigated. The complexity of the environments was changed by varying the number of maskers and the amount of reverberation. Speech served as the main target in all conditions, while the effect of reverberation was additionally examined for a guitar and pink-noise target. The following conclusions can be drawn:

- Detection thresholds for distorted speech targets in anechoic and mild reverberation showed no significant differences, suggesting that findings in anechoic conditions are transferable to conditions with mild reverberation. Conversely, a significant effect of moderate reverberation on detection thresholds for spectral ripples, non-linear saturation, and spatial distortion was found, indicating the

relevance of additional measurements with moderate reverberation when assessing performance in CAEs.

- Reverberation showed only a small effect on quality ratings for distorted speech, but had a stronger effect on guitar and noise signals. This effect is presumably based on differences in the signal properties of the fluctuating speech, guitar music and the stationary noise, that changes the sound character of the distortions.
- Increased detection thresholds for distorted speech in the presence of two maskers were measured compared to the situations without masker. The effect of maskers on quality depended on the type of distortions. In connection with the model analysis, it appears that quality ratings were based on unmasked temporal speech segments.
- A significant correlation between listeners' individual detection thresholds and their quality ratings for spectral ripples in speech targets was found. Sensitive listeners with low detection thresholds also provided lower quality scores than listeners with higher detection thresholds.
- The GPSM (Biberger and Ewert, 2017) and the GPSM$^q$ (Biberger et al., 2018), captured the main effects of CAEs on detection thresholds and quality ratings in different room- and masker configurations, indicated by Pearson linear-correlation coefficient values of 0.8 and 0.87, respectively. For accurate quality predictions in the presence of maskers, a preprocessing that only provided "reliable" speech segments to GPSM$^q$ was required.

## Data availability statement

The raw data supporting the conclusions of this article will be made available (www.faame4u.com) by the authors without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the University of Oldenburg. The participants provided their written informed consent to participate in this study.

## Author contributions

TB and SE co-conceived the ideas. SE supervised the project. TB carried out the experiments and model simulations. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG – 352015383 – SFB1330 A2 and DFG –390895286 – EXC 2177/1).

## Acknowledgments

The authors would like to thank the members of Medizinische Physik and Birger Kollmeier for continued support. English-language editing was provided by www.stels-ol.de.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Allen, J. B., and Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *J. Acoust. Soc. Am.* 65, 943–950. doi: 10.1121/1.382599

Beerends, J. G., Hekstra, A. P., Rix, A. W., and Hollier, M. P. (2002). Perceptual evaluation of speech quality (PESQ) the new ITU standard for end-to-end speech quality assessment part ii: psychoacoustic model. *J. Audio Eng. Soc.* 50, 765–778.

Best, V., Mason, C. R., Kidd, G. Jr., Iyer, N., and Brungart, D. S. (2015). Better-ear glimpsing in hearing-impaired listeners. *J. Acoust. Soc. Am.* 137, EL213–EL219. doi: 10.1121/1.4907737

Biberger, T., and Ewert, S. D. (2016). Envelope and intensity based prediction of psychoacoustic masking and speech intelligibility. *J. Acoust. Soc. Am.* 140, 1023–1038. doi: 10.1121/1.4960574

Biberger, T., and Ewert, S. D. (2017). The role of short-time intensity and envelope power for speech intelligibility and psychoacoustic masking. *J. Acoust. Soc. Am.* 142, 1098–1111. doi: 10.1121/1.4999059

Biberger, T., and Ewert, S. D. (2019). The effect of room acoustical parameters on speech reception thresholds and spatial release from masking. *J. Acoust. Soc. Am.* 146, 2188–2200. doi: 10.1121/1.5126694

Biberger, T., Fleßner, J. H., Huber, R., and Ewert, S. D. (2018). An objective audio quality measure based on power and envelope power cues. *J. Audio Eng. Soc.* 66, 578–593. doi: 10.17743/jaes.2018.0031

Biberger, T., Schepker, H., Denk, F., and Ewert, S. D. (2021). Instrumental quality predictions and analysis of auditory cues for algorithms in modern headphone technology. *Trend. Hear.* 25, 1–22. doi: 10.1177/23312165211001219

Blauert, J. (1971). Localization and the law of the first wavefront in the median plane. *J. Acoust. Soc. Am.* 50, 466–470. doi: 10.1121/1.1912663

Bolt, R. H., and MacDonald, A. D. (1949). Theory of speech masking by reverberation. *J. Acoust. Soc. Am.* 21, 577–580. doi: 10.1121/1.1906551

Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., and Weinzierl, S. (2019). A round robin on room acoustical simulation and auralization. *J. Acoust. Soc. Am.* 145, 2746–2760. doi: 10.1121/1.5096178

Brungart, D. S., Chang, P. S., Simpson, B. D., and Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *J. Acoust. Soc. Am.* 120, 4007–4018. doi: 10.1121/1.2363929

Cauchi, B., Siedenburg, K., Santos, J. F., Falk, T. H., Doclo, S., and Goetze, S. (2019). Non-intrusive speech quality prediction using modulation energies and LSTM-network. *IEEE Trans. Audio Speech Lang. Proces.* 27, 1151–1163. doi: 10.1109/TASLP.2019.2912123

Chalupper, J., and Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acust. Unit. Acust.* 88, 378–386.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Dau, T., Kollmeier, B., and Kohlrausch, A. (1997). Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *J. Acoust. Soc. Am.* 102, 2892–2905. doi: 10.1121/1.420344

Díaz, C., and Pedrero, A. (2005). The reverberation time of furnished rooms in dwellings. *Appl. Acoust.* 66, 945–956. doi: 10.1016/j.apacoust.2004.12.002

Dietz, M., Ewert, S. D., and Hohmann, V. (2011). Auditory model based direction estimation of concurrent speakers from binaural signals. *Speech Comm.* 53, 592–605. doi: 10.1016/j.specom.2010.05.006

Ewert, S. D. (2013). "AFC—A modular framework for running psychoacoustic experiments and computational perception models," in *Proceedings of the international conference on acoustics AIA-DAGA*. 1326–1329.

Ewert, S. D. (2020). "Defining the proper stimulus and its ecology-mammals," in *The senses: A comprehensive reference*. ed. B. Fritsch (Amsterdam, Netherlands: Elsevier).

Ewert, S. D., Paraouty, N., and Lorenzi, C. (2020). A two-path model of auditory modulation detection using temporal fine structure and envelope cues. *Eur. J. Neurosci.* 51, 1265–1278. doi: 10.1111/ejn.13846

Ewert, S. D., Schubotz, W., Brand, T., and Kollmeier, B. (2017). Binaural masking release in symmetric listening conditions with spectro-temporally modulated maskers. *J. Acoust. Soc. Am.* 142, 12–28. doi: 10.1121/1.4990019

Fastl, H., and Zwicker, E. (2006). *Psychoacoustics: Facts and models*. Springer Science & Business Media, Berlin Heidelberg New York.

Fichna, S., Biberger, T., Seeber, B. U., and Ewert, S. D. (2021). "Effect of acoustic scene complexity and visual scene representation on auditory perception in virtual audio-visual environments," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 1–9.

Fleßner, J. H., Biberger, T., and Ewert, S. D. (2019). Subjective and objective assessment of monaural and binaural aspects of audio quality. *IEEE Trans. Audio Speech Lang. Proces.* 27, 1112–1125. doi: 10.1109/TASLP.2019.2904850

Fleßner, J. H., Huber, R., and Ewert, S. D. (2017). Assessment and prediction of binaural aspects of audio quality. *J. Audio Eng. Soc.* 65, 929–942. doi: 10.17743/jaes.2017.0037

Haas, H. (1972). The influence of a single echo on the audibility of speech. *J. Audio Eng. Soc.* 20, 146–159.

Harlander, N., Huber, R., and Ewert, S. D. (2014). Sound quality assessment using auditory models. *J. Audio Eng. Soc.* 62, 324–336. doi: 10.17743/jaes.2014.0020

Holube, I., Fredelake, S., Vlaming, M., and Kollmeier, B. (2010). Development and analysis of an international speech test signal (ISTS). *Int. J. Audiol.* 49, 891–903. doi: 10.3109/14992027.2010.506889

Houtgast, T., Steeneken, H. J. M., and Plomp, R. (1980). Predicting speech intelligibility in rooms from the modulation transfer function I. General room acoustics. *Acta Acust.* 46, 60–72.

Huber, R., and Kollmeier, B. (2006). PEMO-Q—A new method for objective audio quality assessment using a model of auditory perception. *IEEE Trans. Audio Speech Lang. Process.* 14, 1902–1911. doi: 10.1109/TASL.2006.883259

ITU-R (2014). "Method for the subjective assessment of intermediate quality levels of coding systems," in *Series BS: Broadcast Services Recommendation BS*. 1534. International Telecommunications Union, Geneva, Switzerland.

Jot, J.-M., and Chaigne, A. (1991). "Digital delay networks for designing artificial reverberators," in *Proceedings of the 90th Convention on the Audio Engineering Society*, October 4–8, Paris, France.

Kates, J. M., and Arehart, K. H. (2010). The hearing-aid speech quality index (HASQI). *J. Audio Eng. Soc.* 58, 363–381.

Kopčo, N., and Shinn-Cunningham, B. G. (2003). Spatial unmasking of nearby pure-tone targets in a simulated anechoic environment. *J. Acoust. Soc. Am.* 114, 2856–2870. doi: 10.1121/1.1616577

Lavandier, M., and Culling, J. F. (2010). Prediction of binaural speech intelligibility against noise in rooms. *J. Acoust. Soc. Am.* 127, 387–399. doi: 10.1121/1.3268612

Leclere, T., Lavandier, M., and Culling, J. F. (2015). Speech intelligibility prediction in reverberation: towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation. *J. Acoust. Soc. Am.* 137, 3335–3345. doi: 10.1121/1.4921028

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49:467. doi: 10.1121/1.1912375

Lochner, J., and Burger, J. (1964). The influence of reflections on auditorium acoustics. *J. Sound Vib.* 1, 426–454. doi: 10.1016/0022-460X(64)90057-4

Moore, B. C., and Tan, C.-T. (2003). Perceived naturalness of spectrally distorted speech and music. *J. Acoust. Soc. Am.* 114, 408–419. doi: 10.1121/1.1577552

Moore, B. C., and Tan, C.-T. (2004). Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion. *J. Audio Eng. Soc.* 52, 900–914.

Pieper, I., Mauermann, M., Oetting, D., Kollmeier, B., and Ewert, S. D. (2018). Physiologically motivated individual loudness model for normal hearing and hearing impaired listeners. *J. Acoust. Soc. Am.* 144, 917–930. doi: 10.1121/1.5050518

Plomp, R. (1976). Binaural and monaural speech intelligibility of connected discourse in reverberation as a function of azimuth of a single competing sound source (speech or noise). *Acta Acust. Unit. Acust.* 34, 200–211.

Rakerd, B., and Hartmann, W. (1985). Localization of sound in rooms, II: the effects of a single reflecting surface. *J. Acoust. Soc. Am.* 78, 524–533. doi: 10.1121/1.392474

Rakerd, B., and Hartmann, W. M. (1986). Localization of sound in rooms, III: onset and duration effects. *J. Acoust. Soc. Am.* 80, 1695–1706. doi: 10.1121/1.394282

Rennies, J., Warzybok, A., Brand, T., and Kollmeier, B. (2014). Modeling the effects of a single reflection on binaural speech intelligibility. *J. Acoust. Soc. Am.* 135, 1556–1567. doi: 10.1121/1.4863197

Schepker, H., Denk, F., Kollmeier, B., and Doclo, S. (2019). "Subjective sound quality evaluation of an acoustically transparent hearing device," in: *Audio Engineering Society Conference: 2019 AES INTERNATIONAL CONFERENCE ON HEADPHONE TECHNOLOGY, San Francisco, USA*.

Schobben, D., and van de Par, S. (2004). "The effect of room acoustics on mp3 audio quality evaluation," in *Audio Engineering Society Convention 117, San Francisco, USA*.

Schubotz, W., Brand, T., Kollmeier, B., and Ewert, S. D. (2016). Monaural speech intelligibility and detection in maskers with varying amounts of spectro-temporal speech features. *J. Acoust. Soc. Am.* 140, 524–540. doi: 10.1121/1.4955079

Stecker, G. C., and Moore, T. M. (2018). Reverberation enhances onset dominance in sound localization. *J. Acoust. Soc. Am.* 143, 786–793. doi: 10.1121/1.5023221

Tan, C.-T., Moore, B. C., and Zacharov, N. (2003). The effect of nonlinear distortion on the perceived quality of music and speech signals. *J. Audio Eng. Soc.* 51, 1012–1031.

Toole, F. E., and Olive, S. E. (1988). The modification of timbre by resonances: perception and measurement. *J. Audio Eng. Soc.* 36, 122–142.

van Buuren, R. A., Festen, J. M., and Houtgast, T. (1999). Compression and expansion of the temporal envelope: evaluation of speech intelligibility and sound quality. *J. Acoust. Soc. Am.* 105, 2903–2913. doi: 10.1121/1.426943

Wallach, H., Newman, E. B., and Rosenzweig, M. R. (1949). A precedence effect in sound localization. *J. Acoust. Soc. Am.* 21, 468–468. doi: 10.1121/1.1917119

Wallaert, N., Moore, B. C., Ewert, S. D., and Lorenzi, C. (2017). Sensorineural hearing loss enhances auditory sensitivity and temporal integration for amplitude modulation. *J. Acoust. Soc. Am.* 141, 971–980. doi: 10.1121/1.4976080

Wang, D. L. (2005). "On ideal binary mask as the computational goal of auditory scene analysis," in *Speech separation by humans and machines*. ed. P. Divenyi (Norwell, MA: Kluwer Academic).

Weisser, A., Buchholz, J. M., and Keidser, G. (2019). Complex acoustic environments: review, framework, and subjective model. *Trend. Hear.* 23, 1–20. doi: 10.1177/2331216519881346

Wendt, T., van de Par, S., and Ewert, S. D. (2014). A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation. *J. Audio Eng. Soc.* 62, 748–766. doi: 10.17743/jaes.2014.0042

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 01 December 2022
DOI 10.3389/fnins.2022.997656



OPEN ACCESS

EDITED BY
Yi Zhou,
Arizona State University, United States

REVIEWED BY
Edward Lee Bartlett,
Purdue University, United States
Merri J. Rosen,
Northeast Ohio Medical University,
United States

*CORRESPONDENCE
Laurel H. Carney
Laurel_Carney@URMC.Rochester.edu

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 19 July 2022
ACCEPTED 12 October 2022
PUBLISHED 01 December 2022

CITATION
Fan L, Henry KS and Carney LH (2022) Responses to dichotic tone-in-noise stimuli in the inferior colliculus. *Front. Neurosci.* 16:997656. doi: 10.3389/fnins.2022.997656



# Responses to dichotic tone-in-noise stimuli in the inferior colliculus

Langchen Fan[1], Kenneth S. Henry[1,2,3] and Laurel H. Carney[1,2]*

[1]Department of Biomedical Engineering, University of Rochester, Rochester, NY, United States, [2]Department of Neuroscience, University of Rochester, Rochester, NY, United States, [3]Department of Otolaryngology, University of Rochester, Rochester, NY, United States

Human listeners are more sensitive to tones embedded in diotic noise when the tones are out-of-phase at the two ears ($N_0S_\pi$) than when they are in-phase ($N_0S_0$). The difference between the tone-detection thresholds for these two conditions is referred to as the binaural masking level difference (BMLD) and reflects a benefit of binaural processing. Detection in the $N_0S_\pi$ condition has been explained in modeling studies by changes in interaural correlation (IAC), but this model has only been directly tested physiologically for low frequencies. Here, the IAC-based hypothesis for binaural detection was examined across a wide range of frequencies and masker levels using recordings in the awake rabbit inferior colliculus (IC). IAC-based cues were strongly correlated with neural responses to $N_0S_\pi$ stimuli. Additionally, average rate-based thresholds were calculated for both $N_0S_0$ and $N_0S_\pi$ conditions. The rate-based neural BMLD at 500 Hz matched rabbit behavioral data, but the trend of neural BMLDs across frequency differed from that of humans.

KEYWORDS

**binaural masking level difference, binaural cues, binaural detection, interaural correlation, midbrain**

## Introduction

Human listeners benefit from binaural hearing in detection tasks. For example, in the tone-in-noise (TIN) detection task, the threshold for detection of out-of-phase tone in identical noise at the two ears ($N_0S_\pi$) is lower (i.e., better) than that for detection of an in-phase tone ($N_0S_0$) (e.g., Hirsh, 1948; Hawley et al., 2004). The difference in detection thresholds between the $N_0S_0$ and $N_0S_\pi$ conditions is referred to as the binaural masking level difference (BMLD).

In $N_0S_\pi$ stimuli, the difference between the tone-plus-noise waveforms at the two ears results in differences in interaural time or phase and level differences (ITDs, IPDs, or ILDs), as well as changes in the interaural correlation (IAC) (e.g., Domnitz and Colburn, 1976; Bernstein and Trahiotis, 1996). The statistics of the interaural phase and level cues, and their distributions for different signal-to-noise ratios (SNRs) for stimuli used in

binaural detection experiments are described in Zurek (1991). Experiments designed to distinguish the relative importance of dynamic ITD vs. IAC cues have suggested that ITD is most important for 500-Hz binaural detection (van der Heijden and Joris, 2010). Furthermore, a psychophysical study that manipulated ITD and IAC cues over a wide range of frequencies showed that predictions for tone detection differ for ITD and IAC cues (Culling, 2011), and as expected, the role of the ITD cue is diminished at higher target frequencies. The challenge of discriminating between models based on these cues, which co-vary in stimuli used for binaural detection, was described by Domnitz and Colburn (1976), who stressed the importance of testing these models over a range of frequencies or other stimulus parameters in order to distinguish the models. Several subsequent models for binaural detection have focused on detection of a decrease in IAC upon addition of a tone in the $N_0S_\pi$ condition and have tested this class of model across a wide range of stimulus conditions (e.g., Colburn, 1977; Bernstein and Trahiotis, 1997, 2017).

Human listeners can have substantial BMLDs (>3 dB) up to at least 8 kHz (van de Par and Kohlrausch, 1999; Goupell, 2012), yet physiological studies have mainly focused on low frequencies, for which the BMLD is typically larger (up to 20 dB, depending on bandwidth) (e.g., van de Par and Kohlrausch, 1999). Early physiological studies of detection of tones in $N_0S_\pi$ stimuli focused on sensitivity of low-frequency neurons in the auditory midbrain (inferior colliculus, IC) to ITDs (e.g., Caird et al., 1991; McAlpine et al., 1996; Jiang et al., 1997a,b). Later physiological studies analyzed low-frequency IC responses in terms of the IAC cue (Palmer et al., 1999; Lane and Delgutte, 2005), and the effect of decorrelation was estimated over a wider frequency range in the owl (Asadollahi et al., 2010). The current study extends this work by applying an analysis of IAC cues to responses in the IC of awake rabbit across a wide range of frequencies. If interaural decorrelation explains neural responses to $N_0S_\pi$ stimuli, then the difference in average rate between IC responses to diotic noise and binaurally uncorrelated noise should be correlated to the rate difference between responses to the noise-alone condition and the noise-plus-dichotic-tone condition. This correlation was directly tested in this study.

Additionally, human psychophysical studies have shown that BMLDs are robust across a range of noise levels (Buss et al., 2003) and in a roving-level paradigm, in which stimulus level was randomly varied from interval to interval (Henning et al., 2005). Therefore, in the current study neural responses were recorded over a wide range of noise levels to explore trends across sound level.

The IC is a nearly obligatory synapse along the ascending auditory pathway, thus all information available for perception must be encoded at this level. This fact makes the IC an interesting place to examine the relationship between neural and behavioral response properties in tasks such as masked detection. The IC receives afferent inputs from nearly all of the auditory brainstem nuclei (Cant and Oliver, 2018). IC neurons are sensitive to several features of stimuli, including ITDs and ILDs (Reviewed in Yin et al., 2019) and envelope frequency and depth (e.g., Langner and Schreiner, 1988; Krishna and Semple, 2000; Nelson and Carney, 2007; Zheng and Escabi, 2013). Addition of a dichotic tone to a diotic noise masker influences all of these cues. However, individual IC responses are complex in that each neuron responds to different cues with different sizes and directions of rate changes. In the current study, the sensitivities of individual neurons were evaluated using standard physiological characterizations, such as modulation transfer functions and responses to noise with ITDs and ILDs. Responses were then tested for their correlation to the IAC cue. Consistent with previous physiological and psychophysical studies, our results support the importance of the IAC in shaping IC responses to stimuli used to estimate BMLDs, and extend these results by illustrating that this correlation extends across a wide range of noise levels and frequencies.

The current study also computed rate-based IC neural thresholds for comparison with published detection thresholds for human listeners (van de Par and Kohlrausch, 1999; Buss et al., 2003; Goupell, 2012) and rabbits (Zheng et al., 2002).

## Materials and methods

All neurophysiological procedures were approved by the University of Rochester Committee on Animal Resources. Recordings were from four awake, female Dutch-belted rabbits with normal hearing. Distortion product otoacoustic emissions (Whitehead et al., 1992) were used to monitor hearing over the timecourse of the study. Two of the rabbits were studied from 17 to 55 months of age, and two rabbits from age 13 to 23 months.

### Procedures

Surgical and recording procedures are described in detail in Fan et al. (2021). Briefly, rabbits were anesthetized with an intramuscular injection of ketamine (66 mg/kg) and xylazine (2 mg/kg) for both headbar placement and microdrive (five-drive, Neuralynx, Inc., Bozeman, MT, USA) implantation surgeries. The headbar was custom-designed, 3D-printed hard plastic, with a chamber that held the microdrive. The headbar was permanently mounted on the rabbit skull with stainless-steel screws and dental acrylic. After the rabbit recovered from the headbar surgery, a craniotomy was made to allow insertion of guidetubes from the microdrive through the dura. One microdrive held four guidetubes and tetrodes and allowed for independently advancing and retracting each tetrode. Each tetrode consisted of four twisted 18-μm platinum iridium wires, coated in epoxy (California Fine Wire Co., Grover Beach, CA,

USA). The microdrive was replaced as needed, with guidetube positions varied across placements, to search for new neurons.

During recording sessions, the rabbit was placed in a double-walled, sound-proof chamber (Acoustic Systems, Austin, TX, USA), with head fixed using the headbar. Sound was delivered using Beyerdynamic DT990 (Beyerdynamic GmbH & Co., Heilbronn, Germany) or Etymotic ER2 earphones (Etymotic Research, Inc., Elk Grove Village, IL, USA) with custom ear molds for each rabbit. Ear molds were positioned deep in the concha and included an Etymotic probe tube for calibration. The stimulus system included an audio interface (16A, MOTU, Cambridge, MA, USA), a digital-to-analog converter (DAC3 HGC, Benchmark Media Systems, Inc., Syracuse, NY, USA), and earphones (Beyerdynamic DT990, Beyerdynamic GmbH and Co., Heilbronn, Germany or ER2, Etymotic Research). Wideband noise bursts were presented to search for auditory responses. Recordings were made with a multi-channel system (RHD, Intan Technologies, LLC., Los Angeles, CA, USA). When the characteristic frequencies (CFs) increased with tetrode depth, the tetrodes were determined to be in the central nucleus of the IC (ICC). Action potentials were identified offline using spike-sorting techniques applied to the tetrode recordings (Schwarz et al., 2012; Fan et al., 2021). After the termination of recording sessions in each animal, post-mortem histology was applied to verify tetrode locations in the IC.

## Stimuli

Speakers were calibrated with ER-7C or ER-10B+ microphones (Etymotic Research) at the beginning of each recording session. The neurons were characterized in several ways before presenting TIN stimuli. Binaural sensitivity was determined by responses to contralateral, ipsilateral, and binaural wideband noise (0.1–19 kHz) at several sound levels. Responses to contralateral pure tones between 0.25 and 20 kHz from 10 to 70 dB SPL were used to identify CF, the frequency at which the neuron responded at the lowest sound level. Noise delay functions (NDFs) described rate responses to noise stimuli as a function of ITD; NDFs were recorded with wideband noise (0.1–19 kHz), 1-sec duration, 30-dB SPL spectrum level, and ITDs from –2,000 to 2,000 μs with a 200-μs stepsize. Responses to ILDs were recorded with the same noise bandwidth and duration as for the NDF. ILDs ranged from –15 to 15 dB with a 5-dB stepsize; the stimulus on the contralateral side had a fixed spectrum level of 30 dB re 20 μPa. Responses to contralateral sinusoidally-amplitude-modulated (SAM) wideband noise (0.1–19 kHz), with 1-sec duration, were collected to identify the shape of the modulation transfer function (MTF). SAM noises were described by:

$$s = \left[1 + \sin\left(2\pi f_m t\right)\right] n(t)$$

where $n(t)$ is the wideband noise with a spectrum level of 30 dB SPL, and $f_m$ is the modulation frequency. Modulation frequencies were logarithmically spaced between 2 and 350 Hz, with three steps/octave. Responses to contralateral unmodulated noise were also recorded. For all of the above characterizations, three repetitions of each stimulus condition were presented, in random sequence.

For TIN stimuli, the tone frequency and the center frequency of 1/3-oct gaussian noise maskers were chosen to be approximately equal to CF. Noise maskers were simultaneously gated with tone signals and generated by filtering wideband noise with a 5,000th-order FIR band-pass filter. TIN stimuli had 0.3-sec duration with 10-msec $\cos^2$ on/off ramps. Overall noise levels ranged from 35 to 75 dB SPL, with a10-dB stepsize. Signal-to-noise ratio (SNR) ranged from –12 to 8 dB, with a 4-dB stepsize; a noise-alone condition was also included. Tone levels and noise levels were presented in random order, and the order was shuffled for each of the 30 repetitions of the stimulus set. Responses were collected for sets of random noise, or reproducible noise (for the temporal analyses in Fan et al., 2021), or both. If more than one dataset was recorded, the dataset with responses to random noise waveforms was used for the analyses presented here. Among all neurons reported here, there were 55 neurons studied with random noise and 81 neurons studied with reproducible noise. No qualitative differences were observed between these two types of datasets, although the use of random noise would be expected to reduce the potential effect of external noise on neural responses.

To test the influence of IAC on IC neurons, responses to diotic ($N_0$) and binaurally uncorrelated ($N_u$) noise were recorded. For both $N_0$ and $N_u$ conditions, the stimuli were 1/3-octave random gaussian noise, with 2-sec duration, at 65 dB SPL. Five repetitions of five $N_0$ and ten $N_u$ noise were presented, in random sequence.

## Noise delay function shape classification

The shape of the NDF, the best ITD ($d_{BITD}$), and the frequency of ITD tuning ($f_{ITD}$) were determined by fitting the NDF with a Gabor function (Lane and Delgutte, 2005), a sinusoid modulated by a gaussian function:

$$G_1 = \left| Ae^{-\frac{(d_{ITD}-d_{BITD})^2}{2\sigma^2}} \cos\left[2\pi f_{ITD}\left(d_{ITD} - d_{BITD}\right)\right] + B \right|,$$

where $d_{ITD}$ is the interaural delay, A, B, and σ are parameters for the amplitude, DC offset, and standard deviation of the gaussian function, respectively, and $|\bullet|$ refers to half-wave rectification. If a neuron's CF was more than twice $f_{ITD}$ (i.e., a high-frequency neuron), indicating that the neuron did not have fine-structure-based ITD sensitivity, then $f_{ITD}$ was set to zero, and the NDF was

refitted with the following gaussian function:

$$G_2 = \left| Ae^{-\frac{(d_{ITD}-d_{BITD})^2}{2\sigma^2}} + B \right|.$$

The function was fit to an NDF using a least-square fit, obtained with a trust-region-reflective algorithm (*lsqcurvefit* in MATLAB).

Each NDF was classified as peak-like, trough-like, or ITD-insensitive. In the following cases, the neuron was considered sensitive to ITDs: (1) for NDFs fitted with $G_1$, if the absolute value of the amplitude (A) was more than 5 spikes/sec; (2) for NDFs fitted with $G_2$, if the absolute value of the prominence (A/B) was more than 0.25; (3) for NDFs fitted with $G_2$, for a fit with σ between 60 and 1,000 μs. If the amplitude (A) was positive, the neuron was classified as having a peak-like NDF; otherwise, the neuron was classified as having a trough-like NDF. Other neurons were classified as ITD-insensitive. The classification of each NDF generally agreed with a qualitative assessment (**Figure 1**).

## Modulation transfer function shape classification

The MTF shape was classified with rules designed to be simple and to agree with qualitative descriptions of the functions. Enhancement or suppression was identified with the Mann-Whitney test as significantly higher or lower rates at two or more neighboring modulation frequencies than the rates in response to unmodulated noise. The presence or absence of enhancement or suppression was used to classify the MTF into the following four types: all-pass (AP, no enhancement or suppression), band-enhanced (BE, only enhancement), band-suppressed (BS, only suppression), and hybrid (both enhancement and suppression, over different ranges of modulation frequency).

## Rate analysis

Average rates, excluding 20-ms onset responses, were calculated for responses to all stimuli. For TIN stimuli, at each noise and tone level (i.e., SNR), a rate-based receiver-operating-characteristic (ROC, Egan, 1975) was calculated using average rate responses for all 30 noise-alone and tone-plus-noise presentations. The percent-correct performance was estimated from the area under the ROC curve. Note that rates in response to tone-plus-noise stimuli could be either higher or lower than rates in response to noise-alone stimuli, so the minimum percent correct was limited to 50%, regardless of the direction of change in rate. The neural threshold was estimated using linear interpolation to find the lowest SNR with 70.7% correct, which corresponds to a threshold estimated with a two-down, one-up tracking procedure (Levitt, 1971).

# Results

Responses to both $N_0S_0$ and $N_0S_\pi$ stimuli were recorded from 136 isolated single units; responses to $N_0S_0$ of 111 of these units were presented in Fan et al. (2021). Responses to Nu were recorded for 68 units. The distribution of CFs is shown in **Figure 2**. All units were tested using a tone frequency within 1/3-octave of the neuron's CF. Based on the MTF categorization criteria described above, there were 40 BE units (29.4%), 62 BS units (45.6%), 12 hybrid units (8.8%) and 22 AP units (16.2%). Distribution of MTF types across CFs is shown in **Figure 2**.

## Examples of single-neuron responses

Responses of several example units illustrate the complexity of response properties of the IC responses that were analyzed to test the IAC hypothesis. IC neurons have rates that vary with both ITD and ILD, and the interaction of these cues in the $N_0S_0$ and $N_0S_\pi$ stimuli are complex (Zurek, 1991). Additionally, IC neurons are sensitive to periodicity in the stimulus as conveyed in their neural inputs. Adding a tone to narrowband gaussian noise flattens the stimulus envelope (Richards, 1992) and also reduces the amplitudes of neural fluctuations in peripheral responses (Carney, 2018). Therefore, the MTFs of IC neurons are interesting to consider, as well as sensitivity to the classical interaural cues. Neurons with BE MTFs (**Figure 3A**) are excited by fluctuations and therefore expected to have decreasing rate with increasing SNR for TIN stimuli. On the contrary, neurons with BS MTFs (**Figure 3E**) are suppressed by fluctuations and therefore expected to have increasing rate with increasing SNR. As expected, Neuron 1, with a BE MTF, had decreasing rate versus SNR at all noise levels (**Figures 3C,D**), and Neuron 2, with a BS MTF, had increasing rate versus SNR at all noise levels (**Figures 3G,H**). Both of these examples responded as predicted by their MTF types. Note that for both neurons in **Figure 3**, the average rate changed at lower SNRs for the $N_0S_\pi$ condition than for the $N_0S_0$ condition, for all noise levels tested, indicating lower neural thresholds, consistent with psychophysical results (e.g., van de Par and Kohlrausch, 1999).

Neural responses to $N_0S_0$ stimuli have previously been described as having increasing rate as a function of tone level (Jiang et al., 1997a; Ramachandran et al., 2000), possibly based on the assumption that neurons respond more strongly to increasing stimulus energy (i.e., upon addition of a tone). Note that Neuron 1 in **Figure 3** is an example of a neuron that had decreasing rate as tone level increased at each masker level, whereas it had increasing rate versus masker level for the noise-alone stimuli (SNR = –inf); these responses cannot be explained

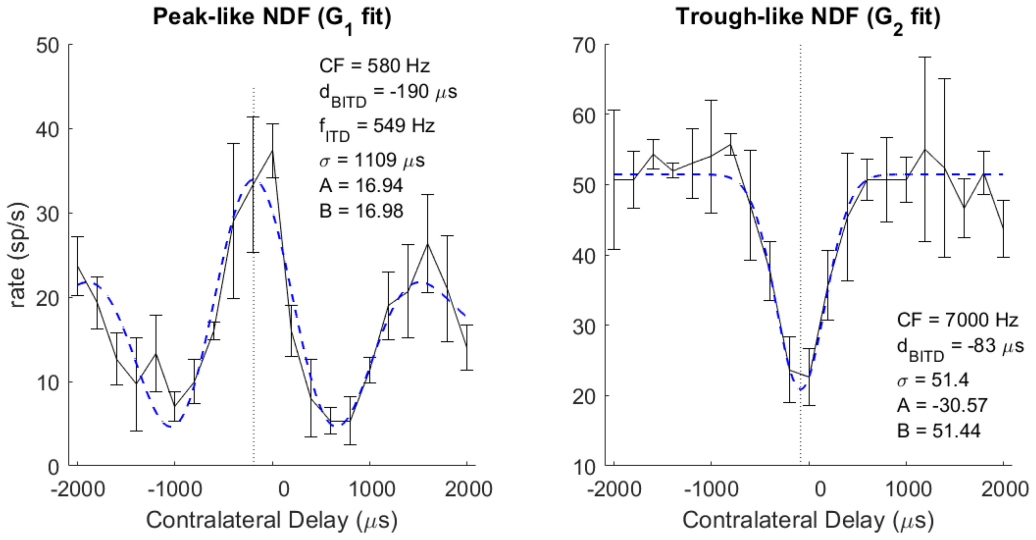



FIGURE 1
Example neural ITD responses (black solid curve) and fitted Gabor function (blue dashed curve) for peak-like **(left)** and trough-like **(right)** NDFs. Vertical dotted line indicates the best ITD. The neuron's CF, Best or Worst ITD ($d_{BITD}$), and for cyclic ITD curves, the ITD tuning frequency ($f_{ITD}$) are described in the text.

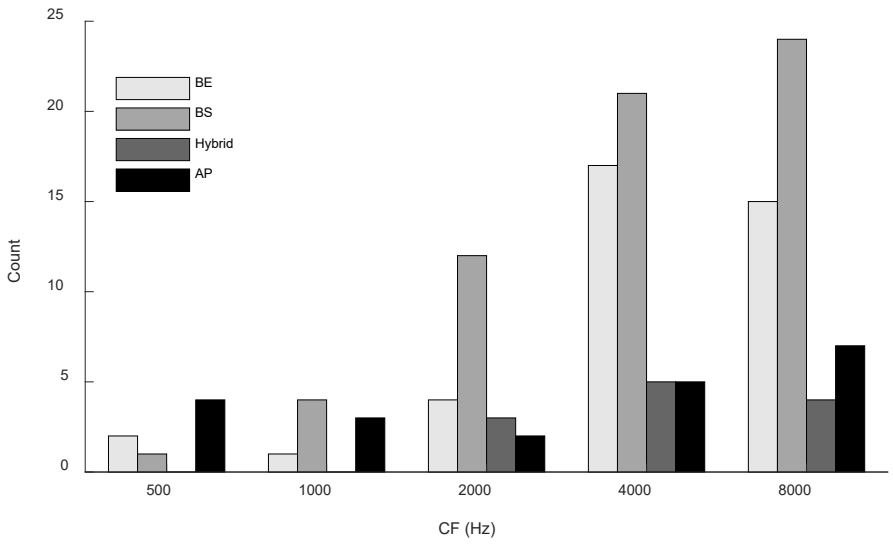



FIGURE 2
Distribution of MTFs across CF (in one-octave bins) for the units presented in this study. Gray shades from light to dark indicate units with band-enhanced (BE), band-suppressed (BS), hybrid and all-pass (AP) MTF shapes. Two neurons with CF of 12.1k were included in the last bin for simplicity. Most MTF types were represented across the range of CFs, although hybrid MTFs were not observed at the lower CFs.

based on stimulus energy. The shape of NDF has been used to explain changes in neural responses for the $N_0S_\pi$ condition (Jiang et al., 1997a,b): a diotic noise masker has zero ITD; adding a dichotic tone introduces non-zero ITDs. Neurons with peak-like NDFs respond most strongly to near-zero ITDs, and thus would be expected to have decreasing rate with increasing SNR based on the ITD hypothesis. In contrast, neurons with trough-like NDFs would be expected to have increasing rate with increasing SNR. Responses to $N_0S_\pi$ stimuli of Neurons 1 and 2 can also be explained by their NDF shapes: Neuron 1 had a peak-like NDF shape (**Figure 3B**) and decreasing rate versus SNR for the $N_0S_\pi$ condition; Neuron 2 had a trough-like NDF (**Figure 3F**) and increasing rate versus SNR.

Single-unit responses to $N_0S_0$ and $N_0S_\pi$ stimuli were analyzed based on MTF properties and responses to ITDs and ILDs. In general, the directions and sizes of rate differences to

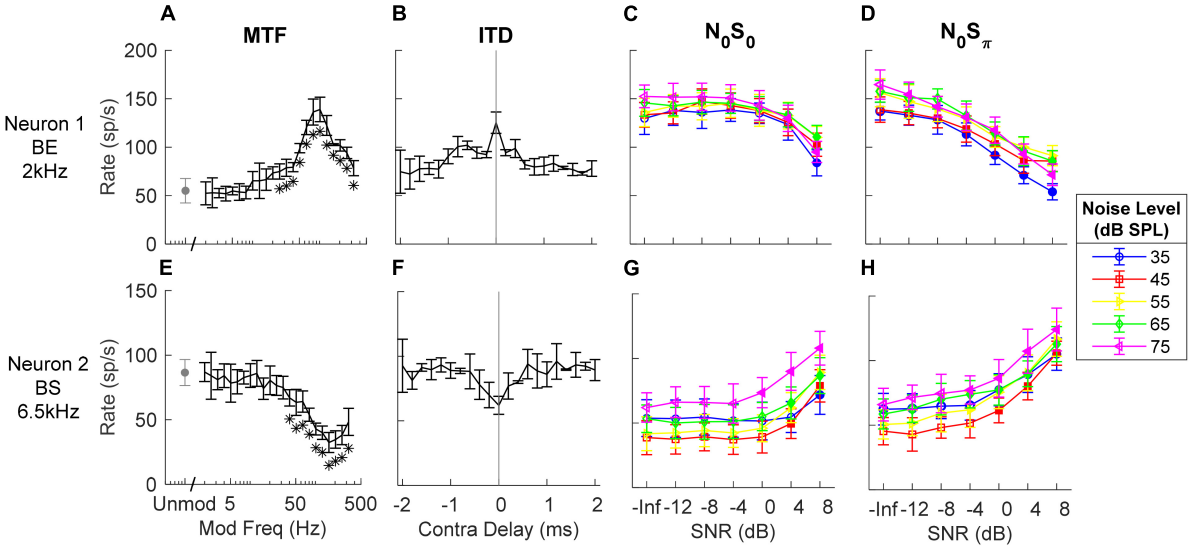



FIGURE 3
Responses of two example neurons (**top** and **bottom** row respectively). **(A,E)** MTF, response rates to amplitude-modulated noise; stars indicate modulation frequencies that had rates significantly different from the unmodulated condition. **(B,F)** ITD sensitivity, response rates vs. time delay in contralateral side (negative indicates ipsilateral side has delay). **(C,D,G,H)** Responses to $N_0S_0$ and $N_0S_\pi$ stimuli at different noise levels (different symbols) vs. SNR (from left to right); filled symbols indicate supra-thresholds. Errorbars indicate standard deviation. MTF shape and tone frequency for TIN stimuli (close to CF) are shown on the left. The example BE neuron had decreasing rate upon addition of a tone for both $N_0S_0$ and $N_0S_\pi$, while the example BS neuron had increasing rate for both conditions.

$N_0S_0$ stimuli can be predicted based on MTF properties (Fan et al., 2021), but in response to $N_0S_\pi$ stimuli, predictions of changes in rate based on MTF properties were only significant at the highest noise level tested (Fan, 2020). Rate differences were also weakly but significantly correlated to rate differences in the NDF, but the correlations decreased as stimulus level increased (Fan, 2020).

In general, IC responses to dichotic TIN stimuli are not easily explained by characterizations based on MTFs, ITDs, or ILDs (see below), likely because of the interaction of these cues in $N_0S_0$ and $N_0S_\pi$ stimuli and because of the different types of sensitivity of IC neurons to these cues (**Figure 4**). For example, Neurons 3 and 4 both had BE MTFs and decreasing rate versus SNR for the $N_0S_0$ condition at most noise levels, as expected. However, for the $N_0S_\pi$ condition, Neuron 3 had decreasing rate versus SNR that could be explained by its MTF shape, but not its trough-like NDF. In contrast, Neuron 4 had an increasing rate versus SNR that could be explained by its NDF shape, but not by its MTF shape. Neurons 5, 6, and 7 all had BS MTFs, and thus were expected to have increasing rates versus SNR, but the responses of these neurons differ. Neuron 5 had increasing rate versus SNR for both $N_0S_0$ and $N_0S_\pi$ conditions, which could be explained by its BS MTF, but not by its peak-like NDF. The MTF of Neuron 6 did not explain responses to either $N_0S_0$ or $N_0S_\pi$ stimuli, but responses to $N_0S_\pi$ stimuli (decreasing rate) could be explained by its peak-like NDF. Neuron 7 also had decreasing rate versus SNR, which could not be explained by either MTF or NDF shape. Neuron 8 had an all-pass MTF, and responses to $N_0S_\pi$ stimuli that could be explained by its peak-like NDF.

## Rate differences in response to $N_0S_\pi$ stimuli and binaural cues

The rate differences in response to ITDs or ILDs were quantified by the difference between the maximum and minimum response rates over the range of stimuli tested. The maximum change in rate in response to $N_0S_\pi$ stimuli, for both directions of rate change as a function of SNR, was significantly correlated to the maximum rate differences in both ITD and ILD responses (**Figures 5A,B**), explaining a small but significant proportion of the variance (i.e., $r^2$). There was not an obvious difference between results shown in **Figure 5** for lower-CF neurons ($<$ 1.5 kHz, filled triangles) vs. higher-CF neurons (open circles). The significant correlation between the maximum rate differences for $N_0S_\pi$ responses and rate differences for both ITD and ILD responses could be because (1) adding a dichotic tone not only introduces ITDs, but also ILDs; and/or (2) the dynamic ranges of ITD and ILD responses were significantly correlated (**Figure 5C**). Changes in neural responses to $N_0S_\pi$ are likely due to a combination of ITD and ILD sensitivities and to the co-variation of these cues. The standard deviations of interaural phase and interaural level cues as a function of SNR have been previously described [see Figures 9 and 10 in Zurek (1991)].

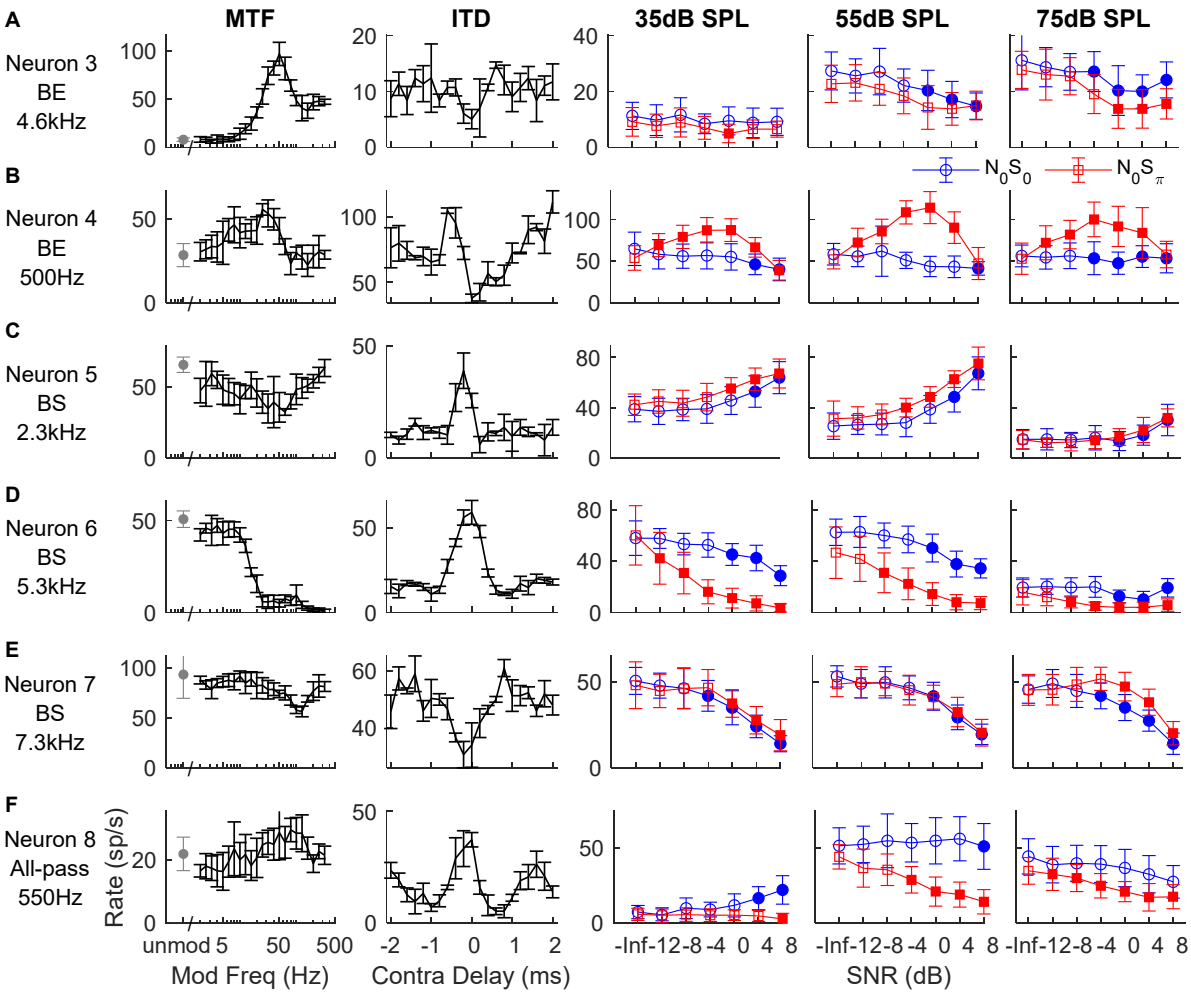



FIGURE 4
Responses of six example neurons **(A–F)**. The left two columns show the neuron's MTF and ITD sensitivity, respectively. The right three columns show the neuron's response to $N_0S_0$ (blue circles) and $N_0S_\pi$ (red squares) TIN stimuli at noise levels of 35, 55, and 75 dB SPL, respectively; filled symbols indicate supra-threshold responses. MTF shape and tone frequency of TIN stimuli (close to CF) are shown on the left.

## Inferior colliculus responses to interaural correlation

Adding a dichotic tone to diotic noise introduces both ITD and ILD cues, as well as interaural decorrelation, but the changes in these cues differ for different tokens of noise waveform as well as for different SNRs. For example, the ITD of a $N_0S_\pi$ stimulus is dominated by the ITD of the added tone with increasing tone level, but the effective ITD of a $N_0S_\pi$ stimulus with a low-SNR tone (e.g., at threshold) is hard to estimate, and varies with the noise token due to the phase interaction between the noise and tone. Additionally, unlike a pure tone, the instantaneous ITD of $N_0S_\pi$ stimuli varies throughout the duration of the stimulus waveform. Therefore, prediction of the rate-change direction upon addition of a tone at threshold based on sensitivity to static ITDs and ILDs is not simple. On the other hand, the effect of interaural decorrelation can be studied with a more straightforward method. To examine the effect of decorrelation, average rates were recorded in response to 1/3-octave diotic ($N_0$) and binaurally uncorrelated ($N_u$) gaussian noise for 68 neurons. The $N_u$ noises presented at the two ears were simply independent narrowband noise tokens. The correlation between the difference in average rate in response to the $N_0S_\pi$ condition (the difference between average rates in response to noise-alone and to $N_0S_\pi$ at 0-dB SNR) and the difference in average rates in response to the $N_0$ and $N_u$ conditions was significant at all noise levels (Figure 6), supporting the hypothesis that IC rates are influenced by IAC. The correlation was strongest for TIN stimuli with a masker level of 65 dB SPL, the level at which the $N_0$ and $N_u$ noise were presented. At 65 dB SPL, additional analyses of the rate differences in responses to $N_0S_\pi$ stimuli at SNRs of –8 to 8 dB relative to the noise-alone condition were all significantly

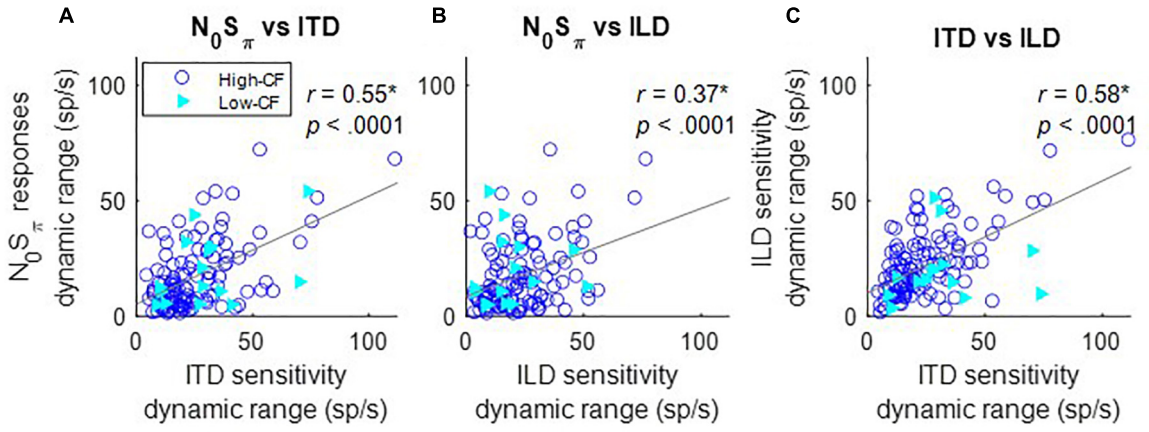



FIGURE 5
Correlation between dynamic ranges of responses to $N_0S_\pi$ and ITD **(A)**, $N_0S_\pi$ and ILD **(B)**, and ILD and ITD **(C)** at 65 dB SPL (as indicated in titles). Correlation coefficients and *p*-values are shown at the top right of each panel; a star indicates that the correlation coefficient was significant after Bonferroni correction ($p < 0.017$). Neurons with CF below 1.5 kHz (low–CF) are shown with filled triangles, whereas neurons with CF above 1.5 kHz (high–CF) are shown with open circles. Solid gray lines indicate linear regressions.

correlated to the rate difference between the responses to $N_u$ and $N_0$ noise, with correlation coefficients ranging from 0.71 to 0.84, and *p* values all less than 0.0001 (significant after Bonferroni correction, not shown). The significant correlation coefficients at *all* SNRs and noise levels indicated that, in general, the direction and size of the changes in rate in response to $N_0S_\pi$ stimuli were explained by the change in the stimulus from $N_0$ toward $N_u$. Note that there were only a few low-CF (<1.5 kHz, filled triangles) in this dataset, so it is clear that the correlations illustrated in **Figure 6** applied to the much larger group of high-CF neurons (open circles).

## Rate-based neural thresholds

Rate-based thresholds of all units for the $N_0S_0$ and $N_0S_\pi$ conditions at five noise levels were computed and compared with behavioral data from previous studies (**Figure 7**). There was no clear trend in the numbers of units with increasing or decreasing rate-change direction across frequency, for either the $N_0S_0$ or $N_0S_\pi$ condition, except a weak trend of more units with increasing rate at the lowest noise level tested (bottom row). The lowest rate thresholds across frequency were lower for the $N_0S_\pi$ condition than for the $N_0S_0$ condition, as expected.

The lowest rate thresholds at 500 Hz matched the mean rabbit behavioral detection threshold at the same frequency (Zheng et al., 2002). Compared with human thresholds, the lowest rate thresholds for the $N_0S_0$ condition were close to human thresholds across frequencies, but the lowest rate thresholds for the $N_0S_\pi$ condition only matched human thresholds at high frequencies (note that the lower limit of SNRs tested limited this comparison, see below). Human thresholds from Goupell (2012) are slightly lower than van de Par and Kohlrausch (1999) at some frequencies, possibly due to differences in paradigm and stimulus bandwidths. Note that stimuli used in previous studies have slightly different parameters from this study: stimuli in Zheng et al. (2002) had 200-Hz bandwidth (vs. 116 Hz in this study) and an overall level of 63 dB SPL; stimuli in van de Par and Kohlrausch (1999) had bandwidths of 100, 250, 500 Hz and 1 kHz (vs. 116 Hz, 232, 463, and 926 Hz in this study) for center frequencies of 500 Hz, 1, 2, and 4 kHz, and with overall level of 70 dB SPL; stimuli in Buss et al. (2003) had 50-Hz bandwidth and overall noise levels of 42, 57, and 72 dB SPL; stimuli in Goupell (2012) had bandwidths of 78, 240, 456, and 888 Hz (vs. 116, 463, 926, and 1,852 Hz in this study) for center frequencies of 500 Hz, 2, 4, and 8 kHz. However, despite the discrepancies among stimuli, in general, the lowest rate-based thresholds could explain human thresholds for the $N_0S_0$ condition across all frequencies tested and for the $N_0S_\pi$ condition at high frequencies. Note that the thresholds of most sensitive neurons across frequencies did not vary qualitatively across noise levels, consistent with human thresholds tested at multiple noise levels (Buss et al., 2003) and with a roving-level paradigm (Henning et al., 2005).

## Rate-based neural binaural masking level differences

Neural BMLDs were evaluated in two ways: using the BMLDs of individual neurons, and using the BMLDs calculated from the $N_0S_0$ and $N_0S_\pi$ thresholds of the neural population. For BMLDs of single neurons (**Figure 8**), only neurons with measurable thresholds for both $N_0S_0$ and $N_0S_\pi$ conditions are plotted, together with human BMLDs (van de Par and

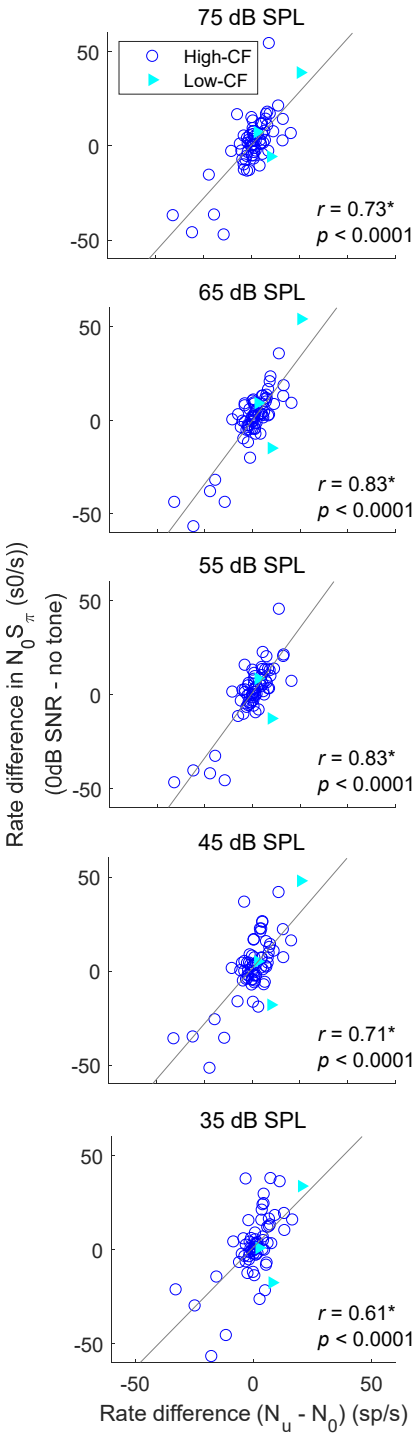



FIGURE 6
Correlation between the rate difference elicited by addition of a dichotic tone ($N_0S_\pi$) at 0 dB SNR and the rate difference between responses to $N_0$ and $N_u$ conditions. Correlation coefficients and *p*-values are shown; a star indicates that the correlation coefficient was significant after Bonferroni correction ($p < 0.0014$). Neurons with CF below 1.5 kHz (low–CF) are shown in filled triangles, whereas neurons with CF above 1.5 kHz (high–CF) are shown in open circles. Solid lines show linear regressions.

Kohlrausch, 1999; Buss et al., 2003; Goupell, 2012). BMLDs were typically positive, indicating greater TIN sensitivity for $N_0S_\pi$ compared to $N_0S_0$. There was no clear association observed between small or negative.

BMLDs and rate-change direction for either $N_0S_0$ or $N_0S_\pi$ conditions, in contrast to a previous report (Jiang et al., 1997a). There was also no clear pattern of same (open symbols) or opposite (filled symbols) rate-change directions for $N_0S_0$ and $N_0S_\pi$ conditions across frequency (i.e., thresholds were similar for upward and downward triangles). Overall, there were more neurons with the same rate-change directions than with opposite rate-change directions (more open symbols than filled symbols) between $N_0S_0$ and $N_0S_\pi$ conditions. Among neurons with opposite rate-change directions across conditions, more neurons had decreasing rate at threshold for the $N_0S_0$ condition (more filled downward than upward triangles). At 500 Hz, single-neuron BMLDs were close to human BMLDs at noise levels of 45 and 65 dB SPL, but not at other noise levels. At 1 kHz and above, the maximum single-neuron BMLDs were larger than human BMLDs. The maximum BMLDs were similar across noise levels, as well as across frequencies, unlike human BMLDs that decrease substantially with increasing frequency (van de Par and Kohlrausch, 1999; Goupell, 2012).

To calculate BMLDs of the neural population, neural thresholds for the most sensitive subset of neurons were calculated for 0.5, 1, 2, 4, and 8 kHz for the $N_0S_0$ or $N_0S_\pi$ conditions. The decision to focus on the most sensitive units for this analysis, as proposed by the lower-envelope principle (Barlow et al., 1971), was based on the fact that many of the neural thresholds were significantly higher than behavioral thresholds (**Figure 7**). Due to the limited SNR range that was tested, many sensitive neurons were suprathreshold (greater than 70.7% correct) at the lowest tested SNR, especially for the $N_0S_\pi$ condition. To reduce the number of neurons for which the BMLD estimate was limited in this way, individual thresholds were recalculated using a criterion of 79.1% correct for the population-threshold results shown in **Figure 9** (squares and diamonds). Individual symbols in **Figure 9** represent all neurons that had thresholds above the lowest SNR tested. For each frequency, the population threshold was based on the neurons with thresholds in the lowest 10th percentile within a one-octave range centered at that frequency. Thresholds at 55–75 dB SPL had similar patterns and were plotted together in **Figure 9**, which shows that neural population thresholds for both $N_0S_0$ (blue solid line) and $N_0S_\pi$ conditions (red dashed line) did not vary across frequency. Human thresholds were moved up by 4 dB to align the means of the human and $N_0S_0$ thresholds of the population, to better compare the trend across frequency (**Figure 9**). Human $N_0S_\pi$ thresholds increase as a function of frequency, whereas thresholds of the neural population did not. Therefore, human and neural BMLDs had different trends across frequency: human BMLDs decrease with increasing frequency, whereas neural BMLDs did not.

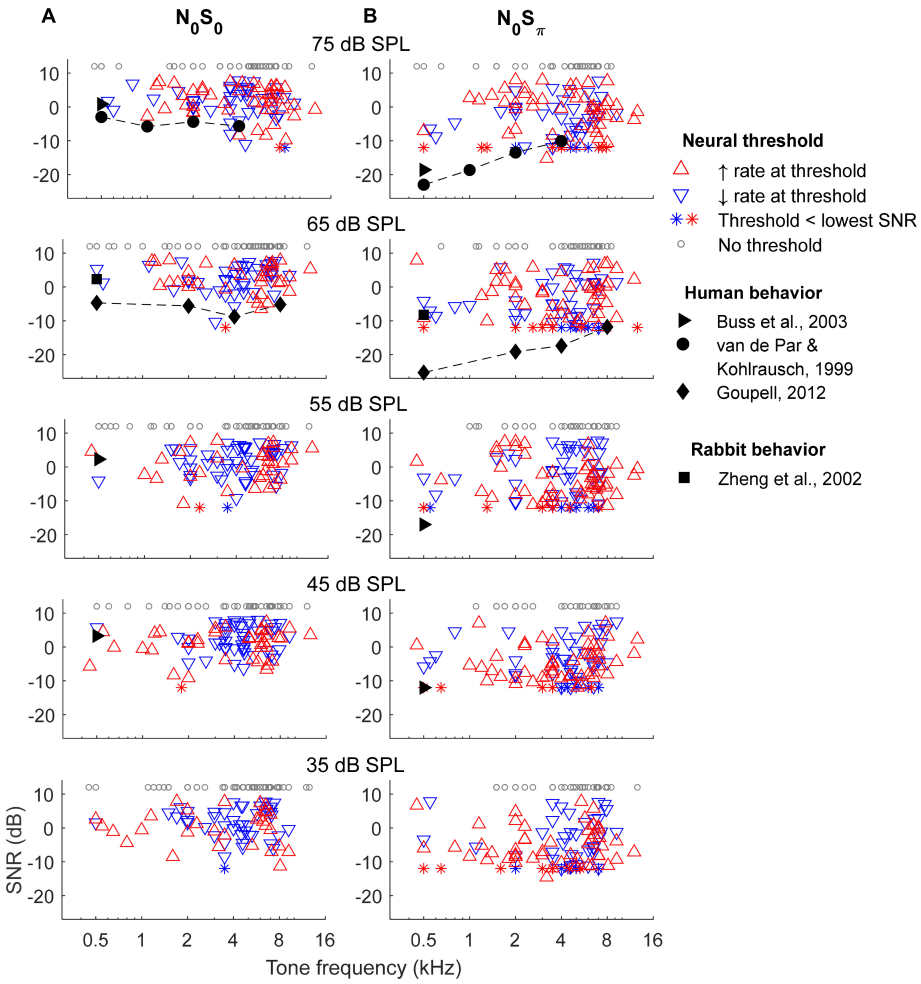



FIGURE 7
Rate-based threshold for $N_0S_0$ **(A)** and $N_0S_\pi$ **(B)** conditions. Thresholds of most sensitive neurons across frequencies matched human behavioral data for the $N_0S_0$ condition, but had a trend different from human for the $N_0S_\pi$ condition. Neural thresholds at 500 Hz matched rabbit behavioral data for both conditions.

The BMLDs based on the neural population thresholds were smaller than the maximum single-neuron BMLDs, as expected due to averaging across the subsets of sensitive neurons for calculation of the population thresholds. However, the BMLDs based on either neural-population or single-neuron thresholds had similar trends across frequency.

## Discussion

In the current study, single-neuron responses to TIN stimuli were recorded in the IC for both $N_0S_0$ and $N_0S_\pi$ conditions over a wide range of target frequencies, as well as noise and tone levels. For the population of neurons, changes in rate due to interaural decorrelation were strongly correlated with changes in rate upon addition of an out-of-phase tone to identical noise at all noise levels.

### Comparison with previous physiological studies

There have been a limited number of physiological studies of neural responses to both $N_0S_0$ stimuli and $N_0S_\pi$ stimuli in the IC (Jiang et al., 1997a,b; Lane and Delgutte, 2005). The results here were most comparable to those of Jiang et al. (1997a,b), who used a tone target, as opposed to the chirp target used in Lane and Delgutte (2005). There were a few differences between the stimuli used in the current study and in Jiang et al. (1997a) that may explain differences in the results between the two studies. First, responses were only recorded for a tone frequency of 500 Hz in Jiang et al. (1997a), for neurons with a range of CFs, up to 1.5 kHz. Large differences between the tone frequency and CF would be expected to affect response properties. For example, the response of a model auditory-nerve (AN) fiber (Zilany et al., 2014) is saturated in response to a CF tone at 65 dB

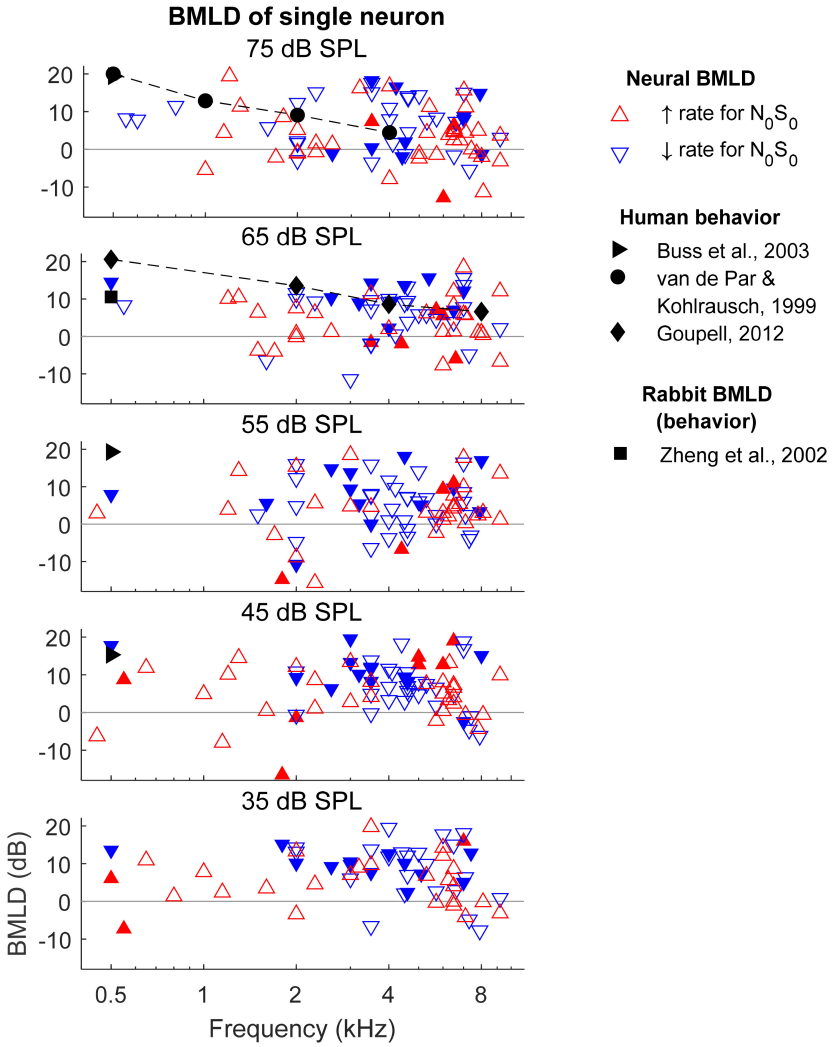



FIGURE 8
Binaural masking level differences (BMLDs) calculated based on single-neuron thresholds for both $N_0S_0$ and $N_0S_\pi$ conditions. Open triangles indicate that the direction of change in rate vs. SNR at threshold for the $N_0S_\pi$ condition was the *same* as for the $N_0S_0$ condition, whereas filled triangles indicate *opposite* direction of change in rate at threshold for the $N_0S_0$ and $N_0S_\pi$ conditions. Only neurons that had measurable thresholds in both $N_0S_0$ and $N_0S_\pi$ conditions are shown here.

SPL, but not in response to a 65-dB-SPL tone one octave below CF. Therefore, when the tone frequency is far from CF, AN rates would vary with stimulus sound level, possibly a stronger cue than the relatively small change in neural fluctuations that would result from an off-CF tone. Thus, the difference between CF and target-tone frequency could explain the finding that the majority of neurons in Jiang et al. (1997a) had increasing rate with increasing SNR for the $N_0S_0$ condition, whereas many neurons in the current study had decreasing rate versus SNR.

Second, many neurons in the current study did not have measurable thresholds due to the limited range of SNRs tested, but finer steps and a wider range of SNRs were used in Jiang et al. (1997a), so thresholds were measurable for almost all neurons. However, it is worth noting that a 20-dB range of SNRs were tested in this study; thus, neurons without a measurable threshold over this SNR range were largely insensitive to addition of a tone. Thresholds for more neurons might have been measured if the SNR had been increased further, but such thresholds would likely reflect changes in response to tone levels high above behavioral thresholds, and would thus not be relevant to tone-in-noise detection.

Third, the masker in Jiang et al. (1997a) had a bandwidth from 50 Hz to 5 kHz and a level of 65 dB SPL, whereas the current study used 1/3-octave noise centered at the tone frequency, presented over a wide range of noise levels, including 65 dB SPL. The difference in masker bandwidth between studies represents a large difference in noise spectrum level: e.g., 28 dB SPL for Jiang et al. (1997a) 65 dB SPL overall level noise, versus a spectrum level of 44 dB SPL for the 500-Hz target tone tested at the overall noise level of 65 dB SPL in the

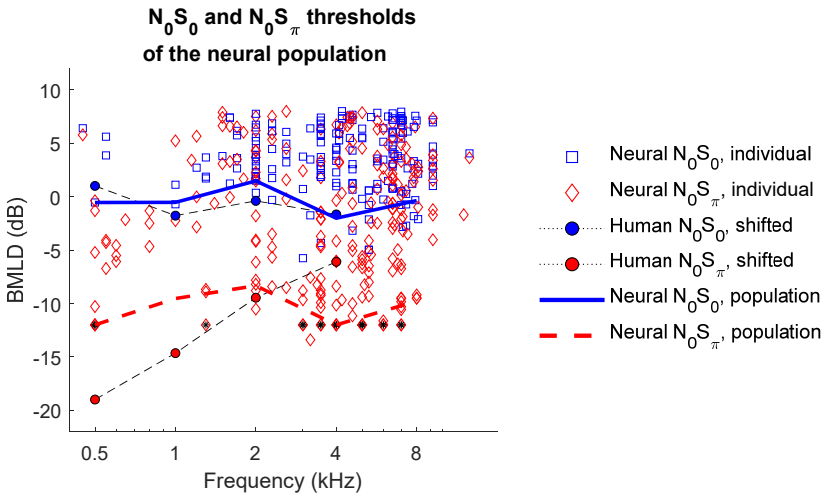



FIGURE 9
$N_0S_0$ (solid blue line) and $N_0S_\pi$ thresholds (dashed red line) of the neural population across frequency. Individual neural thresholds at 79.1% correct for $N_0S_0$ (blue square) and $N_0S_\pi$ (red diamond) conditions, for noise levels of 55–75 dB SPL are shown for all neurons with measurable thresholds above the lowest SNR tested. Symbols with a black star indicate that the threshold was lower than the lowest measured SNR. Human detection thresholds are from van de Par and Kohlrausch (1999) and shifted up by 4 dB for comparison with neural thresholds, which were computed using a higher criterion. Neural binaural masking level differences (BMLDs) had a different trend across frequency compare to human BMLDs.

current study. This difference in spectrum level would have elicited different responses in the periphery, especially at low stimulus frequencies. Even though peripheral neurons respond to a wide frequency range at high sound levels (Ruggero, 1992), the tuning is usually asymmetric and spreads more toward lower frequencies (Schmiedt, 1989). Therefore, for low-CF neurons (e.g., 1 kHz), possibly only the low frequency components of the noise masker used in Jiang et al. (1997a) effectively masked the tone. Additionally, due to non-linear cochlear compression (Robles and Ruggero, 2001), neural responses would differ for maskers having different spectral levels, though the overall level may be matched.

## The role of interaural correlation in $N_0S_\pi$ responses and relationship to other binaural cues

Adding an out-of-phase tone reduces the IAC (e.g., Bernstein and Trahiotis, 2017). The change in rate elicited by an out-of-phase tone was significantly correlated with the rate difference between responses to $N_0$ and $N_u$ noise (**Figure 6**); the large proportion of variance explained (37–69%) suggested an important role of the IAC in physiological $N_0S_\pi$ responses.

Results showed that both ITD-, and IAC-based cues explained a proportion of neural responses to $N_0S_\pi$ stimuli (maximum 34 and 69%, respectively) (Fan, 2020). The ITD-based hypothesis explained responses at low-to-medium noise levels, whereas the IAC-based hypothesis explained TIN responses at all noise levels. The IAC-based hypothesis explained a larger proportion of variance in rate responses at 65 dB SPL, at which $N_0$ and $N_u$ noise responses were collected. However, these cues are not independent. For example, the decreasing trend in the proportion of results explained by the ITD-based hypothesis as noise level increased could be due to the fact that envelope ITDs dominated responses of the high-CF neurons, which were the majority of the neurons in the population studied here. However, the fluctuation amplitudes in AN responses saturate (i.e., flatten) at higher sound levels, and thus binaural differences in the neural representations of the stimulus envelope would also decrease with increasing sound level, which would explain a weaker effect of envelope ITDs at high sound levels. Also, at high frequencies, IAC-cues have been proposed to be envelope-based (Durlach, 1964; Bernstein and Trahiotis, 1996).

Some effort has been made to separate the role of IAC and ITD in binaural detection (van der Heijden and Joris, 2010; Culling, 2011). Based on results from these studies, both ITD and ILD cues are proposed to contribute to interaural decorrelation. Adding an out-of-phase tone not only introduces ITDs, but also ILDs; additionally, the added binaural cues are time-varying. The dynamic range of neural responses to ILD was correlated not only to that of $N_0S_\pi$ responses, but also to the dynamic range of ITD responses (**Figure 5**). Fluctuations of ITD in an $N_0S_\pi$ stimulus increase with increasing tone level, whereas fluctuations of ILD first increase and then decrease as tone level increases (Zurek, 1991). Therefore, interaural decorrelation involves a nonlinear combination of ITD and ILDs cues: both ITD and ILD cues affect IAC at low tone levels, whereas at high tone levels (e.g., above 4 dB SNR), ITD cues

dominate IAC. This proposed idea is consistent with a previous modeling study (Mao and Carney, 2014) in which ITD cues are shown to dominate in stimuli with low modulation depths (e.g., tone-plus-noise), and the combination of ITD and ILD cues dominate in stimuli with high modulation depths (e.g., noise). In that study, the nonlinear combination of ITD and ILD cues is described by the slope of the interaural envelope difference (SIED), whereas detection in the $N_0S_\pi$ condition at high frequencies has been proposed to be explained by the envelope-based IAC (Durlach, 1964; Bernstein and Trahiotis, 1996). Thus, the SIED cue is hypothesized to be a specific implementation of an envelope-based IAC in explaining $N_0S_\pi$ responses.

### Neural binaural masking level differences vs. human binaural masking level differences

Rate-based thresholds were estimated for both $N_0S_0$ and $N_0S_\pi$ conditions in order to estimate neural BMLDs over a range of frequencies and noise levels. For the $N_0S_0$ condition, the lowest rate-based thresholds across frequency could explain human detection thresholds. For the $N_0S_\pi$ condition, the lowest rate-based thresholds across frequency had a different trend from human detection thresholds: neural thresholds were higher (i.e., worse) than human thresholds at low frequencies, and lower (i.e., better) than human thresholds at high frequencies. Many neurons had BMLDs as large as 20 dB. BMLDs estimated based on the most sensitive units in the neural population and estimates of maximum BMLDs for single neurons only varied slightly across frequency, whereas human BMLDs decrease substantially with increasing frequency. BMLDs estimated for the neural population were shown to be slightly lower than maximum single-neuron BMLDs across all frequencies, because individual neurons with the lowest thresholds in either the $N_0S_0$ or $N_0S_\pi$ condition did not always have the lowest thresholds in the other condition.

Rate-based neural thresholds were similar across noise levels, consistent with human psychophysical studies (Buss et al., 2003). Human BMLDs have been shown to be minimally affected by the roving-level paradigm, in which stimulus levels randomly vary from interval to interval (Henning et al., 2005). Similar patterns of rate-based neural BMLDs across noise levels could explain the level-resistance of human listeners.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://osf.io/kbrnw/.

## Ethics statement

This animal study was reviewed and approved by University of Rochester, University Committee on Animal Resources.

## Author contributions

LF designed and conducted the experiment, analyzed data, and wrote the manuscript. KH was involved in data analysis and provided feedback on the manuscript. LC proposed the general hypothesis, was involved in experimental design, data analysis, and edited the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This study was supported by NIH-DC-010813.

## Acknowledgments

Thanks to Kristina Abrams for help with animal husbandry and surgical procedures, to Margaret Youngman and Meron Abate for assistance with data collection, and to Douglas Schwarz for help with software and hardware development. Thanks to Emily Buss and Matthew Goupell for providing the human threshold values from their studies. Also thanks to Goupell for discussions about binaural cues. Ralf Haefner, Marc Schieber, and Steven McAleavey provided useful suggestions for analysis and presentation. We remember Armin Kohlrausch, who discussed this work with us and provided extensive comments on a previous version of the manuscript.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

# References


Asadollahi, A., Endler, F., Nelken, I., and Wagner, H. (2010). Neural correlates of binaural masking level difference in the inferior colliculus of the barn owl (*Tyto alba*). *Eur. J. Neurosci.* 32, 606–618. doi: 10.1111/j.1460-9568.2010.07313.x

Barlow, H. B., Levick, W. R., and Yoon, M. (1971). Responses to single quanta of light in retinal ganglion cells of the cat. *Vision Res.* 11, 87–101. doi: 10.1016/0042-6989(71)90033-2

Bernstein, L. R., and Trahiotis, C. (1996). On the use of the normalized correlation as an index of interaural envelope correlation. *J. Acoust. Soc. Am.* 100, 1754–1763. doi: 10.1121/1.416072

Bernstein, L. R., and Trahiotis, C. (1997). The effects of randomizing values of interaural disparities on binaural detection and on discrimination of interaural correlation. *J. Acoust. Soc. Am.* 102, 1113–1120. doi: 10.1121/1.419863

Bernstein, L. R., and Trahiotis, C. (2017). An interaural-correlation-based approach that accounts for a wide variety of binaural detection data. *J. Acoust. Soc. Am.* 141, 1150–1160. doi: 10.1121/1.4976098

Buss, E., Hall, J. W. III, and Grose, J. H. (2003). The masking level difference for signals placed in masker envelope minima and maxima. *J. Acoust. Soc. Am.* 114, 1557–1564. doi: 10.1121/1.1598199

Caird, D. M., Palmer, A. R., and Rees, A. (1991). Binaural masking level difference effects in single units of the Guinea pig inferior colliculus. *Hear. Res.* 57, 91–106. doi: 10.1016/0378-5955(91)90078-N

Cant, N. B., and Oliver, D. L. (2018). "Overview of auditory projection pathways and intrinsic microcircuits," in *The mammalian auditory pathways*, eds D. L. Oliver, N. B. Cant, R. R. Fay, and A. N. Popper (Cham: Springer), 7–39. doi: 10.1007/978-3-319-71798-2_2

Carney, L. H. (2018). Supra-threshold hearing and fluctuation profiles: Implications for sensorineural and hidden hearing loss. *J. Assoc. Res. Otolaryngol.* 19, 331–352. doi: 10.1007/s10162-018-0669-5

Colburn, H. S. (1977). Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. *J. Acoust. Soc. Am.* 61, 525–533. doi: 10.1121/1.381294

Culling, J. F. (2011). Subcomponent cues in binaural unmasking. *J. Acoust. Soc. Am.* 129, 3846–3855. doi: 10.1121/1.3560944

Domnitz, R. H., and Colburn, H. S. (1976). Analysis of binaural detection models for dependence on interaural target parameters. *J. Acoust. Soc. Am.* 59, 598–601. doi: 10.1121/1.380904

Durlach, N. I. (1964). Note on binaural masking-level differences at high frequencies. *J. Acoust. Soc. Am.* 36, 576–581. doi: 10.1121/1.1919006

Egan, J. P. (1975). *Signal detection theory and ROC-analysis*. Cambridge, MA: Academic press.

Fan, L. (2020). *Physiological studies of binaural tone-in-noise detection in the inferior colliculus: The role of stimulus envelope and neural fluctuations*. Unpublished dissertation. Rochester, NY: University of Rochester.

Fan, L., Henry, K. S., and Carney, L. H. (2021). Responses to diotic tone-in-noise stimuli in the inferior colliculus: Stimulus envelope and neural fluctuation cues. *Hear. Res.* 409:108328. doi: 10.1016/j.heares.2021.108328

Goupell, M. J. (2012). The role of envelope statistics in detecting changes in interaural correlation. *J. Acoust. Soc. Am.* 132, 1561–1572. doi: 10.1121/1.4740498

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *J. Acoust. Soc. Am.* 115, 833–843. doi: 10.1121/1.1639908

Henning, G. B., Richards, V. M., and Lentz, J. J. (2005). The effect of diotic and dichotic level-randomization on the binaural masking-level difference. *J. Acoust. Soc. Am.* 118, 3229–3240. doi: 10.1121/1.2047167

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *J. Acoust. Soc. Am.* 20, 536–544. doi: 10.1121/1.1906407

Jiang, D., McAlpine, D., and Palmer, A. R. (1997a). Responses of neurons in the inferior colliculus to binaural masking level difference stimuli measured by rate-versus-level functions. *J. Neurophysiol.* 77, 3085–3106. doi: 10.1152/jn.1997.77.6.3085

Jiang, D., McAlpine, D., and Palmer, A. R. (1997b). Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. *J. Neurosci.* 17, 9331–9339. doi: 10.1523/JNEUROSCI.17-23-09331.1997

Krishna, B. S., and Semple, M. N. (2000). Auditory temporal processing: Responses to sinusoidally amplitude-modulated tones in the inferior colliculus. *J. Neurophysiol.* 84, 255–273. doi: 10.1152/jn.2000.84.1.255

Lane, C. C., and Delgutte, B. (2005). Neural correlates and mechanisms of spatial release from masking: Single-unit and population responses in the inferior colliculus. *J. Neurophysiol.* 94, 1180–1198. doi: 10.1152/jn.01112.2004

Langner, G., and Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. *J. Neurophysiol.* 60, 1799–1822. doi: 10.1152/jn.1988.60.6.1799

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Mao, J., and Carney, L. H. (2014). Binaural detection with narrowband and wideband reproducible noise maskers. IV. Models using interaural time, level, and envelope differences. *J. Acoust. Soc. Am.* 135, 824–837. doi: 10.1121/1.4861848

McAlpine, D., Jiang, D., and Palmer, A. R. (1996). Binaural masking level differences in the inferior colliculus of the Guinea pig. *J. Acoust. Soc. Am.* 100, 490–503. doi: 10.1121/1.415862

Nelson, P. C., and Carney, L. H. (2007). Neural rate and timing cues for detection and discrimination of amplitude-modulated tones in the awake rabbit inferior colliculus. *J. Neurophysiol.* 97, 522–539. doi: 10.1152/jn.00776.2006

Palmer, A. R., Jiang, D., and McAlpine, D. (1999). Desynchronizing responses to correlated noise: A mechanism for binaural masking level differences at the inferior colliculus. *J. Neurophysiol.* 81, 722–734. doi: 10.1152/jn.1999.81.2.722

Ramachandran, R., Davis, K. A., and May, B. J. (2000). Rate representation of tones in noise in the inferior colliculus of decerebrate cats. *J. Assoc. Res. Otolaryngol.* 1, 144–160. doi: 10.1007/s101620010029

Richards, V. M. (1992). The detectability of a tone added to narrow bands of equal-energy noise. *J. Acoust. Soc. Am.* 91, 3424–3435. doi: 10.1121/1.402831

Robles, L., and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiol. Rev.* 81, 1305–1352. doi: 10.1152/physrev.2001.81.3.1305

Ruggero, M. A. (1992). Responses to sound of the basilar membrane of the mammalian cochlea. *Curr. Opin. Neurobiol.* 2, 449–456. doi: 10.1016/0959-4388(92)90179-O

Schmiedt, R. A. (1989). Spontaneous rates, thresholds and tuning of auditory-nerve fibers in the gerbil: Comparisons to cat data. *Hear. Res.* 42, 23–35. doi: 10.1016/0378-5955(89)90115-9

Schwarz, D. M., Zilany, M. S., Skevington, M., Huang, N. J., Flynn, B. C., and Carney, L. H. (2012). Semi-supervised spike sorting using pattern matching and a scaled Mahalanobis distance metric. *J. Neurosci. Methods* 206, 120–131. doi: 10.1016/j.jneumeth.2012.02.013

van de Par, S., and Kohlrausch, A. (1999). Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. *J. Acoust. Soc. Am.* 106, 1940–1947. doi: 10.1121/1.427942

van der Heijden, M., and Joris, P. X. (2010). Interaural correlation fails to account for detection in a classic binaural task: Dynamic ITDs dominate N0Spi detection. *J. Assoc. Res. Otolaryngol.* 11, 113–131. doi: 10.1007/s10162-009-0185-8

Whitehead, M. L., Lonsbury-Martin, B. L., and Martin, G. K. (1992). Evidence for 2 discrete sources of 2f1-F2 distortion-product otoacoustic emission in rabbit. 1. Differential dependence on stimulus parameters. *J. Acoust. Soc. Am.* 91, 1587–1607. doi: 10.1121/1.404382

Yin, T. C., Smith, P. H., and Joris, P. X. (2019). Neural mechanisms of binaural processing in the auditory brainstem. *Compr. Physiol.* 9, 1503–1575. doi: 10.1002/cphy.c180036

Zheng, L., Early, S. J., Mason, C. R., Idrobo, F., Harrison, J. M., and Carney, L. H. (2002). Binaural detection with narrowband and wideband reproducible noise maskers: II. Results for rabbit. *J. Acoust. Soc. Am.* 111, 346–356. doi: 10.1121/1.1423930

Zheng, Y., and Escabi, M. A. (2013). Proportional spike-timing precision and firing reliability underlie efficient temporal processing of periodicity and envelope shape cues. *J. Neurophysiol.* 110, 587–606. doi: 10.1152/jn.01080.2010

Zilany, M. S., Bruce, I. C., and Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *J. Acoust. Soc. Am.* 135, 283–286. doi: 10.1121/1.4837815

Zurek, P. M. (1991). Probability distributions of interaural phase and level differences in binaural detection stimuli. *J. Acoust. Soc. Am.* 90, 1927–1932. doi: 10.1121/1.401672

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 07 December 2022
DOI 10.3389/fnins.2022.1038856



OPEN ACCESS

EDITED BY
Huiming Zhang,
University of Windsor, Canada

REVIEWED BY
Kristina DeRoy Milvae,
University at Buffalo, United States
Ann Perreau,
Augustana College, United States

*CORRESPONDENCE
Emily A. Burg
emily.burg@wisc.edu

†PRESENT ADDRESS
Emily A. Burg,
Stanford Ear Institute, Stanford University, Stanford, CA, United States

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 07 September 2022
ACCEPTED 21 November 2022
PUBLISHED 07 December 2022

CITATION
Burg EA, Thakkar TD and Litovsky RY (2022) Interaural speech asymmetry predicts bilateral speech intelligibility but not listening effort in adults with bilateral cochlear implants. *Front. Neurosci.* 16:1038856. doi: 10.3389/fnins.2022.1038856



# Interaural speech asymmetry predicts bilateral speech intelligibility but not listening effort in adults with bilateral cochlear implants

Emily A. Burg[1,2*†], Tanvi D. Thakkar[3] and Ruth Y. Litovsky[1,2,4]

[1]Waisman Center, University of Wisconsin-Madison, Madison, WI, United States, [2]Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, United States, [3]Department of Psychology, University of Wisconsin-La Crosse, La Crosse, WI, United States, [4]Division of Otolaryngology, Department of Surgery, University of Wisconsin-Madison, Madison, WI, United States

**Introduction:** Bilateral cochlear implants (BiCIs) can facilitate improved speech intelligibility in noise and sound localization abilities compared to a unilateral implant in individuals with bilateral severe to profound hearing loss. Still, many individuals with BiCIs do not benefit from binaural hearing to the same extent that normal hearing (NH) listeners do. For example, binaural redundancy, a speech intelligibility benefit derived from having access to duplicate copies of a signal, is highly variable among BiCI users. Additionally, patients with hearing loss commonly report elevated listening effort compared to NH listeners. There is some evidence to suggest that BiCIs may reduce listening effort compared to a unilateral CI, but the limited existing literature has not shown this consistently. Critically, no studies to date have investigated this question using pupillometry to quantify listening effort, where large pupil sizes indicate high effort and small pupil sizes indicate low effort. Thus, the present study aimed to build on existing literature by investigating the potential benefits of BiCIs for both speech intelligibility and listening effort.

**Methods:** Twelve BiCI adults were tested in three listening conditions: Better Ear, Poorer Ear, and Bilateral. Stimuli were IEEE sentences presented from a loudspeaker at 0° azimuth in quiet. Participants were asked to repeat back the sentences, and responses were scored by an experimenter while changes in pupil dilation were measured.

**Results:** On average, participants demonstrated similar speech intelligibility in the Better Ear and Bilateral conditions, and significantly worse speech intelligibility in the Poorer Ear condition. Despite similar speech intelligibility in the Better Ear and Bilateral conditions, pupil dilation was significantly larger in the Bilateral condition.

**Discussion:** These results suggest that the BiCI users tested in this study did not demonstrate binaural redundancy in quiet. The large interaural speech asymmetries demonstrated by participants may have precluded them from

obtaining binaural redundancy, as shown by the inverse relationship between the two variables. Further, participants did not obtain a release from effort when listening with two ears versus their better ear only. Instead, results indicate that bilateral listening elicited increased effort compared to better ear listening, which may be due to poor integration of asymmetric inputs.

KEYWORDS

**listening effort, binaural hearing, pupillometry, speech intelligibility, bilateral cochlear implants, binaural redundancy, interaural asymmetry**

## Introduction

Patients with cochlear implants (CIs) commonly report that listening is exhausting. This is because listening requires effort, defined as the intentional focus of cognitive resources to perform listening tasks (Pichora-Fuller et al., 2016). The amount of mental resources allocated can be influenced by many different variables, including the environment (e.g., quiet versus noisy) and individual factors such as linguistic skills, working memory capacity, and audibility (Wendt et al., 2016; Winn et al., 2018). Additionally, the amount of effort a listener expends is thought to be influenced by their motivation to perform the task (Pichora-Fuller et al., 2016; Hughes et al., 2018). Thus, two individuals listening to the same conversation may exert different amounts of effort depending on how motivated they are to pay attention and understand what is being said (Winn et al., 2018). Listening effort is an important aspect of communication to investigate because elevated effort is associated with fatigue and stress, especially for individuals who must overcome additional listening obstacles like hearing loss. Compared to individuals with normal hearing (NH), studies have found that individuals with hearing loss report higher levels of effort and fatigue, are more likely to require recovery after work, and are more inclined to take sick-leave due to stress-related factors (Kramer et al., 2006; Kramer, 2008; Nachtegaal et al., 2009; Alhanbali et al., 2017). Additionally, the subjective feeling that one needs to exert elevated effort in complex listening situations has been associated with feelings of social isolation and anxiety in individuals with hearing loss (Hughes et al., 2018).

Patients with severe-to-profound hearing loss who struggle to understand speech with a hearing aid can receive a cochlear implant (CI). An increasing number of patients with hearing loss in both ears are now being bilaterally implanted to maximize speech perception and improve spatial hearing abilities. Compared to hearing aids or a unilateral CI, most individuals with bilateral CIs (BiCIs) demonstrate improvements in sound localization (Gantz et al., 2002; van Hoesel and Tyler, 2003; Laszig et al., 2004; Litovsky et al., 2004, 2009; Nopp et al., 2004; Grantham et al., 2007; Tyler et al., 2007) and speech understanding in noise (Gantz et al., 2002; van Hoesel et al., 2002; van Hoesel and Tyler, 2003; Litovsky et al., 2004, 2009; Nopp et al., 2004; Schleich et al., 2004; Tyler et al., 2007; Loizou et al., 2009). Further, advantages of BiCIs have also been documented using subjective questionnaires. Tyler et al. (2009) administered the Spatial Hearing Questionnaire (SHQ) to bilateral and unilateral cochlear implantees and found that BiCIs users rated their localization, speech understanding in quiet, and music perception abilities significantly higher than unilateral CI users. Similarly, using the SHQ, Perreau et al. (2014) found that BiCI users reported better subjective hearing performance on individual spatial hearing items as well as sound localization, music, and speech understanding in quiet subscales compared to unilateral CI or bimodal CI users. Together, these findings suggest that bilateral implantation provides both objective and subjective benefit on a variety of listening tasks compared to unilateral implantation.

Binaural redundancy is another benefit that can be derived from having access to sound in both ears. This phenomenon arises from access to duplicate copies of a signal that can be combined centrally, resulting in improved speech intelligibility and an increase in perceptual loudness (Litovsky et al., 2006; Avan et al., 2015). Mosnier et al. (2009) found a binaural redundancy benefit of 10% in quiet for BiCI listeners using disyllabic word stimuli. Similarly, BiCI users in Laszig et al. (2004) demonstrated a binaural redundancy benefit of 4% using open-set sentence stimuli. In contrast, the same group of listeners in Laszig et al. (2004) did not show a significant binaural benefit using a different open-set sentence corpus, and BiCI users in Goupell et al. (2018) also did not demonstrate a binaural redundancy benefit using the IEEE sentence corpus. At least some of the variability in binaural redundancy benefit appears to be related to interaural asymmetry (either in speech intelligibility or hearing history). When Mosnier et al. (2009) split listeners into symmetric and asymmetric groups based on the difference in speech scores across ears, symmetric listeners ($<$ 20% difference in percent correct across ears) demonstrated a significant binaural redundancy benefit, whereas asymmetric listeners did not. Yoon et al. (2011) used this same asymmetry criterion and measured binaural redundancy in quiet using sentences, consonants, and vowels. When averaging binaural redundancy for all three stimuli together, they observed

significant benefit in quiet in symmetric BiCI users, but not asymmetric BiCI users. Likewise, listeners in Goupell et al. (2018) were recruited based on their asymmetric hearing history or early onset of deafness and late implantation. Together these results suggest that interaural asymmetry may preclude binaural redundancy benefits in quiet. However, due to methodological differences between studies (i.e., definition of "asymmetry," stimuli used) this relationship warrants further investigation. We aim to examine this in the present study.

Historically, the primary measures of success regarding bilateral implantation have been bilateral speech intelligibility scores and spatial hearing abilities. There has been significantly less attention given to the potential impact of bilateral implantation on listening effort. Litovsky et al. (2006) administered a subjective questionnaire known as the Abbreviated Profile of Hearing Aid Benefit (APHAB) to BiCI users during a "bilateral deprivation" period in which participants only wore the CI of their better performing ear, and again several months later after participants had access to both of their CIs (Cox and Alexander, 1995; Litovsky et al., 2006). The APHAB contains 24 statements about everyday communication abilities or sound perception and asks participants to rate how often each statement is true. Statements are split into four subscales: Ease of Communication, Reverberation, Background Noise, and Aversiveness (Cox and Alexander, 1995). They found that participants perceived bilateral listening to be beneficial in background noise and reverberant environments and experienced increased ease of communication for bilateral compared to unilateral listening (Litovsky et al., 2006). Another study employed the Speech, Spatial, and Qualities of Hearing Scale, and found that individuals with two CIs expressed higher ability ratings on the spatial hearing domain, as well as segregation, naturalness, and listening effort aspects, compared to individuals with one CI (Noble et al., 2008). Together, these studies demonstrate that many patients subjectively experience reduced listening effort from BiCIs compared to a unilateral CI.

Another common method for quantifying listening effort is the behavioral dual-task paradigm. Hughes and Galvin (2013) used this method to assess listening effort during a speech-in-noise task in eight young BiCI users (aged 10–22 years) in unilateral and bilateral listening conditions. These listeners all had an early onset of deafness (before 1 year of age) and long inter-implant delays (mean = 7.8 years). They found that, on average, BiCI users demonstrated a significant reduction in listening effort when using two implants compared to one, however, on an individual level, this effect was only significant for three of the eight listeners (Hughes and Galvin, 2013). Another study asked 16 adult CI participants to repeat monosyllabic words in noise and found no difference in the dual-task or subjective measure of listening effort between unilateral CI and bimodal/bilateral CI listening (Sladen et al., 2018). Similarly, Perreau et al. (2017) found no difference in dual-task or subjective measures of listening effort between 10 unilateral CI users, 12 BiCI users, and 12 unilateral hybrid CI users. Due to the dearth of literature combined with the inconsistent results using either dual-task or subjective measures, we aimed to investigate listening effort with each CI alone and with BiCIs by measuring changes in pupil dilation. We chose this approach because pupillometry is considered to be an objective physiological measure of listening effort (Kramer et al., 1997; Zekveld et al., 2010; McGarrigle et al., 2014). To our knowledge, this is the first study to date that has examined this question using pupillometry.

Pupil dilation is modulated by cognitive load, increasing for difficult tasks that require more processing demand, and decreasing for tasks that are less challenging (Beatty, 1982). Mechanisms underlying the task-evoked pupil response include the activity of noradrenergic neurons in the locus coeruleus (Aston-Jones and Cohen, 2005). When the task becomes so difficult that listeners may feel that additional effort would not benefit performance, motivation declines, and pupil dilation decreases (Pichora-Fuller et al., 2016; Ohlenforst et al., 2017; Wendt et al., 2018). This effect has been shown for listening tasks that measure speech intelligibility. Pupil dilation increases as performance decreases to ∼30% correct, after which pupil dilation then decreases, presumably due to a decline in motivation and engagement (Ohlenforst et al., 2017; Wendt et al., 2018). Pupillometry is an ideal technique for studying listening effort in the hard of hearing population because it has the advantage of being compatible with assistive devices like hearing aids and CIs (Gilley et al., 2006; Friesen and Picton, 2010; Wagner et al., 2019). Additionally, unlike a dual-task paradigm, which is subject to behavioral bias and relies on a single metric such as response time (McGarrigle et al., 2014; Gagné et al., 2017), pupil dilation is completely objective and can be measured throughout the duration of a behavioral listening task to capture mental effort as it unfolds over time (Winn et al., 2018).

In short, the purpose of this study was to investigate the potential benefits of bilateral listening in performance and effort domains, both of which are important for successful communication. To do this, we measured speech intelligibility and listening effort in adults with BiCIs in three conditions: with their poorer ear only, better ear only, and bilaterally. Based on previous work that has shown binaural redundancy benefit in quiet (Laszig et al., 2004; Mosnier et al., 2009) and a reduction in listening effort for bilateral compared to unilateral CI listening (Litovsky et al., 2006; Noble et al., 2008; Hughes and Galvin, 2013), we predicted that speech intelligibility would be better (binaural redundancy) and pupil dilation would be smaller (release from effort) for BiCI users listening with both implants compared to their better ear only. Further, due to the accumulating evidence indicating an association between asymmetry and binaural benefits, we predicted that interaural

speech asymmetry would be negatively related to binaural redundancy.

# Materials and methods

## Participants

Twelve native English-speaking adults with BiCIs were recruited to participate in this experiment (age range 25–78 years). **Table 1** provides demographic information for these participants; 11 were implanted with Cochlear Ltd., devices, and one (IDI) was implanted with Advanced Bionics devices. Participants traveled to Madison, Wisconsin to participate in multiple studies over the course of several days. Testing for the present study took place over the course of one 2-h session. This study was approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board.

## Experimental setup

Testing took place in a standard sound booth (IAC Acoustics, IL, USA). Participants were seated at a table with their chin and forehead supported in a headrest to keep their head stable during testing; the table and chair position and height were adjusted for each participant. A computer monitor was attached to the table and positioned approximately 65 cm away from the headrest. The eyetracker camera was secured to the table using a desktop mount 8 cm in front of the monitor. Illumination of the test room was controlled for all participants (93 lux). Stimuli were played to a loudspeaker (Tannoy, Coatbridge, Scotland) positioned at 0° azimuth. Pupil size was measured in pixels using the "Area" setting on an eyetracker (Eyelink 1000 Plus; SR Research, Ontario, Canada) and a sampling rate of 1,000 Hz.

## Stimuli

Stimuli were drawn from the Institute of Electrical and Electronics Engineers sentence corpus (IEEE, 1969) and were recorded by a female talker. All stimuli were scaled to 65 dB SPL-A and played to the loudspeaker through a USB high-speed audio interface (RME Fireface, Haimhausen, Germany). Duration of sentences ranged from 4,000 to 6,000 ms. Custom software written in MATLAB (The MathWorks, Natick, MA, USA) with PsychToolbox 3 was used to deliver stimuli and collect data (Brainard, 1997; Kleiner et al., 2007).

## Procedure

Participants were tested in three listening conditions: better ear CI only ("Better Ear"), poorer ear CI only ("Poorer Ear"), and both CIs ("Bilateral"). Prior to testing, the better ear was classified as the ear with the higher word recognition score measured in the audiology clinic. If there was no difference in word recognition score between the two ears, the participant's preferred ear according to subjective reporting was labeled the "better" ear. Participants were tested using their clinical programs with noise reduction and beamforming settings disabled. Before beginning the experiment, an informal interaural loudness balance check was completed with participants wearing both CI processors together to verify that they were equal in loudness. An experimenter stood directly in front of participants at the same distance as the loudspeaker and asked participants whether the ears were equally loud and sound was centered between the two ears. If participants perceived one CI to be noticeably louder than the other, the volume settings were adjusted so that the ears were balanced. Participants completed a familiarization procedure in which they listened to and repeated 10 sentences in each condition. Stimuli for

TABLE 1 Participant demographics.

| Subject ID | Sex | Age (years) | First implant | Better Ear | Inter-implant delay (years) | Bilateral CI experience (years) |
|---|---|---|---|---|---|---|
| ICW | F | 25 | Right | Right | 18.6 | 4.9 |
| IBZ | F | 51 | Right | Right | 1.3 | 11.0 |
| IDI | F | 52 | Right | Right | 0.6 | 4.6 |
| IBY | F | 55 | Left | Right | 4.2 | 7.3 |
| ICP | M | 56 | Left | Left | 3 | 7.3 |
| ICD | F | 61 | Right | Left | 6.0 | 10.0 |
| ICB | F | 67 | Right | Left | 2.8 | 12.9 |
| ICJ | F | 69 | N/A | Right | 0.0 | 8.8 |
| IDG | F | 70 | Left | Right | 2.0 | 7.7 |
| IBL | F | 72 | Left | Right | 4.8 | 12.8 |
| ICK | M | 75 | Right | Left | 1.0 | 7.2 |
| IBK | M | 78 | Left | Left | 6.0 | 9.8 |

practice trials were randomly selected and then excluded from the test corpus.

During testing, participants were asked to fixate their gaze on a small gray cross in the center of the computer screen and attend to open-set target sentences presented by a loudspeaker positioned directly in front of them (0° azimuth). Participants were instructed to repeat the sentence that was heard. Prior to the start of each trial, the gray cross turned white to indicate that the trial was about to commence. This was followed by a 2,000 ms pre-trial interval and then the trial began with a 1,000 ms baseline pupil measurement in silence before the stimulus (IEEE sentence) was presented. Following stimulus offset, participants were given a 2,000 ms silent period before the cross turned green and two beeps were presented, prompting participants to repeat what they heard. Each sentence contained five key words that were scored by an experimenter. The experimenter waited 10–15 s between trials to allow the pupil to return to baseline before beginning the next trial. Participants completed 30 trials per listening condition (30 sentences × 3 conditions = 90 sentences total). Trials were blocked into two runs per listening condition (15 sentences/run) and condition order was randomized for each participant. Target sentences were randomly selected from the corpus without replacement. Participants were given regular breaks during testing to avoid fatigue.

## Data analysis

Prior to data analysis, pupil data were pre-processed to reduce artifacts and discard noisy trials. First, pupil tracks with greater than 45% blinks were discarded from analysis (Burg et al., 2021). This blink criterion was chosen because it is more inclusive compared to other commonly used criteria (e.g., 15%, 30%). Previous work has shown a positive association between task difficulty and blink percentage; therefore, an overly conservative blink criterion like 15% could result in a higher number of difficult trials being excluded from analysis, potentially confounding results (Burg et al., 2021). When calculating the percentage of blinks in a track, samples from the response period were not considered since this part of the pupil track is influenced by the motor response (Privitera et al., 2010; Winn et al., 2015). Blinks were detected by tagging samples that fell below three standard deviations (SDs) from the mean (Zekveld et al., 2010). Consistent with best-practices described by Winn et al. (2018) tracks with irregular baselines, extreme distortions, or atypically large growth that is not consistent with task-evoked changes in pupil dilation were also discarded. In total, 1.4, 1.9, and 2.5% of trials were discarded due to these kinds of contamination for the Better Ear, Bilateral, and Poorer Ear condition, respectively.

The second step in pre-processing was an interpolation process, whereby individual tracks were "de-blinked" by linearly interpolating 80 ms before a blink and 160 ms following a blink to account for eyelid disturbances, and low-pass filtered using the "smooth" function in MATLAB (Zekveld et al., 2010). Next, raw pupil dilation was transformed to proportional change from baseline by subtracting the baseline value (average of first 1,000 ms of each trial) and then dividing by the baseline value. Baseline pupil dilation was compared across conditions to ensure that there were no systematic differences that would influence results. Divisive baseline correction was chosen over subtractive baseline correction because the former accounts for differences in pupil reactivity across participants and across trials for individuals (Winn et al., 2018). Finally, remaining tracks were time-aligned to stimulus offset and averaged together by listening condition for each participant. From the averaged trials, maximum pupil dilation and percentage of correctly repeated words were calculated and extracted for each condition. Maximum pupil dilation was extracted from the "silent period" (i.e., 2,000 ms period after stimulus offset and prior to response prompt), because this processing window has consistently been shown to elicit the largest pupil size during the trial for sentence recognition tasks (Zekveld et al., 2010; Winn et al., 2015, 2018).

## Statistical analysis

Speech intelligibility scores were transformed from percent correct to rationalized arcsine units (RAU) to alleviate ceiling effects and normalize variance (Studebaker, 1985). The effect of listening condition on speech intelligibility and listening effort were each evaluated separately using one-way repeated measure analysis of variance (ANOVA) tests with listening condition (three levels: Better Ear, Poorer Ear, Bilateral) as the independent variable. For these ANOVAs, dependent variables were either speech intelligibility (RAU) or maximum proportional change in pupil dilation (peak pupil size during the silent period). *Post hoc* pairwise comparisons were completed using paired *t*-tests. Benjamini-Hochberg corrections were employed to control false discovery rate (Benjamini and Hochberg, 1995). Correlational analyses were conducted to examine potential relationships between interaural speech asymmetry, change in speech intelligibility (RAU) from Better Ear to Bilateral conditions, and change in listening effort (pupil dilation) from Better Ear to Bilateral conditions. Assumptions for omnibus, *post hoc* tests, and correlations were statistically evaluated using Mauchly's Test of Sphericity and Shapiro–Wilk normality tests. Due to our directional hypothesis that interaural speech asymmetry would be inversely related to change in speech intelligibility from the Better Ear to Bilateral condition (i.e., binaural redundancy), a one-sided test was used to evaluate this relationship. The relationship between interaural speech asymmetry and change in listening effort from Better Ear to Bilateral conditions was evaluated with a two-sided test. An

alpha of 0.05 was used to determine whether results were statistically significant.

# Results

## Speech intelligibility

Mean speech intelligibility (RAU) for each listening condition is shown in Figure 1. Speech intelligibility was higher for the Better Ear and Bilateral conditions than the Poorer Ear condition (Better Ear mean ± SD = 83.2 ± 24.4; Bilateral = 84.8 ± 26.9; Poorer Ear = 62.2 ± 35.2). Notably, there was substantial inter-subject variability in performance, as demonstrated by the wide range of performance (Better Ear = 21–105 RAUs; Bilateral = 19–114 RAUs; Poorer Ear = –5–104 RAUs) and large standard deviation for all conditions. There was also considerable variability in the amount of interaural asymmetry demonstrated by participants, which ranged from 2 to 65 RAUs (Table 2). A one-way repeated measures ANOVA using a Greenhouse–Geisser correction revealed a significant main effect of listening condition on speech intelligibility [$F(2,22) = 13.4$, $p < 0.01$]. *Post hoc* pairwise comparisons revealed that speech intelligibility did not significantly differ between Better Ear and Bilateral conditions ($p = 0.55$), indicating that, on average, participants did not demonstrate a binaural redundancy benefit. Further, speech intelligibility was significantly worse for the Poorer Ear condition compared to the Better Ear ($p < 0.01$) and Bilateral conditions ($p < 0.01$). Finally, Figure 2 plots binaural redundancy as a function of interaural speech asymmetry, with a higher positive value indicating greater binaural redundancy. Consistent with previous work reporting an association between asymmetry and binaural redundancy benefit (Litovsky et al., 2006; Mosnier et al., 2009; Yoon et al., 2011), a Pearson correlation revealed a significant negative relationship between the two variables, indicating that less speech asymmetry was associated with greater binaural redundancy benefit ($r = -0.61$, $p < 0.05$, one-tailed).

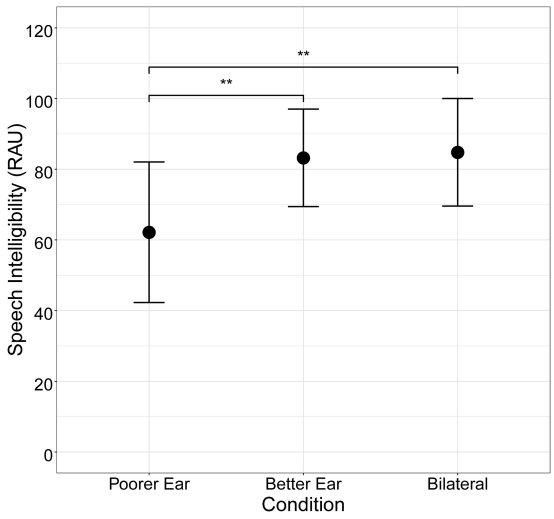



FIGURE 1
Mean speech intelligibility (RAU; $n = 12$) for each listening condition. Error bars represent ± 1.96 SE (95% confidence interval). Asterisks indicate the significance level of pairwise comparison results (* for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$).

## Listening effort

Grand average pupil tracks for each condition (with 95% confidence intervals) are shown in Figure 3. In general, average pupil dilation during the silent period was largest for the Poorer Ear condition, followed by the Bilateral condition, and finally the Better Ear condition. Maximum pupil dilation was extracted from this period and is plotted in Figure 4. Maximum pupil dilation was smallest for the Better Ear condition and similar for Poorer Ear and Bilateral conditions (Better Ear = 0.23 ± 0.15; Poorer Ear = 0.27 ± 0.12; Bilateral = 0.28 ± 0.15). A one-way repeated measures ANOVA revealed that the main effect of listening condition was not significant [$F(2,22) = 2.4$, $p = 0.1$]. However, *F*-tests have the potential to lead to either false positives or false negatives; thus, the pairwise comparisons can be informative regardless of the omnibus result (Chen et al., 2018). Indeed, *post hoc* testing revealed that pupil dilation was significantly larger for the Bilateral condition compared to the Better Ear condition ($p < 0.05$). Contrary to our prediction, this indicates that participants exerted greater effort or engagement when listening bilaterally than with their better ear only. There were no significant differences between the Poorer Ear and Better Ear conditions ($p = 0.24$), or between the Poorer Ear and Bilateral conditions ($p = 0.74$).

TABLE 2 Interaural speech asymmetry for each participant, defined as the difference in RAU scores between the Better Ear and Poorer Ear conditions.

| | Subject ID | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ICW | ICJ | IBL | ICP | ICK | ICB | IDG | IBK | IDI | ICD | IBY | IBZ |
| Interaural speech asymmetry | 64.5 | 51.4 | 43.4 | 26.3 | 22.7 | 15.7 | 10.7 | 9.1 | 7.0 | 5.2 | 2.0 | –5.4 |

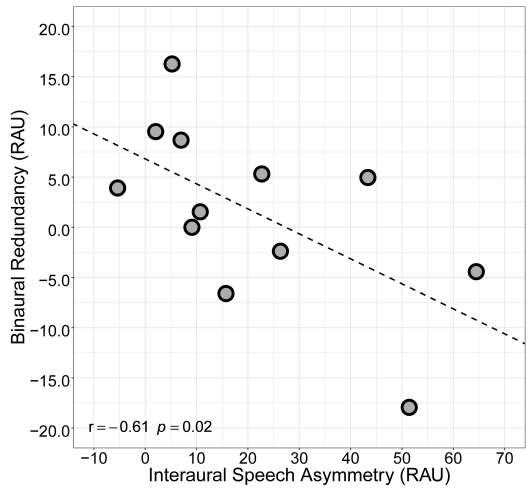



FIGURE 2
Relationship between interaural speech asymmetry (RAU) and binaural redundancy, defined as the difference in speech intelligibility (RAU) between Bilateral and Better Ear conditions.

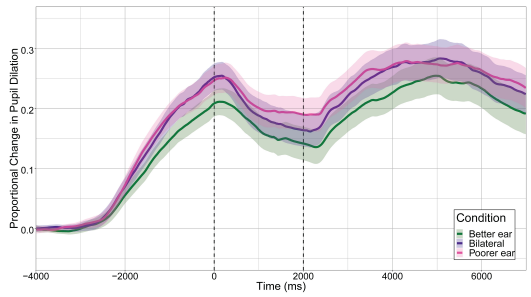



FIGURE 3
Grand average pupil tracks ($n$ = 12) for each listening condition. Maximum proportional change in pupil dilation was extracted from the silent period, indicated by the vertical dashed lines (0–2,000 ms). Shaded regions represent ± 1.96 SE (95% confidence interval).

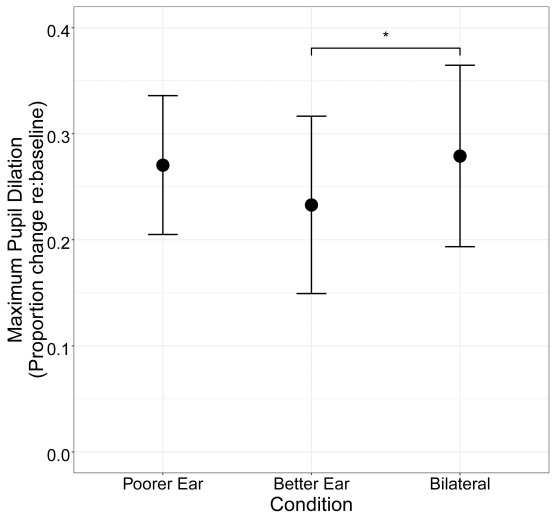



FIGURE 4
Mean maximum proportional change in pupil dilation ($n$ = 12) for each listening condition. Error bars represent ± 1.96 SE (95% confidence interval). Asterisks indicate the significance level of pairwise comparison results (* for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$).

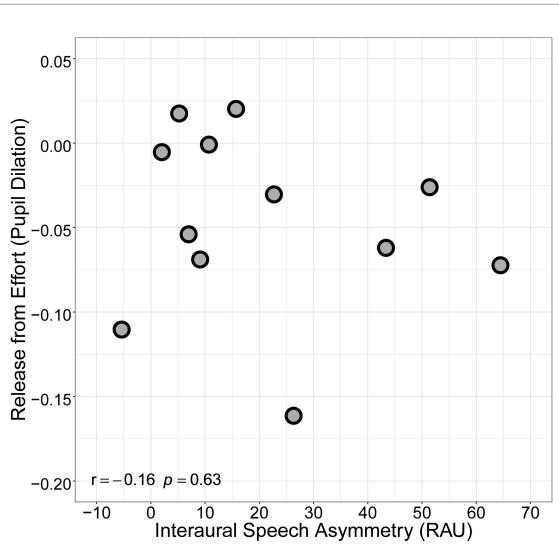



FIGURE 5
Relationship between interaural speech asymmetry (RAU) and release from listening effort, defined as the difference in maximum pupil dilation between Better Ear and Bilateral conditions.

Finally, we examined whether interaural speech asymmetry was related to release from effort (Figure 5). Release from effort was calculated as the difference in maximum pupil dilation between the Better Ear and Bilateral conditions, with a higher positive value indicating a greater reduction in pupil dilation (and effort) when listening bilaterally. A Pearson correlation indicated that interaural speech asymmetry was not related to release from effort ($r$ = −0.16, $p$ = 0.63, two-tailed).

## Discussion

This study measured speech intelligibility and listening effort in adults with BiCIs to examine whether bilateral listening provides a benefit above the better ear alone. Speech intelligibility was significantly worse in the Poorer Ear condition compared to the Better Ear and Bilateral conditions, and there was no significant difference between performance in the Better Ear and Bilateral conditions. This indicates that, on average, the BiCI users in the present study had significant asymmetry in speech intelligibility

across ears, but this asymmetry did not negatively affect performance in the Bilateral condition since performance was similar to the Better Ear condition. This is not surprising considering that listeners were tested in quiet and could rely on their better ear for speech intelligibility in the Bilateral condition. Further, pupil dilation was significantly larger in the Bilateral compared to Better Ear condition, and there was no significant difference between either of these and the Poorer Ear condition. This suggests that, on average, the BiCI users tested in this study did not obtain a performance benefit from binaural redundancy, nor did they obtain a release from effort when listening with two CIs versus their better ear alone.

## Interaural speech asymmetry predicts binaural redundancy benefit

The lack of measurable binaural redundancy benefit in the present study contrasts with results from Laszig et al. (2004) and Mosnier et al. (2009), which both reported significant binaural redundancy benefit for their BiCI listeners. However, there are noteworthy demographic differences between their participants and participants in the present study. Mosnier et al. (2009) required that their BiCI participants had less than a 5-year difference in duration of deafness between the two ears and were simultaneously implanted. Listeners in the present study, on the other hand, had variable differences in duration of deafness across ears, and inter-implant delays ranging from 0 to 18 years. Thus, our group of BiCI listeners was more heterogeneous and included listeners with asymmetric hearing histories. Similarly, participants in Laszig et al. (2004) did not demonstrate significant interaural speech asymmetry, whereas our BiCI participants exhibited large interaural speech asymmetries, with an average of $21.0 \pm 22$ RAU difference across ears. These observations indicate that interaural asymmetry may be key to understanding why our BiCI users, on average, did not demonstrate binaural redundancy. Consistent with this theory, Mosnier et al. (2009) and Yoon et al. (2011) split their participants into symmetric and asymmetric groups based on the difference in speech intelligibility across ears and found that only the symmetric groups demonstrated significant binaural redundancy benefit. Additionally, Goupell et al. (2018) failed to find a significant binaural redundancy benefit in BiCI listeners with asymmetric hearing histories or early onset of deafness and late implantation. These findings suggest that interaural asymmetries in hearing history and speech intelligibility may limit listeners' ability to benefit from binaural redundancy. Indeed, we found that interaural speech asymmetry was inversely related to binaural redundancy in the present study (**Figure 2**), suggesting that the relatively large speech asymmetries demonstrated by our BiCI listeners (as compared to listeners in Laszig et al., 2004) may have limited their ability to successfully combine input from both ears and benefit from binaural redundancy.

While binaural redundancy was not observed at the group level, the majority of listeners demonstrated improved performance in the Bilateral condition compared to the Better Ear condition. The largest binaural redundancy benefits were demonstrated by ICD (16 RAUs), IBY (10 RAUs), and IDI (9 RAUs; **Figure 6**). These listeners all demonstrated relatively small interaural asymmetries of 7 RAUs or less (**Table 2**). ICD had the second longest inter-implant delay of 6 years but also had 10 years of bilateral experience prior to testing, while IDI had the second shortest inter-implant delay of 0.6 years but only 5 years of bilateral experience prior to testing. In contrast, four listeners (ICW, ICJ, ICP, ICB) demonstrated worse performance in the Bilateral condition compared to the Better Ear condition. Three of these listeners demonstrated interaural asymmetries greater than 20 RAUs (**Table 2**), which was the percent correct criterion used by Mosnier et al. (2009) to categorize listeners into symmetric and asymmetric groups. The greatest decrement in performance from the Better Ear to Bilateral condition (18 RAUs) was shown by ICJ (**Figure 6B**). This participant had the second largest interaural asymmetry (51 RAUs). Interestingly, ICJ was simultaneously implanted, and had almost 9 years of bilateral CI experience prior to testing. In contrast, the participant with the largest interaural speech asymmetry (ICW: 65 RAUs; **Figure 6A**) and the longest inter-implant delay (18.6 years) only demonstrated a 4 RAU decrease in performance from the Better Ear to Bilateral condition. These are prime examples of the extreme variability that exists among BiCI users, and how difficult it can be to predict outcomes due to the vast number of variables that contribute to performance in each ear and across ears (Gantz et al., 2002; Litovsky et al., 2006; Mosnier et al., 2009).

## Bilateral listening is more effortful than better ear listening

Unlike previous studies that have shown that BiCIs may facilitate reduced listening effort compared to a unilateral CI (e.g., Litovsky et al., 2006; Noble et al., 2008; Hughes and Galvin, 2013), results from the present study indicate that, on average, bilateral listening elicited increased listening effort compared to better ear listening. In fact, out of the 12 BiCI participants tested, only two demonstrated a reduction in pupil dilation from the Better Ear to Bilateral condition (ICB and ICD, **Figure 6**, panels F and J). This is the first study to date that has shown this effect. Further, our results indicate that this increase in listening effort cannot be explained by a change in speech intelligibility, since there was no significant difference in performance between the Better Ear and Bilateral conditions. This is further supported by our correlation analysis that found no relationship between binaural redundancy and release from effort ($r = 0.15$, $p = 0.65$,

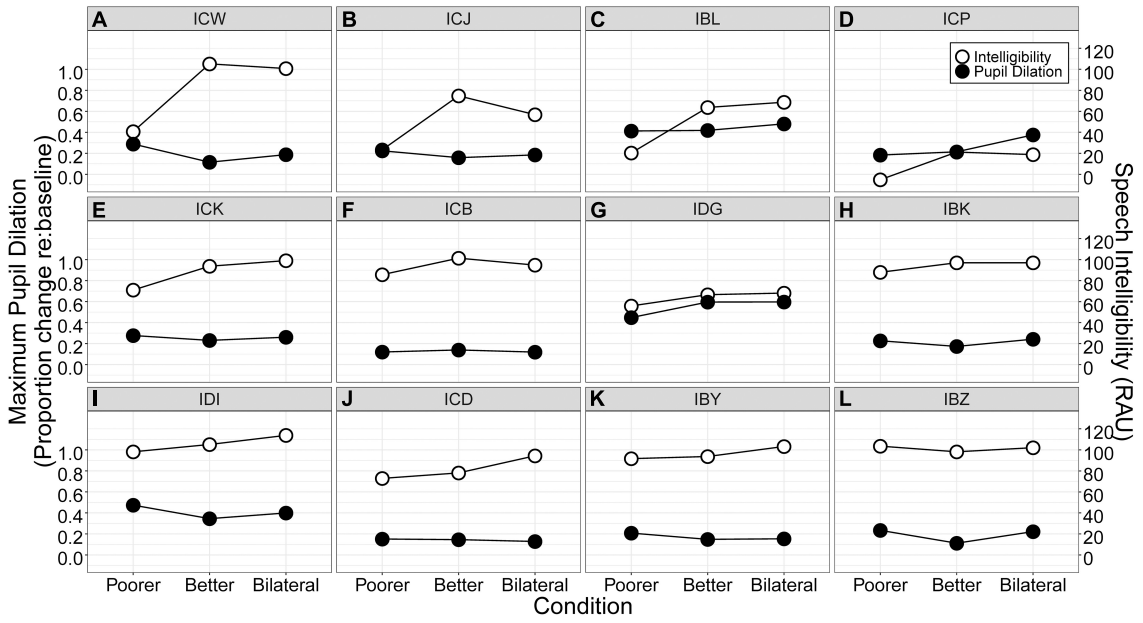



FIGURE 6
Speech intelligibility (RAU; open circles) and maximum proportional change in pupil dilation (closed circles) for each participant. Participants **(A–L)** are ordered from largest to smallest interaural speech asymmetry.

two-tailed). Indeed, previous studies have shown differences in listening effort across conditions when speech intelligibility is held constant (e.g., Koelewijn et al., 2012). This underscores the value of measuring listening effort in studies examining speech intelligibility, as it can reveal additional information that is not apparent from performance alone.

To obtain a binaural redundancy benefit, listeners must be able to centrally combine information across ears (Litovsky et al., 2006). Results of the present study indicate that this ability was largely inaccessible to our group of BiCI users due to the large degree of interaural speech asymmetry observed. One reason that asymmetries may preclude binaural redundancy is that it is difficult to combine disparate signals into one coherent sound, which may in turn result in increased listening effort. In other words, increased effort in the Bilateral condition may be explained by a lack of binaural fusion. Steel et al. (2015) examined binaural fusion and listening effort in children with BiCIs. They found that poorer binaural fusion was associated with greater pupil dilation and longer reaction times. Further, larger brainstem asymmetries, classified by mismatched electrically evoked auditory brainstem latencies, were associated with worse binaural fusion abilities (Steel et al., 2015). Indeed, we also found a relationship between our measures of asymmetry (interaural difference in speech intelligibility) and binaural integration (binaural redundancy) (**Figure 2**). This suggests that increased listening effort in the Bilateral condition may be related to poor binaural fusion due to the relatively large interaural speech asymmetries demonstrated by our BiCI listeners. Pragmatically, it makes sense that attempting to integrate two disparate signals, or ignore an impoverished signal from the poorer ear, would require more effort than simply attending to the better ear alone. This theory is supported by previous work that has demonstrated impaired binaural fusion in BiCIs users (Fitzgerald et al., 2015) that is exacerbated by asymmetries, such as interaural place-of-stimulation mismatch (Kan et al., 2013). While degree of speech asymmetry was not significantly correlated with release from effort, this does not disqualify the possibility that the two are related in some way since the relationship was assessed using a simple linear correlation, and pupil dilation does not always scale linearly with task difficulty (Koelewijn et al., 2012; Ohlenforst et al., 2017; Wendt et al., 2018). Further, there is also evidence that BiCI users have abnormally broad pitch fusion ranges and that bimodal CI users with a hearing aid in the contralateral ear can experience interference, a decrease in performance when listening with two ears versus one. This may arise from involuntary fusion of disparate inputs (Reiss et al., 2016, 2018). Thus, it is also possible that BiCI users in the present study experienced unfavorable fusion, making it more difficult to understand the target speech. This effect might not have been reflected by speech intelligibility scores because listeners may have been able to compensate by using context clues to repair missing or ambiguous information, ultimately requiring more effort (Winn, 2016).

Alternatively, it is also possible that increased pupil dilation in the Bilateral condition represents increased engagement

in the speech intelligibility task. Previous work has shown that pupil dilation increases with increasing task performance until the task becomes so difficult that increased effort is unlikely to improve performance (Ohlenforst et al., 2017; Wendt et al., 2018). In other words, listeners will continue to be engaged in a task so long as they perceive a potential benefit. Additionally, stimulus or task value to the participant can modulate engagement and pupil dilation even when speech is equally intelligible (Eckert et al., 2016; Winn et al., 2018). Because participants are accustomed to listening with both CIs in daily life, they may have expected to perform best in the Bilateral condition, resulting in increased pupil dilation due to greater engagement or motivation. Since we did not explicitly measure task engagement, we cannot disentangle engagement or motivation from effort in the present study. Another possibility is that increased loudness due to binaural summation contributed to greater pupil dilation in the Bilateral condition compared to either monaural condition. Indeed, Legris et al. (2022) demonstrated increasing maximum pupil diameter with increasing tone burst level (40 dBA, 60 dBA, 80 dBA) in both NH participants and hearing aid users. For NH participants, pupil dilation was significantly larger for all increases in level, but for hearing aid users, pupil dilation was only significantly different when comparing the 40 dBA condition to 80 dBA condition, regardless of whether or not participants were using their hearing aids. The 20 dB step size used by Legris et al. (2022) corresponds to a fourfold increase in loudness, whereas an increase of about 3 dB, as is typical for binaural summation, only corresponds to a 1.2-fold increase in loudness (Epstein and Florentine, 2012). Thus, the need to use large step sizes, especially in the hearing aid user group, indicates that the potential 3 dB of binaural summation experienced by BiCI users in the present study is very unlikely to have caused any significant change in pupil dilation. This is further supported by Nunnally et al. (1967) who only saw a significant effect of intensity on pupil dilation for very loud levels above 90 dB.

As mentioned previously, this is the first study to find bilateral listening to be more effortful than unilateral listening in BiCI users. One reason for the discrepancy between the present results and previous work may be the method used to gauge listening effort, as this was also the first study to investigate this question using pupillometry. In general, studies that have employed both subjective rating and pupillometry to measure effort have found that the two measures are typically uncorrelated (e.g., Zekveld et al., 2011; Zekveld and Kramer, 2014; Wendt et al., 2016). Lack of correspondence between these measures is likely related to participants' subjective interpretation of what is "effortful" (Colby and McMurray, 2021). For example, some participants may base their effort rating on their performance accuracy rather than mental effort, resulting in a linear relationship between accuracy and subjective effort, whereas the relationship between accuracy and objective effort measured by pupil dilation has been shown to be non-monotonic (Koelewijn et al., 2012; Ohlenforst et al., 2017; Wendt et al., 2018; Winn and Teece, 2021). Finally, another important factor to consider is the unilateral comparison condition (i.e., Better versus Poorer Ear). The comparison between participants' best possible unilateral listening condition (i.e., the Better Ear condition) and the Bilateral condition reveals changes explicitly due to bilateral listening. In contrast, if one were to compare the Poorer Ear condition to the Bilateral condition, it would be unclear whether changes in effort are simply due to the addition of the better ear or are explicitly related to bilateral listening. Litovsky et al. (2006) and Sladen et al. (2018) compared better ear listening to bilateral listening, but Hughes and Galvin (2013) did not report which ear their unilateral condition represented. If the unilateral condition represented the poorer performing ear or a mixture of poorer and better performing ears, a comparison of their results to the present study would be invalid.

## Limitations

The present study tested participants in quiet, which resulted in near-ceiling level performance for some listeners. While previous work has measured significant binaural redundancy benefit in BiCI users in quiet conditions (e.g., Laszig et al., 2004; Mosnier et al., 2009; Yoon et al., 2011) this benefit can be larger in noise conditions (e.g., Yoon et al., 2011). Measuring differences in speech intelligibility and pupil dilation from better ear to bilateral listening in both quiet and noise would elicit a wider range of performance and ultimately help elucidate whether bilateral CI listening is more effortful or more engaging than unilateral CI listening. Further, while subjective reports of listening effort do not always correlate with pupillometry results (e.g., Zekveld et al., 2011; Zekveld and Kramer, 2014; Wendt et al., 2016), it could nonetheless be interesting to compare the two metrics in future studies. Finally, a subjective measure that attempts to disentangle engagement/motivation from effort could be very useful for virtually any future study using pupillometry to gauge listening effort.

## Summary and conclusion

The present study measured speech intelligibility and pupil dilation to quantify differences in performance and listening effort in adults with BiCIs when listening with their poorer ear only, better ear only, or bilaterally in quiet. Previous studies have shown that some BiCI users demonstrate an increase in performance from better ear to bilateral listening. This was not observed in the present study, as BiCI users performed similarly when listening with their better ear only and bilaterally.

The large interaural speech asymmetries demonstrated by our BiCI users may have precluded them from obtaining binaural redundancy benefit, as shown by the significant negative relationship between the two factors. Additionally, listeners exhibited an increase in pupil dilation for bilateral compared to better ear listening, indicating that bilateral listening was more effortful. Due to the substantial interaural asymmetries demonstrated by our participants (in speech intelligibility and hearing history) we propose that increased listening effort may be due to difficulty combining two disparate signals. In conclusion, these results indicate that interaural speech asymmetries can impede BiCI patients' ability to access binaural redundancy and may provoke increased listening effort for bilateral compared to better ear listening. Therefore, investigating methods for reducing interaural asymmetries seems to be a promising direction for future research seeking to improve binaural hearing outcomes in BiCI patients.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by the University of Wisconsin-Madison Health Sciences Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

EB, TT, and RL conceived the original idea and designed the study. RL supervised all aspects of the study. TT created custom software to collect and analyze the data. EB and TT tested the participants. EB processed the experimental data, performed the analysis, drafted the manuscript, and designed the figures with support and input from TT and RL. All authors discussed the results and provided critical feedback on the manuscript.

## Funding

This research was supported by NIH-NIDCD Grant No. RO1DC03083 (to RL) and NIH-NICHD Grant No. U54HD090256 (to Waisman Center).

## Acknowledgments

We thank Shelly Godar and members of the Binaural Hearing and Speech Lab at the University of Wisconsin-Madison for assistance with participant recruitment, scheduling, and testing. We also thank the individuals who participated in this study. Portions of this work were presented at the 2019 American Academy of Audiology Conference, 2019 Association for Research in Otolaryngology Conference, and 2019 Conference on Implantable Auditory Prostheses.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Alhanbali, S., Dawes, P., Lloyd, S., and Munro, K. J. (2017). Self-reported listening-related effort and fatigue in hearing-impaired adults. *Ear Hear.* 38, e39–e48. doi: 10.1097/AUD.0000000000000361

Aston-Jones, G., and Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu. Rev. Neurosci.* 28, 403–450. doi: 10.1146/annurev.neuro.28.061604.135709

Avan, P., Giraudet, F., and Büki, B. (2015). Importance of binaural hearing. *Audiol. Neurotol.* 20(Suppl. 1), 3–6. doi: 10.1159/000380741

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91, 276–292. doi: 10.1037/0033-2909.91.2.276

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357

Burg, E. A., Thakkar, T., Fields, T., Misurelli, S. M., Kuchinsky, S. E., Roche, J., et al. (2021). Systematic comparison of trial exclusion criteria for pupillometry data analysis in individuals with single-sided deafness and normal hearing. *Trends Hear.* 5, 1–17. doi: 10.1177/23312165211013256

Chen, T., Xu, M., Tu, J., Wang, H., and Niu, X. (2018). Relationship between omnibus and post-hoc tests: an investigation of performance of the F test in

ANOVA. *Shanghai Arch Psychiatry* 30, 60–64. doi: 10.11919/j.issn.1002-0829.218014

Colby, S., and McMurray, B. (2021). Cognitive and physiological measures of listening effort during degraded speech perception: relating dual-task and pupillometry paradigms. *J. Speech Lang. Hear. Res.* 64, 3627–3652. doi: 10.1044/2021_JSLHR-20-00583

Cox, R. M., and Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear Hear.* 16, 176–186. doi: 10.1097/00003446-199504000-00005

Eckert, M. A., Teubner-Rhodes, S., and Vaden, K. I. (2016). Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. *Ear Hear.* 37, 101S–110S. doi: 10.1097/AUD.0000000000000300

Epstein, M., and Florentine, M. (2012). Binaural loudness summation for speech presented via earphones and loudspeaker with and without visual cues. *J. Acoust. Soc. Am.* 131, 3981–3988. doi: 10.1121/1.3701984

Fitzgerald, M. B., Kan, A., and Goupell, M. J. (2015). bilateral loudness balancing and distorted spatial perception in recipients of bilateral cochlear implants. *Ear Hear.* 36, e225–e236. doi: 10.1097/AUD.0000000000000174

Friesen, L. M., and Picton, T. W. (2010). A method for removing cochlear implant artifact. *Hear. Res.* 259, 95–106. doi: 10.1016/j.heares.2009.10.012

Gagné, J. P., Besser, J., and Lemke, U. (2017). Behavioral assessment of listening effort using a dual-task paradigm: a review. *Trends Hear.* 21:233121651668728. doi: 10.1177/2331216516687287

Gantz, B. J., Tyler, R. S., Rubinstein, J. T., Wolaver, A., Lowder, M., Abbas, P., et al. (2002). Binaural cochlear implants placed during the same operation. *Otol. Neurotol.* 23, 169–180. doi: 10.1097/00129492-200203000-00012

Gilley, P. M., Sharma, A., Dorman, M. F., Finley, C. C., Panch, A. S., and Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clin. Neurophysiol.* 117, 1772–1782. doi: 10.1016/j.clinph.2006.04.018

Goupell, M. J., Stakhovskaya, O. A., and Bernstein, J. G. W. (2018). Contralateral interference caused by binaurally presented competing speech in adult bilateral cochlear-implant users. *Ear Hear.* 39, 110–123. doi: 10.1097/AUD.0000000000000470

Grantham, D. W., Ashmead, D. H., Ricketts, T. A., Labadie, R. F., and Haynes, D. S. (2007). Horizontal-plane localization of noise and speech signals by postlingually deafened adults fitted with bilateral cochlear implants. *Ear Hear.* 28, 524–541. doi: 10.1097/AUD.0b013e31806dc21a

Hughes, K., and Galvin, K. L. (2013). Measuring listening effort expended by adolescents and young adults with unilateral or bilateral cochlear implants or normal hearing. *Cochlear Implants Int.* 14, 121–129. doi: 10.1179/1754762812Y.0000000009

Hughes, S., Hutchings, H. A., Rapport, F. L., McMahon, C. M., and Boisvert, I. (2018). Social connectedness and perceived listening effort in adult cochlear implant users: a grounded theory to establish content validity for a new patient-reported outcome measure. *Ear Hear.* 39, 922–934. doi: 10.1097/AUD.0000000000000553

IEEE (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans. Audio Electroacoust.* 17, 225–246. doi: 10.1109/TAU.1969.1162058

Kan, A., Stoelb, C., Litovsky, R. Y., and Goupell, M. J. (2013). Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users. *J. Acoust. Soc. Am.* 134, 2923–2936. doi: 10.1121/1.4820889

Kleiner, M., Brainard, D. H., Pelli, D. G., Broussard, C., Wolf, T., and Niehorster, D. (2007). What's new in Psychtoolbox-3? *Perception* 36, 1–16.

Koelewijn, T., Zekveld, A., Festen, J. M., and Kramer, S. E. (2012). Pupil dilation uncovers extra listening effort in the presence of a single-talker masker. *Ear Hear.* 33, 291–300. doi: 10.1097/AUD.0b013e3182310019

Kramer, S. E. (2008). Hearing impairment, work, and vocational enablement. *Int. J. Audiol.* 47(Suppl. 2), S124–S130. doi: 10.1080/14992020802310887

Kramer, S. E., Kapteyn, T. S., and Houtgast, T. (2006). Occupational performance: comparing normally-hearing and hearing-impaired employees using the Amsterdam checklist for hearing and work. *Int. J. Audiol.* 45, 503–512. doi: 10.1080/14992020600754583

Kramer, S. E., Kapteyn, T. S., Festen, J. M., and Kuik, D. J. (1997). Assessing aspects of auditory handicap by means of pupil dilatation. *Int. J. Audiol.* 36, 155–164. doi: 10.3109/00206099709071969

Laszig, R., Aschendorff, A., Stecker, M., Müller-Deile, J., Maune, S., Dillier, N., et al. (2004). Benefits of bilateral electrical stimulation with the nucleus cochlear implant in adults: 6-Month postoperative results. *Otol. Neurotol.* 25, 958–968. doi: 10.1097/00129492-200411000-00016

Legris, E., Galvin, J., Mofid, Y., Aguillon-Hernandez, N., Roux, S., Aoustin, J. M., et al. (2022). Relationship between behavioral and objective measures of sound intensity in normal-hearing listeners and hearing-aid users: a pilot study. *Brain Sci.* 12:392. doi: 10.3390/brainsci12030392

Litovsky, R. Y., Parkinson, A., and Arcaroli, J. (2009). Spatial hearing and speech intelligibility in bilateral cochlear implant users. *Ear Hear.* 30, 419–431. doi: 10.1097/AUD.0b013e3181a165be

Litovsky, R. Y., Parkinson, A., Arcaroli, J., and Sammeth, C. (2006). Simultaneous bilateral cochlear implantation in adults: a multicenter clinical study. *Ear Hear.* 27, 714–731. doi: 10.1097/01.aud.0000246816.50820.42. Simultaneous

Litovsky, R. Y., Parkinson, A., Arcaroli, J., Peters, R., Lake, J., Johnstone, P., et al. (2004). Bilateral cochlear implants in adults and children. *Arch. Otolaryngol. Head Neck Surg.* 130, 648–655. doi: 10.1001/archotol.130.5.648

Loizou, P. C., Hu, Y., Litovsky, R. Y., Yu, G., Peters, R., Lake, J., et al. (2009). Speech recognition by bilateral cochlear implant users in a cocktail-party setting. *J. Acoust. Soc. Am.* 125, 372–383. doi: 10.1121/1.3036175

McGarrigle, R., Munro, K. J., Dawes, P., Stewart, A. J., Moore, D. R., Barry, J. G., et al. (2014). Listening effort and fatigue: what exactly are we measuring? A British Society of Audiology Cognition in Hearing Special Interest Group "white paper." *Int. J. Audiol.* 53, 433–445. doi: 10.3109/14992027.2014.890296

Mosnier, I., Sterkers, O., Bebear, J. P., Godey, B., Robier, A., Deguine, O., et al. (2009). Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients. *Audiol. Neurotol.* 14, 106–114. doi: 10.1159/000159121

Nachtegaal, J., Kuik, D. J., Anema, J. R., Goverts, S. T., Festen, J. M., and Kramer, S. E. (2009). Hearing status, need for recovery after work, and psychosocial work characteristics: results from an internet-based national survey on hearing. *Int. J. Audiol.* 48, 684–691. doi: 10.1080/14992020902962421

Noble, W., Tyler, R., Dunn, C., and Bhullar, N. (2008). Unilateral and bilateral cochlear implants and the implant-plus-hearing-aid profile: comparing self-assessed and measured abilities. *Int. J. Audiol.* 47, 505–514. doi: 10.1080/14992020802070770

Nopp, P., Schleich, P., and D'Haese, P. (2004). Sound localization in bilateral users of MED-EL COMBI 40/40+ cochlear implants. *Ear Hear.* 25, 205–214. doi: 10.1097/01.AUD.0000130793.20444.50

Nunnally, J. C., Knott, P. D., Duchnowski, A., and Parker, R. (1967). Pupillary response as a general measure of activation. *Percept. Psychophys.* 2, 149–155. doi: 10.3758/BF03210310

Ohlenforst, B., Zekveld, A. A., Lunner, T., Wendt, D., Naylor, G., Wang, Y., et al. (2017). Impact of stimulus-related factors and hearing impairment on listening effort as indicated by pupil dilation. *Hear. Res.* 351, 68–79. doi: 10.1016/j.heares.2017.05.012

Perreau, A. E., Ou, H., Tyler, R., and Dunn, C. (2014). Self-reported spatial hearing abilities across different cochlear implant profiles. *Am. J. Audiol.* 23, 374–384. doi: 10.1044/2014_AJA-14-0015

Perreau, A. E., Wu, Y. H., Tatge, B., Irwin, D., and Corts, D. (2017). Listening effort measured in adults with normal hearing and cochlear implants. *J. Am. Acad. Audiol.* 28, 685–697. doi: 10.3766/jaaa.16014

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Privitera, C. M., Renninger, L. W., Carney, T., Klein, S., and Aguilar, M. (2010). Pupil dilation during visual target detection. *J. Vis.* 10:3. doi: 10.1167/10.10.3

Reiss, L. A. J., Eggleston, J. L., Walker, E. P., and Oh, Y. (2016). Two ears are not always better than one: mandatory vowel fusion across spectrally mismatched ears in hearing-impaired listeners. *JARO* 17, 341–356. doi: 10.1007/s10162-016-0570-z

Reiss, L. A. J., Fowler, J. R., Hartling, C. L., and Oh, Y. (2018). Binaural pitch fusion in bilateral cochlear implant users. *Ear Hear.* 39, 390–397. doi: 10.1097/AUD.0000000000000497

Schleich, P., Nopp, P., and D'Haese, P. (2004). Head shadow, squelch, and summation effects in bilateral users of the MED-EL COMBI 40/40+ cochlear implant. *Ear Hear.* 25, 197–204. doi: 10.1097/01.AUD.0000130792.43315.97

Sladen, D. P., Nie, Y., and Berg, K. (2018). Investigating Speech Recognition and listening effort with different device configurations in adult cochlear implant users. *Cochlear Implants Int.* 19, 119–130. doi: 10.1080/14670100.2018.1424513

Steel, M. M., Papsin, B. C., and Gordon, K. A. (2015). Binaural fusion and listening effort in children who use bilateral cochlear implants: a psychoacoustic

and pupillometric study. *PLoS One* 10:e0117611. doi: 10.1371/journal.pone.0117611

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J. Speech Hear. Res.* 28, 455–462. doi: 10.1044/jshr.2803.455

Tyler, R. S., Dunn, C. C., Witt, S. A., and Noble, W. G. (2007). Speech perception and localization with adults with bilateral sequential cochlear implants. *Ear Hear.* 28(2 Suppl.), 86S–90S. doi: 10.1097/AUD.0b013e31803153e2

Tyler, R. S., Perreau, A. E., and Ji, H. (2009). Validation of the spatial hearing questionnaire. *Ear Hear.* 30, 466–474. doi: 10.1097/AUD.0b013e3181a61efe

van Hoesel, R., and Tyler, R. S. (2003). Speech perception, localization, and lateralization with bilateral cochlear implants. *J. Acoust. Soc. Am.* 113, 1617–1630. doi: 10.1121/1.1539520

van Hoesel, R., Ramsden, R., and O'Driscoll, M. (2002). Sound-direction identification, interaural time delay discrimination, and speech intelligibility advantages in noise for a bilateral cochlear implant user. *Ear Hear.* 23, 137–149. doi: 10.1097/00003446-200204000-00006

Wagner, A. E., Nagels, L., Toffanin, P., Opie, J. M., and Başkent, D. (2019). Individual variations in effort: assessing pupillometry for the hearing impaired. *Trends Hear.* 23:2331216519845596. doi: 10.1177/2331216519845596

Wendt, D., Dau, T., and Hjortkjær, J. (2016). Impact of background noise and sentence complexity on processing demands during sentence comprehension. *Front. Psychol.* 7:345. doi: 10.3389/fpsyg.2016.00345

Wendt, D., Koelewijn, T., Książek, P., Kramer, S. E., and Lunner, T. (2018). Toward a more comprehensive understanding of the impact of masker type and signal-to-noise ratio on the pupillary response while performing a speech-in-noise test. *Hear. Res.* 68, 255–278. doi: 10.1016/j.heares.2018.05.006

Winn, M. B. (2016). Rapid release from listening effort resulting from semantic context, and effects of spectral degradation and cochlear implants. *Trends Hear.* 20:2331216516669723. doi: 10.1177/2331216516669723

Winn, M. B., and Teece, K. H. (2021). Listening effort is not the same as speech intelligibility score. *Trends Hear.* 25:23312165211027688. doi: 10.1177/23312165211027688

Winn, M. B., Edwards, J. R., and Litovsky, R. Y. (2015). The impact of auditory spectral resolution on listening effort revealed by pupil dilation. *Ear Hear.* 36, e153–e165. doi: 10.1097/AUD.0000000000000145

Winn, M. B., Wendt, D., Koelewijn, T., and Kuchinsky, S. E. (2018). Best practices and advice for using pupillometry to measure listening effort: an introduction for those who want to get started. *Trends Hear.* 22, 1–32. doi: 10.1177/2331216518800869

Yoon, Y. S., Li, Y., Kang, H. Y., and Fu, Q. J. (2011). The relationship between binaural benefit and difference in unilateral speech recognition performance for bilateral cochlear implant users. *Int. J. Audiol.* 50, 554–565. doi: 10.3109/14992027.2011.580785

Zekveld, A., and Kramer, S. E. (2014). Cognitive processing load across a wide range of listening conditions: insights from pupillometry. *Psychophysiology* 51, 277–284. doi: 10.1111/psyp.12151

Zekveld, A., Kramer, S. E., and Festen, J. M. (2010). Pupil response as an indication of effortful listening: the influence of sentence intelligibility. *Ear Hear.* 31, 480–490. doi: 10.1097/AUD.0b013e3181d4f251

Zekveld, A., Kramer, S. E., and Festen, J. M. (2011). Cognitive load during speech perception in noise: the influence of age, hearing loss, and cognition on the pupil response. *Ear Hear.* 32, 498–510. doi: 10.1097/AUD.0b013e31820512bb

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 22 December 2022
DOI 10.3389/fnins.2022.1022354



OPEN ACCESS

EDITED BY
Yi Zhou,
Arizona State University, United States

REVIEWED BY
Frederick Jerome Gallun,
Oregon Health and Science University, United States
John C. Middlebrooks,
University of California, Irvine, United States

*CORRESPONDENCE
Anna Dietze
anna.dietze@uni-oldenburg.de

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 18 August 2022
ACCEPTED 28 November 2022
PUBLISHED 22 December 2022

CITATION
Dietze A, Sörös P, Bröer M, Methner A, Pöntynen H, Sundermann B, Witt K and Dietz M (2022) Effects of acute ischemic stroke on binaural perception.
*Front. Neurosci.* 16:1022354.
doi: 10.3389/fnins.2022.1022354



# Effects of acute ischemic stroke on binaural perception

Anna Dietze[1,2]*, Peter Sörös[3,4], Matthias Bröer[3], Anna Methner[3], Henri Pöntynen[1,2], Benedikt Sundermann[4,5], Karsten Witt[3,4] and Mathias Dietz[1,2,4]

[1]Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany, [2]Cluster of Excellence "Hearing4all", University of Oldenburg, Oldenburg, Germany, [3]Department of Neurology, School of Medicine and Health Sciences, University of Oldenburg, Oldenburg, Germany, [4]Research Center Neurosensory Science, University of Oldenburg, Oldenburg, Germany, [5]Institute of Radiology and Neuroradiology, Evangelisches Krankenhaus Oldenburg, Oldenburg, Germany

Stroke-induced lesions at different locations in the brain can affect various aspects of binaural hearing, including spatial perception. Previous studies found impairments in binaural hearing, especially in patients with temporal lobe tumors or lesions, but also resulting from lesions all along the auditory pathway from brainstem nuclei up to the auditory cortex. Currently, structural magnetic resonance imaging (MRI) is used in the clinical treatment routine of stroke patients. In combination with structural imaging, an analysis of binaural hearing enables a better understanding of hearing-related signaling pathways and of clinical disorders of binaural processing after a stroke. However, little data are currently available on binaural hearing in stroke patients, particularly for the acute phase of stroke. Here, we sought to address this gap in an exploratory study of patients in the acute phase of ischemic stroke. We conducted psychoacoustic measurements using two tasks of binaural hearing: binaural tone-in-noise detection, and lateralization of stimuli with interaural time- or level differences. The location of the stroke lesion was established by previously acquired MRI data. An additional general assessment included three-frequency audiometry, cognitive assessments, and depression screening. Fifty-five patients participated in the experiments, on average 5 days after their stroke onset. Patients whose lesions were in different locations were tested, including lesions in brainstem areas, basal ganglia, thalamus, temporal lobe, and other cortical and subcortical areas. Lateralization impairments were found in most patients with lesions within the auditory pathway. Lesioned areas at brainstem levels led to distortions of lateralization in both hemifields, thalamus lesions were correlated with a shift of the whole auditory space, whereas some cortical lesions predominantly affected the lateralization of stimuli contralateral to the lesion and resulted in more variable responses. Lateralization performance was also found to be affected by lesions of the right, but not the left, basal ganglia, as well as by lesions in non-auditory cortical areas. In general, altered lateralization was

common in the stroke group. In contrast, deficits in tone-in-noise detection were relatively scarce in our sample of lesion patients, although a significant number of patients with multiple lesion sites were not able to complete the task.

KEYWORDS

binaural hearing, psychoacoustics, brain lesions, lateralization, binaural masking level difference, magnetic resonance imaging, stroke

## 1. Introduction

The interaural level differences (ILD) and interaural time differences (ITD) provide the basis for localizing sound sources in the horizontal plane. This ability informs the listener about the spatial location of an approaching vehicle, for instance, but is also crucial for segregating different auditory streams in more complex listening environments, such as multiple talkers in a crowded restaurant. Especially the latter ability is clearly compromised in listeners with sensorineural hearing loss (e.g., Gatehouse, 2004; Shinn-Cunningham and Best, 2008). However, spatial hearing can also be impaired by damage to the central nervous system. The consequences of such damage for spatial hearing and binaural perception are arguably less well understood (Gallun, 2021).

One relatively prevalent type of central nervous system damage is stroke. For instance, the GEDA 2014/2015-EHIS study found that, in Germany, 1.6% of adults suffered a stroke or chronic consequences of a stroke during the past 12 months (Robert Koch-Institut, 2017). Central stroke lesions do not usually affect hearing thresholds, but they can affect binaural hearing (Häusler and Levine, 2000). This is also reflected in patient-reported difficulties in sound localization in the chronic phase after stroke, as shown in Bamiou et al. (2012). Given the relatively high prevalence of stroke in the general population, an improved understanding of its effects on spatial hearing would be desirable.

Previous studies have revealed deficits in binaural hearing in patients with different stroke lesion locations. Furst et al. (2000) investigated the binaural performance of patients with brainstem lesions using a test of interaural difference discrimination and with a lateralization task. Binaural performance was affected whenever the lesion overlapped the auditory pathway. Lesions of the caudal pons led to center-oriented lateralization, whereas lesions rostral to the superior olivary complex led to side-oriented lateralization results. Just-noticeable differences in ILD and ITD were affected in some patients with pontine lesions.

Comparable methods were used by Spierer et al. (2009), who studied the effects of cortical lesions on ITD- and ILD-based lateralization. The findings suggested a dominance of the right hemisphere in auditory spatial representation. More frequent and more severe deficits were observed after right-sided, compared to left-sided, damage. Lesions of the right hemisphere influenced contralesional as well as ipsilesional lateralization, whereas the effect of left-sided damage was restricted mainly to the contralesional hemifield.

Along the same lines, the effect of auditory neglect (impaired perception of auditory stimuli in one hemispace) is also more frequently observed for right-hemispheric lesions, especially when the temporal lobe is damaged (Gokhale et al., 2013). The term neglect is used for various impairments and different modalities (Heilman et al., 2000). As reviewed in Gokhale et al. (2013), language-related stimuli are mainly associated with the left temporal cortex, whereas non-language stimuli are predominantly processed in the right hemisphere. As a result, processing of non-language stimuli is often impaired, and in some cases, neglected after damage to the right hemisphere.

Two separate processing streams are suspected to be responsible for the 'where' and 'what' of auditory perception. This hypothesis is supported by the fact that binaural hearing performance of the centrally impaired auditory system depends not only on the location of the damaged area, but also on the task to be performed (Bellmann et al., 2001). For instance, a case report of a patient with lesions in the right hemisphere showed a difference between using binaural cues implicitly or explicitly (Thiran and Clarke, 2003). The patient was able to implicitly use binaural cues for stream segregation in a spatial-release-from-masking task, but had no explicit lateralized perception at all when presented with stimuli with ITDs. The implicit and explicit use of binaural cues was also investigated by Tissieres et al. (2019), with a larger number of participants. They concluded that the implicit use of auditory spatial cues relies on a distinct, left-dominated network.

In general, previous studies on the effect of lesions of the central nervous system on binaural perception were mainly investigated in the chronic phase of stroke in subgroups of stroke populations. Based on the results of, e.g., Trapeau and Schönwiesner (2015), who showed that relearning of localization with altered ITDs is possible within a few days, we assume that stroke-induced lateralization impairments will be strongest in the acute phase and at least partially recovered in the chronic phase of stroke. The existing studies revealed a plethora of deficits that vary significantly across lesion location, stimulus

material and patients. The great variability and individual nature of the findings indicate that further large-scale research is needed to move closer to a complete understanding of the effects of stroke on binaural hearing performance. By studying the disturbed system shortly after stroke onset, the patients' responses may give novel insights into the role of the affected areas in spatial hearing, including its relevance for the healthy system.

In addition to studies with stroke patients, neuroscientific experiments with healthy adults revealed different mechanisms of ITD processing along the auditory pathway. Thompson et al. (2006) presented large ITDs (±1500 μs), well outside the range of ITDs of ±700 μs, that are usually experienced under natural listening conditions. Using functional magnetic resonance imaging (MRI) neural activity was measured by means of the blood oxygenation level dependent response. For these large ITDs, they found higher neural activity in the ipsilateral, compared to the contralateral, side of the mid-brain, which is the opposite of findings for smaller ITDs. A related study by von Kriegstein et al. (2008) revealed that at the level of the cortex, both hemispheres were activated for these large ITDs. For the small ITDs, predominantly the primary auditory cortex in the contralateral hemisphere was active. These data show that coding of ITD in the cortex is fundamentally different from the mid-brain representation of ITD, but it remains unclear how such large ITDs are perceived if lesions impair the encoding or decoding at different stages of the auditory pathway.

Studying clinical populations has shaped our understanding of binaural processing, and is still useful to supplement studies in different animal models (Gallun, 2021). Currently, structural MRI is used in standard clinical routine for stroke patients. The combination of the information on the precise lesion location, and the patients' performance in behavioral tasks, could lead to insights into individual problems in binaural processing and possible ways to individualize therapies.

The detrimental effects of stroke lesions on binaural hearing tasks vary not only for different lesion locations and lesion sizes, but can also be shaped by factors such as age, conductive or sensorineural hearing loss, cognitive abilities, and other non-auditory characteristics. Therefore, in addition to group analyses that are compared to age-matched control subjects, focusing on individual patients with all their confounding influences case by case remains unavoidable.

The objective of the current exploratory study was to investigate the binaural perception of individuals in the acute phase of stroke, compared to an age-matched control group in a quantitative, yet individual manner. Since binaural deficits have been observed for lesions across multiple brain areas that are not directly related to audition, we did not limit our study to predefined regions of interest. This choice was further motivated by our aim to conduct a relatively large-scale study with potential to reveal patterns that would remain unnoticed or ambiguous with smaller patient cohorts. We conducted two binaural experiments using headphone stimulation. Performance in both experiments relied on using interaural differences. In the first experiment, a binaural tone-in-noise detection task, the implicit use of interaural cues was sufficient to detect differences to the reference stimulus. In the second experiment, a lateralization task, listeners had to explicitly use interaural cues to judge the perceived intracranial position of the stimulus. These experiments, and an additional general assessment, were completed by patients that had rather small lesions in different brain areas. The location of the lesion was established based on previously acquired MRI data.

# 2. Materials and methods

## 2.1 Participants

In total, 50 stroke patients (mean age of 63 years, SD: 14 years, 20 female, 30 male) and 12 control subjects (mean age of 61 years, SD: 14 years, 9 female, 3 male) participated after passing audiometric and cognitive assessments (see Sections "2.2 General assessment" and "2.4.1 Audiometry" for details) and providing written informed consent. Participants that had a stroke will be referred to as patients, whereas those participating in the control group will be referred to as control subjects. The study was approved by the Medical Research Ethics Board of the University of Oldenburg, Germany. The stroke patients were recruited in the stroke unit of the Evangelisches Krankenhaus, Oldenburg, Germany and tested in a quiet room. Only those patients participated who could understand and produce speech, who were mobile and in a general stable condition, and able to complete the different tasks despite their recent stroke. Exclusion criteria were additional neurological diseases or a pure-tone average of 40 dB HL or more (see Section "2.4.1 Audiometry"). The stroke patients participated in the experiments on average 5 days (range: 1—9 days, 16 days for one patient, SD: 2 days) after stroke onset. The symptoms of stroke, as measured by the National Institute of Health stroke scale (see Section "2.2 General assessment"), ranged from 0 to 6 points, except for one patient with a score of 20 points. The median of the scores was one point, thus representing a stroke cohort suffering from minor stroke. The control group was age-matched and followed the same exclusion criteria.

## 2.2 General assessment

Preceding the psychoacoustic experiments, the Montreal Cognitive Assessment (MoCA, Nasreddine et al., 2005) was used to screen for mild cognitive impairment or dementia. The test contains 30 tasks targeting different cognitive abilities, and is scored with a maximum of 30 points. Scores below 26 points suggest mild cognitive impairment. Three patients with

a performance score of 17 or lower were excluded from the subsequent experiments.

The National Institute of Health stroke score (National Institute of Neurological Disorders and Stroke [NIHSS], 2019) was obtained as part of the clinical routine 24 h after the patients came to the hospital. It consists of several measures judging the severity of the symptoms of stroke, with a maximum score of 42 points. Scores below 5 are classified as minor stroke, below 15 as moderate stroke, and above this as moderate to severe and severe stroke. The score includes several items related to motor functions, but no item explicitly targeting auditory impairments.

To quantitatively assess the intensity of possible depression, we used the short version of the Beck's Depression Inventory (BDI, Beck et al., 2013). It contains 7 sets of statements from which are chosen those that best describe the patient's current state. To be compatible with the full version, the results are scaled to fall within the ranges of the full test. Scores below 9 indicate no or minimal depression, those between 9 and 13 indicate mild depression. Moderate depression is indicated by scores between 20 and 28, and severe depression by scores in the range between 29 and 63.

The multiple-choice vocabulary intelligence test, the German MWT-B (Lehrl, 2005), was used as an estimator for the premorbid intelligence (unaffected by the stroke). It consists of 37 items, each containing five words. Only one of the five words is an established word that must be recognized, whereas the others are neologisms. The higher the number of correctly detected words, the higher the estimated crystallized intelligence (part of a person's intelligence that consists of knowledge that comes from prior learning and past experiences).

## 2.3 Magnetic resonance imaging

Magnetic Resonance Imaging (MRI) was obtained as part of the clinical routine for all patients. Two different systems were used: a Siemens Magnetom Symphony (1.5 T) and a Magnetom Sola (1.5 T). Lesion location and lesion volume were extracted from these images based on the combined information of diffusion weighted imaging (DWI), apparent diffusion coefficient (ADC) mapping, and the fluid-attenuated inversion recovery (FLAIR) sequence. All areas that were hyperintense in DWI (and had a low signal in the ADC map, thus representing restricted diffusion) were outlined on the FLAIR images using the visualization tool MRIcroGL (Rorden and Brett, 2000), and the volume of the lesions were calculated. The analyses of the images was done using FSL (Jenkinson et al., 2012), a library of analysis tools for FMRI, MRI and DTI brain imaging data. Brain extraction was carried out using FSL BET (Smith, 2002) based on the FLAIR images, since they allowed better extraction than the available T1-weighted images. The fractional intensity threshold for BET was chosen case by case, to obtain the best extraction. The MR images were obtained in the standard clinical routine. Thus, for a majority of patients, only 2D MR images were available. Only in some cases 3D T1 and/or 3D FLAIR data were acquired. Linear registration of the brain-extracted FLAIR images to MNI 152 space, a structural template, provided by the Montreal Neurological Institute, was carried out using FSL FLIRT (Jenkinson et al., 2002). The quality of the resulting images was visually controlled for every subject. The same transformations were applied to the lesion masks. The lesion location was then estimated based on the AICHA atlas (Joliot et al., 2015).

Overlap of the MNI-registered stroke lesions with brain areas that belong to the auditory pathway were estimated as follows: The main nuclei of the primary auditory pathway were defined by the mask provided by Sitek et al. (2019) for the subcortical areas. The auditory cortex was defined by the term-based meta-analyses for the term 'auditory cortex' on the website neurosynth.org (Yarkoni et al., 2011), which created a mask using data from 279 MRI studies.

## 2.4 Psychoacoustic measurements

For all of the psychoacoustic experiments, closed headphones with high passive sound attenuation (HDA300, Sennheiser electronic GmbH, Wedemark, Germany) and driven by an external soundcard (UR22mkII, Steinberg Media Technologies GmbH, Hamburg, Germany) were used. The stimuli were generated and reproduced by custom-made MATLAB scripts using the psychophysical measurement package AFC (Ewert, 2013). The sampling rate was 48 kHz.

### 2.4.1 Audiometry

Pure-tone audiometric testing for a restricted set of frequencies (500, 1000, and 3000 Hz) was conducted preceding the psychoacoustic experiments using a one-interval two-alternative forced-choice procedure controlled by the experimenter. The testing followed a one-up, one-down adaptive procedure. The tracks ended after eight reversals (initial step size was 20 dB, after the second reversal 10 dB, after the fourth reversal 5 dB) and the thresholds were computed from the mean of the last four reversals. The pure-tone average over the three measured frequencies was calculated for the left and right ear individually (PTA3 L and PTA3 R, respectively), and averaged over the two ears (PTA3). In addition, the absolute difference between the left and right PTA3 (PTA3 asymmetry) was calculated. Two patients with a PTA3 L and/or a PTA3 R of more than 40 dB HL were excluded, leading to a total of 50 patients for further study.

### 2.4.2 Tone-in-noise detection

In the tone-in-noise detection experiments, the participants were presented with three intervals containing 500-ms bursts of octave-wide white noise centered around 500 Hz (333–666 Hz).

The stimuli were gated with 20-ms raised cosine onset and offset ramps. The intervals were separated by 300-ms silent gaps. In one of the three intervals, an additional 500-Hz pure-tone of 420 ms duration was added and temporally centered in the noise. The tone had the same ramp parameters as the noise, but its onset was 40 ms later than the noise. Similarly, the tone offset was 40 ms before the noise offset. The participants' task was to detect the deviating interval (the one containing the tone) and to press key number '1,' '2,' or '3' on a computer keyboard, indicating whether the first, second, or third interval was the odd one. The tone was either interaurally in phase with the noise (condition $N_0S_0$), or had an interaural phase difference of π (condition $N_0S_\pi$). The experiment started without any training and with two runs of the $N_0S_\pi$ condition. This was followed by one run of the $N_0S_0$ condition. The noise was presented with 60 dB sound-pressure level (SPL). The level of the tone was initially 65 dB SPL in the $N_0S_0$ condition and 50 dB SPL in the $N_0S_\pi$ condition. The level varied according to a one-up, three-down procedure, with a step size of 4 dB up to the second reversal, and a step size of 2 dB for the remaining 8 reversals, converging to 79.4% correct thresholds. Thresholds are calculated as the average of the last 8 reversals. If the staircase track hit the maximum tone level of 80 dB SPL during a measurement, re-instructions on how to perform the task were provided. If this did not lead to improvements in task performance, the run was stopped and marked as invalid. No feedback was given during the runs. The binaural masking level difference (BMLD) was calculated from the threshold difference between $N_0S_0$ and the better of the two $N_0S_\pi$ runs.

### 2.4.3 Lateralization

For the lateralization task, again a one-octave wide white noise, centered around 500 Hz with an interaural difference in either level or time was presented. The stimuli were generated by copying the same noise sample to both channels and then applying the interaural difference in time or level. The task was to indicate where the sound was perceived inside the head. Responses were given by pressing one of the horizontally aligned numbers '1' to '9' on a computer keyboard, above the letter keys. The participants were instructed to press '1' when the sound was heard on the very left side of their head, '5' for sounds perceived in the center of the head and '9' for the very right side. For possible intracranial positions between the center and the two extremes, the participants were asked to press the respective number '2,' '3,' '4,' '6,' '7,' or '8' on the keyboard. For visual guidance, a template with a schematic drawing of a head indicated the positions of the ears and the center relative to the response buttons. The template covered all of the keyboard except for the numbers '1' to '9.' The duration of the stimuli was 1 s, gated with cosine ramps of 10 ms duration and presented at 70 dB SPL. ITDs ranging from −600 to 600 μs in steps of 200 μs, and two ITDs outside the physiological range (−1500 and 1500 μs), were presented. The ILDs ranged from −12 to 12 dB in steps of 4 dB. The level of the left- and right-ear signals was changed without changing the overall energy by applying the formula presented in Dietz et al. (2013). In addition, monaural stimulation of the left ear and right ear was tested. Each stimulus was presented six times in random order. The diotic stimulus (zero ITD/ILD) was presented eight times. To ensure one common reference system for both types of interaural differences, ILD and ITD stimuli were presented interleaved. In contrast to the investigations by Furst et al. (2000), no training and no center reference were provided in our study. The response to the first trial of each stimulus was not used in further analyses.

Several variables for quantitative description of the lateralization pattern were calculated:

A linear fit to the three left-favoring and right-favoring stimuli, individually for ILD stimuli (−12, −8, −4 dB and 4, 8, and 12 dB) and ITD stimuli (−600, −400, and −200 μs; 200, 400, and 600 μs) was used to describe the steepness of the participants' lateralization percept (*ILD L slope, ILD R slope, ITD L slope*, and *ITD R slope*). The logarithmic ratio of the left and right slope [*ILD slope ratio, ITD slope ratio, e.g., ILD slope ratio = log(ITD slope L/ILD slope R)]* indicates an asymmetric steepness of the two sides.

Variables that inform about side biases in the responses were calculated: The mean of the responses to all ITD or all ILD stimuli (*ITD mean* and *ILD mean*) and the mean of the fit to left-favoring and right-favoring stimuli (*ITD L fit, ITD R fit, ILD L fit*, and *ILD R fit*) were calculated. Furthermore, the mean of those stimuli that were perceived as being in the center of the head (when key '5' was pressed), was calculated for ILD and for ITD stimuli (*ITD center* and *ILD center*). The so-called *diotic percept* was the mean of the responses given for the zero ILD/ITD stimuli.

Another feature of the lateralization data is its variability. For this, the standard deviation for zero ILD/ITD was calculated (*diotic std.*), as well as the mean of the standard deviations of the responses to each ILD stimulus (excluding the monaural stimulation, *ILD std.*), each ITD stimulus (*ITD std.*) and the mean standard deviation of the left-favoring and right-favoring stimuli independently (*ITD L std., ITD R std., ILD L std.,* and *ILD R std.*). Their logarithmic ratios (*ITD std. ratio* and *ILD std. ratio)* can indicate differences in the variability of left-favoring and right-favoring stimuli.

The maximal range of lateralization was calculated by the difference of the maximally lateralized responses given for ITDs within the physiological range (*ITD range*), and for all ILDs excluding monaural stimulation (*ILD range*). The logarithmic ratio of the ranges obtained with ILD and ITD stimuli (*range ratio*) informs about differences in the ranges perceived using the two types of stimuli.

The perception of the monaural left and right (*mon left* and *mon right*), and the ITDs of ±1500 μs (*neg 1500* and

*pos 1500*) was only evaluated in terms of the mean response to these stimuli.

For all the variables, values within the interval of 1.5 times the standard deviation around the control group mean were considered to be normal. As we did not want to overemphasize possible asymmetries of the left and right side of individual control subjects, we also added the mirrored control data before calculating the mean and standard deviation. With this, the mean was not biased by individual asymmetries and the standard deviation remained unaffected. We verified that adding the mirrored data did not change the results substantially from those obtained without adding the mirrored responses to the data set. Whenever values of the calculated variables are reported, they are in the unit of response keys (a difference of one response button corresponds to 1/8 of the distance between the two ears), except for the variables describing the goodness of fit and the ratios.

# 3. Results

Analyses will be presented grouped by the presence of stroke (control vs. stroke group) and grouped by the anatomical location of the lesion (lesion groups). In addition, a selected set of individual stroke patients will be shown throughout the results section. These patients are chosen to highlight the individual character of each stroke patient's performance. The color-coding of the eight selected patients is consistent across Figures, allowing for comparison of their measurement results across experiments. In the last subsection, deviations from the control group are shown for individual cases and for lesion groups.

## 3.1 General assessment

Mean values and standard deviations of the non-auditory testing of the stroke and the control group are shown in Table 1. According to statistical tests (two-sample *t*-tests), the two groups did not differ in age, not in their pure-tone average over the three tested frequencies, and also not in the absolute asymmetry of their left and right PTA3. The scores for the multiple-choice vocabulary intelligence test (MWT-B) and the short form of Beck Depression Inventory (BDI) also did not differ significantly between the groups. In the cognitive screening test (MoCA) however, significantly lower scores were obtained for the stroke group compared to the control group.

## 3.2 Audiometry

The pure-tone audiometry thresholds for 500, 1000, and 3000 Hz revealed that 35% of the participants had a PTA3 (hearing thresholds averaged over the three frequencies and the two ears) of 20 dB HL or higher, indicating a slight hearing loss (Figure 1). Increased hearing thresholds were especially prevalent at the highest tested frequency of 3000 Hz. Similar PTA3 thresholds were found for the control subjects and for the stroke patients (Table 1), indicating that the pure tone hearing thresholds were not stroke-specific. The selected set of eight stroke patients, indicated by the colored dots, and the two selected control subjects, indicated by the filled gray boxes, span the range of hearing thresholds.

## 3.3 Correlation analyses

We computed correlations between age and PTA3 and the scores obtained from the general assessment (MoCA, BDI, NIHSS, and MWTB). All correlations were computed for the stroke group and the control group together, because the mean values for the two groups did not differ significantly, except for the MoCA scores (see Table 1). The correlation between age and PTA3 was statistically significant ($\rho = 0.59$, $p < 0.01$). With this, one cannot clearly distinguish between age effects and effects of hearing loss on the other outcome measures. Age and the MoCA score ($\rho = -0.36$, $p < 0.01$) and PTA3 and the MoCA score ($\rho = -0.35$, $p = 0.01$) were negatively correlated. The negative correlation between age and the BDI score ($\rho = -0.28$, $p = 0.03$) was statistically significant, as well. None of the other correlations were statistically significant with the alpha level set to 0.05 (see Table 2 and Supplementary Figure 1).

## 3.4 Tone-in-noise detection

The majority of participants (44 of the 50 stroke patients, and 11 of the 12 control subjects) produced converging tracks

TABLE 1 General assessment results.

| | Stroke $N = 50$ | Control $N = 12$ | Test statistics |
|---|---|---|---|
| Age [years] | 63 (14) | 61 (14) | $t(61) = -0.46$, $p = 0.647$ |
| PTA3 [dB HL] | 18 (8) | 14 (9) | $t(61) = -1.54$, $p = 0.129$ |
| PTA3 asymm. [dB] | 4 (4) | 5 (4) | $t(61) = 0.65$, $p = 0.518$ |
| MoCA score | 23.90 (4.68) | 28.36 (1.63) | $t(60) = 3.10$, $p = 0.003$ |
| MWT-B score | 29.72 (4.07) | 31.37 (4.15) | $t(59) = 1.21$, $p = 0.231$ |
| BDI short score | 7.60 (5.04) | 6.30 (3.15) | $t(61) = -0.86$, $p = 0.394$ |

Mean, standard deviation, and *t*-test results for stroke and control groups. Values are given in the form 'mean (standard deviation)'.

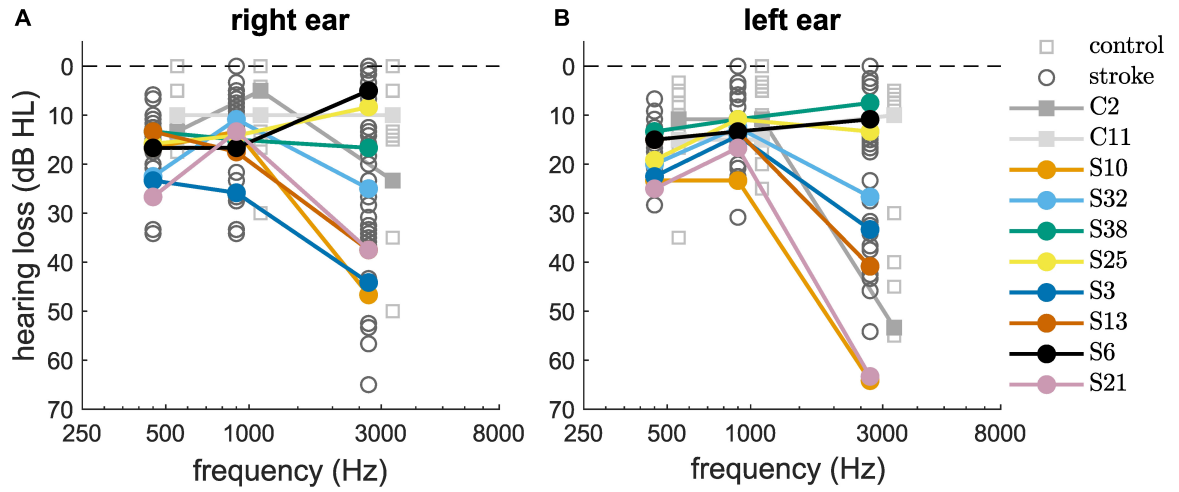



FIGURE 1
Pure-tone audiometric thresholds of the control group (squares) and the stroke group (circles) measured at the right ear **(A)** and left ear **(B)**. Selected participants are highlighted by the color coding used throughout the figures.

TABLE 2 Correlations between age and PTA3 thresholds and the results of the non-auditory measurements (MoCA, BDI, NIHSS, and MWT-B) for stroke and control group together.

| | Age [years] | PTA3 [dB HL] |
|---|---|---|
| Age [years] | – | $\rho = 0.59, p < 0.001$ |
| PTA3 [dB HL] | $\rho = 0.59, p < 0.001$ | – |
| MoCA score | $\rho = -0.36, p < 0.001$ | $\rho = -0.35, p = 0.01$ |
| BDI score | $\rho = -0.28, p = 0.03$ | $\rho = 0, p = 0.97$ |
| NIHSS score | $\rho = 0.11, p = 0.44$ | $\rho = 0.20, p = 0.17$ |
| MWT-B score | $\rho = 0.14, p = 0.29$ | $\rho = -0.18, p = 0.17$ |

in both conditions of the tone-in-noise detection task. The BMLD was calculated from the difference between $N_0S_\pi$ and $N_0S_0$ thresholds (see Supplementary Figure 2). Four patients (S6, S18, S20, and S24) and one control subject (C3) produced a convergent track only in the $N_0S_0$-condition, preventing the estimation of the BMLD. S2 and S44 did not produce any converging tracks. It is not known why these participants were not able to perform the task. Due to restricted measurement time, the tasks were not repeated. The normal values of BMLD, as defined by the mean ±1.5 times the standard deviation of the control group results, ranged from 7.5 to 20.1 dB. Of those participants that produced convergent tracks, 93% of the stroke group (41 of 44 patients) showed a BMLD of 7.5 dB or more. This result is comparable to the result from the control group, with 91% of the subjects demonstrating a BMLD of 7.5 dB or more. As shown in Figure 2 there was a significant negative correlation of the BMLD with age ($\rho = -0.36$, $p = 0.02$), but not with PTA3 ($\rho = -0.22$, $p = 0.11$).

## 3.5 Lateralization

In general, all participants were able to complete the lateralization task and almost all reported that the monaural stimuli were perceptually different from the binaural stimuli, and that they were the easiest stimuli to lateralize. For many patients, visual inspection of the data did not reveal any impairments in lateralization. Selected group analyses (averages over lesion groups) are presented in Table 3. In the following paragraphs, the observed lateralization patterns of the control group and the lesion groups will be discussed in terms of group averages and examples of individual patients.

In particular, data from eight patients with different lesion locations and volumes (see Figure 3) were selected for individual presentation. The results of the lateralization task (perceived intracranial position for the presented ILDs and ITDs) are shown in Figure 4 for two example control subjects (panel A and B) and the eight selected stroke patients (panels C-J). These patients are not fully representative of the average patient for their respective lesion group, but rather display distinct response patterns. The lateralization results of all other participants can be found in the Supplementary Figures 3–8.

### 3.5.1 Control group

Physically left-favoring, to consecutively more right-favoring stimuli, were perceived from the left to the right inside the participants' heads for the ILD and ITD stimuli for all control subjects, with only slight deviations. Apparently, the chosen ILDs, ranging from −12 to 12 dB did not lead to strongly lateralized auditory images (responses close to response keys 1 = left and 9 = right). Previous studies already demonstrated that the extent of perceived lateralization for ILDs

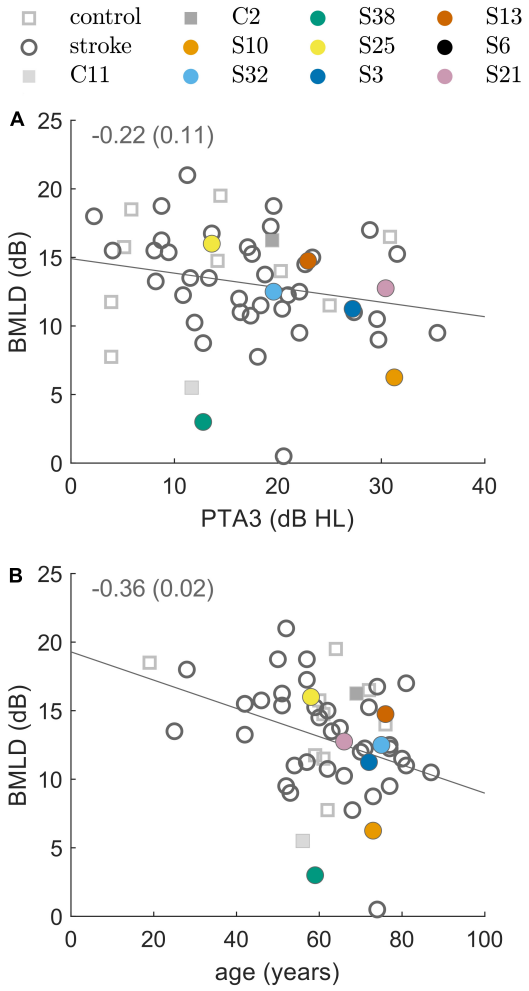



FIGURE 2
Binaural masking level difference (BMLD) resulting from the binaural tone-in-noise detection experiment. BMLD over PTA3 in panel **(A)** and BMLD over age in panel **(B)** for control subjects (squares) and stroke patients (circles). The line represents a linear fit and the inset represents the correlation coefficient and the respective *p*-value. Selected participants are highlighted by the color coding used throughout the figures.

of this magnitude varies across subjects (Baumgärtel and Dietz, 2018). It also depends on frequency, with stronger lateralization perceived for the same ILD magnitude and higher-frequency signals (Bernstein and Trahiotis, 2011). Auditory space was distributed roughly symmetrically around zero ITD/ILD, being reflected in the average perceived position over all ILD and ITD stimuli (*mean*) of 5.2 in the control group. Even in the control group, the perceived intracranial positions were not perfectly distributed around the center (5.0). Monaural left or right stimulation was perceived close to the most lateralized intracranial positions (*mon left*: 1.5 and *mon right*: 8.6) with almost no intra-individual variability. For all ILDs and all absolute ITDs = 600 μs, a small variability in single trials can be seen. The standard deviation of given responses was for all stimuli approximately in the range of one response key for the control subjects (e.g., 1.1 for *diotic std.*, the standard deviation of zero ILD/ITD). Only one person of the control group produced much more variable data. The variability of ITDs of ±1500 μs was larger than for smaller ITDs in most control subjects. This unnaturally large ITD was perceived less lateralized compared to smaller absolute ITDs. Based only on the center frequency (500 Hz), one cannot distinguish between a time shift of −500 or +1500 μs, as the period at this frequency is 2000 μs. However, since the stimulus is a white noise of 333 Hz bandwidth centered around 500 Hz, the auditory system can partially resolve this ambiguity, by exploiting either the interaural correlation at other frequencies or the envelope ITD. The *range of lateralization* was larger for ITDs (5.5) compared to ILDs (3.7) and for both interaural differences was much smaller than the maximal possible range of 8.

In the Figures 4A, B, examples of data from two typical control subjects (C2 and C11) show the main trends described above. The colored symbols represent the responses to individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively. If no asymmetry was present in a participant's responses, they should have the same slope on both sides. Completely symmetrical responses to left-favoring and right-favoring stimuli were obtained only by a small number of control subjects. Obviously, for some individual trials the participants' responses differed from the expected pattern, as for example in one trial with subject C2, the response to monaural-right stimulation was the left-most response key. This intra-individual variability can occur for various reasons. For example, it may be due to perceptual variability *per se*, but could also depend on the state of attention, or lack of concentration when reporting the percept. In Figures 4C–J, the general trends observed in the control group are visualized with the gray line and shaded area indicating the mean and the 1.5 times standard deviation interval around the mean response of the control subjects.

Despite the reduced range of lateralization in most participants, different lesion groups were found to be associated with altered lateralization percepts.

### 3.5.2 Brainstem lesions

In only one of the seven patients with a brainstem lesion (S42) did the lateralization results visually resemble the control group. All the others showed obvious deviations from the control group. In four of the seven patients of the brainstem lesion group (S7, S10, S12, S22, and S32), a reduced set of response keys was used. The responses were given in the categories left–center–right or only left–right. This is partially reflected in the *diotic std.* of 1.8 for this lesion group. Lesions in the brainstem (medulla, pons, or midbrain) did not alter

TABLE 3 Quantification of the lateralization results for the lesion groups.

| | *Mean* | *diotic std.* | *ITD range* | *ILD range* | *mon left* | *mon right* |
|---|---|---|---|---|---|---|
| Control (12) | 5.2 | 1.0 | 5.5 | 3.7 | 1.5 | 8.6 |
| bs l (3) | 5.9 | 1.2 | 5.3 | 5.0 | 1.5 | 8.7 |
| bs r (4) | 4.8 | 1.4 | 4.8 | 4.0 | 1.3 | 8.9 |
| thal l (4) | 4.7 | 1.4 | 4.3 | 3.4 | 2.1 | 8.3 |
| bg l (7) | 4.9 | 1.3 | 4.8 | 3.2 | 1.9 | 8.7 |
| bg r (4) | 5.6 | 1.6 | 4.2 | 2.8 | 3.1 | 7.8 |
| multi l (7) | 5.3 | 1.8 | 5.5 | 4.4 | 1.5 | 7.9 |
| multi r (9) | 5.5 | 1.3 | 4.8 | 3.5 | 2.3 | 8.5 |
| occi l (3) | 5.1 | 0.8 | 4.3 | 4.0 | 1.1 | 8.9 |
| cereb l (2) | 5.1 | 1.4 | 5.1 | 3.4 | 1.3 | 8.5 |
| cereb r (2) | 5.7 | 1.0 | 3.8 | 3.2 | 1.4 | 8.7 |
| multi b (5) | 5.1 | 2.3 | 6 | 4.4 | 2.0 | 8.3 |

Bs, brainstem; thal, thalamus; bg, basal ganglia; multi, multiple lesion sites; occi, occipital lobe; cereb, cerebellum; l, left; r, right. All values are in the unit of response keys (1 = left, 5 = center, and 9 = right).

the perception of monaural stimuli (average of the left-sided and right-sided lesions for the *monaural left* stimulus: 1.4 and 8.8 for *monaural right* stimulation), except in patient S7. Two examples of this group (S10 and S32) are presented in Figure 4 and discussed below.

Patient S10 (73 years) had a lesion in the caudal medulla to rostral pons on the left side. All stimuli, except for the monaural left stimulus, were perceived in the right hemifield (see Figure 4C). This patient gave no responses between center (key 5) and right (key 9). Especially in the case of ITD, right-favoring stimuli were mainly perceived on the right side, whereas left-favoring stimuli were perceived in the center or at the right ear. For the monaural-left stimulus, however, the patient consistently reported the left-most position. The patient had the maximal possible score in the MoCa, but, with a PTA3 of 31 dB, a mild hearing loss and also a PTA3 asymmetry of 11 dB, with a higher threshold in the left ear. The patient was not using a hearing aid. In the tone-in-noise detection task, the track of the binaural condition ($N_0S_\pi$) was initially approximately 10 dB below the monaural condition ($N_0S_0$). The track finally converged to the monaural threshold, as the interaural information was no longer exploited, leading to a BMLD below the normal range.

A lesion comparable to the case described above, but on the right side, was found in the patient S32 (75 years), and is shown in Figure 4D. The patient never reported a centralized percept (answer keys 4, 5, and 6 were never used) and all stimuli were perceived very close to either ear. The ILD/ITD = 0 stimulus was more often perceived on the left side. Also, both of the supranatural 1.5 ms ITD stimuli were perceived on the left side. This patient had a BMLD of 12.5 dB (within the normal range) and a MoCA score of 22.

Patient S26 (77 years) who was not in the pure-brainstem lesion group, but had multiple lesion sites in both hemispheres, including the left brainstem, also only responded in two categories, but never reported a stimulus to be in the center.

### 3.5.3 Thalamus lesions

We observed a shift of the auditory space in all patients with a thalamic lesion. However, one left-sided stroke patient showed a shift toward the right side, the other three to the left side. Therefore, the mean responses in this lesion group were only

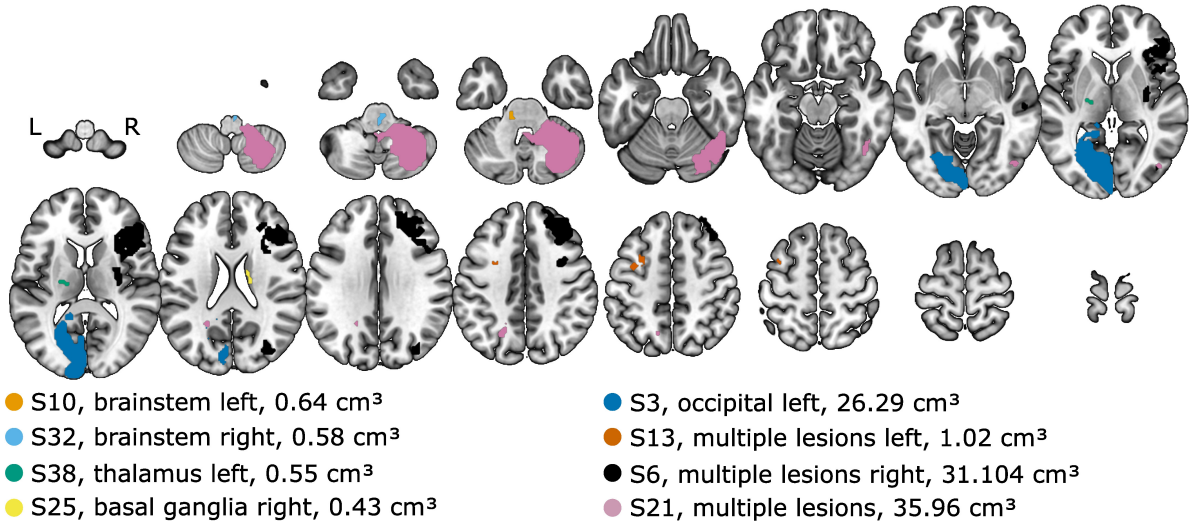



FIGURE 3
Lesion locations overlaid on axial slices of the MNI152 template. The lesion group and the lesion volume is given in the legend. Selected patients are highlighted by the color coding used throughout the figures.

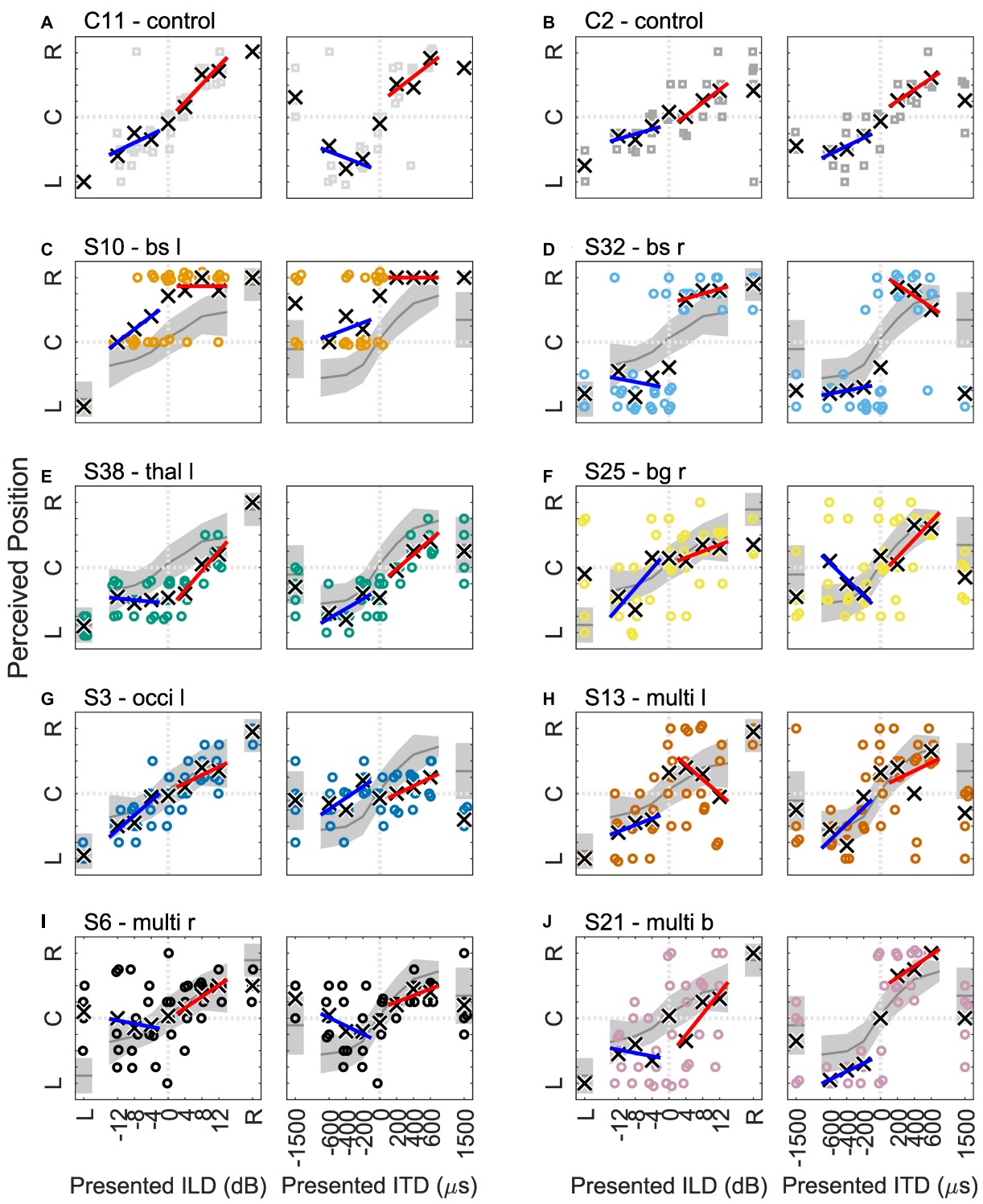



FIGURE 4
Results of the lateralization task for two example control subjects in panels **(A,B)** and eight selected stroke patients in panels **(C–J)**. The colored symbols represent the responses given to the individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively. The gray line and shaded area in panels **(C–J)** indicate the mean and the 1.5 times standard deviation interval around the mean response of the control subjects.

slightly shifted toward the left side (*mean* of 4.7). This also led to a smaller *ITD range* (4.3) and *ILD range* (3.4) than in the control group. In this group, on average, the monaural stimuli were not perceived as much lateralized as in the control group (monaural left: 2.1, monaural right: 8.3). However, this group finding resulted mainly from one patient (S36) that had a high rate of left-right confusions, this was not observed in any other patient of the group.

The patient S38 (59 years) chosen as an example for this group and shown in Figure 4E had a very small lesion in the left lateral thalamus (calculated lesion volume = 278 $mm^3$). This damage seems to have led to a shift in the auditory space toward the left side and a reduced range of lateralization. All left-favoring ILD stimuli and the diotic stimulus were perceived at the same position on the right side, indicating that they were indistinguishable by this patient (see Figure 4E). Unlike the other patients with a thalamic stroke, this patient had almost no benefit from binaural listening in the tone-in-noise detection task (BMLD of 3 dB, below the normal range), even though small changes in ITD led to more lateralized percepts. It is unclear, however, if this patient would have improved with more training, as the second run of the $N_0S_\pi$ condition converged to a lower threshold compared to the first run.

### 3.5.4 Basal ganglia lesions

Due to the small number of patients in the previously presented groups, comparisons between left-sided and right-sided lesions were not feasible. The basal ganglia lesion group, however, consisted of a larger number of patients (11) with 7 left-sided and 4 right-sided lesions. Comparison of the results between the left- and right-sided lesion cases revealed clear differences in lateralization results. Left-sided basal ganglia lesions resulted in lateralization patterns comparable to the control group. Also, the BMLD for these patients was 10 dB to 19 dB and was thus within the normal range. Patients with right-sided lesions however, showed a higher trial-to-trial variability, compared to the left-sided lesion group. On average, the auditory space of the right-sided stroke group was shifted toward the right side (*mean* of 5.6). Two patients in the right-sided basal ganglia lesion group (S19 and S25) perceived the monaural stimuli more centralized than the control group. One patient (S2) of the right-sided lesion group was not able to carry out the tone-in-noise experiment, while the other three had BMLDs of 11 to 16 dB, within the normal range.

The lateralization results of one of the patients with right-sided basal ganglia damage (patient S25, 58 years) is shown in Figure 4F. In this selected patient, the patterns described above (high variability and shift) are also present. The patient had a BMLD within the normal range (16 dB).

### 3.5.5 Cerebellar lesions

Four patients had lesions in the cerebellum. By visual inspection, in two of them (S4 and S37) the lateralization performance differed from the control group. In patient S4, with a right-sided lesion, almost no change in lateralization for different ITDs could be observed, but a smooth, though flat, transition for ILD-based lateralization. Patient S37 showed no impairments in ITD-based lateralization, but the variability of left-favoring ILD stimuli was larger than for right-favoring ILD stimuli. All BMLDs of this group were within the normal range.

### 3.5.6 Multiple lesions in one hemisphere

In many cases, stroke lesions were distributed over several cortical and subcortical areas (see, e.g., patient S6 in Figure 3). Therefore, this group is especially heterogeneous. Almost all patterns described in the previous groups can be found in some of the patients in this group. In more than half of the cases, large differences to the control group can be observed by visual inspection. The trial-to-trial variability of the given responses was increased in a large number of patients with multiple cortical lesions, even if the auditory cortex was not directly affected (e.g., S23 and S48). Especially contralesional difficulties, as shown by highly variable lateralization responses or a less steep slope in the contralesional hemifield, can be found (e.g., S13 and S20). Interestingly, only in two patients (S6 and S48) was a neglect reported with the NIHSS tests. Both had increased variability on the left (contralesional) side and reported some of the left-favoring stimuli on the right side. For some patients (e.g., S6, S20, and S41) with right-sided cortical and subcortical lesions, the left-favoring and the right-favoring stimuli with an ITD of $\pm$1500 $\mu$s were both perceived on the right (the ipsilesional) side. With multiple lesion sites in the left hemisphere, only one patient (S13) had this ipsilesional shift, whereas two others (S29 and S45) also had a shift toward the right—in this case contralesional side.

Two of three patients with damage to the occipital lobe showed almost normal patterns of lateralization, and BMLDs of 11 dB and 18 dB (within the normal range). The third member of this lesion group (patient S3, 72 years, lateralization results shown in Figure 4G) showed almost no sensitivity to ITD-based stimuli, whereas ILDs led to lateralization within the normal range, very similar to the cerebellar stroke patient S4 described above. Compared to the other group members, patient S3 had a slightly reduced BMLD of 8 dB, just within the normal range.

In patient S13 (76 years, presented in Figure 4H) damage mainly to the superior frontal lobe on the left side led to an almost normal lateralization performance in the ipsilesional hemifield, but increased variability for the zero ILD/ITD stimulus and right-favoring stimuli. The monaural left and right stimuli and the BMLD were unremarkable and within the normal range.

Patient S6 (62 years) had widespread lesions in the right hemisphere, including temporal and frontal cortex areas, the insula and basal ganglia. This patient showed high variability for the left monaural stimulus, whereas responses to the right monaural stimulus did not vary much from trial to trial (Figure 4I). In this patient, the difference between the left and the right hemifield was even stronger than in S13. The responses to right-favoring interaural differences varied very little, whereas the left-sided (contralesional) stimuli varied a lot and were even sometimes reported on the other side. This person also showed signs of neglect that were captured with the NIHSS. Again, the BMLD was not affected.

### 3.5.7 Multiple lesions in both hemispheres

Compared to the other lesion groups, data interpretation for the patients with multiple lesions distributed in both hemispheres was very difficult. None of these patients showed results similar to the control group.

Patient S21 (66 years), presented in Figure 4J, had small lesions in the left precuneus cortex and the right occipital cortex, but a large lesion in the right cerebellum, probably also including small portions of the left medulla. This patient had a NIHSS score of 20 points, but the item on neglect was rated with zero points. Patient S21 showed considerable differences in lateralizing ILD- or ITD-based stimuli. ITD-based lateralization appeared mostly unaffected, whereas the ILD-based lateralization results were shifted toward the left with high trial-to-trial variability. However, this could also be related to the PTA3 difference of 9 dB between the two ears (right ear more sensitive). Nevertheless, responses to stimuli without any interaural difference varied strongly, but adding an ITD of 200 μs or −200 μs already led to strong and reliable right-lateralized or left-lateralized percepts. The BMLD of 13 dB was within the normal range.

### 3.5.8 Lesions on the primary auditory pathway

This lesion group contains patients for which the MNI-registered lesion outline overlaps to some extent with areas of the primary auditory pathway (subcortically or cortically). These patients are already included in the previous lesion groups. Many patients of this group show distinct lateralization patterns. Three of the selected subjects shown in Figure 4 had lesions of the auditory pathway. S10 (Figure 4C) had a lesion of the left superior olivary complex (SOC), S32 (Figure 4D) a lesion of the right SOC, and S6 (Figure 4I) a lesion of the right auditory cortex. Altered lateralization patterns were also found for S7 (lesion close to the left cochlear nucleus and SOC) and S31 (lesion between SOC and inferior colliculus). This indicates that direct involvement of the auditory pathway does affect the lateralization in almost all cases. However, in S14 (multiple lesions close to the left SOC and dorsal of the right AC) and S16 (partial overlap with left AC) parts of the auditory pathway seem to be affected without leading to obvious influences on these patients' lateralization performance.

## 3.6 Differences to the control group

Verbal description of the performance in the two experiments as given above fails to reveal some of the general patterns within specific lesion groups. An attempt to quantify the results of both experiments relative to the control group is shown in Figure 5, showing divergences from the control group for all individual patients for different variables calculated from the results of the tone-in-noise detection experiment and the lateralization experiment. For each variable, the upward and downward triangles indicate higher or lower values compared to the normal range (mean ± 1.5 standard deviation) of the control group. The patients are grouped according to the lesion locations. The variables are clustered in group *A* to group *G*, describing different response characteristics. The gray shadings indicate the percentage of deviations from the control results within each specific subgroup (lesion group and variable cluster). For the lesion group 'brainstem left' for example, the percentage of divergences in cluster *A* is approximately 11 percent (one out of nine).

The variables in cluster *A* are the thresholds of the tone-in-noise experiment. For these variables, the strongest divergences were found in the 'thalamus left' lesion group. Cluster *B* consists of variables describing a shift of auditory space. Again, the 'thalamus left' lesion group showed the most divergences for these variables, followed by the groups 'brainstem left,' 'multiple lesions left,' and 'basal ganglia right.' The highest percentage of divergences in cluster *C* (variability of the data) can be observed for the 'basal ganglia right' group, followed by the groups 'brainstem left,' 'multiple lesions bilateral,' and 'multiple lesions left.' The highest percentage of divergences from the control group in variables of cluster *D* are found in the 'brainstem left' lesion group. Cluster *D* is a collection of variables that describe the slopes of the fits. Cluster *E* describes the ranges of ITD- and ILD-based stimuli, as well as the difference between the ranges. Again, the most divergences are found for the group 'brainstem left.' The perception of monaural stimuli (cluster *F*) differed from the control group most for the lesion group 'basal ganglia right,' whereas the large ITDs outside the physiological range (cluster *G*) were perceived differently to the control group by the groups 'brainstem left' and 'brainstem right.'

From the data presented in Figure 5, it becomes apparent that lesions in the left basal ganglia, the occipital lobe and the cerebellum did not lead to lateralization patterns that differ from the control group to any great extent (no more than 33 percent), whereas divergences in many variable clusters are found for patients with damage in the brainstem, the thalamus, and right basal ganglia, and for those individuals with multiple lesions in one or both hemispheres. Much stronger differences,

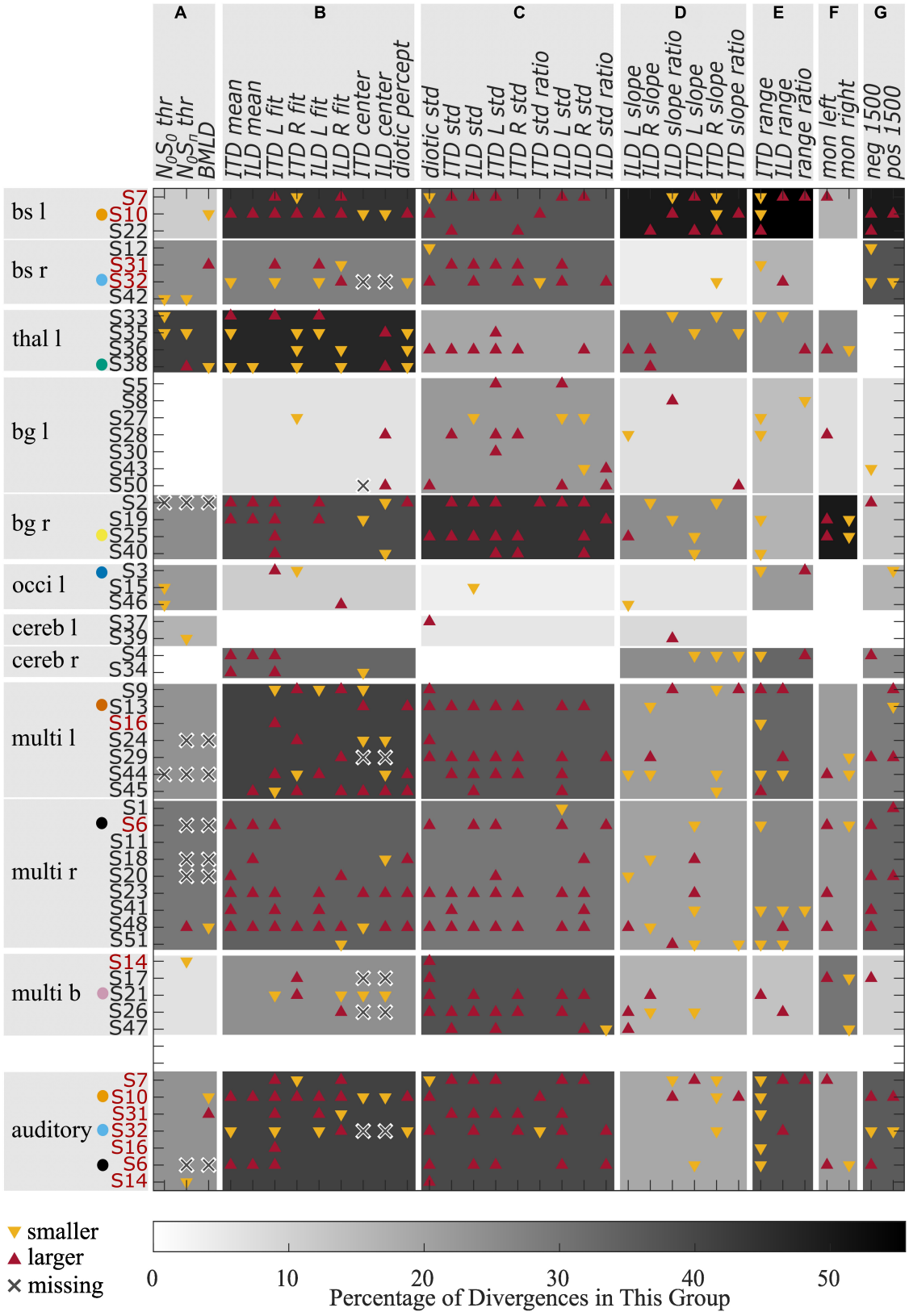



FIGURE 5
Divergences from the control group for different variables. Red upward triangles indicate values of individuals that are larger and yellow downward triangles values that are smaller than the normal values (mean ± 1.5 times standard deviation) calculated from the control group. Crosses indicate missing values. The gray shading indicates the percentage of deviations found within one lesion group for one of the variable groups **(A–G)**. The red font is used for those patients who had a lesion on the primary auditory pathway. Selected patients are highlighted by the color coding used throughout the figures.

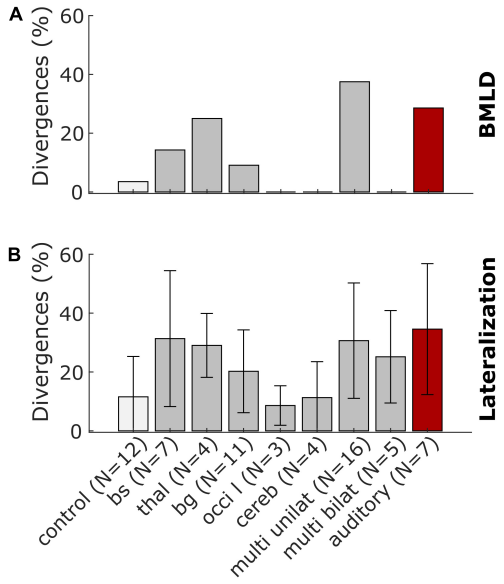



FIGURE 6
Summary representations of the stroke-related binaural hearing deficits across patient groups. Panel **(A)** shows the percentage of patients who had BMLD-values that were worse *(smaller)* than the control group mean by more than 1.5 standard deviations or could not complete the task. Panel **(B)** shows the percentage of possible divergences of all values calculated for the lateralization patterns with error bars denoting the standard deviation across participants for the control and lesion groups.

especially in clusters *B, C,* and *F* are present in those patients with lesions to the right basal ganglia compared to left basal ganglia. Furthermore, all but one of the seven patients who were not able to complete the tone-in-noise detection experiment had multiple lesion sites.

The data presented in Figure 5 was condensed to a simpler representation by extracting the percentage of divergences of the BMLD and a general measure of lateralization performance by averaging over the number of divergences of all variables in clusters *B*-*G*. Lesion groups were pooled over left-sided and right-sided groups. The value for these simplified BMLD and lateralization measures are shown in Figures 6A, B, respectively. Note, that panel A represents the percentage of patients showing smaller than normal or non-convergent tracks in the BMLD task, whereas panel B represents the mean percentage of possible deviations in a given group with the error bars denoting the standard deviation across participants in the group.

For two of eleven patients with lesions in the brainstem or the thalamus, the BMLD diverged from the normal values. One patient of each lesion group had a BMLD of less than 7.5 dB. One patient with a lesion of the basal ganglia and five of 16 patients with unilateral cortical lesions did not produce converging tracks in the task, representing the most remarkable divergence in this task. No divergences were observed for the other lesion groups (see Figure 6A). Two of seven patients with a lesion on the primary auditory pathway diverged from the normal values. One of these two patients produced a BMLD of 6.25 dB, the other one did not produce converging tracks in the dichotic condition of the task. In general, deviations of the BMLD were not frequently observed in the stroke group.

In contrast, for all lesion groups, divergences in terms of the lateralization pattern are found (see Figure 6B). Both measures have the highest percentage of divergences for the patients with a lesion on the primary auditory pathway as shown with the red bars in Figure 6.

## 4. Discussion

In the present exploratory study, our aim was to investigate the binaural perception of individuals in the acute phase of stroke. The performance of the stroke patients in two binaural headphone experiments and the results of the general assessment were compared to an age-matched control group. To our knowledge, this was the first time that the same binaural hearing tasks were conducted in acute-phase stroke patients with various lesions, ranging from the brainstem up to cortical areas. Interpreting these data is a challenging endeavor, especially for the results of the lateralization task, where several metrics are possible and necessary. Using various approaches of comparing patients on a group level and individually with the control group, we found impaired binaural hearing in the majority of stroke patients as shown in Figures 5, 6.

One of the most prominent results was that some of the brainstem-lesion patients lateralized ITD and ILD stimuli in a categorical manner, as suggested by the fact that only a reduced set of response keys was used. For instance, some of these patients commonly gave responses in the categories left-center-right or only left-right, with no responses at intermediate positions. As the information from the left and right ear is integrated in brainstem nuclei for the first time, strongly altered lateralization patterns were expected for the patients who had suffered a stroke to these structures. Accordingly, some of the most prominent distortions in spatial perception were found for brainstem lesion patients. For instance, the cases without responses in the center position were almost exclusively associated with damage of the brainstem (e.g., S32). For this lateralization pattern, at least two interpretations are possible. First, it is possible that a fused image was perceived, but it was lateralized very much toward the sides. An alternative explanation would be that binaural fusion failed for these subjects. As a result, they might have perceived split auditory images (two separate sound sources rather than a single fused image) and reported the position of the dominant image. This ambiguity could be resolved by asking for the number of perceived sound sources in any subsequent studies. The described pattern of side-oriented lateralization was also reported by Furst et al. (2000) for lesions in rostral parts of

the brainstem. In contrast to their findings, we did not observe center-oriented patterns (consistently no lateralized percept) in any of the patients from the brainstem lesion group. Despite the many differences between the brainstem patients and the control group as seen in **Figure 5**, both ILD and ITD stimuli evoked lateralized percepts in all but one patient (S31) of this lesion group. The mean responses of this patient were close to center for all ILD stimuli and left-sided ITD stimuli. The patient had a reduced ITD range, but also a larger standard deviation than the control group. This pattern of responses is suggestive of reduced sensitivity to interaural cues rather than of a center bias.

Left-sided thalamic lesions were, in all cases, correlated with a shift in the lateralization results for both ILD and ITD stimuli. This becomes clear from the high prevalence of deviations in cluster *B* of this group shown in **Figure 5**. Three out of four patients of this lesion group showed a shift toward the ipsilesional side. No conclusion on the effects of left- vs. right-sided lesions can be drawn, because none of the patients in this study had a lesion of the right thalamus. In addition, one subject with a thalamic lesion displayed remarkably high trial-to-trial variability in their lateralization responses. The medial geniculate nucleus (MGN) located in the auditory thalamus receives projections from the ipsilateral and contralateral inferior colliculus and projects to the ipsilateral auditory cortex (Pickles, 2013). Assuming that these projections are damaged by the stroke, one possible explanation for the lateralization shifts could be that corrupted inputs reach the MGN. Higher variability could be related to altered inputs to the cortical representation stages (outputs of the MGN). Besides damage to auditory nuclei, shifts in auditory space could also be related to asymmetrical hearing thresholds (Florentine, 1976). The hearing thresholds at 500 Hz were symmetrical between the two ears in all patients of this lesion group (except for the one with the increased trial-to-trial variability), but the PTA3 asymmetry was in a range of 4 to 13 dB and pointed toward the direction of the shift. This is in agreement with a significant correlation of PTA3 asymmetry with both ILD mean ($\rho = 0.34$, $p = 0.017$) and ITD mean ($\rho = 0.34$, $p = 0.016$) when including all patients of the stroke group. Even though the PTA3 asymmetry can influence the results of the lateralization task, the finding of shifted auditory space for all thalamic-lesion patients indicates an influence of the left thalamus on lateralization. This is in line with previous studies that have found a connection between thalamus lesions and visuospatial neglect (Karnath et al., 2002).

A biased auditory egocentric space in cases with inferior parietal and frontal dysfunction was reported by Bellmann et al. (2001). Further, they found an imbalance of attentional load allocated to the left and right hemispaces (hemispatial inattention) following lesions of basal ganglia and insular cortex. Both mechanisms (biased spatial perception and unbalanced spatial attention across hemifields) come into play for our lateralization task, but their effects are difficult to distinguish in our data. Shifted auditory space and altered lateralization slopes (steepness of the lateralization function of ITD/ILD, see, e.g., S38) indicate distortions of spatial representation. Increased trial-to-trial variability, on the other hand (e.g., S13 and S25), may be indicative of attentional or cognitive impairments, or both. Also, Gutschalk and Dykstra (2015) concluded that more work is needed to develop clinical protocols that can clearly distinguish localization deficits from disorders of spatial cognition. The effects of the right basal ganglia on the lateralization patterns that we observed, could be attributed to attentional deficits. In contrast to Bellmann et al. (2001), our results show that the perception of both left-favoring, as well as right-favoring stimuli was affected in some patients (see **Figure 5**). Given the supra-modal nature of the neglect syndrome, a basal ganglia lesion may affect auditory and visual hemispatial attention. Influences of right basal ganglia lesions on the visuospatial perception of both, ipsi- and more frequently contralesional stimuli were already reported by Karnath et al. (2002).

For almost all patients with multiple lesions in one or both hemispheres, we found lateralization patterns that differed from the control group in terms of increased variability and decreased slopes, as shown by the high number of divergences in the clusters *C* and *D* of **Figure 5**. Besides contralesional deficits as in patient S6 with multiple lesion sites, including the right temporal lobe, many patients also displayed ipsilesional deficits for both left- and right-sided lesions. This is only partially in line with previous literature (see Häusler and Levine, 2000 for a review) that suggests a dominance of the right hemisphere in auditory spatial representation. In our study, a comparison of left-sided and right-sided cortical lesions might not be meaningful, because of the unequal distribution of lesion sites. Since the inability to understand and produce speech is mainly observed after damage to left-hemispheric language areas, and was one of the exclusion criteria, left-sided and right-sided groups differed in terms of their lesion locations. For basal ganglia lesions however, strong differences between the left and right side were observed, with more frequent and more severe deficits after right-sided lesions than for left-sided lesions. This result is similar to the results presented in Karnath et al. (2002) for the visual modality.

The perception of $\pm 1500$ μs ITDs, i.e., ones that are larger than those usually experienced under natural listening conditions, was only rarely affected. In the brainstem-lesion patients S10, S22, and S32, the left-favoring and right-favoring stimuli were both perceived on the contralesional side. The ambiguity of this stimulus stems from the conflicting interaural cues conveyed by the envelope (indicating the position on the leading side) and the temporal fine structure (indicating a stimulus on the opposite site). With damage in one side of the brainstem, the ipsilesional cue may not be accessible to the next processing stage or less weight might be given to it. With multiple cortical and subcortical lesions, the outcomes are more diverse. While some patients (e.g., S13 and S20) perceived

both of these stimuli only on the ipsilesional side, other patients (e.g., S29 and S45) perceived them exclusively on the contralesional side. These findings point to the interpretation that disturbances at different levels of ITD representation stages can lead to stimuli with unnaturally large ITDs being perceived at different intracranial positions. Coding of such large ITDs was already found to differ at midbrain, compared to cortical, levels (Thompson et al., 2006; Kriegstein et al., 2008). While the exact combination of computational processes by which the auditory system encodes ITDs remains elusive, stroke lesion studies such as the present one could potentially aid in their elucidation. However, due to the rarity of psychoacoustic data from stroke survivors, combined with the highly individual nature of stroke lesions, more data is needed before meaningful interpretations are possible.

The dichotic tone-in-noise detection task is a better test for the implicit use of interaural differences compared to the more commonly used measurements of just-noticeable differences in ILD and ITD cues. In many cases, the performance in these tasks depends on the explicit perception of intracranial positions rather than on the general ability to exploit binaural cues for unmasking. To be able to directly compare the results of the implicit tone-in-noise detection task with those of the explicit lateralization task, we refrained from using speech-related tasks such as the one used in, e.g., Tissieres et al. (2019). The results of our lateralization task revealed that five of the six patients with lesions in the right basal ganglia showed remarkable impairments in ITD-based lateralization, which requires the explicit use of interaural differences. Four of these five patients had a BMLD in the normal range (and one only slightly below the normal range), indicating that they had access to implicit ITD information, despite the fact that they could not exploit ITDs explicitly in the lateralization task. This reveals that altered ITD-based lateralization is not necessarily related to dysfunctional encoding at the primary stage in the superior olivary complex. Instead, it seems that damage to the explicit representation stages can impair lateralization even if the primary encoding stages remain unaffected. In general, few patients had smaller than normal BMLDs. Similarly, also Lynn et al. (1981) reported that the speech BMLD was not affected in patients with lesions on cerebral, thalamic, midbrain or rostral pontine levels. In their study, only patients with lesions at the ponto-medullary level showed a reduced BMLD. In our study, two patients had a lesion at the ponto-medullary junction. One of these two patients had a reduced BMLD. Only two of the remaining 48 patients with lesions at other areas had a reduced BMLD. Due to these low numbers, no clear supporting or contradicting conclusions can be drawn. On the other hand, the inability to do the tone-in-noise detection task (missing values due to non-convergent tracks, indicated by crosses in **Figure 5**) was observed in some patients in which, among other areas, the basal ganglia were damaged and in some patients also frontal cortical areas. Cortico-striatal loops have been shown to be involved in auditory discrimination learning (Znamenskiy and Zador, 2013), which is a necessary ability for this experiment. This implies that the slightly more complex tone-in-noise detection task needs to be learned first, and may therefore not be an optimal measure of the accessibility of implicit interaural information for participants with learning difficulties. Besides the theoretical implications, the deviations in the BMLD as shown in **Figure 6**, and in particular the inability to complete the task, could be of clinical interest. The BMLD is correlated with age, but the occurrence of stroke does appear to constitute an additional factor affecting binaural tone-in-noise detection performance for some stroke patients. As such, the BMLD could potentially be used clinically to detect effects of stroke on binaural hearing.

Due to the heterogeneous group of participants and the highly individual nature of stroke lesions, the present study is affected by a number of confounding factors. We sought to capture some of these by additional auditory and non-auditory measures such as the audiometry and the MoCA. To paraphrase Gallun (2021), the perturbations caused by nature and not manipulated in the laboratory are never uniform and not easily documented.

In the present study, the selection of patients could not control for the influence of age and hearing loss, but the control group was age-matched and did not differ significantly in their hearing thresholds or in the results of the general assessment. Only the results of the MoCA differed significantly between the stroke and control groups (see **Table 1**). Almost all non-stroke-related difficulties should be rather equally present in both groups. We therefore concluded that the observed effects on a group level, though not on an individual level, can be attributed to the stroke and possible comorbidities, rather than on hearing loss. The selection of those cases presented in **Figures 3**, **4** was based on the results of the lateralization task. The selected stroke patients span the whole range for all measured variables (see **Supplementary Figure 1**). For the stroke patients, of course, the premorbid performance is not known.

The stimuli of both experiments were chosen to be centered around 500 Hz, which is usually spared by age-related hearing loss. The threshold for this frequency was on average 16 dB HL and did not exceed 35 dB HL for any participant. No more than a 10 dB difference between the left and right side was measured at this frequency for any of the participants. We therefore did not expect large influences of hearing loss or asymmetrical hearing abilities on our results. Nevertheless, as discussed above, a correlation of PTA3 asymmetry and shifted auditory space was found.

We focused only on those lesions that had a high signal on the DWI and a low signal in the ADC map, thus representing restricted diffusion. In many cases, older lesions and other damage to brain tissue were present that could have influenced performance in the different tasks. However, improvements from diaschisis or functional reorganization is

known to drive neurologic recovery already in the acute phase (Sang-Bae and Byung-Woo, 2013). In addition, in healthy subjects, reorganization of lateralization with altered ITD cues occurs within few days (Trapeau and Schönwiesner, 2015). This suggests, that binaural hearing impairments are dominated by the acute damage and less by old lesions. Complete lesions of specific parts of the brain are used to study the system in ablation studies in animals. In our patients however, the damage does not necessarily include entire brain structures and may leave some functioning neuronal processing. Furthermore, as pointed out by Neff et al. (1975), experiments in well controlled ablation studies in animals measure the functioning of the remaining system and not necessarily the functioning of the damaged part. In contrast to such ablation studies, the general state of brain structures that were not damaged by the acute stroke varied widely in our population. The observed variability in performance must therefore be partially attributed to differences in the damages as well as to differences in the remaining brain structures, rather than solely to the acute stroke lesion.

Not only did individual characteristics of the patients affect the data, but also external constraints such as the restricted time for the behavioral experiments. The short time we had with the patients did not allow for dedicated training runs nor for repetitions of any task. One example where more time would have been necessary was when patients were not able to do the tone-in-noise experiment. In retrospect, from the trend in these patients' adaptive tracks, it appeared as if some of these patients would have learned to do the task had there been more runs of the same experiment. In addition, the hospital room in which the study was conducted was comparably quiet, but had no sound booth. Finally, the fact that some lesion groups contained only two patients, allowed only limited interpretations. Differentiation between the effects of lesions of a particular anatomical structure as opposed to differences between left-sided and right-sided lesions of that brain area is restricted.

From the data obtained in our experiments, we do not know if these patients also had difficulties in free-field-localization tasks, in which spectral cues are available in addition to natural combinations of ILDs and ITDs. However, as both cues are often perceived with a similar bias and spectral cues are less salient in elderly listeners, we assume that some patients will have localization biases, at least during the acute phase. If a bias remains in the chronic phase of stroke, individualized ILD- and ITD-manipulating algorithms could potentially be exploited to improve localization performance (e.g., Brown, 2018).

## 5. Conclusion

This exploratory study revealed some expected divergences in binaural perception between the results of patients with acute ischemic stroke lesions and the results of the control group: Impaired contralesional lateralization was found after right cortical and brainstem lesions, which is consistent with previous reports. Other findings could be expected, based on today's understanding of binaural processing and decoding of spatial cues: The perception of binaural stimuli with unnaturally large ITDs is affected differently based on the lesion location. Other findings were less expected, such as the shift in auditory space in all patients with thalamic lesions or the large difference induced by left and right basal ganglia lesions. In contrast to previous reports, no apparent hemispheric difference from cortical lesions regarding the variability of lateralization data were found, and the binaural benefit in the tone-in-noise detection task was unaffected in most patients, although many patients with multiple lesion sites could not complete this task. While it may be too early to suggest any revisions to our understanding of interaural cue encoding or decoding, the outcomes may nevertheless foster more focused future investigations in selected groups of patients with specific lesions, or in animal models. Investigating acute-phase stroke patients may even be an additional avenue to deepen our understanding of the healthy auditory system in a way that is difficult when studying the healthy system in isolation.

## Data availability statement

The dataset analyzed for this study can be found at https://doi.org/10.5281/zenodo.7415436. Due to ethical restrictions, only the lesion masks, but no raw MRI images, are provided. Further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by the Medical Research Ethics Board of the University of Oldenburg, Germany. The participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## Author contributions

AD, MD, and PS contributed to the conception and design of the study. AD, MD, PS, and KW planned the experimental procedures. AD organized the data base, performed the analysis, and wrote the first draft of the manuscript. MB and AM recruited participants and acquired data. AD, PS, and BS did MRI analyses. AD and MD interpreted the data. MD and HP wrote further sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

## Funding

This work was supported by the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme grant agreement no. 716800 (ERC Starting Grant to MD).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1022354/full#supplementary-material

A figure with the correlations between age and PTA and the general assessment scores and the tone-in-noise-detection thresholds, and the lateralization results, such as those presented in **Figure 4** for all control subjects and all stroke patients, are available as supplementary data.

SUPPLEMENTARY FIGURE 1
Scatter plots representing correlations between age and PTA3 thresholds and the results of the non-auditory measurements (MoCA, NIHSS, BDI, and MWT-B) for the control group (squares) and the stroke group (circles). In each subpanel, linear-regression lines, the Pearson correlation coefficient $\rho$, and the respective $p$-value are shown in the form "$\rho$ ($p$-value)". Selected participants are highlighted by the color coding used throughout the figures.

SUPPLEMENTARY FIGURE 2
Results of the binaural tone-in-noise detection experiment. Tone-in-noise detection thresholds for $N0S_\pi$ condition (up- and downward triangles for stroke and control subjects, respectively) and N0S0 condition (left- and right-pointing triangles) over PTA3 **(panel A)** and over age **(panel B)**. In each subpanel, linear-regression lines, the Pearson correlation coefficient $\rho$, and the respective $p$-value are shown in the form "$\rho$ ($p$-value)". Selected participants are highlighted by the color coding used throughout the figures.

SUPPLEMENTARY FIGURE 3
Results of the lateralization task for patients S1–S10. The circles represent the responses given to the individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively. The gray line and shaded area indicate the mean and the 1.5 times standard deviation interval around the mean response of the control subjects.

SUPPLEMENTARY FIGURE 4
Results of the lateralization task for patients S11–S20. The circles represent the responses given to the individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively. The gray line and shaded area indicate the mean and the 1.5 times standard deviation interval around the mean response of the control subjects.

SUPPLEMENTARY FIGURE 5
Results of the lateralization task for patients S21–S30. The circles represent the responses given to the individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively. The gray line and shaded area indicate the mean and the 1.5 times standard deviation interval around the mean response of the control subjects.

SUPPLEMENTARY FIGURE 6
Results of the lateralization task for patients S31–S40. The circles represent the responses given to the individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively. The gray line and shaded area indicate the mean and the 1.5 times standard deviation interval around the mean response of the control subjects.

SUPPLEMENTARY FIGURE 7
Results of the lateralization task for patients S41–S50. The circles represent the responses given to the individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively. The gray line and shaded area indicate the mean and the 1.5 times standard deviation interval around the mean response of the control subjects.

SUPPLEMENTARY FIGURE 8
Results of the lateralization task for control subjects C1–C12. The squares represent the responses given to the individual trials of the same stimulus, except for the discarded first trial. The black crosses indicate the means of the given responses. The red and blue lines represent linear fits to right-favoring and left-favoring stimuli, respectively.

## References

Bamiou, D.-E., Werring, D., Cox, K., Stevens, J., Musiek, F. E., Brown, M. M., et al. (2012). Patient-reported auditory functions after stroke of the central auditory pathway. *Stroke* 43, 1285–1289. doi: 10.1161/STROKEAHA.111.644039

Baumgärtel, R. M., and Dietz, M. (2018). Extent of sound image lateralization: Influence of measurement method. *Acta Acust. United Acust.* 104, 748–752. doi: 10.3813/AAA.919215

Beck, A. T., Steer, R. A., and Brown, G. K. (2013). *Beck depressions-inventar - FS (BDI-FS): Deutsche bearbeitung*. London: Pearson.

Bellmann, A., Meuli, R., and Clarke, S. (2001). Two types of auditory neglect. *Brain* 124, 676–687.

Bernstein, L. R., and Trahiotis, C. (2011). Lateralization produced by interaural intensitive disparities appears to be larger for high- vs low-frequency stimuli. *J. Acoust. Soc. Am.* 129, 15–20. doi: 10.1121/1.3528756

Brown, C. A. (2018). Corrective binaural processing for bilateral cochlear implant patients. *PLoS One* 13:e0187965. doi: 10.1371/journal.pone.0187965

Dietz, M., Bernstein, L. R., Trahiotis, C., Ewert, S. D., and Hohmann, V. (2013). The effect of overall level on sensitivity to interaural differences of time and level at high frequencies. *J. Acoust. Soc. Am.* 134, 494–502. doi: 10.1121/1.4807827

Ewert, S. (2013). "AFC - A modular framework for running psychoacoustic experiments and computational perception models," in *Proceedings of the international conference on acoustics AIA-DAGA*, Merano, 1326–1329.

Florentine, M. (1976). Relation between Lateralization and loudness in asymmetrical hearing losses. *J. Am. Audiol. Soc.* 1, 243–251.

Furst, M., Aharonson, V., Levine, R., Fullerton, B., Tadmor, R., Pratt, H., et al. (2000). Sound lateralization and interaural discrimination. Effects of brainstem infarcts and multiple sclerosis lesions. *Hear Res.* 143, 29–42. doi: 10.1016/s0378-5955(00)00019-8

Gallun, F. J. (2021). Impaired binaural hearing in adults: A selected review of the literature. *Front. Neurosci.* 15:610957. doi: 10.3389/fnins.2021.610957

Gatehouse (2004). *The speech, spatial and qualities of hearing scale (SSQ)*. Perinton, NY: Gatehouse.

Gokhale, S., Lahoti, S., and Caplan, L. R. (2013). The neglected neglect: Auditory neglect. *JAMA Neurol.* 70, 1065–1069. doi: 10.1001/jamaneurol.2013.155

Gutschalk, A., and Dykstra, A. (2015). Auditory neglect and related disorders. *Handb. Clin. Neurol.* 129, 557–571. doi: 10.1016/B978-0-444-62630-1.00031-7

Häusler, R., and Levine, R. A. (2000). Auditory dysfunction in stroke. *Acta Otolaryngol.* 120, 689–703. doi: 10.1080/000164800750000207

Heilman, K. M., Valenstein, E., and Watson, R. T. (2000). Neglect and related disorders. *Semin. Neurol.* 20, 463–470. doi: 10.1055/s-2000-13179

Jenkinson, M., Bannister, P., Brady, M., and Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 17, 825–841. doi: 10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E. J., Woolrich, M. W., and Smith, S. M. (2012). FSL. *Neuroimage* 62, 782–790. doi: 10.1016/j.neuroimage.2011.09.015

Joliot, M., Jobard, G., Naveau, M., Delcroix, N., Petit, L., Zago, L., et al. (2015). AICHA: An atlas of intrinsic connectivity of homotopic areas. *J. Neurosci. Methods* 254, 46–59. doi: 10.1016/j.jneumeth.2015.07.013

Karnath, H. O., Himmelbach, M., and Rorden, C. (2002). The subcortical anatomy of human spatial neglect: Putamen, caudate nucleus and pulvinar. *Brain* 125, 350–360. doi: 10.1093/brain/awf032

Kriegstein, K., Griffiths, T. D., Thompson, S. K., and McAlpine, D. (2008). Responses to interaural time delay in human cortex. *J. Neurophysiol.* 100, 2712–2718. doi: 10.1152/jn.90210.2008

Lehrl, S. (2005). *Mehrfach-wortschatz-intelligenztest: MWT-B*. Frederick, MD: Spitta.

Lynn, G. E., Gilroy, J., Taylor, P. C., and Leiser, R. P. (1981). Binaural masking-level differences in neurological disorders. *Arch. Otolaryngol.* 107, 357–362. doi: 10.1001/archotol.1981.00790420031007

Nasreddine, Z. S., Phillips, N. A., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., et al. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

National Institute of Neurological Disorders and Stroke [NIHSS] (2019). *NIH stroke scale*. Bethesda, MD: National Institute of Neurological Disorders and Stroke. Available online at: https://www.ninds.nih.gov/stroke-scales-and-related-information

Neff, W. D., Diamond, I. T., and Casseday, J. H. (eds) (1975). *Behavioral studies of auditory discrimination: Central nervous system*. Berlin: Springer International Publishing.

Pickles, J. O. (2013). *An introduction to the physiology of hearing*, 4 Edn. Leiden: Brill.

Robert Koch-Institut (2017). *12-Month prevalence of stroke or chronic consequences of stroke in Germany. RKI-Bib1*. Berlin: Robert Koch-Institut.

Rorden, C., and Brett, M. (2000). Stereotaxic display of brain lesions. *Behav. Neurol.* 12, 191–200. doi: 10.1155/2000/421719

Sang-Bae, K., and Byung-Woo, Y. (2013). Mechanisms of Functional Recovery from Stroke. *Front. Neurol. Neurosci.* 32, 1–8. doi: 10.1159/000346405

Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306

Sitek, K. R., Gulban, O. F., Calabrese, E., Johnson, G. A., Lage-Castellanos, A., Moerel, M., et al. (2019). Mapping the human subcortical auditory system using histology, postmortem MRI and in vivo MRI at 7T. *Elife* 8:e48932. doi: 10.7554/eLife.48932

Smith, S. M. (2002). Fast robust automated brain extraction. *Hum. Brain. Mapp.* 17, 143–155. doi: 10.1002/hbm.10062

Spierer, L., Bellmann-Thiran, A., Maeder, P., Murray, M. M., and Clarke, S. (2009). Hemispheric competence for auditory spatial representation. *Brain* 132, 1953–1966. doi: 10.1093/brain/awp127

Thiran, A. B., and Clarke, S. (2003). Preserved use of spatial cues for sound segregation in a case of spatial deafness. *Neuropsychologia* 41, 1254–1261. doi: 10.1016/S0028-3932(03)00014-9

Thompson, S. K., Kriegstein, K., Deane-Pratt, A., Marquardt, T., Deichmann, R., Griffiths, T. D., et al. (2006). Representation of interaural time delay in the human auditory midbrain. *Nat. Neurosci.* 9, 1096–1098. doi: 10.1038/nn1755

Tissieres, I., Crottaz-Herbette, S., and Clarke, S. (2019). Implicit representation of the auditory space: Contribution of the left and right hemispheres. *Brain Struct. Funct.* 224, 1569–1582. doi: 10.1007/s00429-019-01853-5

Trapeau, R., and Schönwiesner, M. (2015). Adaptation to shifted interaural time differences changes encoding of sound location in human auditory cortex. *Neuroimage* 118, 26–38. doi: 10.1016/j.neuroimage.2015.06.006

Yarkoni, T., Poldrack, R. A., Nichols, T. E., van Essen, D. C., and Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nat. Methods* 8, 665–670. doi: 10.1038/nmeth.1635

Znamenskiy, P., and Zador, A. M. (2013). Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. *Nature* 497, 482–485.

frontiers | Frontiers in Neuroscience

TYPE Correction
PUBLISHED 01 February 2023
DOI 10.3389/fnins.2023.1143063



OPEN ACCESS

EDITED AND REVIEWED BY
Yi Zhou,
Arizona State University, United States

*CORRESPONDENCE
Anna Dietze
✉ anna.dietze@uni-oldenburg.de

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 12 January 2023
ACCEPTED 17 January 2023
PUBLISHED 01 February 2023

CITATION
Dietze A, Sörös P, Bröer M, Methner A, Pöntynen H, Sundermann B, Witt K and Dietz M (2023) Corrigendum: Effects of acute ischemic stroke on binaural perception. *Front. Neurosci.* 17:1143063. doi: 10.3389/fnins.2023.1143063



# Corrigendum: Effects of acute ischemic stroke on binaural perception

Anna Dietze[1,2]*, Peter Sörös[3,4], Matthias Bröer[3], Anna Methner[3], Henri Pöntynen[1,2], Benedikt Sundermann[4,5], Karsten Witt[3,4] and Mathias Dietz[1,2,4]

[1]Department of Medical Physics and Acoustics, University of Oldenburg, Oldenburg, Germany, [2]Cluster of Excellence "Hearing4all", University of Oldenburg, Oldenburg, Germany, [3]Department of Neurology, School of Medicine and Health Sciences, University of Oldenburg, Oldenburg, Germany, [4]Research Center Neurosensory Science, University of Oldenburg, Oldenburg, Germany, [5]Institute of Radiology and Neuroradiology, Evangelisches Krankenhaus Oldenburg, Oldenburg, Germany

KEYWORDS

**binaural hearing, psychoacoustics, brain lesions, lateralization, binaural masking level difference, magnetic resonance imaging, stroke**

A corrigendum on
Effects of acute ischemic stroke on binaural perception

by Dietze, A., Sörös, P., Bröer, M., Methner, A., Pöntynen, H., Sundermann, B., Witt, K., and Dietz, M. (2022). *Front. Neurosci.* 16:1022354. doi: 10.3389/fnins.2022.1022354

In the published article, there was a typographical error. In the study by Bernstein and Trahiotis (2011), stronger lateralization was perceived for the same ILD magnitude and higher-frequency signals, not for lower-frequency signals as stated in the original version of the manuscript. A correction has been made to Section 3.5.1 Control group, Paragraph 1. The corrected paragraph is below.

"Physically left-favoring, to consecutively more right-favoring stimuli, were perceived from the left to the right inside the participants' heads for the ILD and ITD stimuli for all control subjects, with only slight deviations. Apparently, the chosen ILDs, ranging from −12 to 12 dB did not lead to strongly lateralized auditory images (responses close to response keys 1 = left and 9 = right). Previous studies already demonstrated that the extent of perceived lateralization for ILDs of this magnitude varies across subjects (Baumgärtel and Dietz, 2018). It also depends on frequency, with stronger lateralization perceived for the same ILD magnitude and higher-frequency signals (Bernstein and Trahiotis, 2011). Auditory space was distributed roughly symmetrically around zero ITD/ILD, being reflected in the average perceived position over all ILD and ITD stimuli (*mean*) of 5.2 in the control group. Even in the control group, the perceived intracranial positions were not perfectly distributed around the center (5.0). Monaural left or right stimulation was perceived close to the most lateralized intracranial positions (*mon left*: 1.5 and *mon right*: 8.6) with almost no intra-individual variability. For all ILDs and all absolute ITDs $\leq$ 600 μs, a small variability in single trials can be seen. The standard deviation of given responses was for all stimuli approximately in the range of one response key for the control subjects (e.g., 1.1 for *diotic std.*, the standard deviation of zero ILD/ITD). Only one person of the control group produced much more variable data. The variability of ITDs of ±1500 μs was larger than for smaller ITDs in most control subjects. This unnaturally large ITD was perceived less lateralized compared to smaller absolute ITDs. Based only on the center frequency (500 Hz), one cannot distinguish between a time shift of −500 or +1500 μs, as the period at this frequency is 2000 μs. However, since the stimulus is a white noise of 333 Hz bandwidth centered around 500 Hz, the auditory system

can partially resolve this ambiguity, by exploiting either the interaural correlation at other frequencies or the envelope ITD. The range of lateralization was larger for ITDs (5.5) compared to ILDs (3.7) and for both interaural differences was much smaller than the maximal possible range of 8".

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

## Publisher's note



## References

Baumgärtel, R. M., and Dietz, M. (2018). Extent of sound image lateralization: Influence of measurement method. *Acta Acust. United Acust.* 104, 748–752. doi: 10.3813/AAA.919215

Bernstein, L. R., and Trahiotis, C. (2011). Lateralization produced by interaural intensitive disparities appears to be larger for high- vs low-frequency stimuli. *J. Acoust. Soc. Am.* 129, 15–20. doi: 10.1121/1.3528756

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 04 January 2023
DOI 10.3389/fnins.2022.1021541



OPEN ACCESS

EDITED BY
Lina Reiss,
Oregon Health and Science University,
United States

REVIEWED BY
Jan W. H. Schnupp,
City University of Hong Kong,
Hong Kong SAR, China
Yoojin Chung,
Decibel Therapeutics, Inc.,
United States

*CORRESPONDENCE
Michael Pecka
✉ pecka@bio.lmu.de

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 17 August 2022
ACCEPTED 08 December 2022
PUBLISHED 04 January 2023

CITATION
Müller M, Hu H, Dietz M, Beiderbeck B, Ferreiro DN and Pecka M (2023) Temporal hyper-precision of brainstem neurons alters spatial sensitivity of binaural auditory processing with cochlear implants.
*Front. Neurosci.* 16:1021541.
doi: 10.3389/fnins.2022.1021541



# Temporal hyper-precision of brainstem neurons alters spatial sensitivity of binaural auditory processing with cochlear implants

Michaela Müller[1], Hongmei Hu[2,3], Mathias Dietz[2,3], Barbara Beiderbeck[1], Dardo N. Ferreiro[4,5] and Michael Pecka[1,4]*

[1]Graduate School of Systemic Neurosciences, Ludwig-Maximilians-Universität, Munich, Germany, [2]Department of Medical Physics and Acoustics, Carl von Ossietzky University of Oldenburg, Oldenburg, Germany, [3]Cluster of Excellence "Hearing4All", Universität Oldenburg, Oldenburg, Germany, [4]Section of Neurobiology, Faculty of Biology, LMU Biocenter, Ludwig-Maximilians-Universität, Munich, Germany, [5]Department of General Psychology and Education, Ludwig-Maximilians-Universität, Munich, Germany

The ability to localize a sound source in complex environments is essential for communication and navigation. Spatial hearing relies predominantly on the comparison of differences in the arrival time of sound between the two ears, the interaural time differences (ITDs). Hearing impairments are highly detrimental to sound localization. While cochlear implants (CIs) have been successful in restoring many crucial hearing capabilities, sound localization *via* ITD detection with bilateral CIs remains poor. The underlying reasons are not well understood. Neuronally, ITD sensitivity is generated by coincidence detection between excitatory and inhibitory inputs from the two ears performed by specialized brainstem neurons. Due to the lack of electrophysiological brainstem recordings during CI stimulation, it is unclear to what extent the apparent deficits are caused by the binaural comparator neurons or arise already on the input level. Here, we use a bottom-up approach to compare response features between electric and acoustic stimulation in an animal model of CI hearing. Conducting extracellular single neuron recordings in gerbils, we find severe hyper-precision and moderate hyper-entrainment of both the excitatory and inhibitory brainstem inputs to the binaural comparator neurons during electrical pulse-train stimulation. This finding establishes conclusively that the binaural processing stage must cope with highly altered input statistics during CI stimulation. To estimate the consequences of these effects on ITD sensitivity, we used a computational model of the auditory brainstem. After tuning the model parameters to match its response properties to our physiological data during either stimulation type, the model predicted that ITD sensitivity to electrical pulses is maintained even for the hyper-precise inputs. However, the model exhibits severely altered spatial sensitivity during electrical stimulation compared to acoustic: while

resolution of ITDs near midline was increased, more lateralized adjacent source locations became inseparable. These results directly resemble recent findings in rodent and human CI listeners. Notably, decreasing the phase-locking precision of inputs during electrical stimulation recovered a wider range of separable ITDs. Together, our findings suggest that a central problem underlying the diminished ITD sensitivity in CI users might be the temporal hyper-precision of inputs to the binaural comparator stage induced by electrical stimulation.

KEYWORDS

**sound localization, hearing, jitter, electrophysiology, computer modeling, electrical hearing**

## Introduction

Spatial hearing is vital to navigate the busy environments of our daily life. The location of a sound source is neuronally determined by binaural comparison of sound parameters between the two ears, namely interaural time and level differences (ITDs and ILDs, respectively). However, sound localization is impacted already by moderate hearing deficits, resulting in difficulties– amongst others–to orient and identify speakers. In recent years, many efforts have been made to improve hearing based on bilateral cochlear implants (CIs). With CIs, the amplitude envelope of sounds reaching the ears is extracted in multiple spectral channels (Wilson et al., 1991). This envelope information is subsequently passed onto the auditory nerve (AN) fibers by modulating the amplitude of an electrical pulse-train stimulation. The AN fibers then provide input to the brainstem nuclei involved in ILD and ITD detection (see below and **Figure 1**). While sensitivity to ILDs is rather well maintained in bilateral CI users, ITD sensitivity is very coarse (compared to normal acoustic listeners) and mostly resembles lateralization, even under laboratory conditions (Laback et al., 2015). Moreover, CI-based ITD sensitivity is limited to carrier frequencies of the electrical pulse trains below 500 pulses per second (pps), which is much lower than the typically used pulse rates of CIs pps (Laback et al., 2015). Since ITDs are the crucial cue for human communication and orientation (Wightman and Kistler, 1998; Macpherson and Middlebrooks, 2002), overcoming this lack of electrical ITD sensitivity is desirable. However, the underlying physiological reasons are not well understood. In particular, little is known about potential differences in the neuronal computations between acoustic and electrical ITDs.

Acoustic ITDs are primarily detected by neurons in two brainstem nuclei, the medial superior olive (MSO) and the lateral superior olive (LSO). Since MSO is predominately tuned to low frequencies (<2 kHz), it is regarded to primarily detect fine-structure ITDs (Grothe et al., 2010), while the LSO is mostly sensitive to ITDs in the envelope waveforms of high-frequency carriers (Finlayson and Caspary, 1991). Notably, electrical CI stimulation of AN fibers provides envelope ITD information only and mostly activates higher frequency regions of the cochlea, which predominately innervates LSO. Correspondingly, the observed limits of electrical ITD sensitivity match those of acoustic envelope ITD sensitivity reported for the LSO (Bernstein and Trahiotis, 2002, 2009; Klein-Hennig et al., 2011; Ihlefeld et al., 2014; Laback et al., 2015). Hence, it was recently hypothesized that the LSO is the primary site for ITD sensitivity in CI listeners (Dietz et al., 2016; Hu et al., 2022).

The neuronal processing of ITDs is based on μs precise temporal comparison/integration of synaptic inputs from both ears. Specifically, LSO neurons receive excitatory synaptic innervation from the spherical bushy cells in the ipsilateral antero-ventral cochlear nucleus (AVCN) and inhibitory input from the medial nucleus of the trapezoid body (MNTB) (Grothe et al., 2010). The MNTB itself is innervated by the contralateral AVCN. Both the AVCN and MNTB neurons exhibit faithful locking of their action potential (AP) response times to specific phases of the stimulus waveform (fine structure or envelope transients) (Smith et al., 1998). This enables LSO neurons to generate exquisite ITD sensitivity by detecting differences in input timing. With regard to comparisons to electrical hearing, we have recently demonstrated faithful and very precise ITD sensitivity in high frequency LSO neurons using short acoustic click trains (the stimulus that most resembles electrical CI stimulation) for click-rates up to 500 pps (Beiderbeck et al., 2018). Interestingly, the phase-locking of AN, which directly innervates AVCN (and indirectly the MNTB *via* the AVCN), is severely heightened (e.g., up to approx. a factor of 10 in synchronization index at frequencies < 3 kHz; Dynes and Delgutte, 1992) during electrical stimulation compared to acoustic stimulation. Hence, *a priori*, there is no reason that timing-based processing should be degraded with CIs. However,

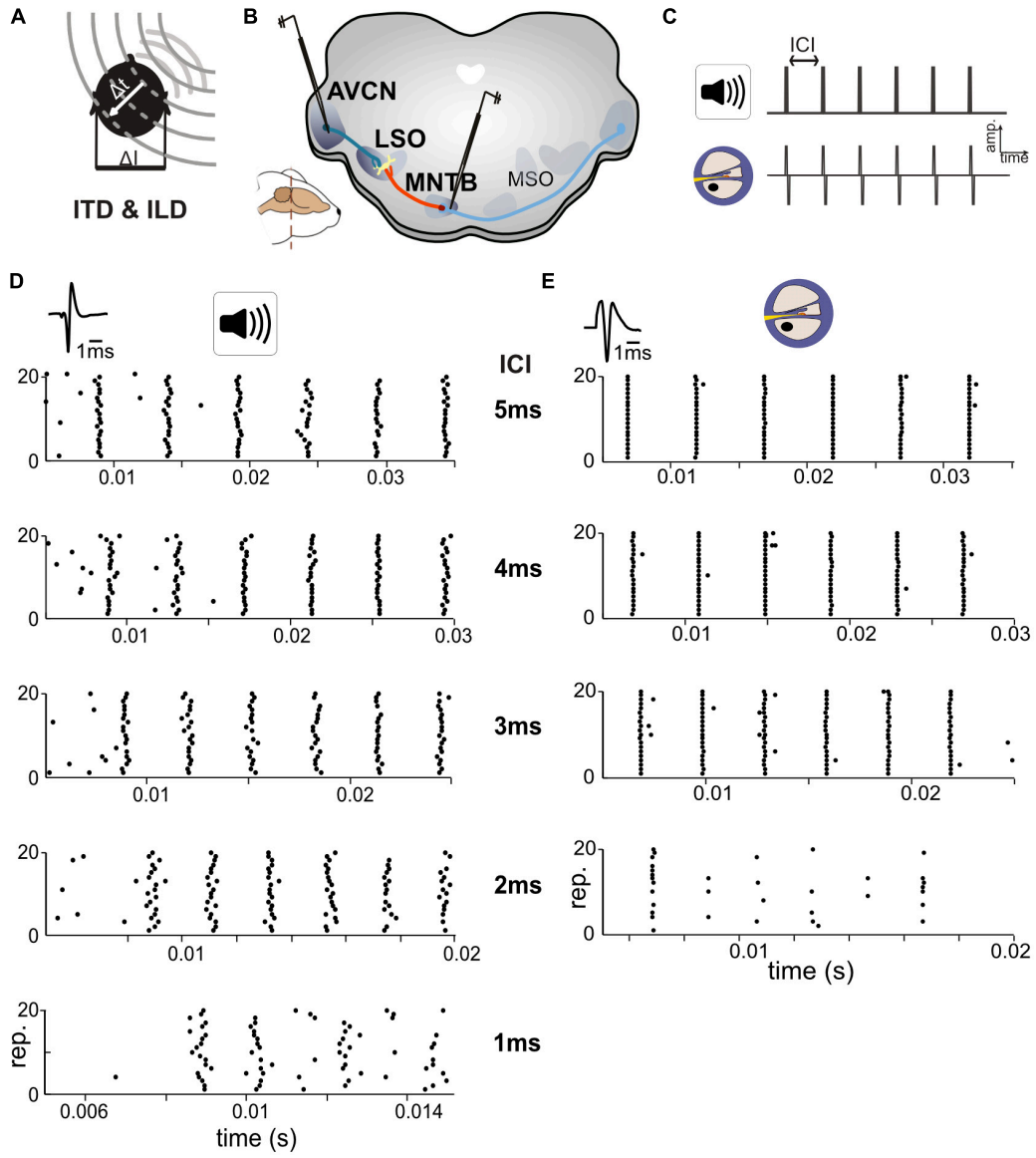



FIGURE 1
Responses in auditory brainstem neurons differ between acoustic and electrical stimulation. **(A)** The egocentric location of a sound source in the horizontal plane generates specific interaural time differences (ITDs) and interaural level differences (ILDs). **(B)** Extracellular single cell recordings with glass pipettes were conducted in either the antero-ventral cochlear nucleus (AVCN), medial nucleus of the trapezoid body (MNTB), or lateral superior olive (LSO). **(C)** Acoustic and electrical [i.e., cochlear implant (CI)-based] stimulation consisted of a train of six clicks with varying inter-click intervals (ICIs) from 5 to 1 ms. **(D)** Exemplary dot-raster displays of single cell recordings from MNTB [action potential (AP) waveforms shown on top] during acoustic in response to different ICIs. **(E)** Same as panel **(D)** but for electrical stimulation. Each ICI was repeated 20 times. Responses to the individual clicks are readily identifiable, particularly at larger ICIs and overall spike probability decreased with smaller ICIs. Note the difference in spike timing variability (jitter) for each click at all ICIs between acoustic and electrical stimulation. Data from ICI = 1 ms for electrical stimulation was not analyzed due to a strong overlap of APs and electrical artifacts.

it is not known to what extent the hyper-precision (phase-locking) found in AN fibers is maintained in the AVCN and MNTB or rather leads to degraded processing, e.g., by failure of transmission. Interestingly, in rats, whose ITD sensitivity is predominately derived from the LSO, lateralization ability to electrical pulse ITDs was shown to be very good (Buck et al., 2021; Rosskothen-Kuhl et al., 2021), suggesting that the LSO could provide behaviorally relevant information during bilateral CI stimulation. Still, it is unclear in what way the information carried by the inputs to LSO, or ITD sensitivity in the LSO itself, is altered during electrical stimulation. More generally speaking, a better understanding of the electrically evoked responses in the brainstem is needed. However, to date an investigation of the LSO pathway during electrical stimulation is missing.

Here, we use a bottom-up approach and ask in what way response features differ between electrical and acoustic stimulation within the initial stages of the LSO circuit. We compare responses to acoustic click and electrical pulses in the AVCN and MNTB by conducting extracellular single neuron recordings in the gerbil, an animal known for good ITD sensitivity (Grothe and Pecka, 2014) both behaviorally and on the level of the LSO. We find that neurons in both of these monaural nuclei exhibit severe hyper-precision and moderate hyper-entrainment during electrical stimulation. This finding establishes conclusively that neuronal ITD detection must cope with highly altered input statistics during CI stimulation compared to acoustic hearing. To determine how ITD computation in LSO might be affected by these changes, we used a previously published model of the LSO circuit that allows using both acoustic and electrical front-ends to compare stimulation types (Ashida et al., 2016, Hu et al., 2022). We carefully adapted the model parameters to match the response characteristics to our physiological recordings in AVCN and MNTB during either stimulation type. The model LSO closely reproduced the ITD sensitivity and its ILD dependency as observed in gerbil LSO. Moreover, the model predicts that ITD sensitivity of LSO to electrical pulses is maintained even for the hyper-precise inputs but exhibits a narrowed dynamic range compared to acoustic stimulation. Calculations of the separability of nearby ITDs suggests heightened resolution at small ITDs during electric compared to acoustic stimulation, while larger ITDs were rendered inseparable. These findings correspond well both with recent reports of high electrical ITD sensitivity in rats, who only experience comparably small ITDs due to the small inter-ear distance (Buck et al., 2021; Rosskothen-Kuhl et al., 2021) as well as with the crude ITD sensitivity found in human CI listeners that often is reduced to lateralization only (Laback et al., 2015). Interestingly, reducing the hyper-precision (phase-locking) of AN fibers during electrical stimulation to acoustically physiological levels recovered a wider dynamic range of ITD coding in the model. Together, our findings suggest that a better understanding of processing of electrical stimuli along the LSO circuit could be crucial for improving ITD sensitivity in bilateral CI users.

## Results

To compare response properties of the excitatory and inhibitory inputs to LSO, we obtained single cell recordings in bushy cells (BCs) in AVCN and principal cells in MNTB (see section "Materials and methods"), in two groups of gerbils (**Figures 1A–C**). One group of animals was presented with acoustic click-train stimuli (six clicks) at various inter-click intervals (ICIs) between 5 and 1 ms (in 1 ms steps), corresponding to click-frequencies of 200, 250, 333, 500, and 1,000 pps. The second group of animals was implanted unilaterally with an intra-cochlear electrode (see section "Materials and methods") to allow delivery of electrical stimulation of the AN. As for the acoustic group, the electrical stimuli consisted of a six-pulse long train with ICIs varying between 5 and 1 ms in 1 ms steps. However, strong interferences by the stimulation artifact during recordings prevented analysis of the 1 ms ICI data and only data for ICIs $>$ 1 ms can be presented.

We obtained recordings from 22 AVCN and 18 MNTB neurons (median characteristic frequencies; AVCN: 16.4 kHz, MNTB: 15.0 kHz) during acoustic stimulation, and 11 AVCN and 9 MNTB neurons with electrical stimulation (**Figures 1D, E**). Stimulus intensity was adjusted for each neuron individually to 30/20 dB above threshold for acoustic and electrical stimuli, respectively (see section "Materials and methods"). We started by analyzing response reliability and timing accuracy of the excitatory and inhibitory inputs to LSO, two crucial factors for the generation of ITD sensitivity. To quantify and compare these parameters for the recorded BC and MNTB neurons, AP responses were analyzed in two ways: first, we determined the "spike probability," i.e., the average percentage of clicks in the train that elicited APs. For instance, an average response of six APs corresponds to 100% spike probability, as there are six clicks in the train. Second, we calculated the "response jitter," i.e., the standard deviation of the AP latency relative to the eliciting click (see section "Materials and methods").

In agreement with previous reports (Joris and Yin, 1998; Smith et al., 1998), we found that spike probabilities in response to acoustic click-trains were similar in BCs and MNTB, with slightly improved spike probabilities found in MNTB at smaller ICIs (**Figure 2A** and **Table 1**; $p = 0.098$ Kruskal–Wallis $H$-test). For BCs, the response probability of all recorded neurons dropped considerably at ICI of 2 ms for electrical pulse-trains (**Figure 2B**). Interestingly, we noted that about half of the MNTB neurons sustained high response probabilities at this IC (**Figure 2B**), suggesting that at least some MNTB neurons maintain coding capacity at 500 pps. The spike timing jitter was also similar between the two nuclei (**Figure 2A** and **Table 1**; $p = 0.949$ Kruskal–Wallis $H$-test). Crucially, electrical pulse-train stimulation resulted in obvious changes in response properties. The spike probability relative to acoustic stimulation tended to be elevated for both BCs and MNTB (**Figure 2B** and **Table 1**).

The largest differences to acoustic stimulation was found in terms of temporal precision, as the jitter during electrical stimulation (**Table 1**) was approx. 10-fold smaller (**Figure 2B** and **Table 1**, all ICIs for MNTB and BCs resulted in $p < 0.001$, Mann–Whitney-$U$-test). These data demonstrate that electrically induced hyper-precision that has been found in AN (Hartmann and Klinke, 1990; Dynes and Delgutte, 1992) is conserved (if not increased) at downstream brainstem nuclei and is likely to influence binaural spatial processing in MSO and LSO.

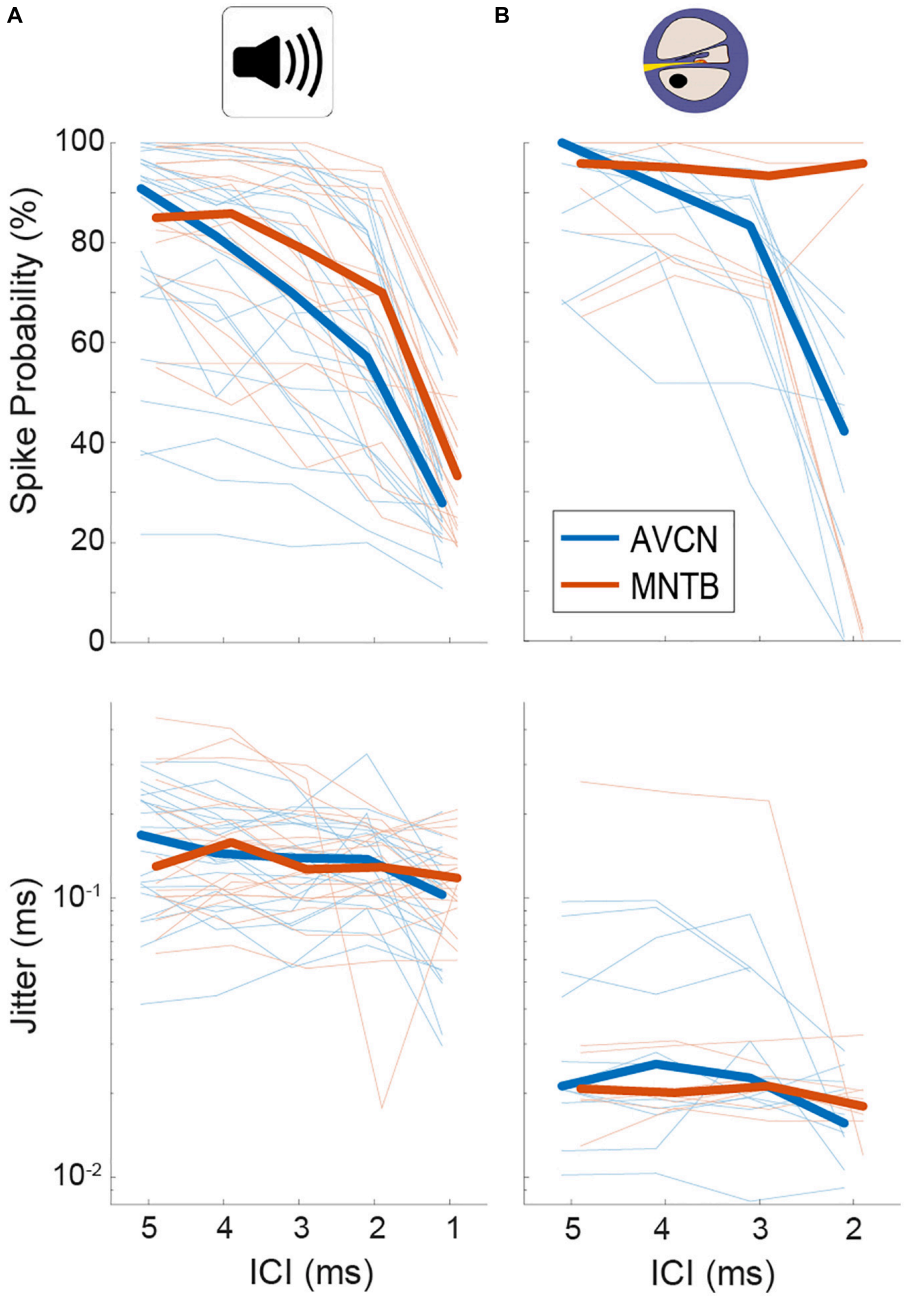



FIGURE 2
Quantification of spiking probability (upper row) and precision (lower row) in gerbil antero-ventral cochlear nucleus (AVCN) and medial nucleus of the trapezoid body (MNTB) in response to acoustic **(A)** and electrical **(B)** click train stimulation. Plotted are average values for each recorded neuron (thin lines) and the sample medians (bold lines). See also **Table 1**.

Yet how exactly could these differences in response properties during electrical stimulation affect spatial processing? For the detection of ITDs, the LSO (as well as the MSO) integrates the inputs from BCs (ipsi-ear) and MNTB (contra-ear) for each click individually. Specifically, using the same acoustic click-train stimuli as for AVCN and MNTB in this study, we had previously determined that LSO neurons exhibit high sensitivity to ITDs of each click in the train, and that the binaural integration of relative strength and timing of inhibition compared to excitation underlies response modulation with changes in ITD in the LSO (Beiderbeck et al., 2018). Hence, we hypothesized that the unusually high precision we found during electrical stimulation for both the excitatory (BCs) and inhibitory (MNTB) input will

TABLE 1 Summary of median values for spike probabilities and jitter in BCs/AVCN and MNTB shown in Figure 2.

| | | ICI 5 ms | ICI 4 ms | ICI 3 ms | ICI 2 ms | ICI 1 ms |
|---|---|---|---|---|---|---|
| Acoustic: Figure 2A, medians | MNTB (probability) | 90.83% | 85.42% | 76.67% | 64.95% | 35.88% |
| | BCs (probability) | 84.58% | 81.25% | 70.00% | 57.08% | 27.92% |
| | MNTB (jitter) | 0.14 ms | 0.15 ms | 0.14 ms | 0.14 ms | 0.11 ms |
| | BCs (jitter) | 0.18 ms | 0.14 ms | 0.13 ms | 0.14 ms | 0.10 ms |
| Electrical: Figure 2B, medians | MNTB (probability) | 95.83% | 95.00% | 93.33% | 95.83% | |
| | AVCN (probability) | 100.00% | 91.67% | 83.33% | 42.11% | |
| | MNTB (jitter) | 0.021 ms | 0.019 ms | 0.023 ms | 0.020 ms | |
| | AVCN (jitter) | 0.031 ms | 0.025 ms | 0.020 ms | 0.017 ms | |

ICI 5 ms: MNTB $p = 0.27$, BC $p = 0.058$; ICI 4 ms: MNTB $p = 0.21$, BC $p = 0.089$; ICI 3 ms: MNTB $p = 0.14$, BC $p = 0.61$; ICI 2 ms: MNTB $p = 0.2$, BC $p = 0.017$; Mann–Whitney $U$-test.

impact the temporal integration process in the LSO. More generally, we wondered to what extent the observed changes in neuronal responsiveness might constitute a mechanistic explanation for the altered sound localization ability of bilateral CI listeners.

To this end, we set out to compare ITD sensitivity in the LSO during acoustic and electrical stimulation. We recorded from 15 LSO neurons in response to the same acoustic click-train stimuli used for AVCN and MNTB, presented binaurally at various ITDs (unpublished subset of data reported in Beiderbeck et al., 2018). Typically, response rates were strongly modulated as a function of ITD for a wide range of tested ITDs (**Figure 3A**). Across the 15 LSO neurons for which we recorded rate-ITD functions at ILD = 0 dB, the dynamic ITD range (range of ITDs between maximal and minimal response rate) reliably covered or exceeded the physiological (i.e., naturally occurring) range of ITD of gerbils (approx. 300 μs, Maki and Furukawa, 2005) for all ICIs (**Figure 3B**, median dynamic ITD ranges for ICIs of 5, 4, 3, 2, and 1 ms: 400, 400, 400, 400, and 600 μs). This wide range allows for linear attributions of response rates to ITDs, and thus provides the encoding basis for reliable sound source localization based on "hemispheric rate-difference coding," that is, the encoding of individual ITDs *via* the relative activity levels between the two LSO populations in each brain hemisphere (Klug et al., 2020; for review see Pecka and Encke, 2020). However, ITD sensitivity in the LSO can be influenced by relative changes in the intensity on the two ears (i.e., ILDs, Park et al., 1996; Beiderbeck et al., 2018). To capture this dependency quantitatively in our recordings, we applied various ILDs and determined their effect on the slope steepness of the rate-ITD functions (change in normalized AP rate over the dynamic ITD range, see section "Materials and methods"). On average, the effect of changing ILD had a small to modest effect across all ICIs [**Figure 3C**, gray bars, medians, and interquartile range (norm. spikes/rep/μs/dB*$10^{-5}$): ICI 5 ms: 6.0, 2.3, 7.1; ICI 4 ms: 6.0, 3.0, 7.1; ICI 3 ms: 2.0, 1.5, 8.4; ICI 2 ms: 4.2, 2.0, 9.4, ICI 1 ms: 3.2, 2.0, 6.6, suggesting that ITD sensitivity in LSO can be maintained over a wide range of binaural conditions.

How does the coding of ITDs change in LSO during electrical stimulation? Unfortunately, recording of single LSO neuron with bilateral electrical CI stimulation proved to be exceedingly difficult. Therefore, to approximate physiological data, we utilized a functional count-comparison model of the LSO (Ashida et al., 2016) as a surrogate for LSO during electrical stimulation. This model can replicate typical response properties of LSO neurons and accompanying spatial perceptions to a wide range of stimulus classes with high precision (Klug et al., 2020; Hu et al., 2022). For example, by connecting this model with an existing acoustically or electrically stimulated AN model, Hu et al. (2022) were able to reproduce most characteristics of acoustically stimulated LSO neurons (Joris and Yin, 1998) and electrically stimulated step-type or trough-type IC neurons (e.g., Smith and Delgutte, 2007, 2008; Chung et al., 2016). Moreover, acoustic and electrical response properties can be readily read-out at various stages of the LSO pathway (Hu et al., 2022). This feature allows comparing model performance to our physiologically recorded AVCN and MNTB data for benchmarking. Specifically, to be able to make informative predictions about changes in ITD sensitivity during electrical stimulation, the model should exhibit similar responsiveness of the excitatory and inhibitory inputs to the LSO during acoustic stimulation. Hence, we determined to what extent the model was able to replicate the response behavior of these LSO inputs.

Using previously published general parameter setting (Ashida et al., 2016), we first determined a suitable acoustic stimulus intensity by analyzing the AP responses of the model input stage (AN, note that no explicit CN, and MNTB stage exists in the model, see section "Materials and methods"). At model intensities of 50/60 dB for the excitatory and inhibitory input, the model displayed similar levels of entrainment and jitter as our physiological data (**Figure 4A**, spike probability for excit./inh.: ICI 5 ms: 84.6/96.1%, ICI 4 ms: 78.6/87.9%, ICI 3 ms: 72.6/79.7%, ICI 2 ms: 63.7/67.8%, ICI 1 ms: 51.9/52.7% and jitter for excit./inh.: 0.23/0.24 ms at all ICIs; compare **Figure 2**).

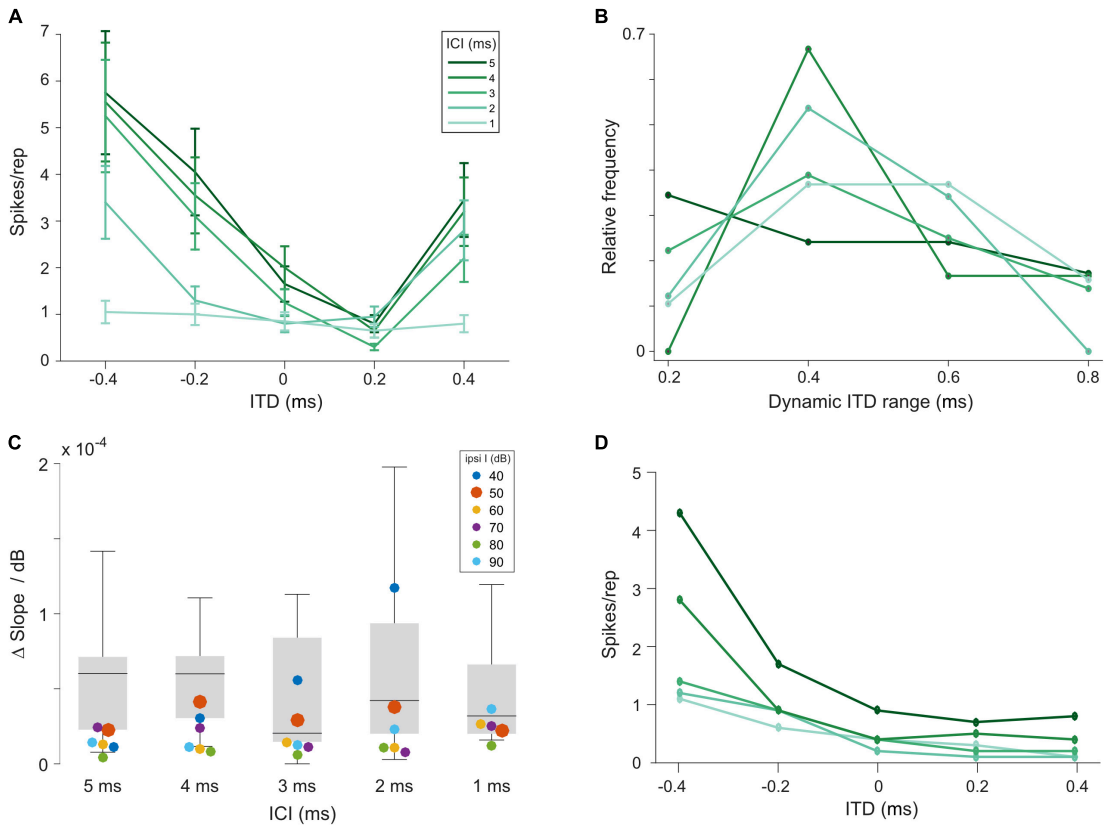



FIGURE 3
Interaural time differences (ITD) sensitivity of lateral superior olive (LSO) neurons to click trains is maintained over a large range of interaural level differences (ILDs). **(A)** Example rate-ITD function from a gerbil LSO neuron. Plotted are mean rates and the standard errors. **(B)** Histogram of dynamic ITD ranges at ILD = 0 dB across all 15 recorded LSO neurons. **(C)** Quantification of magnitude of changes in the slope steepness of rate-ITD functions with changes in ILD for gerbil LSO (boxplots) and model (at various intensities of the excitatory input, colored dots). The $\delta$slope/dB was calculated by first determining the slope of rate-ITD functions (difference in the normalized maximal and minimal spike rates divided by the respective ITD range), and then subtracting these values between the most positive and negative ILD that each LSO neuron was tested for and dividing this difference by the respective difference in ILD. **(D)** Model rate-ITD function during acoustic stimulation (ipsi intensity = 50 dB; ILD = 0 dB).

We further verified that the model also exhibited changes in its responsiveness during electrical stimulation (Figure 4B, spike probability for excit./inh.: ICI 5 ms: 89.4/98.2%, ICI 4 ms: 86.8/97.5%, ICI 3 ms: 80.0/94.0%, ICI 2 ms: 61.6/77.8%) that were comparable to what we had observed experimentally. In particular, the jitter decreased by a similar factor (compare Figure 2) using the electrical front-ends (Figure 4B, excit./inh.: 0.05/0.04 ms at all ICIs).

Based on these identified model parameters we next evaluated the model LSO stage. At 0 dB ILD, the model exhibited ITD tuning with monotonic rate modulation over the range of tested ITDs, resembling the physiological LSO examples qualitatively (Figure 3D). To further test to what extent the model LSO circuit can quantitatively capture the dependency of the slope of ITD tuning on ILD, we applied various ILDs and determined their effect on the slope of the rate-ITD functions (change in AP rate per unit ITD). The previously obtained data set (using the same click-trains) from the gerbil LSO allowed a direct comparison to the model. This comparison demonstrated that the model was adequately capturing the binaural integration process underlying the ITD sensitivity in the LSO. Specifically, the changes in ITD sensitivity with ILD exhibited by the model at 50 dB ipsilateral intensity (Figure 3C, red dots; delta/dB $\times 10^{-5}$: ICI 5 ms: 2.3; ICI 4 ms: 4.1; ICI 3 ms: 2.9; ICI 2 ms: 3.8; ICI 1 ms: 2.2) were within the range of observed changes in the gerbil LSO (Figure 3C, gray bars). Hence, these comparisons demonstrate that the response behavior of the model resembles the physiological recordings from gerbil qualitatively and even quantitatively with high accuracy, both for acoustic and electrical pulse-train stimuli. Thus, the model should serve as a valuable proxy for predicting ITD sensitivity of LSO neurons during bilateral electrical stimulation.

To allow for precise evaluation of changes in ITD coding between acoustic and electric stimulation and its impact on spatial resolution, we determined the informational content

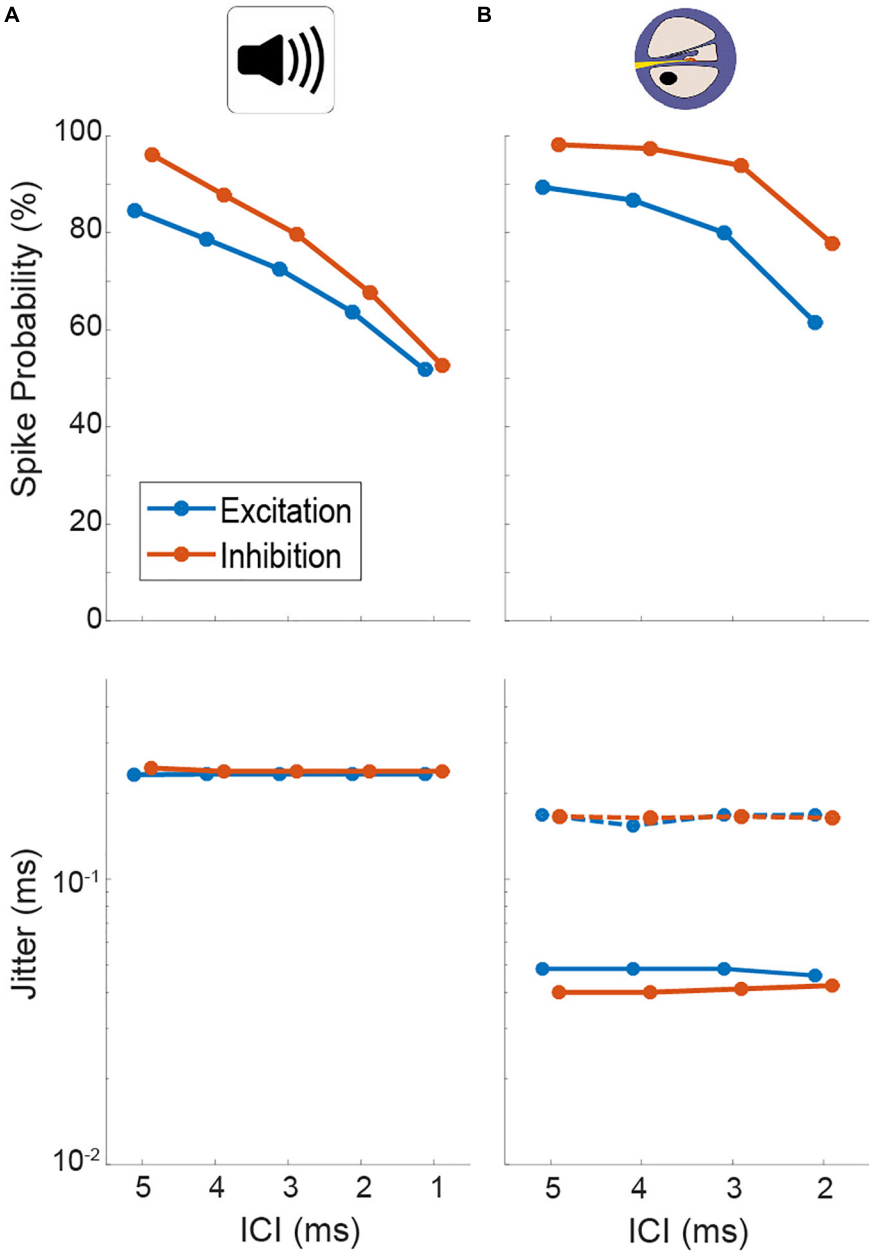



FIGURE 4
Quantification of spiking probability (upper row) and precision (lower row) of model excitatory and inhibitory inputs to lateral superior olive (LSO) [corresponding to antero-ventral cochlear nucleus (AVCN) and medial nucleus of the trapezoid body (MNTB), respectively] in response to acoustic **(A)** and electrical **(B)** click train stimulation. Plotted are medians and 95% confidence intervals. Dotted lines represent model condition with auditory nerve (AN) jitter levels during electrical stimulation that resemble physiological acoustic jitter levels.

of the models' response toward the ability to distinguish adjacent ITDs with 20 μs increments throughout the entire range of ITDs that is generated by the human head (i.e., inter-ear-distance, approx. ±600 μs, Moore, 2013). To this end we calculated the standard separation "D" (Sakitt, 1973), which quantifies the separability of adjacent ITDs based on the ratio of differences in mean rate and response variability. Since the LSO model is highly deterministic, we used Poisson noise as a conservative assumption (see section "Discussion"). In accordance with the hemispheric rate-difference model of spatial coding (Grothe et al., 2010; Pecka and Encke, 2020), we determined D based on the rate difference between the LSOs on either hemisphere (see insets in **Figure 5**; responses in one hemisphere were assumed to be mirrored by the LSO in the other hemisphere and subtracted from each other at each ITD). The distribution of D for the rate-ITD

functions in response to acoustic stimulation (**Figure 5A**) spanned almost the entire physiological ITD range of humans (±600 μs) and were either centered on midline (**Figure 5B**; ICIs of 2, 3, and 4 ms) or peaked at slightly lateralized ITDs (**Figure 5B**; ICIs of 1 and 5 ms). Next, we tested the model LSO during bilateral electrical stimulation and repeated the D measurements. Notably, ITD sensitivity was maintained for electrical ITDs, as the model LSO displayed steep rate modulation as function of ITD for all ICIs (**Figure 5C**). However, compared to acoustic ITDs, this modulation (the slope of the function) extended only over a narrower range of ITDs between approx. ±100 and 0 μs ITD, while response rates were effectively identical for more lateralized ITDs to either hemisphere. This "hemispheric binarization" resulted in a highly narrowed distribution of D (**Figure 5D**), where separability was very high near midline (±100 μs ITD, even higher than for acoustic stimulation) but absent at more lateralized positions on either side. Such an effect can be interpreted as a lateralization effect for human listeners where left can be reliably distinguished from right even for small ITDs around midline, but resolution is almost absent at larger ITDs within each hemisphere. This is contrasted by the more graded distribution of D that we observed during acoustic stimulation, which would allow to also distinguish ITDs within a hemisphere (**Figure 5B**).

What factors underlie this alteration of the dynamic ITD range? The most drastic change that we revealed for the inputs to LSO during electrical stimulation was the increase in response precision (decrease in jitter). To evaluate the effect of this change in response precision, we re-introduced jitter to the inputs of the LSO model to approximate those of the acoustic stimulation (excit./inh.: 0.17/0.16 ms for all ICIs, dotted lines in **Figure 4B**). Remarkably, this modification resulted in a re-installment of a wider dynamic ITD range (**Figure 5E**) and corresponding widening of the range of separable ITDs (**Figure 5F**). Thus, the model suggests that the increase in spiking precision in response due to artificial electrical stimulation directly alters the ITD coding capacity of LSO during CI-based hearing. Specifically, jitter seems to affect the range of ITDs that alter LSO response rates.

## Discussion

This study is the first to directly assess response properties of brainstem neurons to CI-based stimulation. Spike timing (i.e., jitter) of both BCs and MNTB neurons exhibit severe hyper-precision and moderate hyper-entrainment during electrical stimulation. Thus, the response alterations that have been reported for the AN due to CI stimulation are passed on to subsequent synaptic stages. Specifically, increased response rates and entrainment during electrical stimulation compared to acoustical stimulation had been reported previously at stimulation rates < 800 Hz (Hartmann et al., 1984; van den Honert and Stypulkowski, 1987; Javel and Shepherd, 2000; Litvak et al., 2001). Our recordings in the gerbil AVCN indicate that there seems to be no significant failure of transmission nor compensation caused by the integration mechanisms at the endbulb of Held synapses in BCs for rates below 500 pps. Indeed, precision and spike probability might be further elevated by the coincidence mechanism of the endbulbs. Similarly, the calyx of Held synapse at the MNTB is known to be specialized for maintaining exquisite temporal fidelity and seems to pass on the hyper-precise spiking that is generated by electrical stimulation. Indeed, about half of the MNTB neurons we encountered exhibited high spike probabilities even at 500 pps (ICI = 2 ms), suggesting slightly enhanced fidelity of MNTB compared to BC neurons, albeit not evident in all cells. These observations are significant as they establish conclusively that neuronal ITD detection at the next synaptic stage—MSO and LSO—must cope with highly altered input statistics during CI stimulation compared to acoustic hearing. Thus, any diminished ITD sensitivity on the perceptual level is not caused by a lack of temporal information provided to the binaural integration stage.

This finding naturally raises the questions: which of the two nuclei is the likely site of electrical ITD integration and how do hyper-precise inputs affect the integration mechanism? Perceptual and anatomical data indicate strong bias toward LSO processing (similar cut-off frequencies of ITD sensitivity, basal insertion of electrodes, etc.). Moreover, we had previously established that ITD sensitivity of high-frequency LSO neurons to acoustic click trains is exquisite (Beiderbeck et al., 2018). Therefore, we utilized an established LSO model of acoustic and electrical ITD sensitivity to investigate possible consequences of electrical stimulation. Our data indicate that LSO should not only be able to generate ITD sensitivity to electrical pulses, but be sensitive to the heightened temporal precision, which in turn might be contributing to perceptual alterations of CI users compared to normal listeners. While LSO processing and coding of ITD can be assumed to be similar in all mammals and thus can be extrapolated from gerbils to humans (Grothe and Pecka, 2014), these alterations might have differential effects in large-skulled (e.g., humans) and small-skulled (e.g., gerbils or rats) listeners, because of the difference in the respective available ITD range (approx. ±600 and ±150 μs, respectively). That is, the model predicts that the reported limited perceptual electrical ITD sensitivity in human listeners might not be caused by a general lack of ITD tuning in the brainstem, but rather that the hyper-precise inputs to LSO result in unusually steep rate-ITD functions with altered dynamic range. According to these predictions, the resolution of adjacent source locations might even by improved for small ITDs around midline (up to approx. ±100 μs), while it quickly decreases to inseparability at larger ITDs. Current data on ITD thresholds in CI listeners shows great variability ranging from a few tens μs to >1,000 μs (Laback et al., 2015), but there are no

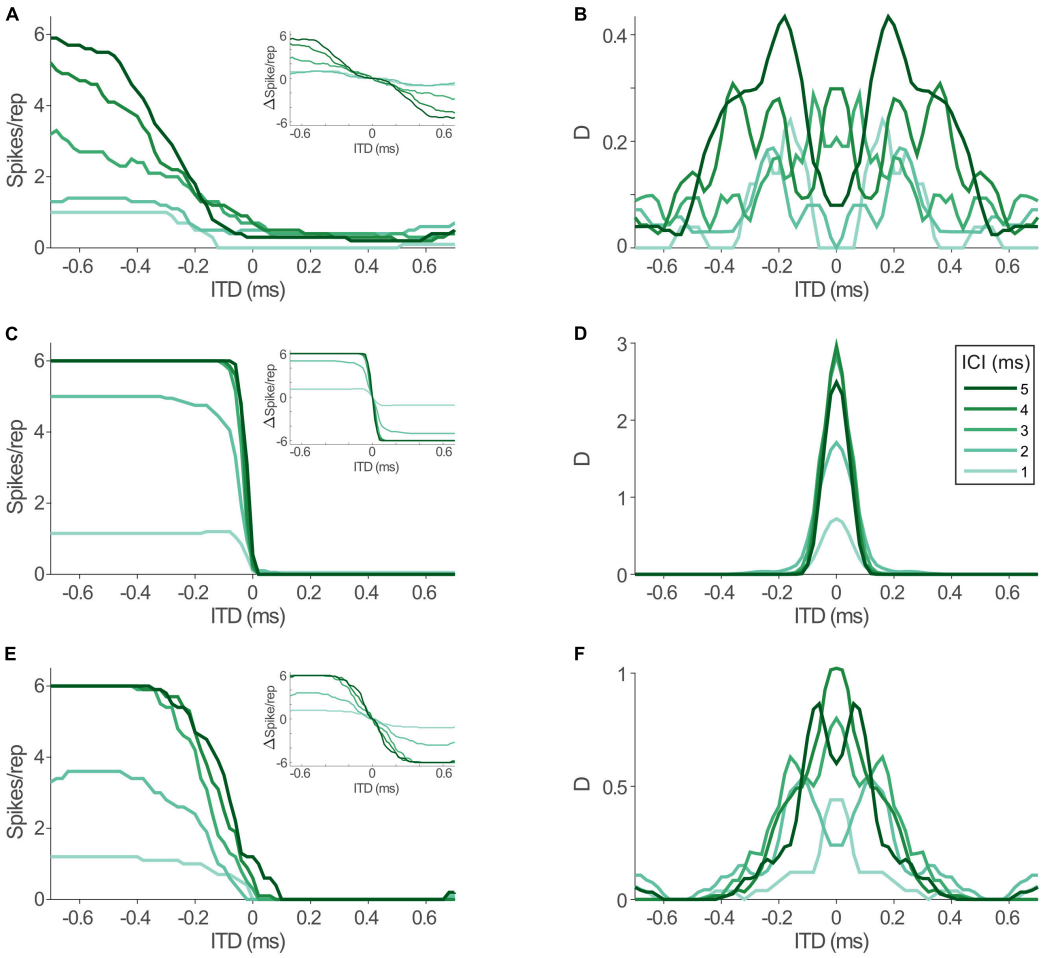



FIGURE 5
Interaural time differences (ITD) sensitivity in model lateral superior olive (LSO) is maintained during electrical stimulation, but altered by jitter level. **(A)** Model rate-ITD functions during acoustic stimulation. Inset shows hemispheric rate differences for all inter-click intervals (ICIs) (assuming mirrored rate-ITD functions in the LSO on the other brain hemisphere). **(B)** Standard separability D of the hemispheric rate differences for the data shown in panel **(A)**. **(C,D)** Same as in panels **(A,B)**, but for electric stimulation. **(E,F)** Same as in panels **(A,B)**, but for electric stimulation with jitter levels of the model inputs increased to resemble acoustic conditions.

indications of "super-resolution," i.e., improvements compared to normal hearing threshold. These data are typically collected using a symmetrical forced-choice lateralization paradigm, yet any "super-resolution" area as suggested by our model might not necessarily be located symmetrically at midline, since many factors could shift this region (e.g., electrode mismatch in amplitude or cochlear location, asymmetric de-innervation, subjective training and adaptation effects, etc.). Hence it is currently unclear and would be intriguing to investigate if a small region of improved separability can be found with bilateral CI listeners and/or contributes to the reported variability of thresholds.

However, it must be stressed that the effect of increased separability as measured by D decreases considerably with increasing levels of variability in the LSO population code. Our LSO model is highly deterministic and since we were ignorant about the actual *in vivo* LSO population rate variability, we were required to assume a level. We chose to follow our previous LSO modeling study (Hu et al., 2022) and used Poisson noise as variance measure during the calculation of D. However, this choice might underestimate the true variability *in vivo*, particularly during electric stimulation (Javel and Viemeister, 2000). Importantly, the loss of spatial resolution at larger ITDs is mostly independent of this estimate.

Another notable conclusion of the model's prediction about improved resolution at small ITDs is that it potentially provides a mechanistic explanation for the recent reports of highly precise ITD sensitivity both on the level of midbrain coding (Buck et al., 2021) and perceptual resolution (Rosskothen-Kuhl et al., 2021) in bilaterally implanted rats. Since these animals have smaller inter-ear-distances compared to human subjects, the predicted region of increased LSO ITD coding capacity would

cover a large fraction of the available ITDs in rats, and hence provide an advantageous effect on spatial resolution for almost the entire frontal field. However, as explained above, the extent of the effect is dependent on the actual magnitude of variability in LSO. Our third key finding emphasized the role of jitter on the shape of rate-ITD functions: by increasing the jitter of the inputs to the electrical LSO model to the physiological (i.e., acoustically induced) level, a wider dynamic range of rate-ITD functions (and accompanying separability) was restored. The model consequently suggests that physiological jitter levels are key in establishing the temporal width of the integration window between excitatory and inhibitory inputs. This window width in turn defines how gradually the LSO spike rate is modulated as a function of ITD. In line with this idea, Myoga et al. (2014) demonstrated that jitter affected the interaction of MNTB-derived inhibition and BC-derived excitation and thereby shaped the overall response to specific ITDs. Even though these results were obtained in the MSO, they emphasize the importance of jitter level on the precision of the ITD integration of MTNB and BC inputs and hence the ITD computation mechanism.

Our findings directly extend on previous work on the role of response precision on ITD perception and might provide a complementary mechanistic explanation. Laback and Majdak (2008) proposed that jittering the inter-pulse intervals would lead to improvement of ITD sensitivity by restarting the adaptation process every time the pulse is randomized. They hypothesized that binaural adaption during ongoing high stimulation leads to degraded ITD detection and demonstrated that binaural jitter enhanced ITD sensitivity at higher stimulation rates ($\geq$800 pps). Correspondingly, midbrain recordings showed increased ITD sensitivity by the introduction of short inter-pulse intervals (Hancock et al., 2012; Buechel et al., 2018). Introducing jitter increased firing, presumably by counteracting adaptation to the prolonged stimulation with high-rate click trains. In contrast, our click trains were much shorter compared to earlier studies and resulted in slight increased responsiveness compared to acoustic stimuli and thus might not be directly comparable. Thus, our data indicates that diminished spatial sensitivity might, at least in part, be caused by hyper-precise input timing that results in hyper-acute ITD sensitivity in the LSO. In the future, it would be interesting to repeat our experiments using longer click-trains to test the influence of rate adaptation of the monaural inputs. Moreover, it is currently unclear to what extent the effects of jittering depend on synchrony between the two ears, i.e., our model could be used to test the differential effects of synchronized and desynchronized temporally jittered inputs on electrical ITD processing in LSO.

We used animals with fully developed hearing prior to the experiment. Hence, any experience-dependent mechanisms for the fine-tuning of inputs strength and timing was left in place. Likewise, no degenerative effects had occurred prior to implantation. However, patients that receive CIs typically underwent prolonged periods of deafness and its degradation of the auditory system could lead to a multitude of complications related to this period of inactivity of the system. For example, Hancock et al. (2010) and Tillein et al. (2010) demonstrated reduced ITD sensitivity recorded from the IC and auditory cortex, respectively, in congenital deaf animals. Likewise, it has been shown that the ITD threshold of neonatally deafened animals were similar to the thresholds of normal hearing rats (Buck et al., 2021, Rosskothen-Kuhl et al., 2021), suggesting that electric ITD information can be exploited by the brain. These findings could be explained by our hypothesis that the LSO and not the MSO is mostly responsible for CI-based ITD detection. Excitatory and inhibitory tuning curves of the LSO and the developmental changes of inhibitory projections to the LSO are largely completed before hearing onset (Sanes and Rubel, 1988; Kim and Kandler, 2003). Hence, the LSO circuits develop to functional maturity even in the absence of auditory experience. Any CI-based activity introduced in the system at later stages of life could thus be readily utilized by the LSO circuits to generate spatial sensitivity. In contrast, ITD sensitivity of MSO requires developmental maturation and is dependent on hearing experience in a short critical period after hearing onset (Kapfer et al., 2002; Seidl and Grothe, 2005). Likewise, the data of Buck et al. (2021) and Rosskothen-Kuhl et al. (2021) were obtained in rats, an animal model with high frequency hearing and well-developed LSO (while ITD detection in the MSO can be neglected). In partial support of these rodent data, it has been found that hearing experience in bilateral CI subjects with post-lingual onset of deafness tended to exhibit sizeable ITD sensitivity, while it was not present in subjects with pre-lingual onset of deafness. In contrast, ILD cue sensitivity was similarly present in both groups (Litovsky et al., 2010), supporting our hypothesis of the LSO as the main binaural detector during CI based stimulation. However, electrophysiological recordings of the cellular integration mechanism of the LSO and MSO during CI based stimulation have not been obtained yet. Furthermore, the experimentally obtained rate-ITD functions from the midbrain of implanted rats (Buck et al., 2021) do not readily hint at originating from excitatory-inhibitory interaction and cannot be reproduced by our model.

In summary, our findings suggest that LSO processing is likely the main site of electrical CI-mediated ITD processing and that a key problem underlying the diminished ITD sensitivity in CI users is the temporal hyper-precision of inputs to the binaural comparator stage.

## Materials and methods

All experiments were conducted according to the German animal welfare law (55.2-1-54-2532-53-2015). A total of 37 Mongolian Gerbils (*Meriones unguiculatus*) of either sex and at

least 3 month of age were used in this study. After 3 month of age the hearing is fully developed (McFadden et al., 1996), and the cardiovascular system of the animal was capable of enduring long anesthesia (Pecka et al., 2007, 2008, 2010). In animals older than 2 years the hearing threshold and capacity declines, and therefore, animals over 2 years of age were not used in this study (Mills et al., 1990). The animals were kept in Tecniplast Typ 4 cages (610 mm × 435 mm × 215 mm) filled with wooden chipping and wooden wool and a house served for withdrawal. Up to five animals were held in one cage. Temperature and humidity were controlled (temperature 23 ± 2°C, 50 ± 10% humidity) and the animals had a 12-h dark/light circle.

## Anesthesia

Thirty minutes ahead of surgery the animal was injected subcutaneously with an analgesic non-steroidal anti-inflammatory drug (Metacam® 1.5 mg/ml suspension, Boehringer Ingelheim Vetmedica GmbH, Ingelheim, Germany, 0.2 mg/kg). The animal was intraperitoneally injected with 0.5–0.6 ml/100 g body weight Ringer solution mixed with ketamine and xylazine (Ketamine 10%, 100 mg/ml, MEDISTAR GmbH, Ascheberg, 50 mg/kg) und xylazine (Rompun® 2%, 20 mg/ml, Bayer AG, Leverkusen, 2 mg/kg). To maintain anesthesia the animal was injected continuously during the whole experiment with a micropump (Univentor 801 Syringe Pump, Univentor, France) with a flow rate of 1.7 μl/70 g per minute. Besides general anesthesia, local anesthetics were needed for the placing of the head pin as well as for the cochlea implantation. After cutting the skin local anesthetics (Xylocaine® Pumpspray Dental, 50 ml, AstraZeneca GmbH, Wedel, Germany) were used to sedate the muscles and the rest of the surrounding tissue. Surgical anesthesia was reached if the animal did have a negative lid reflex, slight rotation of the *Bulbus*, positive corneal reflex and negative leg withdrawal reflex. Anesthesia was monitored by checking the temperature (*via* a rectal probe), heart rate (EKG), breathing rate, pulse, and oxygen (Pulsoxymeter LifeSense® VET Portable Capnography and Pulse Oximetry Monitor, Nonin Medical, Inc., Plymouth, MA, USA). The body temperature was constantly maintained at 37°C.

## Surgical preparation for cochlea implantation and acute deafening

For the cochlea implantation the post-auricular area was opened with a small incision above the bulla. The skin and the temporalis muscle were removed until the bulla is visible and a bullostomy was performed by drilling a small window inside the bulla. The stapes, stapedial artery, and cochlear fenestra were identified and a small dorsal part of the round window niche was removed to facilitate the insertion of the CI. Afterward the round window membrane was withdrawn in preparation of the deafening of the animal.

For the deafening with neomycin, first the perilymph had to be extracted with a GELoader® (20 μl, Eppendorf SE, Hamburg, Germany). The GELoader® was inserted right at the beginning of the scala tympani. To prevent any destruction of cochlea structures the withdrawal of the perilymph was executed very gently and slowly. Next a careful and slow flushing of the scala tympani with neomycin sulphate (60 mg/ml in NaCl) was conducted. In total the scala tympani was flushed 5 min. This was repeated every 10 min for 90 min. After 90 min the neomycin was extracted. To avoid neurotoxic effects on the spiral ganglia cells the scala tympani was afterward flushed with Ringer solution.

## Auditory brainstem recordings

To test the effectiveness of the deafening procedure on the hearing threshold of the animal an auditory brainstem recordings (ABR) recording was carried out. The animal was placed onto a heating pad powered by an ATC 1000 DC Temperature Controller (World Precision Instruments, Sarasota, FL, USA) in a double-walled sound-attenuated chamber (Industrial Acoustics, GmbH, Niederkrüchten, Germany) lined with acoustic foam. It was set at 37°C to maintain stable body temperature. The reference electrode was inserted subdermally at the vertex, the active electrode was inserted over the bulla and the ground electrode was inserted above the hindlimb. A microphone type 4938 and a preamplifier type 2670 (Bruel and Kjaer, Nærum, Denmark) were used to calibrate the loudspeaker (MF1 Tucker Davis Technologies, Alachua, FL, USA). A short plastic tube was used to extend the loudspeaker. This plastic tube was inserted into the ear. A RZ6 Multi I/O Processor (TDT) was used to generate broadband clicks (0.1 ms duration) and tones of 28, 36, and 44 kHz (5 ms duration, 1 ms rise/fall time) which were produced with Spike software (Brandon Warren, University of Washington, Seattle, WA, USA; pre-amp gain of 20) and presented at a repetition rate of 50 Hz. A RA16 PA 16 Channel Medusa preamplifier (TDT) and RZ6 Multi I/O Processor were used to record ABR waveforms. The recordings were averaged over 1,000 repetitions for each frequency and intensity. If the ABR-based hearing threshold was above 70 dB the deafening procedure was regarded to be successful, otherwise the procedure was repeated.

## Cochlea implantation

After the deafening the cochlea implant (MED-EL animal implants for Guinea pigs and Mongolian gerbils, MED-EL, Innsbruck, Austria) was implanted into the scala tympani of

the cochlea. We used bipolar stimulation, where the most apical intra-cochlear electrode was active while the adjacent basal intra-cochlear electrode was used as reference. To ensure the bedding of the cochlea implant throughout the whole recording period a peripheral venous catheter was inserted from the neck muscle to the bulla. The electrode array of the implant was placed within the catheter and lead toward the bulla. After the removal of the catheter, the surrounding muscles kept the implant fixed at this position. The next step was the insertion of the CI. The implant was inserted through the round window and placed within the scala tympani. To ensure the correct insertion depth, the implant was inserted until the black depth marker of the wire. Finally, the implant was fixed with a drop of Histoacryl® and glued to the bulla. This ensured the placing of the cochlea implant within the scala tympani. In all experiments that we had performed the cochlea implant stayed in place.

## Craniotomy and *in vivo* electrophysiology

Recordings for AVCN, MNTB, and LSO were carried out on the same setup using the same hardware. Further details on LSO recordings using acoustic stimuli can be found in Beiderbeck et al. (2018). The animal was placed at a thermostatically controlled heating pad (Fine Science Tools GmbH, Heidelberg, Germany) in a sound-attenuated chamber on a custom-made stereotactic setup. The temperature was monitored using a rectal probe and the head was fixed by a metal rod. The reference electrode was placed in a small craniotomy between bregma and lambda. A second craniotomy and durotomy was drilled behind the sinus transversus lateral to the midline. The lateral position depended on the targeted auditory nuclei. The surface of the brain was covered with physiological NaCl solution (0.9%). To find the correct nuclei the head of the animal was stereotactically aligned.

A glass electrode was lowered into the brain with an angle of 20° by using a motorized micromanipulator (Inchworm controller 8200, EXFO Burleigh Products Group, ON, Canada). APs were recorded using a glass electrode filled with 5 units/μl horseradish peroxidase (HRP) (Sigma-Aldrich Corp., St. Louis, MO, USA) The HRP was diluted in 10% NaCl solution. This resulted in a tip resistance of 8–12 MOhm. The recorded extracellular neuronal signals were pre-amplified (Electro 705, World Precision Instruments, Sarasota, FL, USA), further amplified (TOE 7607, Toellner Electronic, Herdecke, Germany), and filtered (Hum Bug Noise Eliminator, Quest Scientific Instruments Inc., New Delhi, India). A real-time processor (RP2, Tucker Davis Technologies Inc., Alachua, FL, USA) transferred the signal to a computer. Stimulus presentation was controlled in Brain Ware (Jan Schnupp, Tucker Davis Technologies Inc., Alachua, FL, USA) or AudioSpike (Hörtech gGmbH, Oldenburg, Germany) using a sound card interface (Fireface UFX, RME-Audio). Brain Ware and AudioSpike were also used to monitor the recording online and for offline spike sorting.

## Stimulus generation

Two different groups of animals were used: implanted (to record responses in AVCN and MNTB to electrical pulses) and control (to record responses in AVCN and MNTB to acoustic clicks). For the control group, only acoustic stimuli were applied. For the implanted group electrically generated stimuli were applied.

Acoustic stimuli ranging between 0.1 and 90 kHz were generated digitally and altered to an analog signal (RX6, Tucker Davis Technologies Inc., Alachua, FL, USA) at 200 kHz sampling rate, attenuated (PA5; Tucker Davis Technologies Inc., Alachua, FL, USA), and transferred to the headphones (Etymotic ER-4 microPro, Houston, TX, USA or custom-made electrostatic headphones). In a subset of the experiments acoustic stimuli ranging between 0.1 and 90 kHz were generated digitally and sent to an Audio Interface (RME Fireface UFX II, Audio AG, Haimhausen, Germany) at 192 kHz. The audio interface transferred the acoustic stimuli to the headphones (Etymotic ER-4 microPro, Houston, TX, USA). For both auditory nuclei (CN and MNTB) white noise bursts (duration 200 ms; rise/fall times of 5 ms) were presented monaurally. The ipsilateral ear was stimulated for the CN and the contralateral ear was stimulated for MNTB. When a neuron was isolated, its characteristic frequency (CF) and threshold was determined using pure tones at various frequency and intensity combinations. Subsequently, peristimulus time histograms (PSTH) were recorded at CF at 20 dB above threshold. Moreover, the neurons broadband threshold was determined audio-visually using white noise stimuli. Afterward a train of six clicks with a single-click-duration of 50 μs was presented at five different ICIs (ICIs; 5, 4, 3, 2, and 1 ms) in a pseudo-randomized order.

Electrical stimuli were generated digitally, altered to an analog signal (RX6, Tucker Davis Technologies Inc., Alachua, FL, USA) or transferred to an Audio Interface (RME Fireface UFX II, Audio AG, Haimhausen, Germany) and forwarded onto a voltage-current converter (ICS5, Thomas Wulf Elektronik, Frankfurt, Germany) and delivered to the animal *via* the cochlea implant. The voltage used for stimulation varied between 0.2 and 1.2 V. A click train with a single-click-duration of 110 μs (Anodic phase 50 μs, Cathodic phase 50 μs, and Interphase 10 μs) was used at a ICIs of 5 ms as search stimulus. After encountering a neuron, its electrical threshold was determined audio-visually. Next a train of six clicks with a single-click-duration of 110 μs (Anodic phase 50 μs, Cathodic phase 50 μs, and Interphase 10 μs) was presented in a pseudo-randomized order at five different ICIs (5, 4, 3, 2, and 1 ms) 2 dB above threshold.

## Models

The coincidence counting model of LSO was fully described and suggested as a fundamental operation in both ILD-coding and phase-coding of AM sounds in Ashida et al. (2016). Briefly, the model compares the weighted numbers of excitatory and inhibitory inputs within a pre-defined time window and generates an AP when the total number reaches the threshold. More specifically, in the current simulation, the spikes were counted within a coincidence window size of 0.8 ms for the ipsilateral excitatory inputs and an inhibition window size of 1.6 ms for the contralateral inhibitory inputs. If the sum of excitatory spikes (each counting +1), and inhibitory spikes (each counting −2) reaches the response threshold of 8, an AP was generated. Within each refractory period of 1.6 ms, only the first one led to a spike and the others were discarded. The same peripheral models as Hu et al. (2022) were used. Briefly, the periphery model of Bruce et al. (2018) was applied in the same fashion as in Klug et al. (2020) for simulations of acoustic hearing. The AN model parameters were kept unchanged from Zilany et al. (2009, 2014) and Bruce et al. (2018). To simulate electrical hearing, the acoustic auditory periphery model was substituted by the AN model of Hamacher (2004). It includes cell membrane, membrane noise, refractory period, and latency and jitter. Most parameters were same as in Hamacher (2004) and Fredelake and Hohmann (2012), except that the mean and standard deviation of the latency and jitter were adjustable to fit the physiologically recorded AVCN and MNTB data. At the binaural interaction stage, the coincidence-counting LSO model of Ashida et al. (2016) was used. It receives excitatory synaptic inputs from ipsilateral AN fibers and inhibitory inputs from contralateral AN fibers. Different from Klug et al. (2020) and Hu et al. (2022), the default parameters of Ashida et al. (2016) were used with the length of the rectangular excitatory coincidence window and the rectangular inhibitory window ($W_{inh}$) of 0.8 and 1.6 ms, respectively.

A train of six rectangular clicks with a single-click-duration of 50 μs or six biphasic constant-amplitude pulse trains (cathodic/anodic, 50 μs phase duration) were generated digitally with sample rate of 100 kHz and used as the inputs to the acoustic or electrical AN model, respectively. The results of five ICIs (5, 4, 3, 2, and 1 ms) at different presentation levels over 200 repetitions were obtained for each condition.

## Histology and data analysis

MNTB and BCs were positively identified in most recordings by demonstrating a pre potential (PP) in the wave form. Additionally, HRP was used to identify the recording site of the stimulated cells. By applying 1 μA for 8 min at the end of the experiment HRP was administered iontophoretically through the recording electrode. After the administration of HRP a lethal dose of Narcoren (Pentobarbital 500 mg/kg) was injected intraperitoneally. Firstly the animal was perfused for 10 min with Ringer-solution containing NaCl (0.9%), heparin (100 μl), and 5 mM phosphate-buffered saline (PBS) in $H_2O$. Afterward the animal was perfused with 4% paraformaldehyde (PFA in PBS pH 7.4) for another 10–25 min. Finally the brain was removed, placed into 4% PFA for 1–2 days at 4°C, then washed three times for 10 min in PBS (0.02 M) and put into 4% agarose. Using a vibratome the brain was sliced into coronal brain slices of 50–80 μm thickness. A diaminobenzidine (DAB) substrate kit (Vector Laboratories, Inc.) served to stain the brain against HRP. The brain was incubated for 8 min with DAB, then the slices were washed three times for 10 min with aqua dest and PBS and placed onto glass objectives to dry overnight. On the following morning, the slices were counterstained with neutral red solution [1 g neutral red in 4 ml acetate buffer (0.2 M) pH + 4.8 mixed with 100 ml distilled water] and covered with glass objectives slides and DePeX.

Custom-made programs in Matlab (The MathWorks Inc.) were used for data analysis. The physiological data was analyzed for each ICI individually by calculating the median spike probability and jitter for AVCN and MNTB, and dynamic ITD range and delta slope per dB ILD for LSO. The spike probability of each neuron was calculated by dividing the median spike rate per trial by 6 (the number of pulses per trial). The jitter was calculated by the standard deviation of the AP latency relative to the eliciting click. The dynamic ITD range was defined as the range between the ITDs that elicited maximal and minimal spike rates. The effect of ILDs on ITD sensitivity was assessed by calculating the delta slope: the difference in the normalized maximal and minimal spike rates were divided by the respective dynamic ITD range to yield the slope. These values were subtracted between the most positive and negative ILD that each LSO neuron was tested for and divided by the respective difference in ILD. Artifact removal in recordings during electrical stimulation was performed by zeroing the amplitude values of the raw data traces (typically ±10 samples centered on the artifact peak). Spike detection and spike time determination were performed subsequently. A non-parametric ANOVA was used to test for across-group significance. The Mann–Whitney *U*-test served to test for statistical significance of individual ICIs.

The standard separation D is calculated as previously described (Sakitt, 1973):

$$D_n = |mu_{n+1} - mu_n|/(sqrt(sigma_{n+1}{}^*sigma_n)),$$

where $mu_{n+1}$ and $mu_n$ are the mean values of the hemispheric rate differences to two ITD values while $sigma_{n+1}$ and $sigma_n$ are their standard deviation. $D_n$ was subsequently smoothed using a 5-sample moving average filter. Sigma of model responses follow a Poisson noise assumption.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

This animal study was reviewed and approved by the Ethikkommission der Regierung von Oberbayern.

## Author contributions

MP conceived the study, designed the physiological experiments, and wrote the first draft of the manuscript. MM designed and conducted the cochlear implantations and carried out the experiments with recordings in AVCN and MNTB. BB conducted the LSO recordings. HH and MD generated the model. HH delivered the model results. MM and DNF analyzed the results. MP and DNF designed and generated the figures. MM and HH made contributions. All authors provided comments and approved the manuscript.

## Funding

This study was supported by the Deutsche Forschungsgemeinschaft (DFG) (DFG Priority Program 1608 to MP) and European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (716800 - IBiDT, Project to MD).

## Acknowledgments

We thank MED-EL for providing the stimulation electrodes and K. Fischer and H. Wohlfrom from help with histology. We are also thankful to M. Vollmer for technical help with the acute implantation, G. Ashida and S. Anderson for support with the model, and to B. Grothe for financial support.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Ashida, G., Kretzberg, J., and Tollin, D. J. (2016). Roles for coincidence detection in coding amplitude-modulated sounds. *PLoS Comput. Biol.* 12:e1004997. doi: 10.1371/JOURNAL.PCBI.1004997

Beiderbeck, B., Myoga, M. H., Müller, N. I. C., Callan, A. R., Friauf, E., Grothe, B., et al. (2018). Precisely timed inhibition facilitates action potential firing for spatial coding in the auditory brainstem. *Nat. Commun.* 9:1771. doi: 10.1038/S41467-018-04210-Y

Bernstein, L. R., and Trahiotis, C. (2002). Enhancing sensitivity to interaural delays at high frequencies by using "transposed stimuli." *J. Acoust. Soc. Am.* 112(3 Pt 1), 1026–1036. doi: 10.1121/1.1497620

Bernstein, L. R., and Trahiotis, C. (2009). How sensitivity to ongoing interaural temporal disparities is affected by manipulations of temporal features of the envelopes of high-frequency stimuli. *J. Acoust. Soc. Am.* 125, 3234–3242. doi: 10.1121/1.3101454

Bruce, I. C., Erfani, Y., and Zilany, M. S. A. (2018). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hear. Res.* 360, 40–54. doi: 10.1016/j.heares.2017.12.016

Buck, A. N., Rosskothen-Kuhl, N., and Schnupp, J. W. (2021). Sensitivity to interaural time differences in the inferior colliculus of cochlear implanted rats with or without hearing experience. *Hear. Res.* 408:108305. doi: 10.1016/J.HEARES.2021.108305

Buechel, B. D., Hancock, K. E., Chung, Y., and Delgutte, B. (2018). Improved neural coding of ITD with bilateral cochlear implants by introducing short inter-pulse intervals. *J. Assoc. Res. Otolaryngol.* 19, 681–702. doi: 10.1007/S10162-018-00693-0

Chung, Y., Hancock, K. E., and Delgutte, B. (2016). Neural coding of interaural time differences with bilateral cochlear implants in unanesthetized rabbits. *J. Neurosci.* 36, 5520–5531. doi: 10.1523/JNEUROSCI.3795-15.2016

Dietz, M., Wang, L., Greenberg, D., and McAlpine, D. (2016). Sensitivity to interaural time differences conveyed in the stimulus envelope: Estimating inputs of binaural neurons through the temporal analysis of spike trains. *J. Assoc. Res. Otolaryngol.* 17, 313–330. doi: 10.1007/S10162-016-0573-9

Dynes, S. B. C., and Delgutte, B. (1992). Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. *Hear. Res.* 58, 79–90. doi: 10.1016/0378-5955(92)90011-B

Finlayson, P. G., and Caspary, D. M. (1991). Low-frequency neurons in the lateral superior olive exhibit phase-sensitive binaural inhibition. *J. Neurophysiol.* 65, 598–605. doi: 10.1152/JN.1991.65.3.598

Fredelake, S., and Hohmann, V. (2012). Factors affecting predicted speech intelligibility with cochlear implants in an auditory model for electrical stimulation. *Hear. Res.* 287, 76–90. doi: 10.1016/J.HEARES.2012.03.005

Grothe, B., and Pecka, M. (2014). The natural history of sound localization in mammals-a story of neuronal inhibition. *Front. Neural Circuits* 8:116. doi: 10.3389/fncir.2014.00116

Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiol. Rev.* 90, 983–1012. doi: 10.1152/physrev.00026.2009

Hamacher, V. (2004). *Signalverarbeitungsmodelle des elektrisch stimulierten gehörs*, Ph.D. thesis, Vol. 17, ed. P. Vary (Aachen: IND, RWTH Aachen, Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN)).

Hancock, K. E., Chung, Y., Delgutte, B., and Hancock, K. E. (2012). Neural ITD coding with bilateral cochlear implants: Effect of binaurally coherent jitter. *J. Neurophysiol.* 108, 714–728. doi: 10.1152/jn.00269.2012.-Poor

Hancock, K. E., Noel, V., Ryugo, D. K., and Delgutte, B. (2010). Neural coding of interaural time differences with bilateral cochlear implants: Effects of congenital deafness. *J. Neurosci.* 30, 14068–14079. doi: 10.1523/JNEUROSCI.3213-10.2010

Hartmann, R., and Klinke, R. (1990). "Response characteristics of nerve fibers to patterned electrical stimulation," in *Cochlear implants: Models of the electrically stimulated ear*, eds J. M. Miller and F. A. Spelman (New York, NY: Springer), 135–160. doi: 10.1007/978-1-4612-3256-8_10

Hartmann, R., Topp, G., and Klinke, R. (1984). Discharge patterns of cat primary auditory fibers with electrical stimulation of the cochlea. *Hear. Res.* 13, 47–62. doi: 10.1016/0378-5955(84)90094-7

Hu, H., Klug, J., and Dietz, M. (2022). Simulation of ITD-dependent single-neuron responses under electrical stimulation and with amplitude-modulated acoustic stimuli. *J. Assoc. Res. Otolaryngol.* 23, 535–550. doi: 10.1007/s10162-021-00823-1

Ihlefeld, A., Kan, A., and Litovsky, R. Y. (2014). Across-frequency combination of interaural time difference in bilateral cochlear implant listeners. *Front. Syst. Neurosci.* 8:22. doi: 10.3389/FNSYS.2014.00022

Javel, E., and Shepherd, R. K. (2000). Electrical stimulation of the auditory nerve. III. Response initiation sites and temporal fine structure. *Hear. Res.* 140, 45–76. doi: 10.1016/S0378-5955(99)00186-0

Javel, E., and Viemeister, N. F. (2000). Stochastic properties of cat auditory nerve responses to electric and acoustic stimuli and application to intensity discrimination. *J. Acoust. Soc. Am.* 107, 908–921. doi: 10.1121/1.428269

Joris, P. X., and Yin, T. C. T. (1998). Envelope coding in the lateral superior olive. III. Comparison with afferent pathways. *J. Neurophysiol.* 79, 253–269. doi: 10.1152/JN.1998.79.1.253

Kapfer, C., Seidl, A. H., Schweizer, H., and Grothe, B. (2002). Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons. *Nat. Neurosci.* 5, 247–253. doi: 10.1038/nn81

Kim, G., and Kandler, K. (2003). Elimination and strengthening of glycinergic/GABAergic connections during tonotopic map formation. *Nat. Neurosci.* 6, 282–290. doi: 10.1038/nn1015

Klein-Hennig, M., Dietz, M., Hohmann, V., and Ewert, S. D. (2011). The influence of different segments of the ongoing envelope on sensitivity to interaural time delays. *J. Acoust. Soc. Am.* 129, 3856–3872. doi: 10.1121/1.3585847

Klug, J., Schmors, L., Ashida, G., and Dietz, M. (2020). Neural rate difference model can account for lateralization of high-frequency stimuli. *J. Acoust. Soc. Am.* 148, 678–691. doi: 10.1121/10.0001602

Laback, B., Egger, K., and Majdak, P. (2015). Perception and coding of interaural time differences with bilateral cochlear implants. *Hear. Res.* 322, 138–150. doi: 10.1016/J.HEARES.2014.10.004

Laback, B., and Majdak, P. (2008). Binaural jitter improves interaural time-difference sensitivity of cochlear implantees at high pulse rates. *Proc. Natl. Acad. Sci. U.S.A.* 105, 814–817. doi: 10.1073/PNAS.0709199105

Litovsky, R. Y., Jones, G. L., Agrawal, S., and van Hoesel, R. (2010). Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans. *J. Acoust. Soc. Am.* 127, 400–414. doi: 10.1121/1.325754

Litvak, L., Delgutte, B., and Eddington, D. (2001). Auditory nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains. *J. Acoust. Soc. Am.* 110, 368–379. doi: 10.1121/1.1375140

Macpherson, E. A., and Middlebrooks, J. C. (2002). Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *J. Acoust. Soc. Am.* 111 (5 Pt 1), 2219–2236. doi: 10.1121/1.1471898

Maki, K., and Furukawa, S. (2005). Acoustical cues for sound localization by the Mongolian gerbil, *Meriones unguiculatus*. *J. Acoust. Soc. Am.* 118, 872–886. doi: 10.1121/1.1944647

McFadden, S. L., Walsh, E. J., and McGee, J. (1996). Onset and development of auditory brainstem responses in the Mongolian gerbil (*Meriones unguiculatus*). *Hear. Res.* 100, 68–79. doi: 10.1016/0378-5955(96)00108-6

Mills, J. H., Schmiedt, R. A., and Kulish, L. F. (1990). Age-related changes in auditory potentials of Mongolian gerbil. *Hear. Res.* 46, 201–210. doi: 10.1016/0378-5955(90)90002-7

Moore, B. C. J. (2013). *An introduction the psychology of hearing*, 6th Edn. Leiden: Brill.

Myoga, M. H., Lehnert, S., Leibold, C., Felmy, F., and Grothe, B. (2014). Glycinergic inhibition tunes coincidence detection in the auditory brainstem. *Nat. Commun.* 5:3790. doi: 10.1038/ncomms4790

Park, T. J., Grothe, B., Pollak, G. D., Schuller, G., and Koch, U. (1996). Neural delays shape selectivity to interaural intensity differences in the lateral superior olive. *J. Neurosci.* 16, 6554–6566. doi: 10.1523/JNEUROSCI.16-20-06554.1996

Pecka, M., Brand, A., Behrend, O., and Grothe, B. (2008). Interaural time difference processing in the mammalian medial superior olive: The role of glycinergic inhibition. *J. Neurosci.* 28, 6914–6925. doi: 10.1523/JNEUROSCI.1660-08.2008

Pecka, M., and Encke, J. (2020). "Coding of spatial information," in *The senses: A comprehensive reference*, Vol. 2, eds B. Fritzsch and B. Grothe (Amsterdam: Elsevier), 681–691.

Pecka, M., Siveke, I., Grothe, B., and Lesica, N. A. (2010). Enhancement of ITD coding within the initial stages of the auditory pathway. *J. Neurophysiol.* 103, 38–46. doi: 10.1152/jn.00628.2009

Pecka, M., Zahn, T. P., Saunier-Rebori, B., Siveke, I., Felmy, F., Wiegrebe, L., et al. (2007). Inhibiting the inhibition: A neuronal network for sound localization in reverberant environments. *J. Neurosci.* 27, 1782–1790. doi: 10.1523/JNEUROSCI.5335-06.2007

Rosskothen-Kuhl, N., Buck, A. N., Li, K., and Schnupp, J. W. (2021). Microsecond interaural time difference discrimination restored by cochlear implants after neonatal deafness. *Elife* 10:e59300. doi: 10.7554/eLife.59300

Sakitt, B. (1973). Indices of discriminability. *Nature* 241, 133–134. doi: 10.1038/241133a0

Sanes, D. H., and Rubel, E. W. (1988). The ontogeny of inhibition and excitation in the gerbil lateral superior olive. *J. Neurosci.* 8, 682–700. doi: 10.1523/JNEUROSCI.08-02-00682.1988

Seidl, A. H., and Grothe, B. (2005). Development of sound localization mechanisms in the Mongolian gerbil is shaped by early acoustic experience. *J. Neurophysiol.* 94, 1028–1036. doi: 10.1152/JN.01143.2004

Smith, P. H., Joris, P. X., and Yin, T. C. T. (1998). Anatomy and physiology of principal cells of the medial nucleus of the trapezoid body (MNTB) of the cat. *J. Neurophysiol.* 79, 3127–3142. doi: 10.1152/JN.1998.79.6.3127

Smith, Z. M., and Delgutte, B. (2007). Sensitivity to interaural time differences in the inferior colliculus with bilateral cochlear implants. *J. Neurosci.* 27, 6740–6750. doi: 10.1523/JNEUROSCI.0052-07.2007

Smith, Z. M., and Delgutte, B. (2008). Sensitivity of inferior colliculus neurons to interaural time differences in the envelope versus the fine structure with bilateral cochlear implants. *J. Neurophysiol.* 99, 2390–2407. doi: 10.1152/JN.00751.2007

Tillein, J., Hubka, P., Syed, E., Hartmann, R., Engel, A. K., and Kral, A. (2010). Cortical representation of interaural time difference in congenital deafness. *Cereb. Cortex* 20, 492–506. doi: 10.1093/CERCOR/BHP222

van den Honert, C., and Stypulkowski, P. H. (1987). Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli. *Hear. Res.* 29, 207–222. doi: 10.1016/0378-5955(87)90168-7

Wightman, F. L., and Kistler, D. J. (1998). The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* 91:1648. doi: 10.1121/1.402445

Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (1991). Better speech recognition with cochlear implants. *Nature* 352, 236–238. doi: 10.1038/352236a0

Zilany, M. S. A., Bruce, I. C., and Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *J. Acoust. Soc. Am.* 135, 283–286. doi: 10.1121/1.4837815

Zilany, M. S. A., Bruce, I. C., Nelson, P. C., and Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *J. Acoust. Soc. Am.* 126, 2390–2412. doi: 10.1121/1.3238250

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 09 January 2023
DOI 10.3389/fnins.2022.1018190



OPEN ACCESS

EDITED BY
Yi Zhou,
Arizona State University, United States

REVIEWED BY
Tara Davis,
University of South Alabama, United States
Ceren Karacayli,
University of Health Sciences, Türkiye

*CORRESPONDENCE
Sean R. Anderson
✉ sean.hearing@gmail.com

†PRESENT ADDRESS
Sean R. Anderson,
Department of Physiology and Biophysics, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 12 August 2022
ACCEPTED 19 December 2022
PUBLISHED 09 January 2023

CITATION
Anderson SR, Gallun FJ and Litovsky RY (2023) Interaural asymmetry of dynamic range: Abnormal fusion, bilateral interference, and shifts in attention. *Front. Neurosci.* 16:1018190. doi: 10.3389/fnins.2022.1018190



# Interaural asymmetry of dynamic range: Abnormal fusion, bilateral interference, and shifts in attention

Sean R. Anderson[1*†], Frederick J. Gallun[2] and Ruth Y. Litovsky[1,3,4]

[1]Waisman Center, University of Wisconsin-Madison, Madison, WI, United States, [2]Department of Otolaryngology-Head and Neck Surgery, Oregon Health and Science University, Portland, OR, United States, [3]Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, WI, United States, [4]Department of Surgery, Division of Otolaryngology, University of Wisconsin-Madison, Madison, WI, United States

Speech information in the better ear interferes with the poorer ear in patients with bilateral cochlear implants (BiCIs) who have large asymmetries in speech intelligibility between ears. The goal of the present study was to assess how each ear impacts, and whether one dominates, speech perception using simulated CI processing in older and younger normal-hearing (ONH and YNH) listeners. Dynamic range (DR) was manipulated symmetrically or asymmetrically across spectral bands in a vocoder. We hypothesized that if abnormal integration of speech information occurs with asymmetrical speech understanding, listeners would demonstrate an atypical preference in accuracy when reporting speech presented to the better ear and fusion of speech between the ears (i.e., an increased number of one-word responses when two words were presented). Results from three speech conditions showed that: (1) When the same word was presented to both ears, speech identification accuracy decreased if one or both ears decreased in DR, but listeners usually reported hearing one word. (2) When two words with different vowels were presented to both ears, speech identification accuracy and percentage of two-word responses decreased consistently as DR decreased in one or both ears. (3) When two rhyming words (e.g., bed and led) previously shown to phonologically fuse between ears (e.g., bled) were presented, listeners instead demonstrated interference as DR decreased. The word responded in (2) and (3) came from the right (symmetric) or better (asymmetric) ear, especially in (3) and for ONH listeners in (2). These results suggest that the ear with poorer dynamic range is downweighted by the auditory system, resulting in abnormal fusion and interference, especially for older listeners.

KEYWORDS

**interaural asymmetry, speech perception, binaural hearing, phonological fusion, dichotic speech, dynamic range compression, aging**

# 1. Introduction

Patients with bilateral severe to profound hearing loss can receive cochlear implants (CIs) to gain access to hearing. Bilateral CIs (BiCIs) improve sound source localization performance and speech understanding in noise relative to unilateral implantation (e.g., Litovsky et al., 2006). However, the extent of this benefit varies highly across patients (Litovsky et al., 2006; Mosnier et al., 2009; Yoon et al., 2011; Reeder et al., 2014; Goupell et al., 2016, 2018; Bakal et al., 2021).

Models of binaural hearing benefits based on studies completed in listeners with normal-hearing (NH) often assume that the ears act as ideal and independent channels that can be used to cancel out the masking stimulus and attend to a target of interest (e.g., Durlach, 1963). Similarly, studies presenting unrelated maskers to the ear opposite the target have suggested that listeners can ignore one ear without any decrement in performance (Cherry, 1953; Brungart and Simpson, 2002). These assumptions may not apply to patients with BiCIs, who often show marked interaural asymmetry in various aspects of auditory processing, such as speech understanding and spectro-temporal resolution. These asymmetries are likely to be produced by many different sources (Anderson, 2022). Thus, throughout this manuscript we define interaurally asymmetric hearing outcomes as any undesirable difference between the two ears to which one would answer affirmatively to the question "Does listening with your left compared to your right ear sound different?"

## 1.1. Interaural asymmetry

### 1.1.1. Poorer ear or amount of asymmetry?

Studies of patients with BiCIs and simulations in NH suggest that interaurally asymmetric hearing outcomes may limit performance in binaural tasks (Mosnier et al., 2009; Yoon et al., 2011; Ihlefeld et al., 2015; Goupell et al., 2016, 2018; Anderson et al., 2019b, 2022; Bakal et al., 2021). These studies assessed sensitivity to binaural cues, sound source localization, and speech understanding in background noise, and related them to asymmetry in sensitivity to temporal cues or monaural speech understanding. To address interaurally asymmetric hearing more directly, some studies first indexed or manipulated temporal fidelity in both ears, then assessed sensitivity to binaural cues. The ear with poorer temporal fidelity predicted the amount of sensitivity to binaural cues (Ihlefeld et al., 2015; Anderson et al., 2019b, 2022). Simulations in NH used asymmetric dynamic range (i.e., amplitude modulation depth), where smaller dynamic ranges in listeners with BiCIs have resulted in poorer sensitivity to binaural cues (Ihlefeld et al., 2014; Todd et al., 2017). In these studies, performance with the *poorer ear* was predictive of the binaural benefit, suggesting that if one ear is poorly performing, it can act as a bottleneck that limits encoding of information which is used in binaural processing. Other studies evaluating speech understanding suggest that a larger relative difference *between ears* is associated with poorer benefits (Mosnier et al., 2009; Yoon et al., 2011; Goupell et al., 2016, 2018). The discrepancy in the interpretation of these findings may be due to difficultly controlling for the degree of asymmetry in studies with patients who use BiCIs, differences in the complexity of the stimuli and task, or differences between ears and limitations of the poorer ear may be at play.

### 1.1.2. Changes in fusion or attention?

Historically, the term "fusion" has referred to many different phenomena. In the spatial hearing literature, fusion can refer to the report of a singular auditory image when a source and simulated echo are presented (Litovsky et al., 1999). In the dichotic pitch literature, fusion classically refers to the perception of a singular pitch (e.g., van den Brink et al., 1976). These subjective approaches to measuring fusion result in notoriously large amounts of variability. Moreover, spatial and pitch fusion may not always occur at the same time (Scharf, 1974). Other experimental approaches have explored fusion of speech stimuli. In the dichotic vowel literature, fusion has referred to the perception of a new vowel not corresponding to that presented in either ear (e.g., Darwin, 1981) or the reporting of only one vowel (e.g., Reiss and Molis, 2021; Eddolls et al., 2022). Similar observations can be made from the dichotic speech literature (Cutting, 1975, 1976). However, it is commonplace to report the number of items responded and interpret them in a similar way to "fusion" in these studies (e.g., Cutting, 1975; Darwin, 1981). The present experiment defines fusion as the reporting of one word, which may correspond to the left, right, both, or neither ear. We define auditory selective attention as the ability to attend to one ear (reflected in the relative weight of the left and right ear in dichotic studies). We define bilateral interference as decreased identification accuracy relative to baseline when in the presence of another stimulus in the opposite ear.

Studies have reported that, compared to a monaural condition in which both target and masker are presented to the same ear, adding a copy of the masker in the ear opposite the target speech results in improved performance for listeners with NH and BiCIs (e.g., Loizou et al., 2009; Bernstein et al., 2016; Goupell et al., 2016). It is assumed that this occurs because the masking stimuli are fused, resulting in a perceived central location within the head (i.e., spatially fused). The target speech is instead perceived on the side of the ear it is presented, resulting in unmasking. In contrast, patients with BiCIs who have marked asymmetry in speech understanding between the ears demonstrate contralateral *interference* when target speech is presented to their poorer ear (Bernstein et al., 2016; Goupell et al., 2016, 2018; Bakal et al., 2021). Listeners with a CI in one ear and NH in the other ear show the same pattern of

performance (Bernstein et al., 2020). In simulations of BiCIs, contralateral interference occurs when one or both ears have poor spectro-temporal resolution (Gallun et al., 2007; Goupell et al., 2021). Two mechanisms have been proposed to drive contralateral interference in experiments where the target is presented to one ear and the masker is presented to one or both ears. The first is differences in how target and masking stimuli are perceptually segregated from one another, suggesting that they may instead be fused together (Gallun et al., 2007; Reiss and Molis, 2021). This disruptive fusion could therefore occur within the ear containing the target, across ears, or both. The second is a failure of attention, where it is more difficult to ignore the clearer stimulus (Goupell et al., 2021).

An attentional basis of contralateral interference is suggested by the finding that performance remains intact if the target is in the better ear. This has been demonstrated for listeners with BiCIs (Goupell et al., 2016, 2018; Bakal et al., 2021), listeners with one CI and one NH ear (Bernstein et al., 2016, 2020), and simulations of BiCIs in listeners with NH (Goupell et al., 2021). If contralateral interference results purely from an inability to segregate target from masker, then it should not matter whether the target is in the better or poorer ear.

Right-ear advantage has been well-documented in the auditory literature and is suspected to result from an attentional bias toward the right ear for typically developing listeners with NH (Kinsbourne, 1970; Hiscock and Kinsbourne, 2011). Another classical theory of ear advantage relates to a structural difference between the connections of the left and right ear to auditory and language processing centers (Kimura, 1967), which could be relevant for listeners with a difference in the fidelity of information represented in the left versus right ear. This may be especially relevant for listeners who experience prolonged periods of deafness, which are known to cause deterioration of the peripheral and central auditory system (e.g., Shepherd and Hardie, 2001). Interestingly, increasing age is associated with an elevated right-ear advantage (Westerhausen et al., 2015). Since most experiments concerning listeners with BiCIs tend to test older individuals, age is an important variable to account for in experiments concerning auditory spatial attention.

## 1.2. Goals and hypotheses of the present study

It is becoming clearer in the literature that processing of auditory inputs is not truly independent in each ear. Instead, information is integrated by the central auditory system and a highly efficient attentional network can be used to focus on a source of interest (e.g., Shinn-Cunningham et al., 2017). Because of the inherent connection between sound source segregation and attention, it is difficult to disentangle both processes from one another and determine how they might affect patients. While there is a right-ear advantage noted in the literature for listeners with NH, listeners with BiCIs can have considerably different speech outcomes between the ears due to many underlying factors (e.g., Litovsky et al., 2006; Mosnier et al., 2009; Goupell et al., 2018; Bakal et al., 2021). One of the major challenges of studies of interaural asymmetry in listeners with BiCIs is that large sample sizes are required to account for differences between patients. Thus, it is sometimes more practical to simulate particular sources of asymmetry in listeners with NH to determine the impact on perception. The present study simulated interaurally asymmetric dynamic range to explore the effects of degraded temporal representations on speech perception.

Our goal was to use a task that explores both fusion and auditory attention. To meet this goal, we adapted a speech perception experiment exploring a phenomenon called "phonological fusion." In phonological fusion experiments, listeners were presented with two rhyming words to the left and right ear (Cutting, 1975, 1976). One word began with a stop consonant (e.g., /b/) and the other began with a liquid (e.g., /l/). Both words shared the same ending (e.g., /ɛd/), and combining the stop and liquid into a cluster would generate a word in English (e.g., bled). In the original experiments, when words were generated using natural speech productions and presented simultaneously, listeners reported hearing the fused word on approximately 30% of trials. Using synthetic speech, listeners reported hearing one word on approximately 70% of trials, which could correspond to the fused word, the word in the left or right ear, or some other word unrelated to those presented. Thus, using this paradigm, it is possible to assess whether listeners fused the percept into one word, whether listeners weighted the ears equally or unequally, and the relationship between fusion and ear-weighting on speech understanding accuracy.

In the present experiment, we assessed phonological fusion as well as closed-set speech identification of the same word or words with different vowels in each ear. This helped us evaluate a broad range of performance. It is well-known that low-frequency temporal envelope cues are essential to speech understanding in CI processing (Drullman et al., 1994; Shannon et al., 1995). The dynamic range of each electrode varies across listeners (Long et al., 2014). Smaller dynamic ranges result in poorer speech understanding (Firszt et al., 2002; Spahr et al., 2007) and binaural processing (Ihlefeld et al., 2014; Todd et al., 2017) for listeners with BiCIs. We simulated CI processing using a vocoder and manipulated the dynamic range of the speech in each ear symmetrically or asymmetrically.

The criteria used to evaluate responses (accuracy, number of words reported, response categories, and vowels) were chosen in an attempt to shed light on fusion and on the relative weight given to either ear (i.e., auditory spatial attention). In the present study, fusion was assessed primarily by the number of words being reported, consistent with recent studies (Reiss et al., 2016; Reiss and Molis, 2021; Eddolls et al., 2022), and

secondarily, in Section "3.3 Phonological fusion trials," by the proportion of phonological fusion responses consistent with classical studies (Cutting, 1975, 1976). Auditory spatial attention was assessed by whether the word(s) reported corresponded more closely to the left or right ear. Bilateral interference was assessed by the proportion of incorrect responses, most notably those that did not correspond to speech presented in either ear and could therefore not be explained by the better dynamic range of the word(s) presented. Unlike many studies concerning ear advantage, the present study asked listeners to report the word in both ears. Thus, if listeners responded with only one word, it was assumed that listeners only heard one word or they were very uncertain about what was presented to the other ear. When listeners reported two words with only one word correct, it was assumed that greater attention was being allocated toward the correctly reported ear. Finally, when one word was reported corresponding to one ear, it was assumed that the words were fused and attention was allocated to that ear. Critically, the present experiment relied on many repeated presentations to assess this and several "anchoring" conditions where both ears provided small or large dynamic range. All analyses were completed within subjects, meaning each subject acted as their own control. For a graphical description of the interpretations applied to the accuracy and number of words responded, see **Supplementary Figure 1**.

Three different types of trials were tested in the present experiment aiming to address different questions, and these are separated into different sections of the Results. All three kinds of trials included symmetric or asymmetric dynamic ranges. In section "3.1 Same world trials," the same word was presented to both ears. This condition provided data concerning the characteristic errors associated with a decrease in dynamic range. Additionally, this condition allowed for the assessment of the alternative prediction: If experienced *less* fusion as dynamic range decreased in one or both ears, then listeners would report hearing one word less often. In section "3.2 Different vowel trials," words with different vowels were presented to both ears. If listeners experienced fusion as dynamic range decreased in one or both ears and were therefore unable to attend to a single ear, then they would report hearing two words less often in these conditions and a decrease in the accuracy of correctly reporting at least one word. The latter result would occur because fusion of degraded words could result in an unintelligible word. If instead listeners were able to attend to one ear and entirely ignore the other ear, then the proportion of at least one word correct would be bounded by the symmetric dynamic range results in section "3.1 Same world trials." In section "3.3 Phonological fusion trials," rhyming words were presented, a subset of which could be phonologically fused to generate a new word as described in the preceding paragraph. Phonological fusion was considered to be a special case of more general fusion. If listeners experienced fusion as dynamic range decreased in one or both ears and listeners were unable to attend to only one ear, the word(s) responded would match the phonologically fused word or an incorrect word. The latter would occur because fusion of degraded words could result in a single, unintelligible word. Alternatively, if listeners were able to attend to one ear and entirely ignore the other ear, the word(s) responded would match the left, right, or both ears. Thus, sections "3.2 Different vowel trials" and "3.3 Phonological fusion trials" shed light onto the role of fusion, attention, and interference, while section "3.1 Same word trials" sheds light on the effects of the vocoder simulation.

We hypothesized that when dynamic range in both ears was decreased, listeners would experience greater fusion of words that are different from one another, decreasing the speech understanding. We further hypothesized that this would occur if the dynamic range was decreased in only one ear, listeners would also experience increased fusion and decreased speech understanding. This would be consistent with previous literature concerning discrimination of binaural cues in listeners with NH and BiCIs (Ihlefeld et al., 2015; Anderson et al., 2019b, 2022). Alternatively, differences between the ears themselves may cause problems for listeners with BiCIs (e.g., Yoon et al., 2011). We therefore alternatively hypothesize that greater differences in dynamic range would lead to increased fusion and decreased speech understanding. We further hypothesized that listeners would exhibit right-ear advantage, weighting speech from the right ear more heavily. Thus, it was predicted that symmetrically smaller dynamic ranges would result in decreased accuracy and increased proportion of one-word responses when two words were presented. It was further predicted that listeners would correctly report more words from the right ear (i.e., right-ear advantage). It was predicted that word identification accuracy would be similar between asymmetric conditions (e.g., 100:60%) and symmetric conditions with the smaller of the asymmetric dynamic ranges (e.g., 60:60%). Alternatively, performance could reflect the difference in dynamic range between ears, where word identification accuracy in asymmetric conditions corresponds to the difference in dynamic range between the left and right ear. It was further predicted that asymmetric dynamic range conditions would bias listeners toward the better ear, where their responses would reflect the word presented to that ear (better-ear advantage).

Two groups of listeners were tested: younger NH (YNH) listeners and older NH (ONH) listeners within a similar age range to the typical CI study cohort, e.g., (Bernstein et al., 2016; Goupell et al., 2016, 2018; Baumgärtel et al., 2017; Reiss et al., 2018; Anderson et al., 2019a, 2022; Bakal et al., 2021). Critically, aging is associated with poorer binaural and monaural temporal processing (Gallun et al., 2014; Baumgärtel et al., 2017; Anderson et al., 2019a), increased aural preference exhibited via right-ear advantage (Westerhausen et al., 2015), and decreased working memory (Roque et al., 2019). We therefore hypothesized that ONH listeners have poorer temporal processing, greater aural preference, and

poorer selective attention than YNH listeners, impairing their ability to accurately identify speech and allocate attention to a better ear. It was predicted that ONH listeners would exhibit lower accuracy compared to YNH listeners across dynamic range due to poorer temporal processing. It was further predicted that ONH listeners would exhibit a higher proportion of one-word responses compared to YNH listeners due to increased aural preference. Finally, it was predicted that ONH listeners would exhibit even less accuracy in trials where two words were presented compared to YNH listeners because of increased cognitive demand.

# 2. Materials and methods

## 2.1. Listeners and equipment

Ten YNH (20–34 years; average 26.8 years) and ten ONH listeners (52–72 years; average 59.4 years) participated in the present study. Because this experiment was completed in listeners' homes via remote testing, experimental software was used to estimate audiometric thresholds. Listeners' estimated audiometric thresholds are presented in **Figure 1**. While it is conventional to report normal hearing for YNH listeners as 20 dB HL or below, some older participants had higher estimated thresholds at 4 and 8 kHz. Audiometric responses were assessed for octave-spaced frequencies between 0.25 and 8.0 kHz using custom software in MATLAB. Sound levels in decibels hearing level (dB HL) for each frequency were determined based on the values in decibels sound pressure level (dB SPL) reported for supra-aural TDH 49/50 headphones (Frank, 1997). A conservative procedure was used to ensure that output during hearing assessment reached the desired level. The lower limit of 20 dB HL was determined based on the noise floor of the sound level meter. Sound levels were confirmed to be within 5 dB(A) of the values in dB HL for all levels except the lowest in some cases. The lowest level confirmed from the output of the sound level meter from lowest to highest frequencies were 20, 25, 30, 25, 25, and 25 dB HL, respectively. For YNH listeners, it was not possible to determine whether there were asymmetric hearing losses because the equipment could not confidently produce sound levels below 20 dB HL. For ONH listeners, participant ONH08 had a 30-dB asymmetry at 8 kHz, where the left ear had an estimated threshold of 50 dB HL and the right ear had an estimated threshold of $\leq$ 20 dB HL. All other ONH listeners with estimated thresholds above 20 dB HL had asymmetries $\leq$ 10 dB. All listeners were included in the results and analysis. All listeners spoke English as their first language. Since individual data are available with the present manuscript, analyses can be re-computed removing listeners from the dataset. All procedures were approved by the University of Wisconsin-Madison Health Sciences Institution Review Board. Listeners completed informed consent online before participation began.

Estimated audiometric thresholds were collected using custom software. The task consisted of a presentation of one, two, or three tone pips with 10-ms cosine onset- and offset-ramps. Each pip had a duration of 300 ms separated by 200-ms inter-stimulus intervals. The listener indicated the number of pips presented (three-alternative forced-choice). Testing followed standard step sizes of 10 dB-down and 5 dB-up, with a one-up, one-down adaptive rule. Levels were initiated at 70 dB HL for each frequency, beginning with 250 Hz in the left ear, increasing in frequency, and then progressing to the right ear. Threshold was estimated by a listener achieving at least two out of three presentations at the same level correct. If responses reached 20 dB HL, listeners were tested until criterion of two out of three correct. Limitations of this approach are addressed in section "4.3 Limitations."

This experiment was conducted after the onset of the COVID-19 pandemic. Thus, testing was completed in listeners' homes via home delivery by the experimenter. Additional applications of this approach, particularly for ONH listeners who may have mobility issues, are addressed in the discussion. Equipment consisted of noise-attenuating Sennheiser HD 280 Pro circumaural headphones, a Microsoft Surface tablet, a sound level data logger, and power supply packaged into a small box. All testing was completed using automated, custom software written in MATLAB with the Microsoft Surface in kiosk mode. Kiosk mode with limited permissions was used to ensure that the listener could not see their data or use other software on the device. Stimuli were presented at a sampling rate of 44.1 kHz. Listeners were given written setup instructions and technical assistance was available from the experimenter via remote conference on video or telephone for the duration of the experiment. Before testing began, the sound level data logger was turned on to record the sound level in the room in 1-min increments during testing, with a noise floor of 40 dB(A). All participants whose sound level data were not lost had median sound level recordings of $\leq$ 50 dB(A) with no more than 10 min of sound above this level during testing.

## 2.2. Stimuli

Stimuli were a subset of monosyllabic words used in previous phonological fusion experiments (Cutting, 1975, 1976). They consisted of three sets of five words. Each set had a word with a stop consonant at the onset only (bed, pay, and go), two possible liquid consonants (/l/ and /r/), and both possible stop-liquid clusters (e.g., bled and bred).

The speech corpus was produced by one male speaker from the Midwest using standard American English. During the recording process, a metronome was used to assist in generating approximately 50 tokens of each word. Two of 50 tokens per word were selected such that the corpus had roughly similar duration and pitch. The duration and pitch were then manipulated in Praat until they were approximately

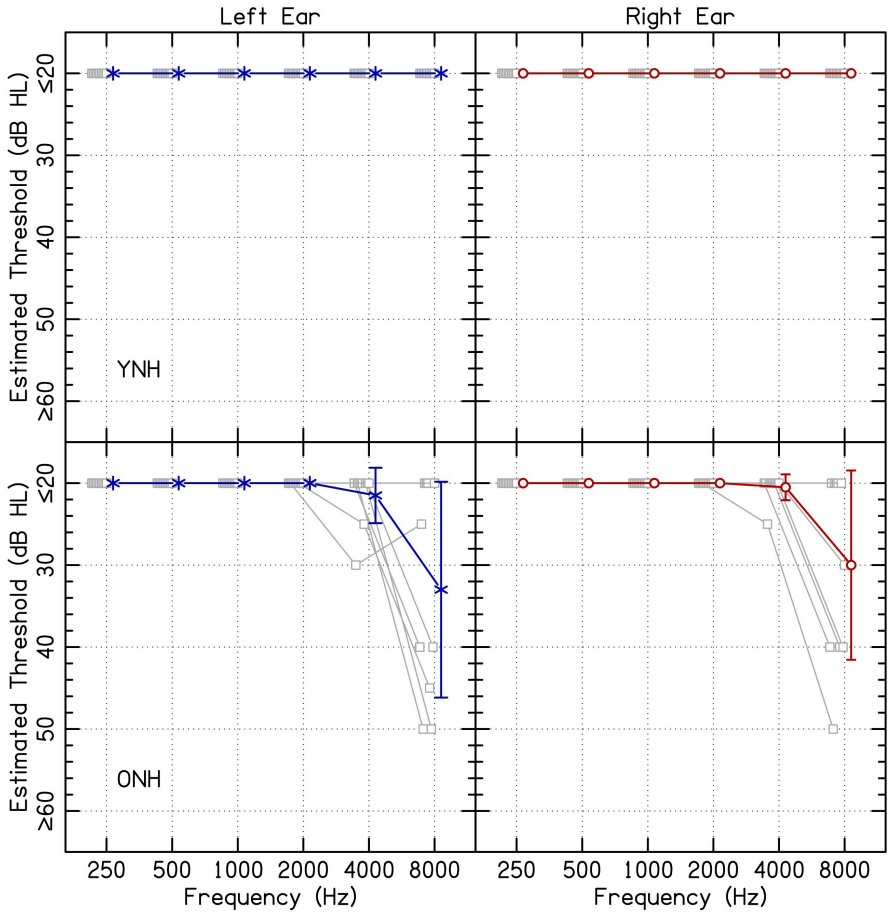



FIGURE 1
Estimated audiometric thresholds for the left and right ear. The panels on the left and right represent responses from the left and right ear, respectively. Average results are shown in blue or red, offset to the right, and error bars represent one standard deviation. Shapes offset to the left represent individual audiometric thresholds. Results from YNH and ONH listeners are shown on the top and bottom rows, respectively.

equal. The resulting mean and standard deviation duration was 558 ± 37 ms. The resulting mean and standard deviation pitch was 101.9 ± 1.1 Hz and all stimuli fell within one semitone. Stimuli were recorded using an M-Audio Fast Track Pro interface and AKG C5900 microphone with pop filter. Stimuli were root-mean-square (RMS) level normalized.

An illustration of stimulus processing is shown in **Figure 2**. Stimuli were vocoded in Praat with software that is available online.[1] Briefly, stimuli were bandpass filtered into 16 frequency bands spaced between 250 and 8,000 Hz. Bandpass filtering was completed by multiplying stimuli in the frequency domain by Hann bands with 12 dB/octave roll-off. Frequency bands were evenly spaced and occupied equivalent cochlear space according to the Greenwood function (Greenwood, 1990). The temporal envelope was extracted using half-wave rectification and a 600-Hz low-pass filter (i.e., a Hann band from 0 to 600 Hz with 12 dB/octave roll-off). The dynamic range of the temporal envelope was manipulated by compressing the extracted envelope to some percentage of its original value in dB. For example, if the dynamic range was equal to 60% and the stimulus normally had a dynamic range of 30 dB (from a minimum level of 40 dB to maximum level of 70 dB), then the new dynamic range would be 18 dB (with a minimum level of 46 dB and a maximum of 64 dB). Therefore, the overall level remained equal between dynamic range conditions while the maximum and minimum levels within each band decreased and increased, respectively. In Praat, this was completed by performing the following procedures on the envelope: (1) adding a small positive value to shift all values above 0 in the envelope amplitude in voltage, (2) converting to dB, (3) subtracting the maximum amplitude to shift the maximum to 0 dB, (4) adding 90 to shift the maximum to 90 dB, (5) filling in the dips in the envelope proportional to one minus the dynamic range (see Eq. 1), (6) subtracting 90 and adding the original maximum to shift the maximum back to the original

1 This code is available for free download at http://mattwinn.com/praat.html#vocoder.

level in dB, (7) converting back to voltage, (8) removing the small positive value shift in voltage, (9) setting low amplitudes to 0 in voltage, (10) low-pass filtering the signal again, and (11) scaling to match the original root-mean-square amplitude of the signal. Equation 1 describes how compression was implemented in the dB domain in step 5:

$$E_{compressed} = E_{full} + \frac{100 - DR}{100}\left(Max - E_{full}\right) \quad (1)$$

where $E_{compressed}$ and $E_{full}$ were the time series of envelope values for the compressed and full dynmic range in dB, respectively, DR was the dynamic range in percent, and *Max* is the maximum value of the envelope in dB. In this case, *Max* was forced to be 90 dB. Thus, compression was implemented by filling in dips in the envelope, then re-scaling the amplitude to match the original level. If dynamic range was 100%, the envelope was unaltered. If dynamic range was 0%, the envelope consisted of only onsets and offsets. For additional details, see the source code available online (see text footnote 1). Low-noise noise bands (Pumplin, 1985) with bandwidths equal to the filter bandwidth were multiplied by resulting envelopes and summed across each frequency band. The resulting stimulus was RMS normalized to be equal in level to pink noise at 65 dB SPL, A-weighted [dB(A)].

## 2.3. Procedures

Listeners first confirmed that their computer set up looked like the instructions provided. Then, they confirmed to the investigator that their headphones were on the correct sides of the head. This was completed with a task measuring the side on which a 250-Hz tone of 1,000-ms duration was presented, with equal probability of being in the left or right headphone (one-interval, two-alternative forced-choice). The stimulus had 10-ms raised cosine onset- and offset-ramps. Two trials were presented at 70 dB(A). If the listener made two errors, they were instructed to reverse the headphones. If the listener made one error, they were instructed to try again. If the listener made no errors, they continued to the testing phase. Listeners completed this check again before each block of experimental trials.

In the second equipment check, listeners confirmed that they were listening via headphones and not loudspeakers using a similar task to that used by Woods et al. (2017). They were presented with three, 250-Hz tone bursts of 1,000-ms duration containing 500-ms inter-stimulus intervals. Their task was to choose the quietest burst (three-interval, two-alternative forced-choice). One tone burst was presented out of phase and at 70 dB(A). The other tones were presented in-phase at levels of 65 and 70 dB(A). Thus, if listeners were using loudspeakers and not headphones, destructive interference from the out-of-phase tone burst would reduce the sound level, making it the quietest. If instead listeners were using headphones, then the in-phase interval presented at 65 dB(A) would be the quietest. Six trials

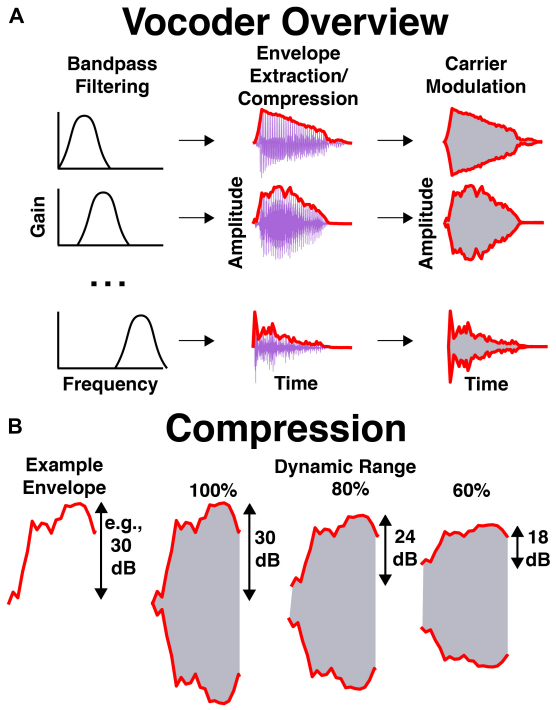



FIGURE 2
Illustration of vocoder processing. **(A)** Processing stages over time, where 16 bandpassed signals are obtained, their envelopes are extracted and compressed, noise-based carrier are modulated in amplitude, and modulated carriers are summed. **(B)** Envelope compression in dB, where the envelope is compressed to some percentage of its original range in dB. This is completed filling in the dips of the envelope, then normalizing to the original envelope level.

were completed. Listeners needed to achieve at least five out of six correct responses in order to progress to the next task. If they did not, they were asked to reconnect the headphones and the test was repeated. Following the second equipment check, audiograms were collected.

Next, listeners completed familiarization and a series of pre-tests. Listeners were first presented with vocoded speech and listened to any word(s) as many times as desired. A grid with the 15 stimuli in the corpus appeared on the screen. Listeners could play any word as many times as desired to the right ear. Stimuli were vocoded with 100% dynamic range. This step was completed in order to gain some familiarity with vocoded speech. Next, listeners were given a test where different tokens of the 15 words were presented one time each simultaneously to both ears and their task was to choose the word presented (i.e., 15 alternative, forced-choice). When the same word was presented to both ears, different tokens (i.e., productions) were used in the left and right ear so that listeners could not capitalize on arbitrary similarities due to using the same speech recording. If the same token had been used, then listeners may have been able to rely on similarities that do not

reflect the typical variability associated with different speech productions present for all other pairs of stimuli. When different words were presented to the two ears, the token was chosen at random. Because they were also processed separately by the vocoder, the interaural correlation of the carriers in each frequency band was 0, resulting in a more diffuse sound image than if the interaural correlation were 1 (Whitmer et al., 2014). The word could not be repeated, and listeners initiated the next trial when they were ready. A minimum criterion of 10 (out of 15) correct was enforced before listeners progressed to the next task. No feedback was provided during or following testing. Next, the same test was given for stimuli vocoded with 40% dynamic range. No minimum criterion was established. Instead, the goal was simply for listeners to gain exposure to the easiest and most challenging stimuli presented during the experiment. The final pre-test consisted of pairs of stimuli with either the same word presented to each ear (10 trials) or words with different vowels presented to each ear (10 trials) using unprocessed (clean) speech. A minimum criterion of four one-word responses and four two-word responses was enforced before listeners progressed. If listeners failed to meet any criteria, they simply repeated the test until they successfully met the criteria.

Finally, listeners began experimental trials. Before each experimental block, listeners were informed that a longer block of testing was about to begin and that they could take a break if necessary. In experimental blocks, three types of trials were presented: (1) the same word using different tokens, (2) two words with different vowels, or (3) two rhyming words. Over the course of the experiment each word was tested 10 times in the "same word" trials ($n$ = 150) and each possible pairing was tested in the different vowel trials ($n$ = 150). When different vowels were presented, every possible combination of words was used (15 words × 10 words with different vowels = 150 combinations). The rhyming word trials consisted of two sub-types: phonological fusion and other trials. Phonological fusion trials consisted of a word beginning with a stop consonant and a word beginning with a liquid consonant, resulting in two pairs per set for the three sets, balanced so that each possible pairing was presented to the left and right ear, and each configuration repeated five times ($n$ = 2 × 3 × 2 × 5 = 60). Other trials consisted of non-fusible pairs of rhyming words, with eight other pairings in each of the three sets, balanced so that each possible pairing was presented to the left and right ear, and each possible configuration repeated two times ($n$ = 8 × 3 × 2 × 2 = 96). Thus, conditions with the same vowel contained a similar number of trials ($n$ = 156). As an example, phonological fusion pairs for the "bed" set were: bed and led; bed and red. Thus, there were six other possible combinations: bed and bled; bed and bred; led and red; led and bled; red and bled; red and bred.

The graphical user interface included the 15 possible words, and listeners chose the word(s) they perceived during the trial. They were required to choose at least one word and were not allowed to choose more than two words. Listeners revised their decision as many times as desired and initiated the next trial by selecting "Submit" on the experiment screen. Listeners were tested with the following stimulus processing conditions: unprocessed, 100, 60, and 40% interaurally symmetric dynamic ranges, and 100:60%, 100:40%, and 60:40% interaurally asymmetric dynamic ranges. The ear with the smaller dynamic range was counterbalanced across participants. Two listeners (one YNH and one ONH) were left-handed. In asymmetric conditions, both were tested with the larger dynamic range in the left ear.

Each block had an equal number of trials from each vocoder and word-pair condition, which consisted of 315 trials for the first nine blocks and 357 trials on the final block, resulting in a total of 3,192 trials. Testing was scheduled over a 4-h period and was able to be completed by most listeners during that time, including equipment assembly and disassembly. Chance performance in the task was 1/120 as there were 120 unique response combinations (105 combinations of two words and 15 single-word responses). One listener (ONH10) had to terminate the experiment during their final block of trials, with approximately 10% of trials remaining in the block. Their data were included and weighted according to the number of trials completed.[2] Another listener (ONH01) reported falling asleep multiple times during testing, so testing was completed over multiple days. On each trial, listeners could enter the reported words before submitting them and initiating the next trial. Thus, listener ONH01 could have entered their responses before falling asleep. Because their performance was not obviously worse than others, their data were also included.

## 2.4. Analysis

All analyses were completed using generalized (logit) linear mixed-effects analysis of variance (ANOVA) models. Version 3.5.1 of R was used with version 1.1–17 of the *lme4* package (Bates et al., 2015) to generate models and version 3.0-1 of the *lmerTest* package (Kuznetsova et al., 2017) to estimate degrees of freedom using the Kenward-Roger approximation (Kenward and Roger, 1997). Each model included a random effect associated with the listener and a fixed-effect of vocoder condition. This random effect allowed variation in mean performance due to difference between listeners to be accounted for in the model without being attributed to residual error. The ear receiving smaller dynamic range was excluded as a factor in the analysis, except in cases

2 Mixed-effects models are robust to some missing data if the data are missing at random, which can be a difficult criterion to meet. Because all trials were completely randomized within- and across-blocks, we can confirm that data were truly missing at random for ONH10 who had to terminate the experiment early.

where ear advantage and ear bias were analyzed. The dependent variable was either: Proportion correct, proportion of one-/two-word responses, or ear advantage. Paired and post-hoc comparisons were completed using estimated marginal means with Tukey adjustments for multiple comparisons using version 1.3.0 of the *emmeans* package in R (Lenth, 2022). For the sake of brevity, z-statistics are omitted and only *p*-values are reported for paired comparisons with significant results, with non-significant pairings noted. In this case, z-tests were used because of the large sample size within each individual, where the t- and standard normal distributions become equivalent. Analyses can be replicated with the data and code provided with the present manuscript. Results were organized according to: (1) Word pairing, (2) symmetric vs. asymmetric dynamic range, (3) age (YNH vs. ONH), (4) accuracy, (5) proportion of one- or two-word responses, and (6) ear advantage (where applicable). There were 182 possible paired comparisons within each sub-section (each vocoder condition for each age group). Thus, the order of paired comparisons was determined post-hoc for aid of readability and does not necessarily reflect hypotheses or predictions, but results are interpreted in terms of predictions. Data were analyzed within the same model for each dependent variable to minimize the risk of Type I error. Analyses were re-completed excluding listeners ONH01 (who fell asleep) and listeners ONH08 (who had measurable, estimated asymmetric hearing thresholds). Any differences from the original models are reported in the results section.

# 3. Results

The goal of the present experiment was to delineate the effects of binaural speech fusion and auditory attention in simulations of BiCIs with YNH and ONH listeners. We created interaurally symmetric and asymmetric conditions with varying dynamic range. We predicted that decreasing dynamic range would result in significantly decreased accuracy, and a significant main effect or interaction showing less accuracy for ONH listeners. We further predicted that in conditions when two words were presented, the proportion of two-word responses would significantly decrease as dynamic range decreased, with a significant main effect or interaction showing fewer two-word responses for ONH listeners. Finally, we predicted that ONH listeners would show significantly more right- or worse-ear responses when dynamic ranges were interaurally symmetric or asymmetric, respectively, compared with YNH listeners. The results are separated into three sections based upon the speech presented to the listener: Same word (section "3.1 Same word trials"), words with different vowels (section "3.2 Different vowel trials"), and phonological fusion pairs (section "3.3 Phonological fusion trials").

## 3.1. Same word trials

Speech identification accuracy is shown in **Figure 3**. Accurate responses were defined as those that included only one word and when the response matched the word presented. When the data were fit with the mixed-effects ANOVA, model diagnostics revealed substantial deviation from the assumption that residuals were normally distributed due to one outlying observation (listener ONH03 in the unprocessed condition). This observation was removed from analysis, which resolved the issue though this resulted in no change in conclusions. The results of the ANOVA demonstrated a significant effect of vocoder condition [$\chi^2(6) = 2705.502$, $p < 0.0001$], with smaller dynamic ranges resulting in less accuracy, consistent with the hypotheses. Age group was not significant [$\chi^2(1) = 0.268$, $p = 0.604$]. However, there was a significant vocoder condition × age group interaction [$\chi^2(6) = 46.421$, $p < 0.0001$] that is investigated further in the sections that follow. Confusion matrices describing errors in interaurally symmetric conditions are shown in **Supplementary Figure 2**. Two patterns are obvious: (1) Few vowel errors occurred, and (2) for 60 and 40% dynamic range, the most common confusion was reporting a stop-liquid cluster when a liquid was presented. For more details, see section "4.2 Ear advantage" of the discussion.

We wanted to rule out the possibility that, as dynamic range was decreased, listeners began to perceive multiple words. The proportion of one-word responses is also shown in **Figure 3**. Results of the ANOVA demonstrated a significant effect of vocoder condition [$\chi^2(6) = 781.616$, $p < 0.0001$] but not age group [$\chi^2(1) = 0.042$, $p = 0.838$] on percent correct. There was a significant vocoder condition × age group interaction [$\chi^2(6) = 31.669$, $p < 0.0001$].

### 3.1.1. Interaurally symmetric conditions

**Figure 3A** shows results from the interaurally symmetric conditions. Consistent with the hypotheses, the percentage of words correctly identified was significantly higher for the larger dynamic range in all pairs of symmetric vocoder conditions for both groups [$p < 0.05$–$0.0001$]. It was of interest to determine whether the vocoder condition × age group interaction was driven by differences between age groups at selected dynamic ranges. Pairwise comparisons with symmetric vocoder conditions showed no significant differences between YNH and ONH listeners in matched vocoder conditions, suggesting that the interaction was driven by the asymmetric conditions or differences in effects within groups.

Similarly, for proportion of one-word responses, post-hoc comparisons showed no significant differences between age groups in matched vocoder conditions. There were significant differences in proportion of one-word responses between all pairs of symmetric vocoder conditions for both groups

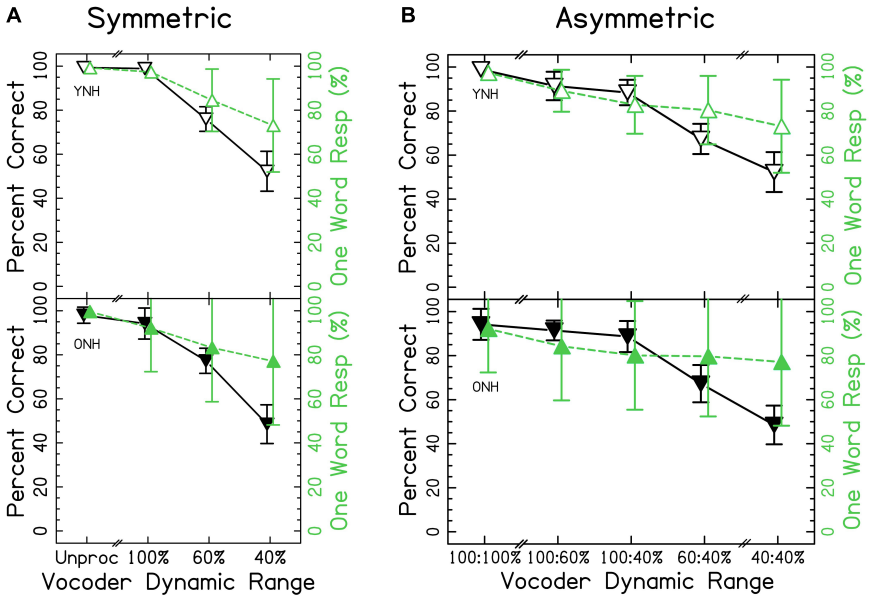



FIGURE 3
Single word accuracy and number of words responded for **(A)** interaurally symmetric, and **(B)** interaurally asymmetric vocoder conditions. The x-axis corresponds to the vocoder condition. The y-axis represents the percentage of trials with accurate, one-word responses (△ shown in black) and the percentage of one-word responses (▽ shown in green). Open and closed shapes represent YNH and ONH listeners, respectively.

[$p$ < 0.01–0.0001]. In all cases, the proportion of one-word responses was higher for the larger dynamic range.

### 3.1.2. Interaurally asymmetric conditions

When analyzing the interaurally asymmetric conditions, it was of interest to determine whether accuracy reflected maximum dynamic range (i.e., better ear), the mean dynamic range, the minimum dynamic range (i.e., worse ear), or the difference in dynamic ranges (i.e., degree of asymmetry). We predicted that accuracy would reflect the poorer ear or degree of asymmetry, with poorer performance for ONH listeners on average. For example, with 100:60% dynamic range, the maximum was 100%, the mean was 80%, the minimum was 60%, and the difference was 40%. **Figure 3B** shows accuracy for the interaurally asymmetric conditions, bounded by interaurally symmetric conditions with the largest and smallest dynamic ranges. Contrary to primary and alternative hypotheses, results support the notion that accuracy reflected the mean dynamic range between ears. Paired comparisons revealed differences in the level of significance between YNH and ONH listeners. In the YNH group, speech identification accuracy was significantly higher for the larger mean dynamic range [$p$ < 0.05–0.0001]. In the ONH group, there was no significant difference between the 100:40% compared to 60:60% [$p$ = 0.446] and the 100:60% compared to the 100:40% [$p$ = 0.128] conditions, but all others were significant. In other words, while the overall patterns were the same, differences between asymmetric conditions tended to be less pronounced for ONH listeners.

### 3.1.3. Summary

These results suggest that speech identification, and to a lesser extent fusion, of the same word reflect the mean dynamic range across-ears (e.g., mean of 100 and 40% is 70%), in disagreement with our predictions that the worse ear or degree of asymmetry would predict accuracy. In further disagreement with our hypotheses, there was no consistent effect of age group when comparing at the same dynamic ranges between groups. By definition, if a listener responded with two words, their response was scored as incorrect. Thus, the highest level of accuracy was defined by the proportion of one-word responses. Based on the analysis, **Figure 3**, and **Supplementary Figure 2**, the results suggest that decreases in accuracy did not strictly reflect reporting more words, rather that listeners made systematic errors.

## 3.2. Different vowel trials

In sections "3.2 Different vowel trials" and "3.3 Phonological fusion trials," different words were presented to each ear. Speech identification accuracy is shown in **Figure 4**. In this case, the accuracy represents the probability of correctly reporting at least one word. We predicted that accuracy would decrease as the dynamic range decreased for both groups. Presenting two words simultaneously was more cognitively demanding.

Accordingly, we predicted that ONH listeners would show poorer accuracy than YNH listeners. This increased difficulty was more likely to elicit ear advantage, where attention to one ear was prioritized. Thus, our definition of accuracy was formulated to determine whether information in at least one ear was preserved. Additionally, if listeners were able to listen to each ear independently, then the accuracy (at least one correct) in section "3.2 Different vowel trials" should be equal to the accuracy in section "3.1 Same word trials." The ANOVA revealed significant effects of vocoder condition [$\chi^2(6) = 2477.006$, $p < 0.0001$] where smaller dynamic ranges resulted in less accuracy, consistent with our hypotheses. There was also a significant effect of age group [$\chi^2(1) = 5.104$, $p < 0.05$], with ONH listeners having a lower percentage of correct responses consistent with our hypotheses. There was also a significant vocoder condition × age group interaction [$\chi^2(6) = 55.975$, $p < 0.0001$] that is explored further in the next sections. Vowel confusion matrices describing errors in interaurally symmetric conditions are shown in **Supplementary Figure 3**. It is important to note that the effect of age changed from $p = 0.040$–$0.085$ if listeners ONH01 (who fell asleep) and listener ONH08 (who had asymmetric, estimated hearing thresholds) were excluded from analysis. No other changes in statistical inference occurred.

We wanted to titrate the types of errors made in each symmetric vocoder condition. There were a total of 75 two-word response combinations as well as 15 one-word response possibilities. Thus, a confusion matrix would be difficult to show with every possible combination. Instead, capitalizing on the small number of vowel errors made with single word trials, **Supplementary Figure 3** shows vowel confusion matrices for interaurally symmetric vocoder trials. There were only three possible vowel combinations on each trial, but single vowel responses were also considered. As can be seen from this figure, vowel errors were very rare. When listeners reported a single vowel, this usually corresponded to one of the vowels presented in one ear. The /ɛ/ and /eI/ pairs were the most likely to result in singular vowel responses. This may have to do with the fact that the /ɛ/ set had an additional /d/ cue at the end of each word.

Compared with single word trials, there was a much higher correspondence between the number of words reported and the accuracy in reporting at least one word correct, consistent with our hypotheses that decreased dynamic range would result in fusion. When percentage of two-word responses was the dependent variable in the ANOVA, there was a significant effect of vocoder condition [$\chi^2(6) = 1658.486$, $p < 0.0001$], with smaller dynamic ranges resulting in greater one-word responses consistent with our hypotheses. There was no effect of age group [$\chi^2(1) = 1.194$, $p = 0.274$], inconsistent with our hypotheses. There was a significant vocoder condition × age group interaction [$\chi^2(6) = 21.421$, $p < 0.01$], which is explored further in the sections "3.2.1 Interaurally symmetric conditions" and "3.2.2 Interaurally asymmetric conditions."

### 3.2.1. Interaurally symmetric conditions

**Figure 4A** shows results from the interaurally symmetric conditions. Consistent with the hypotheses, the percentage of trials with at least one word correctly identified was significantly higher for the greater dynamic range between all pairs of symmetric vocoder conditions for both groups [$p < 0.01$–$0.0001$], except for the unprocessed and 100% dynamic range conditions for ONH listeners [$p = 0.540$]. Further consistent with our hypotheses, pairwise comparisons with symmetric vocoder conditions showed that speech identification accuracy was significantly greater for YNH compared to ONH listeners in the unprocessed [$p < 0.0001$] and 100% [$p < 0.01$] conditions, but not the 60% [$p = 0.975$] and 40% [$p = 0.584$] conditions.

Similar to identification accuracy and consistent with our hypotheses, the proportion of two-word responses was significantly higher for the larger dynamic range in all pairs of symmetric vocoder conditions for both groups [$p < 0.001$–$0.0001$], except for the unprocessed and 100% dynamic range conditions for ONH listeners [$p = 0.563$]. For proportion of two-word responses and inconsistent with our hypotheses, pairwise comparisons showed no significant differences between groups in matched vocoder conditions.

### 3.2.2. Interaurally asymmetric conditions

**Figure 4B** shows accuracy for the interaurally asymmetric conditions, bounded by interaurally symmetric conditions with the largest and smallest dynamic ranges. Results support that accuracy was reflected by the mean dynamic range between ears, inconsistent with our hypotheses that the poorer ear or degree of asymmetry would predict performance. Pairwise comparisons revealed differences in the level of significance between YNH and ONH listeners. In both groups, most speech identification accuracy was significantly greater for the higher mean dynamic range [$p < 0.05$–$0.0001$]. In the ONH group, there was no significant difference between the 100:40% compared to 60:60% [$p = 0.109$] conditions.

The proportion of two-word responses was similar to that observed for speech identification accuracy, with some slight differences between the YNH and ONH groups. In both groups, the proportion of two-word responses was significantly greater for the higher mean dynamic range [$p < 0.05$–$0.0001$], except for the 100:60% compared to 60:60% [$p = 0.288$] in YNH listeners and 100:40% compared to 60:40% [$p = 0.077$] conditions in ONH listeners. There was one interesting exception to this pattern. For both the YNH and ONH groups, the proportion of two-word responses was significantly greater for the 60:60% compared to 100:40% [$p < 0.0001$] conditions. This suggests that, listeners reported one word in the cases with the largest amount of asymmetry (100:40%) compared to when stimuli were symmetric and poorly represented (60:60%), inconsistent with our hypotheses.

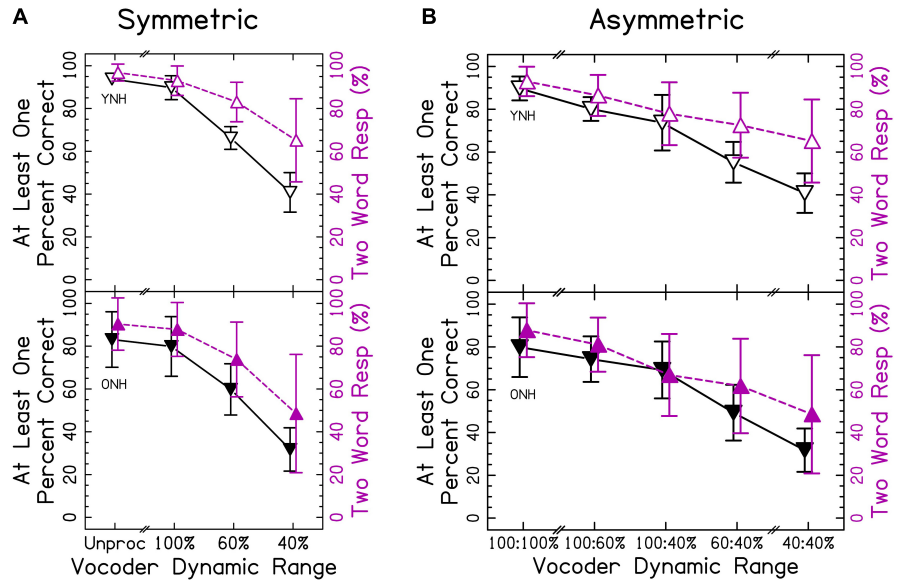



FIGURE 4
Different vowel accuracy and number of words responded for **(A)** interaurally symmetric, and **(B)** interaurally asymmetric vocoding conditions. The x-axis corresponds to the vocoder condition. The y-axis represents the percentage of trials with at least one word accurately identified (△ shown in black) and the percentage of two-word responses (▽ shown in purple). Open and closed shapes represent YNH and ONH listeners, respectively.

### 3.2.3. Ear advantage

Figure 4 suggests that there was a strong correspondence between speech identification accuracy and proportion of two-word responses, despite accuracy being based upon correctly reporting the word presented to either ear. Additionally, Supplementary Figures 2, 3 demonstrate that listeners were unlikely to make a vowel error even with small dynamic ranges. Thus, the vowel reported likely corresponds to the ear to which the listener was allocating attention. The left- and right-ear advantage was explored by evaluating the proportion of vowels reported from the left or right ear when only one word was reported. We hypothesized that listeners would show a right-ear advantage in symmetric conditions that increased with decreasing dynamic range, a better-ear advantage in asymmetric conditions, and that right- or worse-ear advantage would be greater in ONH compared to YNH listeners.

Results from Figure 5A suggest a modest right-ear advantage across interaurally symmetric vocoder conditions, with the smallest dynamic ranges resulting in the greatest ear advantages. A mixed-effects ANOVA revealed significant fixed-effects of ear [$\chi^2(1) = 196.565$, $p < 0.0001$], with a greater proportion of right ear responses consistent with our hypotheses. There was also a significant effect of vocoder condition [$\chi^2(3) = 1329.536$, $p < 0.0001$], with smaller dynamic ranges resulting in increased ear advantage responses consistent with our hypotheses. There was also a significant effect of age group [$\chi^2(1) = 6.018$, $p < 0.05$], with ONH listeners showing greater ear advantage consistent with our hypotheses. There was also a significant vocoder condition × age group interaction [$\chi^2(6) = 19.436$, $p < 0.001$]. Pairwise comparisons showed that ONH listeners had significantly more responses compared to YNH listeners in only the unprocessed condition [$p < 0.0001$]. There was no significant ear × vocoder condition [$\chi^2(3) = 4.938$, $p = 0.176$], ear × age group [$\chi^2(1) = 3.825$, $p = 0.050$], or three-way [$\chi^2(3) = 2.528$, $p = 0.470$] interactions. The model with dependent variable proportion correct did not converge if listeners ONH01 and ONH08 were excluded from analysis.

Figures 5B, C show the same results for interaurally asymmetric vocoder conditions, with the better (i.e., ear with larger dynamic range) or worse ear. In order to analyze the data, the factor "ear" was re-coded from left or right to better or worse. A mixed-effects ANOVA revealed significant fixed-effects of ear [$\chi^2(1) = 856.650$, $p < 0.0001$], with a higher proportion of better-ear responses consistent with the hypotheses. There was also a significant effect of vocoder condition [$\chi^2(2) = 287.988$, $p < 0.0001$], with smaller dynamic ranges resulting in larger ear advantage consistent with our hypotheses. There was not a significant effect of age group [$\chi^2(1) = 3.419$, $p = 0.064$], inconsistent with our hypotheses. There were significant ear × vocoder condition [$\chi^2(2) = 82.685$, $p < 0.0001$] and ear × age group [$\chi^2(1) = 11.016$, $p < 0.001$] interactions, which will be addressed in the next paragraph. Vocoder condition × age group [$\chi^2(2) = 0.730$, $p = 0.694$] and three-way [$\chi^2(2) = 0.196$, $p = 0.906$] interactions were not significant.

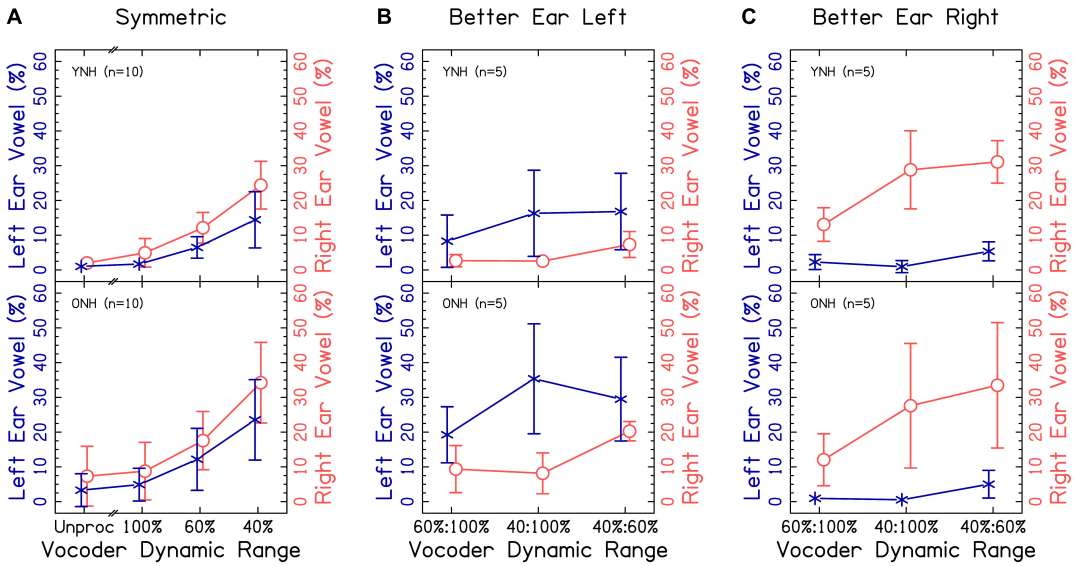



FIGURE 5
Vowel responses by ear for different vowel trials for **(A)** interaurally symmetric or **(B,C)** interaurally asymmetric vocoder conditions. The x-axis corresponds to the vocoder condition. The y-axis corresponds to the percentage of trials where the vowel of the response was one word and came from the left (× shown in blue) or right (◯ shown in orange).

Pairwise comparisons revealed that there were significantly more better-ear responses in the 100:40% compared to 100:60% [$p < 0.0001$] and 60:40% compared to 100:60% [$p < 0.0001$], but not the 60:40% compared to 100:40% [$p = 0.983$] conditions. In contrast, there were significantly more worse-ear responses in the 60:40% compared to 100:60% [$p < 0.0001$] and 60:40% compared to 100:40% [$p < 0.0001$] conditions, but not between the 100:40% and 100:60% [$p = 0.507$] conditions, which we did not predict. The ONH group had significantly more worse-ear responses than the YNH group [$p < 0.05$], but no difference between the better-ear responses [$p = 0.396$], consistent with our hypotheses. There were significantly more better-ear compared to worse ear responses [$p < 0.0001$] in all three vocoder conditions.

### 3.2.4. Summary

The degree of accuracy and proportion of two-word responses decreased in a similar fashion as dynamic range decreased, consistent with our hypotheses. Accuracy was lower for ONH compared to YNH listeners in the unprocessed and 100:100% dynamic range condition, but the proportion of two-word responses was not different between groups. These results suggest that speech identification and fusion of the words with different vowels reflect the average dynamic range across the ears. They further suggest that the accuracy of speech identification might be mediated by fusion in the larger dynamic range conditions. When dynamic range was small in one ear (60 and/or 40%), the correspondence between accuracy and proportion of two-word responses was less strong but still present. This implies that listeners were more likely to report hearing one word, and more likely to have that word be inaccurate, when speech was degraded in one or both ears. This is intuitive, since an effective strategy may be to ignore the poorer ear. The results demonstrate that poor dynamic range in general impairs access to speech in both ears (evidenced by significantly poorer speech understanding for 100:60% and 100:40% compared to 100:100%). Thus, it was of particular interest to explore the probability of responding with a correct response in the left or right ear. In contrast to our hypotheses, there were only significantly more right-ear responses for ONH listeners in the unprocessed condition. Notably, there were significantly more *worse* ear responses for ONH listeners when dynamic range was asymmetric. This result may reflect an inability to ignore the right ear for ONH listeners, even when it has a smaller dynamic range. Additionally, the results suggest that interaurally asymmetric dynamic range interacts with right-ear advantage, with larger right-ear advantage when the right ear has greater dynamic range (**Figures 5B, C** though this was not tested statistically).

## 3.3. Phonological fusion trials

**Figure 6A** shows an example trial from a phonological fusion trial and how responses were scored. Ideal responses required listeners to respond with two words, with both words correct. Phonologically fused responses required listeners to respond with one word containing the stop-liquid cluster of the stop and liquid presented. Biased left required listeners to respond with one or two words, with only the word

from the left ear correct. The same was true for biased right except that the word reported was presented to the right ear. Interference responses required listeners to respond with one or two words corresponding neither to the word presented in the left, right, or phonologically fused word. These possibilities can be viewed as a continuum from best-case (independent channels to which attention can be allocated or ideally linked channels where information is shared) to worse-case scenario (interfering channels to which attention cannot be effectively allocated). Figures 6B–D shows responses from participants. We predicted that the proportion of ideal responses would be highest for the largest dynamic ranges, lowest for the smallest dynamic ranges, and the worse ear would predict ideal and interference responses in asymmetric conditions. We further predicted that the proportion of interference responses would increase as dynamic range decreased. We predicted that asymmetric dynamic ranges would result in increased responses biased toward the better ear. Finally, we predicted that ONH listeners would exhibit fewer ideal, more interference, and more biased responses compared to YNH listeners. Listeners responded with one word on 0–100% of trials, with a mean of 58% and standard deviation of 26%. While the original studies on phonological fusion (Cutting, 1975, 1976) showed phonological fusion responses (i.e., one word with a stop and liquid cluster) in approximately 30% of trials, the relative frequency in the present study was lower. This will be explored further in the discussion. Data were analyzed in two separate sections addressing: (1) the proportion of ideal and interference responses, representing the best- and worst-case scenarios for listeners, and (2) the relative bias toward responding correctly from the left and right (symmetric) or better and worse (asymmetric) ears.

### 3.3.1. Ideal and interference responses

When data were fit with the mixed-effects ANOVAs, residuals were curvilinear and not normally distributed. Diagnostic plots suggested that the residuals may have been Cauchy distributed for the proportion of ideal responses. When a cauchit (rather than logit) ANOVA was used based on the pattern of residuals post-hoc, the model estimating ideal responses improved considerably, resulting in normally distributed residuals and the removal of curvilinearity. Including a random effect of vocoder condition led to normally distributed residuals for the proportion of interference responses. There was still some slight overestimation at the smallest proportions of ideal and interference responses. In both cases, this could have led to decreased power.

The results of the cauchit ANOVA demonstrated significant effects of vocoder condition [$\chi^2(6) = 458.372$, $p < 0.0001$], with smaller dynamic ranges resulting in fewer ideal responses consistent with the hypotheses. There was no significant effect of age group [$\chi^2(1) = 2.490$, $p = 0.115$], inconsistent with our hypotheses. There was a significant vocoder condition × group interaction [$\chi^2(6) = 43.4445$, $p < 0.0001$], which is addressed in the next sections. A logit ANOVA including random effect of vocoder condition showed significant effects of vocoder condition [$\chi^2(6) = 81.024$, $p < 0.0001$], with smaller dynamic ranges resulting in more interference responses consistent with the hypotheses. There was also a significant effect of age group [$\chi^2(1) = 5.717$, $p < 0.05$], with ONH listeners demonstrating more interference responses consistent with the hypotheses. There was no significant vocoder condition × group interaction [$\chi^2(6) = 9.784$, $p = 0.134$]. The model with dependent variable interference did not converge if listeners ONH01 and ONH08 were excluded from analysis.

Interaurally symmetric conditions are shown in Figure 6B. The proportion of ideal responses was not significantly different between YNH or ONH in any of the symmetric vocoder conditions, inconsistent with the hypotheses. For YNH listeners, there were significant differences between all vocoder conditions [$p < 0.01$–$0.0001$] except 60 and 40% dynamic range [$p = 0.211$]. For ONH listeners, there were significant differences between all vocoder conditions [$p < 0.05$–$0.0001$], except unprocessed and 100% dynamic range [$p = 0.527$]. The proportion of interference responses significantly increased as dynamic range decreased in both age groups [$p < 0.01$–$0.0001$], consistent with the hypotheses. There was no difference between the unprocessed and 100% dynamic range condition [$p = 0.346$]. The proportion of interference responses was significantly higher for ONH compared to YNH listeners, consistent with the hypotheses.

Figures 6C, D show the proportion of ideal and interference responses in asymmetric conditions, bounded by interaurally symmetric conditions with the largest and smallest dynamic ranges. In general, the worse ear was predictive of the proportion of ideal responses, consistent with the hypotheses. For YNH listeners, proportion of ideal responses was lowest when at least one ear had 40% dynamic range. There were no significant differences between all conditions where one ear had 40% dynamic range. There was also no significant difference between the 100:40 and 60:60% dynamic range conditions [$p = 1.000$]. For greater dynamic ranges (60–100%), the proportion of ideal responses was highest for conditions with the greatest mean dynamic range [$p < 0.05$–$0.0001$]. For ONH listeners, there no difference between all vocoder conditions with equivalent minimum dynamic ranges, consistent with the hypotheses. There were two exceptions to this, (1) the 100:60% dynamic range led to significantly more ideal responses than the 60:60% [$p < 0.01$] condition, and (2) there was no difference between the 60:60% and 100:40% [$p = 0.464$] conditions. The better ear was predictive of the proportion of interference responses, inconsistent with the hypotheses. Accordingly, there were no significant differences between 100:100% and 100:60% [$p = 0.987$],

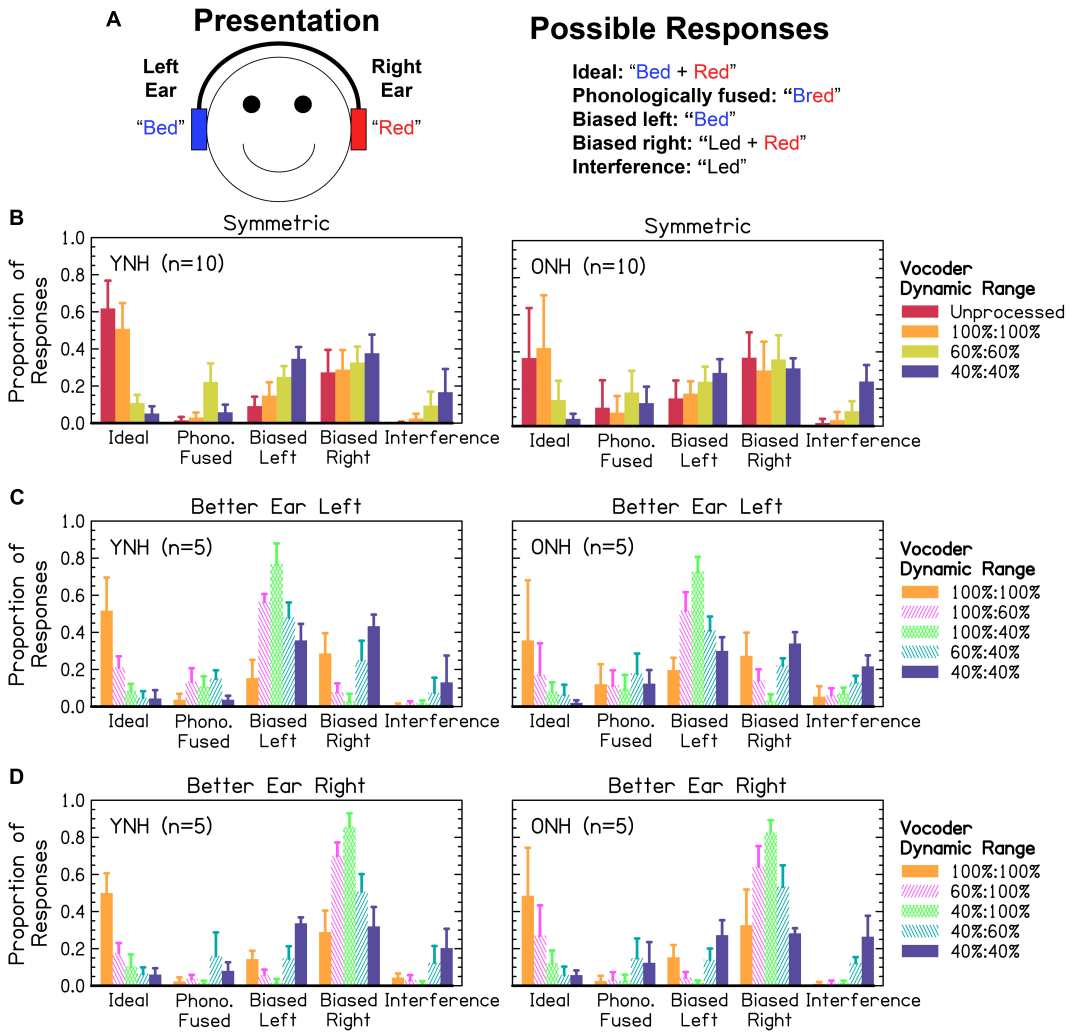



FIGURE 6
**(A)** Response categories for one example phonological fusion trial. **(B–D)** Relative frequency of response categories observed by listeners in **(B)** symmetric and **(C,D)** asymmetric dynamic ranges. The x-axis corresponds to the response category from **(A)**. The y-axis corresponds to the proportion of responses. The color and pattern represent the vocoder condition given in the figure legend and arranged from highest to lowest mean dynamic range. The height of bars represents the mean across listeners. Error bars represent standard deviation of the mean across listeners.

100:100% and 100:40% [$p$ = 1.000], 100:60% and 100:40% [$p$ = 0.994], or 60:60% and 60:40% [$p$ = 0.747]. All other pairs (excluding unprocessed) were significantly different [$p < 0.01$–$0.0001$].

### 3.3.2. Ear bias

Results from Figure 6B show a right ear bias across interaurally symmetric vocoder conditions, with the largest dynamic ranges resulting in the greatest ear bias, inconsistent with the hypotheses. A mixed-effects ANOVA revealed significant fixed-effects of ear [$\chi^2(1) = 150.792$, $p < 0.0001$], with the right ear resulting in a greater proportion of responses consistent with the hypotheses. There was also a significant effect of vocoder condition [$\chi^2(3) = 96.242$, $p < 0.0001$]. The left ear bias increased as dynamic range decreased [$p < 0.05$–$0.0001$], where the right ear bias was not significantly different between any pairs of conditions inconsistent with the hypotheses. This was reflected by a significant ear × vocoder condition interaction [$\chi^2(3) = 69.913$, $p < 0.0001$]. There was not a significant effect of age group [$\chi^2(1) = 0.247$, $p = 0.619$]. However, there was a significant vocoder condition × age group interaction [$\chi^2(3) = 31.953$, $p < 0.0001$]. Older NH compared to YNH listeners had significantly greater bias responses in the unprocessed [$p < 0.05$] but no other conditions, inconsistent the hypotheses. Ear × age group [$\chi^2(1) = 0.002$, $p = 0.962$] and three-way [$\chi^2(3) = 1.904$, $p = 0.593$] interactions were not significant. For YNH listeners, smaller dynamic ranges resulted in significantly greater bias

responses [$p < 0.01$–$0.0001$] except for the unprocessed compared to 100% [$p = 0.098$] condition. For ONH listeners, bias was only significantly greater in the 40% compared to 100% [$p < 0.01$] and 60% compared 100% [$p < 0.01$] conditions.

Results from **Figures 6C, D** show a strong bias toward the better ear, especially in the cases of largest asymmetry. The mixed-effects ANOVA revealed a significant fixed-effect of ear [$\chi^2(1) = 1070.238$, $p < 0.0001$], with a higher proportion of better ear bias consistent with the hypotheses. There was also a significant effect of vocoder condition [$\chi^2(2) = 91.005$, $p < 0.0001$], with greater degree of asymmetry predicting the portion of better ear bias, inconsistent with the hypotheses. There was not a significant effect of age group [$\chi^2(1) = 1.936$, $p = 0.164$], inconsistent with the hypotheses. There was a significant ear × vocoder condition interaction [$\chi^2(2) = 302.531$, $p < 0.0001$], reflected in the increased bias toward the better ear with increasing asymmetry inconsistent with our hypotheses. There were not significant ear × age group [$\chi^2(1) = 2.086$, $p = 0.149$] or vocoder condition × age group [$\chi^2(2) = 3.821$, $p = 0.148$] interactions, further suggesting no differences between age groups and inconsistent with the hypotheses. Paired comparisons revealed that all pairs of ear and vocoder conditions were significantly different [$p < 0.0001$]. Most better-ear responses occurred for the 100:40%, followed by the 100:60%, followed by the 60:40% conditions. The greatest amount of worse-ear responses occurred for the 60:40%, followed by the 100:60%, followed by the 100:40% conditions.

### 3.3.3. Summary

Consistent with our hypotheses, when dynamic range was symmetric, listeners tended to report at least one word incorrectly as dynamic range decreased (i.e., fewer "ideal" responses), shifting their responses to the word presented in the left ear or a word that was not presented (i.e., an "interference" response). Responses from the word in the right ear remained consistent across conditions. When dynamic range was asymmetric, opposite and opposing effects were observed for ideal versus interference and better-ear versus worse-ear responses. The proportion of ideal responses decreased as the dynamic range in the worse ear decreased. The proportion of interference responses decreased as the dynamic range in the better ear increased. Better-ear responses increased and worse-ear responses decreased as the degree of interaural asymmetry increased, consistent with the notion that listeners attended to the better ear when dynamic range was asymmetric. We further hypothesized that ONH listeners would exhibit more interference responses, which was confirmed by the results, and more ear bias, which was refuted by the results. Like different vowel trials, ONH listeners only exhibited greater bias toward the right ear compared to YNH listeners in the unprocessed condition, which was inconsistent with our hypotheses.

## 4. Discussion

Patients with BiCIs often experience substantial differences in hearing outcomes between their ears. While bilateral implantation generally improves speech understanding in noise relative to unilateral implantation, there are some conditions under which bilateral hearing is not beneficial, and listeners experience contralateral interference from the better ear (Bernstein et al., 2016, 2020; Goupell et al., 2016, 2018). Thus, addressing the mechanisms leading to contralateral interference may play a key role in maximizing bilateral outcomes. Two putative mechanisms have been proposed in the literature related to the basis of contralateral interference. The first supposes that it results from a failure of sound source segregation cues to form distinct auditory objects (Gallun et al., 2007; Reiss et al., 2016; Reiss and Molis, 2021). The second supposes that contralateral interference results from a failure to allocate attention away from the source with greatest spectro-temporal fidelity and toward a more degraded sound source (Goupell et al., 2016, 2021).

Results from the present experiment suggest that poor segregation of sound sources and compromised auditory spatial attention both play a role and may interact with one another when temporal information is symmetrically or asymmetrically degraded. Listeners showed an increased number of one-word responses when words with dichotic vowels were presented as stimuli became increasingly symmetrically or asymmetrically degraded. As stimuli became increasingly degraded in one or both ears, listeners also tended to report inaccurate word(s). Inaccuracies occurred despite evidence that listeners shifted attention toward the right or left ear when stimuli were symmetrically degraded and toward the better ear when stimuli were asymmetrically degraded. In particular, the present experiment challenges the assumption of classical theories in binaural hearing that the ears act as independent channels. If both ears were independent, then listeners should have attended to the better ear in asymmetric dynamic conditions and correctly reported that word. This would have resulted in similar accuracy in sections "3.1 Same word trials" and "3.2 Different vowel trials," and no interference responses in section "3.3 Phonological fusion trials." Instead, listeners showed substantially poorer accuracy in section "3.2 Different vowel trials" compared to section "3.1 Same word trials" and monotonically increasing interference as dynamic range decreased in section "3.3 Phonological fusion trials." Moreover, listeners showed marked right-ear advantage when stimuli were symmetrically degraded.

## 4.1. Object-based auditory attention

Auditory attention is thought to be a process with serial and parallel stages (Bregman, 1994; Shinn-Cunningham, 2008).

First, brief and similar spectral components are grouped into auditory streams. These streams compete for attention over longer periods of time, and top-down attention is allocated to a source of interest via source segregation cues. Finally, information (e.g., language content) can be extracted from sources of interest. This is a simpler process for stimuli that occur in quiet with an unobscured onset and offset time. However, real-world listening often occurs in complex auditory environments with competing background noise and ambiguous onset and offset times. It is suspected that listeners maintain an internal perceptual model that can be updated according to new sensory information to efficiently and robustly complete this process (e.g., Rao and Ballard, 1999). Increasingly complex stimulus features are suspected to be extracted at later stages of sensory processing. These features are the product of interactions between internal predictions and incoming sensory input. Thus, for listeners who receive compromised or ambiguous sound source segregation cues, it is likely that the ability to maintain internal predictions, represent sound features, and allocate attention is compromised.

Listeners with BiCIs show patterns of behavior indicating that they might fuse unrelated auditory information. For example, listeners perceive a singular pitch percept over a large disparity of electrodes between ears, corresponding to frequency differences up to one octave (Reiss et al., 2018). Abnormally large fusion ranges for interaural place-of-stimulation differences have been proposed by Reiss et al. (2015) as an adaptive process associated with the large degrees of mismatch due to differences in degree of insertion for listeners with BiCIs (Goupell et al., 2022). Simulations in NH suggest that interaural mismatches in place-of-stimulation result in poorer spatial fusion (Goupell et al., 2013) and speech fusion (Aronoff et al., 2015; Staisloff et al., 2016). Accordingly, listeners with BiCIs perceive a singular spatial image (as opposed to multiple perceived locations) over large interaural electrode disparities, even as the impact of binaural cues on perceived intracranial perception decreases (van Hoesel and Clark, 1995, 1997; Long et al., 2003; Kan et al., 2013, 2019). Listeners with BiCIs will fuse stimuli presented with interaural timing differences up to several milliseconds (van Hoesel and Clark, 1995, 1997) and very large interaural timing differences (∼2 ms) are needed to achieve maximum intracranial lateralization (Litovsky et al., 2010; Baumgärtel et al., 2017; Anderson et al., 2019a).

It has been suggested that listeners with hearing loss struggle in noise because of an inability to accurately segregate speech sounds, leading them to fuse unrelated words (e.g., Reiss et al., 2016; Reiss and Molis, 2021; Eddolls et al., 2022). Results from the present study support these findings. In particular, there was a high correspondence between the proportion of two-word responses and accuracy reporting at least one word correctly when vowels differed (**Figure 4**) as well as a higher proportion of interference (**Figure 6**) when stimuli were temporally degraded. The present study used less conservative criteria for accuracy compared to other fusion studies using synthesized vowels (Reiss and Molis, 2021; Eddolls et al., 2022). While the present study always presented speech bilaterally, we can draw some conclusions from the symmetric versus asymmetric conditions. In particular, results from listeners with hearing loss have shown that stimulating each ear individually results in a different pattern of performance compared to bilateral stimulation for some listeners, presumably because of underlying differences between ears or poor overall representations of spectro-temporal cues (Reiss et al., 2016). The present study showed that small, symmetric or asymmetric dynamic range resulted in a decrease of accuracy in reporting at least one word correct when two were presented compared to symmetric conditions with only one word. This suggests that dichotic speech with poor dynamic range leads to different perception than diotic or monotic (monaurally presented) speech, consistent with the findings in listeners with hearing loss (Reiss et al., 2016).

In comparison to previous work concerning phonological fusion (Cutting, 1975, 1976), our results showed some differences. We showed fewer fusion responses than observed previously [∼10% in the present study and ∼30% with natural speech in (Cutting, 1975)]. This may have to do with differences in the tasks. For example, the present experiment was closed-set and took place over many hours. Additional stimuli were included (e.g., same word and different vowel trials). Independently generated speech tokens were used in same word trials, and the interaural correlation of noise carriers used to generate vocoded stimuli was always zero. In the original phonological fusion experiments, stimuli were played continuously via tape and participants responded aloud in a restricted period of time. It was not described whether other word pairs were used. Subsequent studies used artificially generated speech, which would have maximized similarity between ears. Interestingly, the present results showed similar proportion of one-word responses to previous studies [58 ± 26% compared to ∼70% in (Cutting, 1975)]. Results in sections "3.2 Different vowel trials" and "3.3 Phonological fusion trials" suggest that this may be explained by shifts in attention toward one ear, leading to a one-word response. Additional data containing rhyming words are provided in **Supplementary Figure 4** and suggest that the proportion of one-word responses does not change substantially across any of the vocoder conditions.

Fusion experiments generally ask listeners to report the number of sounds perceived. One alternative approach is to assess a listener's ability to discriminate between sounds suspected to be segregated. In this case, segregation is not necessary to complete the task. This type of procedure may be more sensitive to the continuum between the perception of one versus two clear and coherent objects, where the midpoint between these possibilities is one distorted or perceptually diffuse object (e.g., Suneel et al., 2017). Using a one-interval, two-alternative forced-choice task where listeners were asked

to identify whether envelope fluctuation rates were equal, Anderson et al. (2019b) showed that YNH listeners' sensitivity to rate differences decreased when the dynamic range in one ear was reduced. The authors showed listener-dependent differences in bias of responses toward "same" or "different." This result implies a task-relevant bias of listeners. That is, if listeners are asked to respond with one or two sounds, they might be more biased toward responding one way based upon the task and stimulus statistics. Then the high degrees of fusion in experiments might simply be indicative of poorer perceptual boundaries between features of the target and masking stimulus rather than a likelihood of perceiving one auditory object. Results from the present experiment and Anderson et al. (2019b) in YNH listeners as well as others (Ihlefeld et al., 2015; Todd et al., 2017; Anderson et al., 2022) in listeners with BiCIs suggest that temporal degradation in one ear is sufficient to interfere with cues used to segregate sound sources. In particular, the present study showed that the proportion of one-word responses decreased when the same word was presented and the proportion of two-word responses decreased when words with different vowels were presented. While the latter was a stronger effect, the present study suggests that the boundary between one and two sounds becomes less clear when stimuli are temporally degraded.

The present study investigated the effects of reduced dynamic range on speech identification. Reduced dynamic range is associated with poorer speech understanding in listeners with BiCIs (Firszt et al., 2002; Spahr et al., 2007) and simulations in NH (Loizou et al., 2000). It is also associated with poorer sensitivity to spatial cues in listeners with BiCIs (Ihlefeld et al., 2014; Todd et al., 2017) and listeners with NH (Bernstein and Trahiotis, 2011; Anderson et al., 2019b). Reducing dynamic range is similar in spirit to the spectro-temporal smearing that is thought to occur in patients. Previous experiments addressing bilateral speech understanding and showing contralateral interference in simulations of CI processing have manipulated spectro-temporal fidelity by reducing the number of frequency bands (Gallun et al., 2007; Goupell et al., 2021). Reducing the number of channels simulates spread of current that can occur with a CI, albeit in a less realistic way than vocoders that explicitly simulate current spread (Oxenham and Kreft, 2014; Croghan and Smith, 2018). Increasing the number of maxima in peak-picking, N-of-M processing strategies beyond the eight typically used in clinical practice improves speech understanding (Croghan et al., 2017). Reducing the number of spectral channels below eight is highly unlikely to be used in practice, and it is unlikely that listeners with BiCIs would be presented with such different numbers of frequency channels. It is possible that the number of "effective" channels is different between ears. Recognizing that all vocoder experiments are highly artificial, reduced or asymmetric dynamic range may be an additional problem for patients.

Aging was also associated with poorer speech identification accuracy in the present study, consistent with previous studies concerning speech recognition of older listeners with NH and CIs in background noise (Moberly et al., 2017). Similarly, open-set speech understanding accuracy in noise, but not in quiet, decreases significantly with increasing age for listeners with BiCIs (Shader et al., 2020a). Accuracy in the present study was only worse for ONH listeners when two words were presented, suggesting that effects were driven primarily by increased cognitive demand. Aging and cognition have not been associated with poorer performance in temporally-based speech perception tasks (Roque et al., 2019). Interestingly, another study showed that aging effects were greatest when open-set speech in quiet was spectrally degraded, with no differences in temporal degradation (via lowpass filtering of the envelope) for YNH compared to ONH listeners with CI simulations (Shader et al., 2020b). Older NH listeners also show poorer sensitivity to temporal speech contrasts that are level- and spectral-degradation-dependent (Goupell et al., 2017; Xie et al., 2019), suggesting that some temporal aspects of speech representation vary depending upon age. Our results suggest that task difficulty may play a key role in the emergence of aging effects on temporally based degradations of speech understanding. It was been proposed that poorer speech understanding with increasing age in listeners with CIs may be the result of reduced access to temporal envelope cues (Anderson et al., 2012). Because there was only an age effect in the best dynamic range conditions, the present results suggest that either very temporally degraded envelopes impair perception for younger and older listeners, or that aging effects are primarily driven by well-represented envelopes. Said another way, aging effects on accuracy appeared when the task was sufficiently difficult and listeners had access to the temporal envelope. Because the present study used acute exposure to simulated CI processing, differences between age groups in the 100% dynamic range condition might reflect a similar mechanism involved with poorer speech understanding of listeners with CIs who receive their implant in older age (Blamey et al., 2012).

## 4.2. Ear advantage

One interesting finding in the present study was that ear advantage, or the ear to which listeners attended, was modulated by asymmetries in dynamic range. This is consistent with the results of Goupell et al. (2021). Under symmetric conditions, listeners with NH tend to show modest effects of right-ear advantage, evidenced by greater accuracy or higher probability of reporting speech presented to the right ear compared to the left ear. This effect tends to become exaggerated as listeners get older (Westerhausen et al., 2015), though the present study only replicated this result in a subset of conditions. Thus, peripheral

(e.g., temporal degradation) and central (e.g., attention-based changes with age) mechanisms may play a role in ear advantage.

Two classical hypotheses have been proposed associated with right-ear advantage, based on structural biases in the left hemisphere (Kimura, 1967) or biased attention (Kinsbourne, 1970). For listeners with NH, the attentional hypothesis seems to provide a better explanation of patterns of performance (Hiscock and Kinsbourne, 2011). It may be that shifts in attention also help to explain increased right-ear advantage associated with aging. However, it may be that structures conveying information to either side of the brain are indeed compromised for listeners with detrimental changes to the peripheral and central processing of auditory information like those who have hearing loss. Of particular concern are long periods of auditory deprivation. This conclusion is supported by one study showing that long periods of auditory deprivation in one ear are associated with speech understanding asymmetries and contralateral interference (Goupell et al., 2018). In particular, these listeners demonstrate a pattern of auditory "extinction," where information in the poorer ear is either not perceived or ignored during simultaneous stimulation with the better ear (Deouell and Soroker, 2000). The present study and the study by Goupell et al. (2021) simulated asymmetries in the spectro-temporal fidelity of sounds in listeners with NH, which resulted in a shift in attention toward the better ear. Thus, the conclusions of a structurally based ear advantage framework may be more appropriate for understanding interaural asymmetries in speech understanding for patients with BiCIs.

The results from single word trials (**Supplementary Figure 2**) suggest that ear advantage and interference effects observed in **Figure 6** could be due to consistent substitution errors (e.g., reporting a liquid when a stop and liquid were presented). In phonological fusion trials, this would be scored as a bias toward the left/right ear if the liquid matched the liquid presented, or an interference response if the liquid did not match that presented. Substituting a liquid for a stop-liquid cluster seems especially likely because of the manipulation used in the experiment. That is, decreasing the dynamic range would have smoothed the abrupt onset associated with a stop consonant. To address whether there was a consistent pattern of performance in both sets of trials, **Figure 7** shows the relationship between proportion of liquid responses when stops and liquids were presented in symmetric dynamic ranges in the same word and phonological fusion trials. The results show that smaller dynamic ranges resulted in a greater number of liquid-only responses. Together with **Figure 6**, this suggests that listeners made a similar substitution error, reporting a liquid when both a stop and liquid were presented, especially at 100 and 60% dynamic range for both trial types. Responding with a liquid for a stimulus containing a stop and liquid in same word trials was most common at 40% dynamic range. While chance error was 0.83%, if listeners were able to understand the vowel and responded with one word, there was a 40% chance of guessing a liquid-only response. **Figure 7** shows that several listeners responded with only liquids greater than 40% of the time at small dynamic ranges, suggesting that they demonstrated a consistent bias.

The vocoder manipulation used by Loizou et al. (2000) was similar to that described here, except that they reduced the overall amplitude of the envelope, which would have also decreased audibility. Our approach reduced the amplitude of the envelope and compensated by increasing the minimum, also on dB scale. This would have resulted in similar audibility, but the introduction of more prominent onsets and offsets as dynamic range became smaller. Loizou et al. (2000) had extensive data to examine how their vocoder affected speech from many different speakers, and psychophysical data to evaluate whether their predictions were correct. In our study, we had only two tokens of 15 words spoken by the same individual and accordingly wanted to avoid making generalizations. **Supplementary Figure 3** indicates that listeners made very few vowel errors, suggesting that the cues needed to discriminate between the vowels here remained intact. In contrast, as can be seen in **Supplementary Figure 2** and **Figure 7**, listeners in our study tended to confuse stop-liquid clusters with liquid consonants. The opposite pattern can be seen in the study by Loizou et al. (2000), where listeners tended to make more vowel and fewer consonant errors. This likely reflects the presence of more prominent onsets and offsets in our study.

When symmetric dynamic range was changed from 100 to 60 or 40%, it resulted in an approximately 35 or 45% decrease in the proportion of "ideal" responses, respectively (**Figure 6**). Approximately 10% of this decrease can be explained by an increase in "fusion" responses, corresponding to a combination of the stop and liquid words. Another 10–15% can be explained by an increase in "biased left" responses. Another 10% can be explained by an increase in "biased right" responses. The remaining decrease in "ideal" responses is explained by an increase in "interference" responses, approximately twofold in 60% and eightfold in 40%. Combined with **Figure 7**, these results suggest that during the fusion trials, listeners tended to report only the liquid in the left or right ear slightly more often than only the stop. Most responses from listeners still contained a stop consonant. Thus, while reporting only the liquid was the most common error, this does not mean that listeners were unable to detect the presence of the stop consonant.

Aging also had an effect on ear advantage. Older NH listeners were more likely to report the word in their poorer ear at the expense of task accuracy (**Figures 3**, **4**). This could reflect poorer working memory or selective attention in older listeners (Roque et al., 2019). That is, compared to YNH listeners, ONH listeners were less likely to attend to the better ear but similarly likely to respond with one word when two words were presented. Stronger ear advantage effects have previously been observed (Westerhausen et al., 2015), but they were only replicated in the unprocessed conditions of the present study. Additionally, ONH listeners demonstrated more

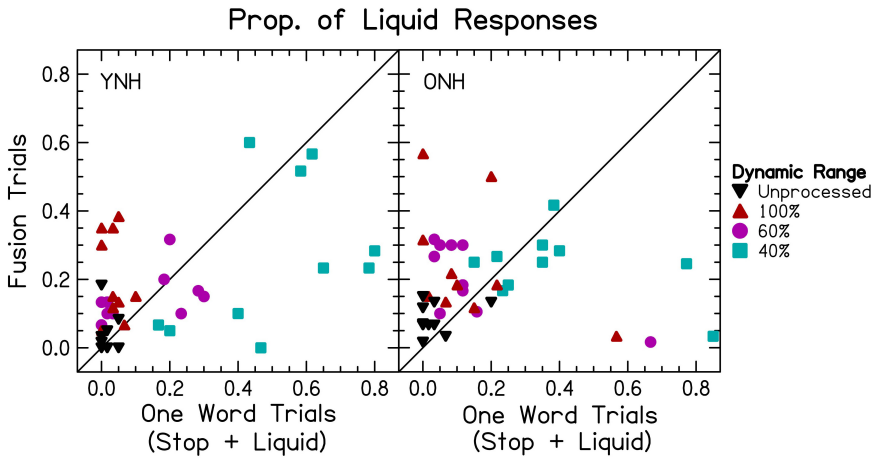



FIGURE 7
Proportion of responses containing only liquid consonants for fusion and single word trials by vocoder condition. The x- and y-axes correspond to the proportion of trials containing only liquid consonants for fusion and single word trials, respectively. The color and shape represent the vocoder condition given in the figure legend. Panels on the **(left)** and **(right)** correspond to YNH and ONH listeners, respectively.

worse ear responses relative to YNH listeners when stimuli were asymmetric. It is worth noting the *p*-value (0.050) of the interaction between age and ear in symmetric dynamic ranges with different vowels and that the present study had a small sample size. Older NH listeners also showed more frequent interference responses (**Figure 6**), suggesting that the task was more difficult overall. Auditory spatial attention has been under investigated in ONH listeners and is a confound in BiCI studies concerning contralateral interference, e.g., (Goupell et al., 2016, 2018). One thing that can be concluded from the present study is that ONH listeners did not demonstrate the auditory extinction (i.e., lack of perceiving the worse ear when the better ear was stimulated) observed in listeners with BiCIs, suggesting that it is either acquired over asymmetric experience or a result of deafness.

## 4.3. Limitations

The greatest limitation to the generalizability of these results is probably the speech stimuli used, which were highly artificial. Speech tokens were presented with the same onset time, spoken by the same individual. It is important to note that in the original phonological fusion experiments, onset time (varied between 0 and 800 ms) played a large role in the perception of listeners. When the stop consonant preceded the liquid, listeners were even more likely to report a fused response (Cutting, 1975). In contrast, when the liquid preceded the stop, listeners were less likely to report a one-word response. When dichotic vowels are offset by as little as 1 ms, and up to 20 ms, listeners with NH and hearing loss are less likely to report a single vowel compared to when they are presented simultaneously (Eddolls et al., 2022). Unfortunately, the vowel reported in the study by Eddolls and colleagues was not included in their results, so it is not possible to determine how onset time affected the vowel being reported. Notably, there was still considerable fusion for vowels with different onset times for both groups of listeners, especially compared to differences in fundamental frequency between ears. The effect of onset time was largest and most consistent when stimuli had the same fundamental frequency, similar to the present study where all stimuli were produced by the same speaker. Stimuli in the present experiment were presented independently to each ear, rather than mixed, as would occur in the free-field. Other experiments exploring the effects of right-ear advantage have varied relative level of the sound in each ear, showing changes in the percentage of correctly reported words (Hugdahl et al., 2008; Westerhausen et al., 2009). Thus, varying the onset time and interaural level difference may help titrate ear advantage. Because there were few words in the present study, so it is possible that some of these effects were driven by specific features. It may be that other phonemes are less or more likely to result in fusion and interference than those used here. It is highly likely that listeners used any features available to distinguish between words (e.g., /d/ at the end of the set with the /ɛ/ vowel). Finally, listeners with CIs take a long period of experience (∼1 year) with CI stimulation before performance saturates (Blamey et al., 2012). Unilateral auditory deprivation changes the representation of auditory cues from preferring the deprived side to preferring the CI-stimulated side from brainstem to cortex (Gordon et al., 2013, 2015; Polonenko et al., 2015, 2018). Like other vocoding studies with acute exposure to CI processed speech, this approach ignores the effects of experience. In particular, because dynamic range conditions were completely randomized, it is likely that listeners would have adjusted their strategy if repeatedly exposed to the same vocoder condition.

Accordingly, the results of the present study should be treated as a proof of concept. With stimuli ideal to cause integration of auditory information, listeners were more likely to report one word that did not correspond to the word presented to either ear. One strength of the phonological fusion procedure is that it provides a spectrum of performance to assess whether the information in either ear is used, integrated, ignored, or interfered with the other ear. Future experiments may be able to use more realistic stimuli to achieve similar goals.

The present experiment successfully collected data in a remote-testing context from older and younger listeners. However, there were several challenges that occurred during testing. It was not possible to assess hearing using traditional audiometry. Instead, experimental audiometric equipment could only reliably produce sound levels of 20 dB HL, meaning that large (10–15 dB) asymmetries could not be detected if they were below 20 dB HL. Audiometric data showed asymmetric hearing responses for at least one ONH listener at high frequencies. Additionally, as monaural speech intelligibility was not measured, it was not possible to determine whether there were asymmetries in speech understanding. Further, because the experiment took so long, it is likely that listeners became tired or bored, which fit the anecdotal reports provided to the experimenter. Another limitation was that boredom or fatigue during the experiment could be misconstrued as attentional difficulties. We attempted to address this by completely randomizing experimental conditions, preventing motivation-driven decreases in performance being confounded with experimental conditions. Additionally, there were technical issues that had to be resolved during testing that would have been easier to address in the laboratory. One individual fell asleep during testing. This could be addressed in future by requiring that listeners remain in contact during the experiment, but in order to prevent burdening older listeners by requiring that they use their own equipment, it may be necessary to provide access to the internet or a phone with laboratory equipment. Remote testing has unique challenges, but it may be an equitable path forward for working with populations who are unable to travel to the laboratory, and was effective for continuing to gather data during the COVID-19 pandemic.

## 5. Summary and conclusion

Several conclusions can be drawn from the present experiment:

1. Listeners demonstrated poorer speech understanding when the dynamic range was reduced, either symmetrically or asymmetrically (**Figures 3**, **5**), consistent with previous experiments in listeners with BiCIs (Firszt et al., 2002; Spahr et al., 2007) and simulations in NH (Loizou et al., 2000).
2. Decreased accuracy in reporting at least one word correctly was related to increased probability of one-word responses when speech with different vowels was presented to each ear (**Figure 4**). Speech identification accuracy was predicted by the *mean* dynamic range across ears, in contrast to previous literature suggesting that bilateral perception is dominated by the worse ear (Ihlefeld et al., 2015; Anderson et al., 2019b, 2022) or degree of asymmetry (Yoon et al., 2011). Both findings suggest that the ears are not independent channels and interact to produce perception of speech.
3. In interaurally symmetric conditions when speech with different vowels was presented to each ear, listeners were more likely to report the stimulus in the right ear, especially when stimuli were temporally degraded (**Figure 5**).
4. In interaurally asymmetric conditions when speech with different vowels were presented to each ear, listeners were more likely to report the stimulus in the better ear. Older NH listeners also showed increased responses from the poorer ear with decreasing dynamic range (**Figure 5**).
5. When phonologically fusible pairs of words were presented to each ear, listeners were more likely to experience interference if both ears were temporally degraded (**Figure 6**). They were more likely to report both words correctly if both ears had a large dynamic range, and likely to report both words incorrectly if both ears had a small dynamic range (especially for older listeners).

## Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by University of Wisconsin-Madison Health Sciences Institution Review Board, Listeners. The patients/participants provided their digital informed consent to participate in this study.

## Author contributions

SA created the stimuli, programmed and tested all equipment, completed data collection, performed the statistical analysis, and wrote the original manuscript. SA and RL acquired funding for the study. All authors contributed to the conception and design of the study, as well as the interpretation of the results

and contributed to the manuscript revision, read, and approved the submitted version.

## Funding

This work was supported by NIH-NIDCD F31 DC018483-01A1 awarded to SA, NIH-NIDCD R01-DC003083 awarded to RL, and NIH-NICHD U54HD090256 awarded to the Waisman Center.

## Acknowledgments

We would like to thank Drs. Ben Parrell, Jasenia Hartman, Jonathan Jibson, and Sara Bakst for their invaluable advice while recording, selecting, and processing speech stimuli. We would also like to thank Matt Winn for sharing his Praat code used for vocoding. SA would like to thank Drs. Lina Reiss, Emily Burg and Lukas Suveg for their helpful feedback and discussions as this project developed. Portions of this work appear in the dissertation of SRA (Anderson, 2022) and were presented at the 2020 Association for Research in Otolaryngology MidWinter meeting in San Jose, CA.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1018190/full#supplementary-material

## References

Anderson, S. R. (2022). *Mechanisms that underlie poorer binaural outcomes in patients with asymmetrical hearing and bilateral cochlear implants*, Dissertation. Madison, WI: University of Wisconsin-Madison.

Anderson, S. R., Kan, A., and Litovsky, R. Y. (2019b). 'Asymmetric temporal envelope encoding: Implications for within- and across-ear envelope comparison'. *J. Acoust. Soc. Am.* 146, 1189–1206. doi: 10.1121/1.5121423

Anderson, S. R., Easter, K., and Goupell, M. J. (2019a). 'Effects of rate and age in processing interaural time and level differences in normal-hearing and bilateral cochlear-implant listeners'. *J. Acoust. Soc. Am.* 146, 3232–3254. doi: 10.1121/1.5130384

Anderson, S. R., Kan, A., and Litovsky, R. Y. (2022). Asymmetric temporal envelope encoding: Within- and across-ear envelope comparisons in listeners with bilateral cochlear implants. *J. Acoust. Soc. Am.* 152, 3294–3312.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., and Kraus, N. (2012). 'Aging affects neural precision of speech encoding'. *J. Neurosci.* 32, 14156–14164. doi: 10.1523/JNEUROSCI.2176-12.2012

Aronoff, J. M., Shayman, C., Prasad, A., Suneel, D., and Stelmach, J. (2015). 'Unilateral spectral and temporal compression reduces binaural fusion for normal hearing listeners with cochlear implant simulations'. *Hear. Res.* 320, 24–29. doi: 10.1016/j.heares.2014.12.005

Bakal, T. A., Milvae, K. D., Chen, C., and Goupell, M. J. (2021). 'Head shadow, summation, and squelch in bilateral cochlear-implant users with linked automatic gain controls'. *Trend. Hear.* 25, 1–17. doi: 10.1177/23312165211018147

Bates, D., Machler, M., Bolker, B., and Walker, S. (2015). 'Fitting linear mixed-effects models using {lme4}'. *J. Stat. Soft.* 67, 1–48. doi: 10.18637/jss.v067.i01

Baumgärtel, R. M., Hu, H., Kollmeier, B., and Dietz, M. (2017). 'Extent of lateralization at large interaural time differences in simulated electric hearing and bilateral cochlear implant users'. *J. Acoust. Soc. Am.* 141, 2338–2352. doi: 10.1121/1.4979114

Bernstein, J. G., Goupell, M. J., Schuchman, G. I., Rivera, A. L., and Brungart, D. S. (2016). 'Having two ears facilitates the perceptual separation of concurrent talkers for bilateral and single-sided deaf cochlear implantees.'. *Ear Hear.* 37, 289–302. doi: 10.1097/AUD.0000000000000284

Bernstein, J. G., Stakhovskaya, O. A., Jensen, K. K., and Goupell, M. J. (2020). 'Acoustic hearing can interfere with single-sided deafness cochlear-implant speech perception'. *Ear Hear.* 41, 747–761. doi: 10.1097/AUD.0000000000000805

Bernstein, L. R., and Trahiotis, C. (2011). 'Lateralization produced by envelope-based interaural temporal disparities of high-frequency, raised-sine stimuli: Empirical data and modeling'. *J. Acoust. Soc. Am.* 129, 1501–1508. doi: 10.1121/1.3552875

Blamey, P., Artieres, F., Başkent, D., Bergeron, F., Beynon, A., Burke, E., et al. (2012). 'Factors affecting auditory performance of postlinguistically deaf adults using cochlear implants: An update with 2251 patients'. *Audiol. Neurootol.* 18, 36–47. doi: 10.1159/000343189

Bregman, A. S. (1994). *Auditory scene analysis: The perceptual organization of sound*. Caimbridge, MA: Bradford Books, MIT Press, doi: 10.1121/1.408434

Brungart, D. S., and Simpson, B. D. (2002). 'Within-ear and across-ear interference in a cocktail-party listening task'. *J. Acoust. Soc. Am.* 112, 2985–2995. doi: 10.1121/1.1512703

Cherry, E. C. (1953). 'Some experiments on the recognition of speech, with one and with two ears'. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Croghan, N. B. H., and Smith, Z. M. (2018). 'Speech understanding with various maskers in cochlear-implant and simulated cochlear-implant hearing: Effects of spectral resolution and implications for masking release'. *Trend. Hear.* 22, 1–13. doi: 10.1177/2331216518787276

Croghan, N. B. H., Duran, S. I., and Smith, Z. M. (2017). 'Re-examining the relationship between number of cochlear implant channels and maximal speech intelligibility'. *J. Acoust. Soc. Am.* 142, EL537–EL543. doi: 10.1121/1.5016044

Cutting, J. E. (1975). 'Aspects of phonological fusion'. *J. Exp. Psychol. Hum. Percept. Perform.* 104, 105–120. doi: 10.1037/0096-1523.1.2.105

Cutting, J. E. (1976). 'Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening'. *Psych. Rev.* 83, 114–140. doi: 10.1037/0033-295X.83.2.114

Darwin, C. J. (1981). 'Perceptual grouping of speech components differing in fundamental frequency and onset-time'. *Q. J. Exp. Psychol.* 33, 185–207. doi: 10.1080/14640748108400785

Deouell, L. Y., and Soroker, N. (2000). 'What is extinguished in auditory extinction?'. *Neuroreport* 11, 3059–3062. doi: 10.1097/00001756-200009110-00046

Drullman, R., Festen, J. M., and Plomp, R. (1994). 'Effect of temporal envelope smearing on speech reception'. *J. Acoust. Soc. Am.* 95, 1053–1064. doi: 10.1121/1.408467

Durlach, N. I. (1963). 'Equalization and cancellation theory of binaural masking-level differences'. *J. Acoust. Soc. Am.* 35, 1206–1218. doi: 10.1121/1.1918675

Eddolls, M. S., Molis, M. R., and Reiss, L. A. J. (2022). 'Onset asynchrony: Cue to aid dichotic vowel segregation in listeners with normal hearing and hearing loss'. *J. Speech Lang. Hear. Res.* 65, 2709–2719. doi: 10.1044/2022_jslhr-21-00411

Firszt, J. B., Chambers, R. D., and Kraus, N. (2002). 'Neurophysiology of cochlear implant users II: Comparison among speech perception, dynamic range, and physiological measures'. *Ear Hear.* 23, 516–531. doi: 10.1097/00003446-200212000-00003

Frank, T. (1997). 'ANSI update: Specification of audiometers'. *Am. J. Audiol.* 6, 29–32.

Gallun, F. J., Mason, C. R., and Kidd, G. J. (2007). 'The ability to listen with independent ears'. *J. Acoust. Soc. Am.* 122, 2814–2825. doi: 10.1121/1.2780143

Gallun, F. J., McMillan, G. P., Molis, M. R., Kampel, S. D., Dann, S. M., and Konrad-Martin, D. L. (2014). 'Relating age and hearing loss to monaural, bilateral, and binaural temporal sensitivity'. *Front. Neuro.* 8:1–14. doi: 10.3389/fnins.2014.00172

Gordon, K. A., Jiwani, S., and Papsin, B. C. (2013). 'Benefits and detriments of unilateral cochlear implant use on bilateral auditory development in children who are deaf'. *Front Psychol* 4:719. doi: 10.3389/fpsyg.2013.00719

Gordon, K., Henkin, Y., and Kral, A. (2015). 'Asymmetric hearing during development: The aural preference syndrome and treatment options'. *Pediatrics* 136, 141–153. doi: 10.1542/peds.2014-3520

Goupell, M. J., Eisenberg, D., and DeRoy Milvae, K. (2021). 'Dichotic listening performance with cochlear-implant simulations of ear asymmetry is consistent with difficulty ignoring clearer speech'. *Atten. Percept. Psychophys.* 83, 2083–2101. doi: 10.3758/s13414-021-02244-x

Goupell, M. J., Gaskins, C. R., Shader, M. J., Walter, E. P., Anderson, S., and Gordon-Salant, S. (2017). 'Age-related differences in the processing of temporal envelope and spectral cues in a speech segment'. *Ear Hear.* 38, e335–e342. doi: 10.1097/AUD.0000000000000447

Goupell, M. J., Kan, A., and Litovsky, R. Y. (2016). 'Spatial attention in bilateral cochlear-implant users'. *J. Acoust. Soc. Am.* 140, 1652–1662. doi: 10.1121/1.4962378

Goupell, M. J., Noble, J. H., Phatak, S. A., Kolberg, E., Cleary, M., Stakhovskaya, O. A., et al. (2022). 'Computed-tomography estimates of interaural mismatch in insertion depth and scalar location in bilateral cochlear-implant users'. *Otol. Neurotol.* 43, 666–675. doi: 10.1097/MAO.0000000000003538

Goupell, M. J., Stakhovskaya, O. A., and Bernstein, J. G. W. (2018). 'Contralateral interference caused by binaurally presented competing speech in adult bilateral cochlear-implant users'. *Ear Hear.* 39, 110–123. doi: 10.1097/AUD.0000000000000470

Goupell, M. J., Stoelb, C., Kan, A., and Litovsky, R. Y. (2013). 'Effect of mismatched place-of-stimulation on the salience of binaural cues in conditions that simulate bilateral cochlear-implant listening'. *J. Acoust. Soc. Am.* 133, 2272–2287. doi: 10.1121/1.4792936

Greenwood, D. D. (1990). 'A cochlear frequency-position function for several species—29 years later'. *J. Acoust. Soc. Am.* 87, 2592–2605. doi: 10.1121/1.399052

Hiscock, M., and Kinsbourne, M. (2011). 'Attention and the right-ear advantage: What is the connection?'. *Brain Cogn.* 76, 263–275. doi: 10.1016/j.bandc.2011.03.016

Hugdahl, K., Westerhausen, R., Alho, K., Medvedev, S., and Hämäläinen, H. (2008). 'The effect of stimulus intensity on the right ear advantage in dichotic listening'. *Neurosci. Lett.* 431, 90–94. doi: 10.1016/j.neulet.2007.11.046

Ihlefeld, A., Carlyon, R. P., Kan, A., Churchill, T. H., and Litovsky, R. Y. (2015). 'Limitations on monaural and binaural temporal processing in bilateral cochlear implant listeners'. *J. Assoc. Res. Otolaryngol.* 16, 641–652. doi: 10.1007/s10162-015-0527-7

Ihlefeld, A., Kan, A., and Litovsky, R. Y. (2014). 'Across-frequency combination of interaural time difference in bilateral cochlear implant listeners'. *Front. Syst. Neurosci.* 8:22. doi: 10.3389/fnsys.2014.00022

Kan, A., Goupell, M. J., and Litovsky, R. Y. (2019). 'Effect of channel separation and interaural mismatch on fusion and lateralization in normal-hearing and cochlear-implant listeners'. *J. Acoust. Soc. Am.* 146, 1448–1463. doi: 10.1121/1.5123464

Kan, A., Stoelb, C., Litovsky, R. Y., and Goupell, M. J. (2013). 'Effect of mismatched place-of-stimulation on binaural fusion and lateralization in bilateral cochlear-implant users'. *J. Acoust. Soc. Am.* 134, 2923–2936. doi: 10.1121/1.4820889

Kenward, M. G., and Roger, J. H. (1997). 'Small sample inference for fixed effects from restricted maximum likelihood'. *Biometrics* 53, 983–997. doi: 10.2307/2533558

Kimura, D. (1967). 'Functional asymmetry of the brain in dichotic listening'. *Cortex* 3, 163–178. doi: 10.1016/s0010-9452(67)80010-8

Kinsbourne, M. (1970). 'The cerebral basis of lateral asymmetries in attention'. *Acta Psychol. (Amst)* 33, 193–201. doi: 10.1016/0001-6918(70)90132-0

Kuznetsova, A., Brockhoff, P. B., and Christensen, R. H. B. (2017). 'lmerTest package: Tests in linear mixed effects models'. *J. Stat. Softw.* 82, 1–26. doi: 10.18637/jss.v082.i13

Lenth, R. V. (2022). *emmeans: Estimated marginal means, aka least-squares means*'. *CRAN*. Available online at: https://cran.r-project.org/package=emmeans (accessed December 24, 2022).

Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J. (1999). 'The precedence effect'. *J. Acoust. Soc. Am.* 106, 1633–1654. doi: 10.1016/0378-5955(83)90002-3

Litovsky, R. Y., Jones, G. L., Agrawal, S., and Hoesel, R v (2010). 'Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans'. *J. Acoust. Soc. Am.* 127, 400–414. doi: 10.1121/1.3257546

Litovsky, R., Parkinson, A., Arcaroli, J., and Sammeth, C. (2006). 'Simultaneous bilateral cochlear implantation in adults: A multicenter clinical study'. *Ear Hear.* 27, 714–731. doi: 10.1097/01.aud.0000246816.50820.42

Loizou, P. C., Dorman, M. F., and Fitzke, J. (2000). 'The effect of reduced dynamic range on speech understanding: Implications for patients with cochlear implants'. *Ear Hear.* 21, 25–31. doi: 10.1097/00003446-200002000-00006

Loizou, P. C., Hu, Y., Litovsky, R., Yu, G., Peters, R., Lake, J., et al. (2009). 'Speech recognition by bilateral cochlear implant users in a cocktail-party setting'. *J. Acoust. Soc. Am.* 125, 372–383. doi: 10.1121/1.3036175

Long, C. J., Eddington, D. K., Colburn, H. S., and Rabinowitz, W. M. (2003). 'Binaural sensitivity as a function of interaural electrode position with a bilateral cochlear implant user'. *J. Acoust. Soc. Am.* 114, 1565–1574. doi: 10.1121/1.1603765

Long, C. J., Holden, T. A., McClelland, G. H., Parkinson, W. S., Shelton, C., Kelsall, D. C., et al. (2014). 'Examining the electro-neural interface of cochlear implant users using psychophysics. CT scans, and speech understanding', *J. Assoc. Res. Otolaryngol.* 15, 293–304. doi: 10.1007/s10162-013-0437-5

Moberly, A. C., Harris, M. S., Boyce, L., and Nittrouer, S. (2017). 'Speech recognition in adults with cochlear implants: The effects of working memory, phonological sensitivity, and aging'. *J. Speech Lang. Hear. Res.* 60, 1046–1061. doi: 10.1044/2016_JSLHR-H-16-0119

Mosnier, I., Sterkers, O., Bebear, J., Godey, B., Robier, A., Deguine, O., et al. (2009). 'Speech performance and sound localization in a complex noisy environment in bilaterally implanted adult patients'. *Audiol. Neurootol.* 14, 106–114. doi: 10.1159/000159121

Oxenham, A. J., and Kreft, H. A. (2014). 'Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing'. *Trend. Hear.* 18, 1–14. doi: 10.1177/2331216514553783

Polonenko, M. J., Papsin, B. C., and Gordon, K. A. (2015). 'The effects of asymmetric hearing on bilateral brainstem function: Findings in children with bimodal (electric and acoustic) hearing'. *Audiol. Neurootol.* 20(Suppl. 1), 13–20. doi: 10.1159/000380743

Polonenko, M. J., Papsin, B. C., and Gordon, K. A. (2018). 'Limiting asymmetric hearing improves benefits of bilateral hearing in children using cochlear implants'. *Sci. Rep.* 8, 1–17. doi: 10.1038/s41598-018-31546-8

Pumplin, J. (1985). 'Low-noise noise'. *J. Acoust. Soc. Am.* 78, 100–104. doi: 10.1121/1.392571

Rao, R. P. N., and Ballard, D. H. (1999). 'Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects'. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580

Reeder, R. M., Firszt, J. B., Holden, L. K., and Strube, M. J. (2014). 'A longitudinal study in adults with sequential bilateral cochlear implants: Time course for individual ear and bilateral performance'. *J. Speech Lang. Hear. Res.* 57, 1108–1126. doi: 10.1044/2014

Reiss, L. A. J., and Molis, M. R. (2021). 'Abnormal fusion of dichotic vowels across different fundamental frequencies in hearing-impaired listeners: An alternative explanation for difficulties with speech in background talkers'. *J. Assoc. Res. Otolaryngol.* 22, 443–461. doi: 10.1007/s10162-021-00790-7

Reiss, L. A. J., Eggleston, J. L., Walker, E. P., and Oh, Y. (2016). 'Two ears are not always better than one: Mandatory vowel fusion across spectrally mismatched ears in hearing-impaired listeners'. *J. Assoc. Res. Otolaryngol.* 17, 341–356. doi: 10.1007/s10162-016-0570-z

Reiss, L. A., Fowler, J. R., Hartling, C. L., and Oh, Y. (2018). 'Binaural pitch fusion in bilateral cochlear implant users'. *Ear Hear.* 39, 390–397. doi: 10.1097/AUD.0000000000000497

Reiss, L. A., Ito, R. A., Eggleston, J. L., Liao, S., Becker, J. J., Lakin, C. E., et al. (2015). 'Pitch adaptation patterns in bimodal cochlear implant users: Over time and after experience'. *Ear Hear.* 36, e23–e34. doi: 10.1097/AUD.0000000000000114

Roque, L., Karawani, H., Gordon-Salant, S., and Anderson, S. (2019). 'Effects of age, cognition, and neural encoding on the perception of temporal speech cues'. *Front. Neuro.* 13:1–15. doi: 10.3389/fnins.2019.00749

Scharf, B. (1974). "Localization of unlike tones from two loudspeakers," in *Sensation and Measurement*, eds H. R. Moskowitz, B. Scharf, and J. C. Stevens (Dordrecht: Springer), 309–314. doi: 10.1007/978-94-010-2245-3_30

Shader, M. J., Nguyen, N., Cleary, M., Hertzano, R., Eisenman, D. J., Anderson, S., et al. (2020a). 'Effect of stimulation rate on speech understanding in older cochlear-implant users'. *Ear Hear.* 41, 640–651. doi: 10.1097/AUD.0000000000000793

Shader, M. J., Yancey, C. M., Gordon-Salant, S., and Goupell, M. J. (2020b). 'Spectral-temporal trade-off in vocoded sentence recognition: Effects of age, hearing thresholds, and working memory'. *Ear Hear.* 41, 1226–1235. doi: 10.1097/AUD.0000000000000840

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). 'Speech recognition with primarily temporal cues'. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303

Shepherd, R. K., and Hardie, N. A. (2001). 'Deafness-induced changes in the auditory pathway: Implications for cochlear implants'. *Audiol. Neurootol.* 6, 305–318. doi: 10.1159/000046843

Shinn-Cunningham, B. G. (2008). 'Object-based auditory and visual attention'. *Trends Cogn. Sci.* 12, 182–186. doi: 10.1016/j.tics.2008.02.003

Shinn-Cunningham, B., Best, V., and Lee, A. K. C. (2017). "Auditory object formation and selection," in *The auditory system at the cocktail party. Springer handbook of auditory research*, 60th Edn, eds J. C. Middlebrooks, et al. (Cham: Springer), 7–40. doi: 10.1007/978-3-319-51662-2_2

Spahr, A. J., Dorman, M. F., and Loiselle, L. H. (2007). 'Performance of patients using different cochlear implant systems: Effects of input dynamic range'. *Ear Hear.* 28, 260–275. doi: 10.1097/AUD.0b013e3180312607

Staisloff, H. E., Lee, D. H., and Aronoff, J. M. (2016). 'Perceptually aligning apical frequency regions leads to more binaural fusion of speech in a cochlear implant simulation'. *Hear. Res.* 337, 59–64. doi: 10.1016/j.heares.2016.05.002

Suneel, D., Staisloff, H., Shayman, C. S., Stelmach, J., and Aronoff, J. M. (2017). 'Localization performance correlates with binaural fusion for interaurally mismatched vocoded speech'. *J. Acoust. Soc. Am.* 142, EL276–EL280. doi: 10.1121/1.5001903

Todd, A. E., Goupell, M. J., and Litovsky, R. Y. (2017). 'The relationship between intensity coding and binaural sensitivity in adults with cochlear implants'. *Ear Hear.* 38, e128–e141. doi: 10.1097/AUD.0000000000000382

van den Brink, G., Sintnicolaas, K., and van Stam, W. S. (1976). 'Dichotic pitch fusion'. *J. Acoust. Soc. Am.* 59, 1471–1476. doi: 10.1121/1.380989

van Hoesel, R. J. M., and Clark, G. M. (1995). 'Fusion and lateralization study with two binaural cochlear implant patients'. *Ann. Otol. Rhinol. Laryngol.* 104, 233–235.

van Hoesel, R. J. M., and Clark, G. M. (1997). 'Psychophysical studies with two binaural cochlear implant subjects'. *J. Acoust. Soc. Am.* 102, 495–507. doi: 10.1121/1.419611

Westerhausen, R., Bless, J., and Kompus, K. (2015). 'Behavioral laterality and aging: The free-recall dichotic-listening right-ear advantage increases with age'. *Dev. Neuropsychol.* 40, 313–327. doi: 10.1080/87565641.2015.1073291

Westerhausen, R., Moosmann, M., Alho, K., Medvedev, S., Hämäläinen, H., and Hugdahl, K. (2009). 'Top-down and bottom-up interaction: Manipulating the dichotic listening ear advantage'. *Brain Res.* 1250, 183–189. doi: 10.1016/j.brainres.2008.10.070

Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2014). 'The perception of apparent auditory source width in hearing-impaired adults'. *J. Acoust. Soc. Am.* 135, 3548–3559. doi: 10.1121/1.4875575

Woods, K. J., Siegel, M. H., Traer, J., and McDermott, J. H. (2017). 'Headphone screening to facilitate web-based auditory experiments'. *Atten. Percept. Psychophys.* 79, 2064–2072. doi: 10.3758/s13414-017-1361-2

Xie, Z., Gaskins, C. R., Shader, M. J., Gordon-Salant, S., Anderson, S., and Goupell, M. J. (2019). 'Age-related temporal processing deficits in word segments in adult cochlear-implant users'. *Trend. Hear.* 23:2331216519886688. doi: 10.1177/2331216519886688

Yoon, Y., Li, Y., Kang, H., and Fu, Q. (2011). 'The relationship between binaural benefit and difference in unilateral speech recognition performance for bilateral cochlear implant users'. *Int. J. Audiol.* 50, 554–565. doi: 10.3109/14992027.2011.580785

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 10 January 2023
DOI 10.3389/fnins.2022.1060172



OPEN ACCESS

EDITED BY
Huiming Zhang,
University of Windsor, Canada

REVIEWED BY
Hongmei Hu,
University of Oldenburg, Germany
Mareike Daeglau,
University of Oldenburg, Germany

*CORRESPONDENCE
Ann Clock Eddins
✉ ann.eddins@ucf.edu

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 02 October 2022
ACCEPTED 16 December 2022
PUBLISHED 10 January 2023

CITATION
Eddins AC, Ozmeral EJ and Eddins DA (2023) Aging alters across-hemisphere cortical dynamics during binaural temporal processing. *Front. Neurosci.* 16:1060172. doi: 10.3389/fnins.2022.1060172



# Aging alters across-hemisphere cortical dynamics during binaural temporal processing

Ann Clock Eddins[1,2]*, Erol J. Ozmeral[1] and David A. Eddins[1]

[1]Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL, United States, [2]School of Communication Sciences and Disorders, University of Central Florida, Orlando, FL, United States

Differences in the timing and intensity of sounds arriving at the two ears provide fundamental binaural cues that help us localize and segregate sounds in the environment. Neural encoding of these cues is commonly represented asymmetrically in the cortex with stronger activation in the hemisphere contralateral to the perceived spatial location. Although advancing age is known to degrade the perception of binaural cues, less is known about how the neural representation of such cues is impacted by age. Here, we use electroencephalography (EEG) to investigate age-related changes in the hemispheric distribution of interaural time difference (ITD) encoding based on cortical auditory evoked potentials (CAEPs) and derived binaural interaction component (BIC) measures in ten younger and ten older normal-hearing adults. Sensor-level analyses of the CAEP and BIC showed age-related differences in global field power, where older listeners had significantly larger responses than younger for both binaural metrics. Source-level analyses showed hemispheric differences in auditory cortex activity for left and right lateralized stimuli in younger adults, consistent with a contralateral activation model for processing ITDs. Older adults, however, showed reduced hemispheric asymmetry across ITDs, despite having overall larger responses than younger adults. Further, when averaged across ITD condition to evaluate changes in cortical asymmetry over time, there was a significant shift in laterality corresponding to the peak components (P1, N1, P2) in the source waveform that also was affected by age. These novel results demonstrate across-hemisphere cortical dynamics during binaural temporal processing that are altered with advancing age.

KEYWORDS

**electrophysiology, cortical auditory evoked potentials, hemispheric asymmetry, interaural time difference, binaural interaction component**

## 1. Introduction

Spatial hearing plays an important role in everyday activities such as driving in noisy traffic, crossing the street at an intersection, and listening to conversations in a crowded restaurant. Not surprisingly, converging evidence indicates that spatial hearing abilities in senescent listeners is impeded by degradations to the binaural auditory system (e.g., Dubno et al., 2008; Eddins and Hall, 2010; Ozmeral et al., 2016; Eddins and Eddins, 2018; Eddins et al., 2018; Gallun and Best, 2020), the key pathway for processing spatial auditory cues. While there is substantial interest in age-related changes in binaural processing and spatial hearing, the nature of those changes and their underlying mechanisms are not fully understood or characterized. Because the power in human communication (i.e., speech) and competing sounds is greatest at low frequencies, it is of value to understand the impact of aging on low-frequency dominant coding of binaural processes, such as coding of interaural time differences (ITD). In avian species, ITDs are topographically encoded *via* cellular arrays tuned to a narrow range of ITDs (Konishi, 2003), as suggested by Jeffress (1948). In mammals, however, converging research points toward a non-topographic, opponent-channel process in the cortex whereby ITDs are deduced from the relative neural activity of opposing channels broadly tuned to the midline and two spatial hemifields (Brand et al., 2002; McAlpine, 2005; Briley et al., 2013; Stecker et al., 2015; McLaughlin et al., 2016; Ozmeral et al., 2016, 2019). Moreover, when stimuli are presented or perceived from one hemifield versus the other, a majority of the cortical activity occurs in the contralateral hemisphere. Although somewhat modest, this contralateral bias has been demonstrated for ITD coding in humans based on both evoked potential (Salminen et al., 2009; Briley et al., 2013; Ozmeral et al., 2016) and functional magnetic resonance imaging (fMRI) BOLD measures (von Kriegstein et al., 2008; Gutschalk and Steinmann, 2015; Stecker et al., 2015; McLaughlin et al., 2016).

Binaural coding and contralateral bias may also be reflected in the binaural interaction component (BIC); a derived measure that can be computed from the auditory brainstem response (ABR), middle latency response (MLR), or the cortical auditory evoked potential (CAEP) (Dobie and Berlin, 1979; McPherson and Starr, 1993; Fowler and Horn, 2012; Van Yper et al., 2015; Dykstra et al., 2016; Laumen et al., 2016; Sammeth et al., 2020). The BIC is a difference waveform obtained by subtracting the algebraic sum of monaural responses to isolated left and right ear stimulation from the binaural response [B–(L + R)] or by computing the converse [(L + R)–B] (McPherson and Starr, 1993; Van Yper et al., 2015). Typically, the binaural response is smaller in amplitude than the summed monaural response giving rise to small difference components, or BIC, at different latencies depending on the measure being analyzed (i.e., ABR, MLR, CAEP). Although some studies have reported an inability to measure an acoustic BIC, even in normal-hearing subjects (Haywood et al., 2015), others suggest that it may serve as useful tool for binaural hearing tests (e.g., Riedel and Kollmeier, 2002; Benichoux et al., 2018; Brown et al., 2019). Further, electrically evoked BIC responses have been recorded in humans with bilateral cochlear implants (CI) (He et al., 2010; Gordon et al., 2012; Hu and Dietz, 2015; Hu et al., 2016) and in bilaterally implanted animals (e.g., in cat, Smith and Delgutte, 2007; Hancock et al., 2010).

The reduced amplitude of the binaural response is not well understood but may originate from a combination of contralateral inhibitory and ipsilateral excitatory neural populations in the superior olivary complex (SOC), as shown in data from cat (Ungan et al., 1997; Ungan and Yagcioglu, 2002) and guinea pig (Goksoy et al., 2005), and similarly modeled data (Gaumond and Psaltikidou, 1991). Such a population-based code for BIC generation is consistent with the population-based opponent channel model for ITD coding in the cortex (Magezi and Krumbholz, 2010; Salminen et al., 2010, 2015; Briley et al., 2013; Ozmeral et al., 2016) and thus may represent related underlying mechanisms of spatial processing. The BIC has been measured from both the ABR and MLR over a range of ITDs where it was shown to decrease in amplitude and increase in latency with increasing ITD (McPherson and Starr, 1995; Riedel and Kollmeier, 2006). Comparable BIC data for CAEPs in humans are limited (e.g., Henkin et al., 2015) and have not been reported across ITDs nor have they been used to assess contralateral bias.

The impact of advancing age on binaural coding in cortical evoked responses or BIC measures may reflect global age-related changes in sensory processing, such as reduced neural inhibition, specifically at the level of the SOC or higher (Willott et al., 1997; Caspary et al., 2008), or a more general reduction in temporal synchrony or increased temporal jitter (Pichora-Fuller et al., 2007; Ozmeral et al., 2016). Recent data indicate that reduced inhibition and reduced temporal synchrony both play a role in age-related changes in ITD processing that are stimulus or context dependent. That is, static or fixed ITDs with strong stimulus onset markers lead to larger evoked response amplitudes (i.e., reduced inhibition; Eddins et al., 2018) while dynamic shifts in ITD, following a different preceding ITD, result in smaller evoked response amplitudes (i.e., reduced temporal synchrony; Ozmeral et al., 2016) in older adults. In the present study we hypothesized that if binaural processing is influenced by a down-regulation in inhibition with age, then neural responses for all ITDs will be larger in older than in younger listeners for all ITDs. Alternatively, if reduced temporal synchrony is a primary age-related factor for binaural coding in older listeners, then neural responses for older listeners would be smaller than those for younger listeners, with the greatest difference occurring for large ITDs and smaller differences for ITDs approaching midline.

Importantly, aging can also influence the distribution of neural activity across the cortex such that it may alter the

expected contralateral bias that occurs with ITD processing. In the context of cognitive aging, Cabeza (2002) proposed the hemispheric asymmetry reduction in older adults' model (HAROLD) based on functional neuroimaging studies in which prefrontal activity during an episodic memory retrieval task was right lateralized in younger adults but showed bilateral activity in both hemispheres in older adults (Cabeza et al., 1997). One hypothesis for the reduced asymmetry is based on compensatory processes, whereby older adults recruit activity from other brain regions to compensate during increased task demands to help enhance performance. As a result, there is broader distribution of activity across hemispheres and reduced asymmetry toward one hemisphere or the other. A second hypothesis for decreased hemispheric lateralization with advancing age is based on the concept of functional dedifferentiation in which neural processes associated with cognitive strategies, and perhaps sensory processing specialization, become less organized or more distributed both regionally and globally across functional networks (Cabeza et al., 1997; Festini et al., 2018). Although development of the HAROLD model was based on asymmetry reductions in prefrontal cortex during cognitive tasks (e.g., episodic and working memory), additional data on visuospatial processing also suggests reduced hemispheric lateralization in older adults (e.g., Learmonth et al., 2017). It remains uncertain whether the model is generalizable to auditory sensory processes, such as ITD coding, that are known to elicit hemispheric bias in neural activation across the cortex. The present study thus serves as an ideal test case of the generality of the HAROLD model. As such, we test the hypotheses that advancing age alters neural encoding of ITD cues and contralateral bias, as indexed by both CAEP and BIC measures, and that such changes follow the HAROLD model whereby hemispheric asymmetry is reduced during sensory processing in older relative to younger listeners with normal hearing.

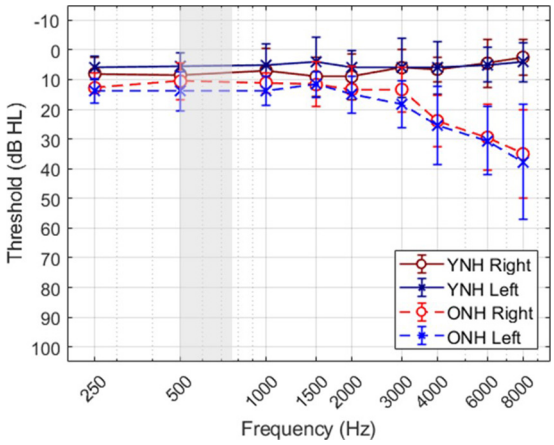



FIGURE 1
Mean and standard deviation of audiometric thresholds for each ear of each group.

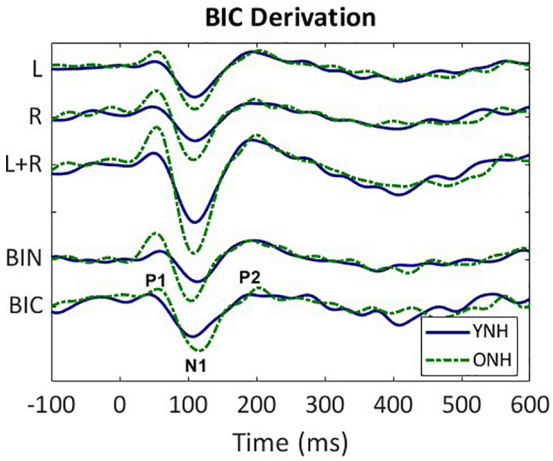



FIGURE 2
Derivation of the binaural interaction component (BIC). Grand average responses (arbitrary units) for each listener group (YNH–blue solid, ONH–green dashed); monaural left (L) and right (R) responses, sum of the monaural responses (L + R), binaural (BIN) response in the ITD Zero condition, and the BIC derived from the summed monaural (L + R) minus binaural (BIN) responses. The BIC is labeled with the three main peak components (P1, N1, P2).

## 2. Materials and methods

### 2.1. Participants

A total of twenty individuals participated in the study; ten younger listeners (mean age ± SD, 24.9 ± 2.5 years; 9 females) and ten older listeners (70.0 ± 2.7 years; 6 females). The sample size was based on a power analysis with an effect size of 0.25, statistical power of 0.95, and alpha of 0.05. **Figure 1** shows the mean and standard deviation of audiometric thresholds for both listener groups, where younger listeners (YNH) had clinically normal pure-tone thresholds ≤25 dB HL at octave frequencies from 250 to 8,000 Hz, and older listeners (ONH) had clinically-normal pure-tone thresholds ≤25 dB HL at octave frequencies from 250 to 4,000 Hz and ≤60 dB above 4,000 Hz. The gray shaded region illustrates the frequency bandwidth of the stimuli used in this study, as described below. The average threshold at 500 Hz (the frequency of focus in this study) across the two ears was 7 dB HL (±4.25) for the YNH group and 12.75 dB HL (±6.29) for the ONH group. All listeners were administered the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) to screen for cognitive impairment and all passed the screening with scores greater than 26. Each participant provided written consent and received hourly compensation for their participation, as approved by the University of South Florida Institutional Review Board.

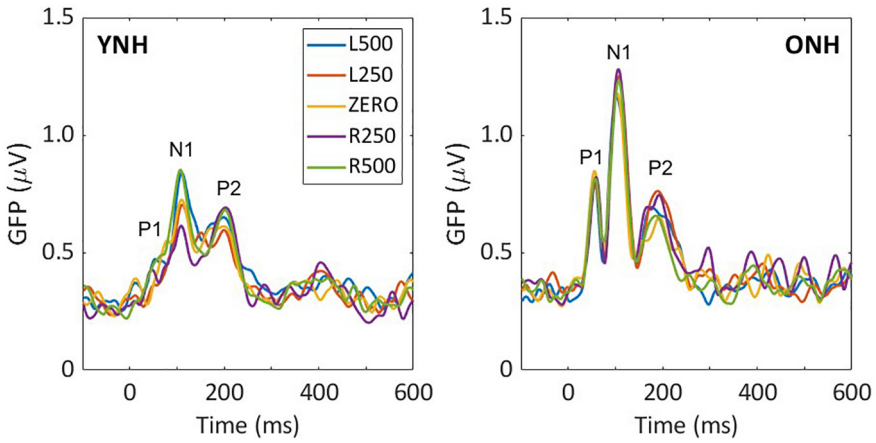



FIGURE 3
Grand average sensor-based global-field power (GFP) responses for each interaural time difference (ITD) condition and each listener group; young **(left)** and older **(right)**. The three main peak components are labeled in each panel (P1, N1, P2).

## 2.2. Stimuli

Stimuli were band-pass Gaussian noise bursts, with lower and upper cutoffs of 500 and 750 Hz. Digital filtering was performed in the frequency domain using MATLAB® (ver. R2018b, The Mathworks, Inc.). A new stimulus token was generated on each trial (sampling rate 24,414 Hz) with a duration of 400-ms, including 10-ms cosine-gated onset and offset ramps, and an inter stimulus interval of 1,600-ms. Stimuli were presented at a fixed level of 80 dB SPL *via* Tucker-Davis Technologies (TDT) RZ6 real-time processor, headphone buffer (HB7) and Etymotic ER-2 insert earphones. The stimuli were calibrated at the output of the earphones using a calibrator (B&K 4230), ear simulator (Knowles Electronics DB-100), 1/2" pressure microphone (B&K 4134), pre-amplifier (B&K 4134), and power conditioner (G.R.A.S. 12AA) routed to a multi-meter (Fluke 45).

Five binaural and two monaural stimulus conditions were run in block format. Binaural conditions included a diotic condition (i.e., ITD = 0 μs), and two left and two right leading ITDs at ±250 μs and ±500 μs (negative to the left, positive to the right). Due to sampling, true ITDs were 246 and 492 μs for the 250 and 500 μs conditions, respectively. Going forward, the conditions are referred to as L500, L250, Zero, R250, R500, with the letter corresponding to left (L) or right (R) leading, and number corresponding to the ITD value in μs. Monaural conditions included both left and right ear presentations. Each recording block consisted of 150 trials and lasted roughly 5 to 6 min, or total of about 45 min per subject with breaks given as needed. During each block, participants listened passively to the stimuli and were instructed to limit eye blinks and body movements while watching a captioned video of their own choosing. The video was used as a perceptual distractor during passive listening, as it has been shown to reduce movement artifacts and neural noise while not degrading response amplitudes or latencies (Pettigrew et al., 2004; Lavoie et al., 2008).

TABLE 1 Repeated-measures analysis of variance (ANOVA) results for cortical auditory evoked potentials (CAEP) sensor-level measures.

| | df | F | *p* |
|---|---|---|---|
| ITD | 4, 232 | 1.129 | 0.342 |
| Age | 1, 58 | **5.330** | **0.025** |
| ITD × Age | 4, 232 | 1.950 | 0.113 |
| P1 | | | |
| ITD | 4, 72 | 1.212 | 0.314 |
| Age | 1, 18 | 2.723 | 0.116 |
| ITD × Age | 4, 72 | 0.662 | 0.576 |
| N1 | | | |
| ITD | 4, 72 | 0.488 | 0.700 |
| Age | 1, 18 | 0.609 | 0.445 |
| ITD × Age | 4, 72 | 0.423 | 0.745 |
| P2 | | | |
| ITD | 4, 72 | 0.716 | 0.543 |
| Age | 1, 18 | 1.971 | 0.177 |
| ITD × Age | 4, 72 | 2.166 | 0.105 |

Significant F-statistic and corresponding *p*-values are shown in bold.

## 2.3. EEG data acquisition

Continuous electroencephalographic (EEG) responses were recorded using an ANT (Advanced Neuro-Technology BV)

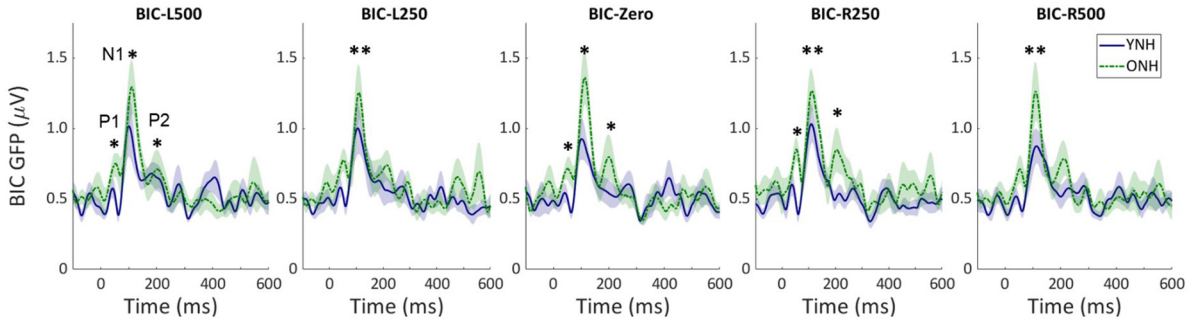



FIGURE 4
Grand average sensor-based global-field power (GFP) responses for the derived binaural interaction component (BIC) responses for each interaural time difference (ITD) condition and each listener group (YNH–blue solid, ONH–green dashed). The shaded region around each curve represents the ±1 standard error (SE) of the mean. The three main peak components are labeled in first panel (P1, N1, P2). Asterisks indicate significant group differences with the following *p*-values: $*p < 0.05$, $**p < 0.01$.

TABLE 2 Repeated-measures analysis of variance (ANOVA) results for binaural interaction component (BIC) sensor-level measures.

| | df | F | *p* |
|---|---|---|---|
| ITD | 4, 232 | 1.076 | 0.369 |
| Age | 1, 58 | **19.629** | **<0.001** |
| ITD × Age | 4, 232 | 1.948 | 0.113 |
| P1 | | | |
| ITD | 4, 72 | 1.192 | 0.322 |
| Age | 1, 18 | **5.418** | **0.032** |
| ITD × Age | 4, 72 | 0.707 | 0.590 |
| N1 | | | |
| ITD | 4, 72 | 0.479 | 0.751 |
| Age | 1, 18 | **8.774** | **0.008** |
| ITD × Age | 4, 72 | 0.433 | 0.739 |
| P2 | | | |
| ITD | 4, 72 | 0.713 | 0.586 |
| Age | 1, 18 | **5.043** | **0.038** |
| ITD × Age | 4, 72 | 2.172 | 0.081 |

Significant F-statistic and corresponding *p*-values are shown in bold.

high-speed amplifier and an active shield, WaveGuard cap with 64 sintered Ag/AgCl electrodes (International 10–20 electrode system). Four additional electrodes were placed at the outer canthus of each eye and on the supra and infra orbital ridges of the left eye to monitor eye movement and blink activity. Electrode impedance was maintained below 10 kΩ across all electrodes. The EEG was recorded at a sampling rate of 512 Hz with 24-bit resolution using asalab™ acquisition software (ANT). Stimulus generation, presentation and event triggering were controlled by custom MATLAB® (ver. R2018b) software scripts paired with asalab™ using activeX controls.

## 2.4. EEG data processing

All EEG data were preprocessed using the software suite Brainstorm (ver. brainstorm3, Tadel et al., 2011) and included the following steps: band-pass filtering (even-order linear phase FIR filter, based on a Kaiser window design) between 0.1 and 100 Hz, notch-filtering at 60 Hz (2nd order IIR notch filter with zero-phase lag, 2 Hz, 3-dB notch bandwidth), artifact detection to identify eye blinks, physical movement, and other extraneous activity (>150 μV), artifact removal *via* principal component analysis (PCA) and signal-space-projection (SSP), detrending to remove the DC signal, baseline correction (−100 to 0 ms), and re-referencing to the average across electrodes. Reponses were then epoched relative to stimulus onset (−200 to 600 ms). For sensor-level processing, epoched responses were averaged across trials (~120 per condition) for each subject and each condition, and global-field power (GFP; Skrandies, 1990) was computed across electrodes. To evaluate peak components of the CAEP, GFP maxima were obtained within predefined temporal windows corresponding to the following components: P1, 40–70 ms; N1, 80–130 ms; and P2, 160–240 ms. CAEP and GFP grand average waveforms were computed for each condition for listeners within each subject group.

The BIC responses were derived for all 64 sensors for each listener and each ITD condition. An exemplar of the derivation based on responses averaged across all listeners in the YNH (blue solid line) and ONH (green dashed line) groups for the ITD Zero condition is illustrated in **Figure 2**. The following steps were completed first for each subject before combining across subject group. First, the CAEP responses were averaged for each of the two monaural conditions (L, R) and were then added together (L + R). Next, the averaged binaural (BIN) response was subtracted from the summed monaural response to obtain the binaural interaction component [BIC = (R + L)–BIN]. These same steps were completed for each of the five ITD conditions. Like the CAEP peak quantification, BIC

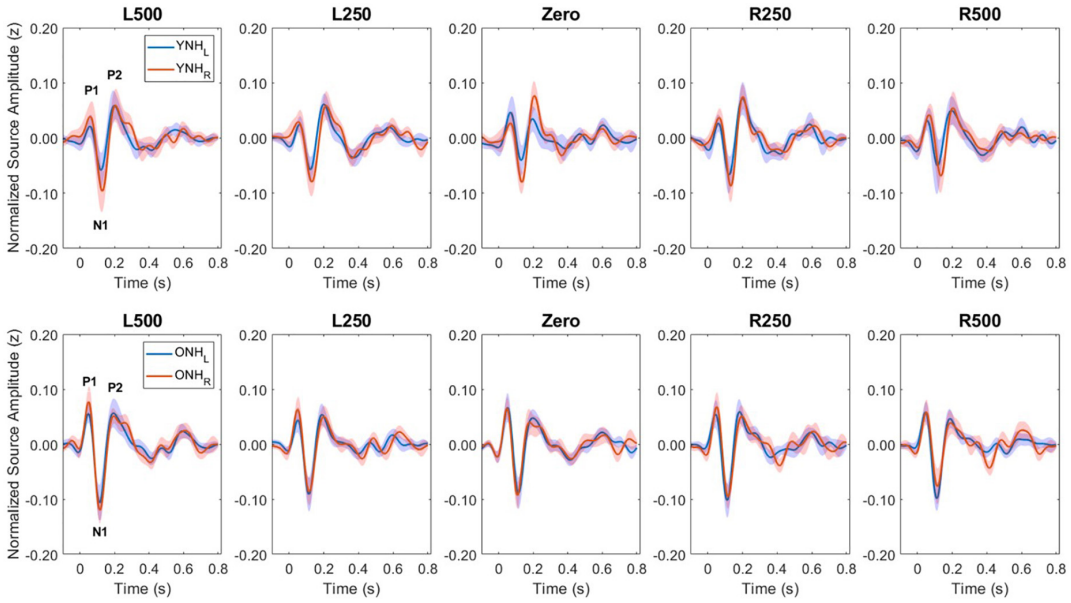



FIGURE 5
Grand average source-localized evoked responses from left (LH) and right (RH) hemisphere regions of interest (ROIs), for each interaural time difference (ITD) condition and each listener group (YNH, **top row**; ONH, **bottom row**). The three main peak components (P1, N1, P2) are labeled in first panel of each row.

maxima were determined during the same temporal windows corresponding to peak components P1, N1, and P2.

## 2.5. EEG source localization analysis

Source localization analyses are designed to make use of scalp-based sensor responses from many electrodes to estimate underlying brain activity from potentially thousands of locations–the so-called inverse problem. Several computationally efficient source localization methods are available that derive brain activity from a linear recombination of sensor recordings. In this study, cortical sources from ongoing EEG responses were estimated using dynamic Statistical Parametric Mapping (dSPM; Dale et al., 2000) as implemented in Brainstorm. dSPM uses minimum norm estimation (MNE) methods to determine current density maps, and then normalizes the maps relative to estimates of the noise covariance in the responses to produce a z-score statistical map. The sources were constrained to the volume of the cortex and mapped to the Montreal Neurological Institute (MNI) Colin27 brain template (Holmes et al., 1998) using a multi-linear registration technique within Brainstorm. This approach uses the open-source software, OpenMEEG (Kybic et al., 2005; Gramfort et al., 2010) and forward models generated with the symmetric boundary element method (BEM). The cortical surface was parcellated into regions of interest (ROIs) defined in the Destrieux structural atlas (Destrieux et al., 2010). The auditory cortex was defined by three ROIs in left and right hemispheres that encompassed Heschl's gyrus (HG, anterior transverse temporal gyrus), planum temporale (PT, temporal plane of the superior temporal gyrus), and the temporal sulcus (TS, transverse temporal sulcus). Similar to sensor-level analyses, source-localized waveforms were used to compute peak component maxima corresponding to P1, N1 and P2 for each listener and each condition. Likewise, source waveforms from monaural and binaural stimulus presentations were used to derive source-level BIC responses for each of the five ITD conditions.

## 2.6. Hemispheric asymmetry analysis

Using source-level data only, differences in hemispheric asymmetry with age and binaural condition for CAEP and BIC responses were quantified using a laterality index (LI) computed with the following equation:

$$LI = (|RH| - |LH|) / (|RH| + |LH|)$$

where, RH was the average response magnitude across the ROI sources in the right hemisphere and LH was the average magnitude of ROI sources in the left hemisphere. If LI = 0, then the magnitude of neural activity was essentially equivalent across hemispheres, whereas if LI > 0, dominant activity would be lateralized to the right hemisphere, and if LI < 0, dominant activity would be lateralized to the left hemisphere. The LI was computed based on the magnitude of the hemispheric activity

TABLE 3 Repeated-measures analysis of variance (ANOVA) results for cortical auditory evoked potentials (CAEP) source-localized measures.

| | df | F | *p* |
|---|---|---|---|
| ITD | 4, 712 | **7.826** | **<0.001** |
| Age | 1, 178 | 0.591 | 0.443 |
| Hemisphere | 1, 178 | 2.960 | 0.087 |
| ITD × Age | 4, 712 | **3.522** | **0.011** |
| ITD × Hemisphere | 4, 712 | **15.854** | **<0.001** |
| Hemisphere × Age | 1, 178 | 0.602 | 0.439 |
| ITD × Age × Hemisphere | 4, 712 | 1.217 | 0.303 |
| P1 | | | |
| ITD | 4, 232 | 0.367 | 0.832 |
| Age | 1, 58 | **8.804** | **0.004** |
| Hemisphere | 1, 58 | 1.378 | 0.245 |
| ITD × Age | 4, 232 | **5.812** | **0.001** |
| ITD × Hemisphere | 4, 232 | **13.121** | **<0.001** |
| Hemisphere × Age | 1, 58 | 0.153 | 0.697 |
| ITD × Age × Hemisphere | 4, 232 | **4.071** | **0.003** |
| N1 | | | |
| ITD | 4, 232 | **8.348** | **<0.001** |
| Age | 1, 58 | 0.025 | 0.874 |
| Hemisphere | 1, 58 | **13.548** | **0.001** |
| ITD × Age | 4, 232 | 1.841 | 0.151 |
| ITD × Hemisphere | 4, 232 | **13.175** | **<0.001** |
| Hemisphere × Age | 1, 58 | 0.644 | 0.426 |
| ITD × Age × Hemisphere | 4, 232 | **2.902** | **0.030** |
| P2 | | | |
| ITD | 4, 232 | **3.080** | **0.023** |
| Age | 1, 58 | 1.248 | 0.269 |
| Hemisphere | 1, 58 | 0.000 | 0.982 |
| ITD × Age | 4, 232 | 0.501 | 0.705 |
| ITD × Hemisphere | 4, 232 | 1.929 | 0.126 |
| Hemisphere × Age | 1, 58 | 0.801 | 0.375 |
| ITD × Age × Hemisphere | 4, 232 | 0.479 | 0.698 |

Significant F-statistic and corresponding *p*-values are shown in bold.

averaged across peak components as well as separately for each peak component (P1, N1, P2).

## 2.7. Statistical analysis

Both sensor- and source-level data were used to evaluate changes in CAEP and BIC component amplitudes (P1, N1, P2) between age groups, across ITD conditions, and for source-level data only, across left and right hemispheres. Statistical analyses were completed on both sensor- and source-level data using SPSS (version 27). A mixed-design analysis of variance (ANOVA) was used to evaluate CAEP and BIC response amplitudes (P1, N1, P2) as a function of within-subject factors of condition (5 ITDs) and hemisphere (left, right), and between-subject factor of age group (YNH, ONH). Additional *post-hoc* analyses were completed as appropriate using pairwise comparisons with Bonferroni correction for multiple comparisons. To reduce Type I errors, all reported *F* values include degrees of freedom adjustments using Greenhouse-Geiser correction when significant deviations from sphericity were observed based on Mauchly's test (Greenhouse and Geisser, 1959).

# 3. Results

## 3.1. Sensor-based measures of ITD processing: CAEP and BIC

Low-frequency noise burst stimuli elicited transient neural responses with peaks corresponding to the P1-N1-P2 complex of the CAEP. The CAEPs from the 64 electrodes were used to compute the GFP for each listener and each stimulus condition. Figure 3 shows the grand average GFP across listeners for each group (YNH–left panel, ONH–right panel), for the five ITD conditions, with peak components labeled (P1-N1-P2). Two clear observations can be made when comparing responses across the two panels. First, older listeners demonstrated larger amplitude responses than younger listeners and little variation in amplitude across ITD conditions. Younger listeners, on the other hand, showed some amplitude variation with ITD conditions, most noticeably for N1, where the two extreme ITDs (L500, R500) produced the most robust responses. To quantify the observed differences, GFP amplitudes (combined across components P1, N1, P2) were submitted to a repeated-measures ANOVA to evaluate the between-subject factor of age group (YNH, ONH) and within-subject factor of ITD condition (L500, L250, Zero, R250, R500). The statistical results are reported in Table 1 and showed a significant main effect of age group [$F(1,58) = 5.33$, $p = 0.025$], supporting the observation that older listeners had larger responses overall than younger listeners. Second, despite the modest variation in response amplitude with changes in ITD for younger listeners, there was no significant main effect of ITD [$F(4,232) = 1.13$, $p = 0.344$] and no significant interaction between ITD and group [$F(4,232) = 1.95$, $p = 0.113$] on response amplitudes. GFP amplitudes were also evaluated independently for each peak component (P1, N1, P2) to assess the effects of age group and ITD condition. As reported in Table 1, repeated-measures ANOVA showed no significant main effect of age group on any of the three peak components (P1, N1, or P2), nor any main effect of ITD condition, and no significant interactions among the two factors.

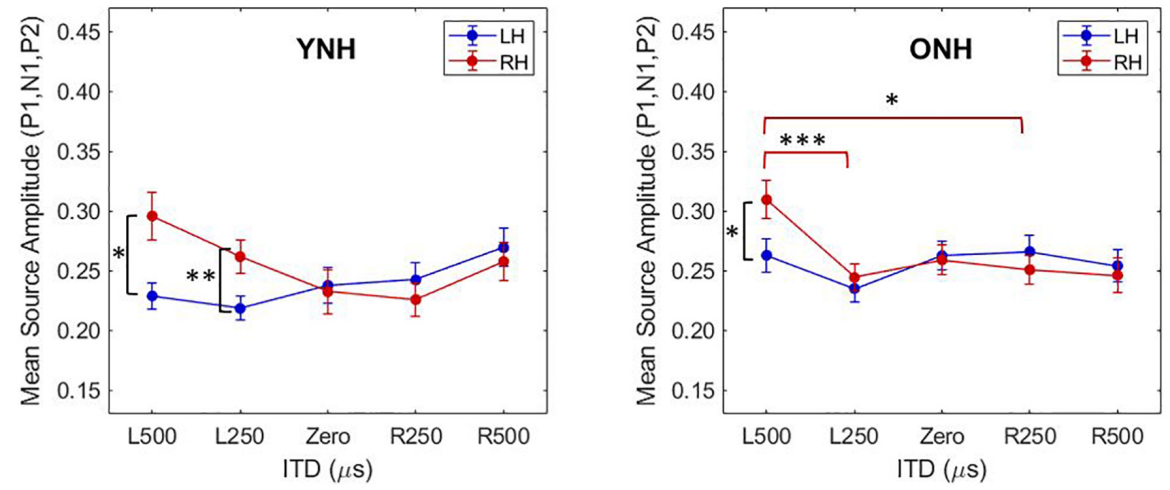



FIGURE 6
Mean source amplitudes averaged across absolute values of peak components (P1, N1, P2) in each hemisphere for each interaural time difference (ITD) condition and listener group (YNH, **left panel**; ONH, **right panel**). Significant differences between hemispheres were observed for both YNH (L500, L250) and ONH (L500) groups, and between conditions within right hemisphere only for ONH. Asterisks indicate significance levels: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

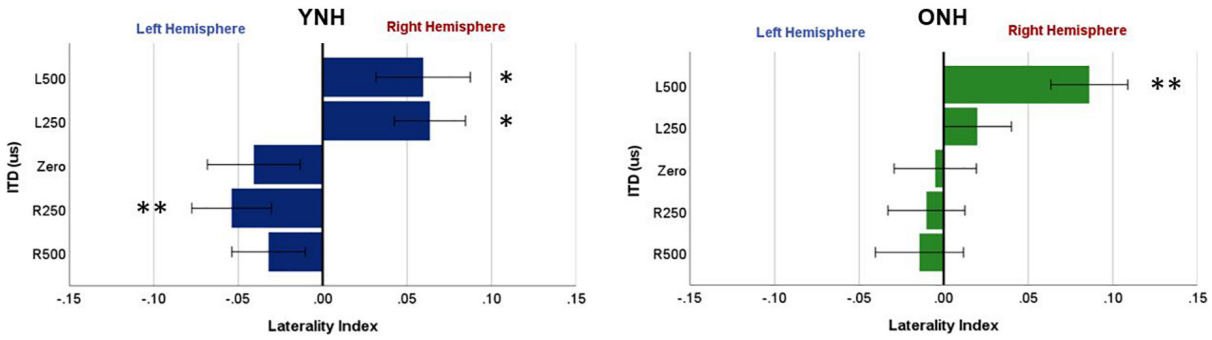



FIGURE 7
Laterality index quantified from source-localized left and right hemisphere regions of interest (ROIs) for each interaural time difference (ITD) condition averaged across each group (YNH, **left panel**; ONH, **right panel**). Significant laterality (relative to zero) for given ITD conditions is indicated for the following $p$-values: $*p < 0.05$, $**p < 0.01$.

Similar to the CAEP results, the BIC derivation, as illustrated in Figure 2, showed that each response contributing to the derivation was larger in amplitude for the ONH than the YNH group, particularly for the summed monaural (L + R) responses. Replotting the BIC responses as GFP, Figure 4 shows a similar comparison between age groups across the five ITD conditions. In each condition, the mean responses for the ONH group (green dashed lines) consistently showed larger BIC amplitudes than the YNH group (blue solid lines). The shaded regions around each mean response function represents ±1 standard error of the mean (SEM). To evaluate the statistical significance of these observed differences, a repeated-measures ANOVA was completed to assess differences in the overall response amplitude (averaged across peak components) as well as differences for each peak component independently. As reported in Table 2, there was a significant main effect of age on overall response amplitude [$F(1,58) = 19.629$, $p < 0.001$], but no significant main effect of ITD condition [$F(4,232) = 1.076$, $p = 0.369$]. When evaluating effects of age and ITD condition on each peak component of the BIC, age had a significant effect on component amplitudes (see Table 2). *Post-hoc* pairwise comparisons, with Bonferroni correction for multiple comparisons, were completed to determine if differential effects of age might be observed for specific peak components across ITD conditions. As shown by the asterisks in Figure 4, significant age effects were observed for N1 across all five ITD conditions (L500, Zero, $p < 0.05$; L250, R250, R500, $p < 0.01$), whereas P1 and P2 showed significant age effects only for L500, Zero, and R250 ($p < 0.05$). Although age was shown to be a factor for sensor-based BIC amplitudes, ITD alone did not have a significant main

TABLE 4 Repeated-measures analysis of variance (ANOVA) results for source-localized laterality index (LI) measures.

| | df | F | *p* |
|---|---|---|---|
| ITD | 4, 232 | **4.843** | **0.002** |
| Age | 1, 58 | 0.024 | 0.877 |
| ITD × Age | 4, 232 | 0.518 | 0.693 |
| P1 | | | |
| ITD | 4, 72 | **3.204** | **0.018** |
| Age | 1, 18 | 0.142 | 0.711 |
| ITD × Age | 4, 72 | 1.598 | 0.184 |
| N1 | | | |
| ITD | 4, 72 | **3.270** | **0.016** |
| Age | 1, 18 | 0.065 | 0.801 |
| ITD × Age | 4, 72 | 1.005 | 0.411 |
| P2 | | | |
| ITD | 4, 72 | 1.235 | 0.304 |
| Age | 1, 18 | 0.005 | 0.945 |
| ITD × Age | 4, 72 | 0.380 | 0.822 |

Significant F-statistic and corresponding *p*-values are shown in bold.

effect on BIC response amplitudes nor was there any significant interaction between age and ITD condition.

## 3.2. Source-localized CAEP measures of ITD processing

A primary goal of this investigation was to evaluate potential age-related differences in hemispheric asymmetry during binaural processing. To do so, neural activity was quantified for source-localized responses derived from scalp-based responses using dSPM methods (Dale et al., 2000). Source responses were computed for three regions of interest (ROI) encompassing the primary auditory cortex in each hemisphere. Given that we did not obtain individual MRI scans from each participant but instead used the MNI Colin27 brain template along with the Destrieux atlas provided in Brainstorm, we chose to average responses across the three ROIs in each hemisphere and compute differences more broadly between left and right hemispheres (LH, RH).

Figure 5 illustrates the normalized source waveforms averaged across the three ROIs for left (blue lines and shading) and right (red lines and shading) hemisphere, with shading around each waveform corresponding to ±1 SEM. Responses are shown for younger (YNH–top panels) and older listeners (ONH–bottom panels) as a function of ITD condition. Unlike the sensor-based CAEP responses, the source-localized response amplitudes are more similar between the two groups and ITD conditions. A mixed model repeated-measures ANOVA was completed to evaluate differences in response amplitudes due to a between-subject factor of age group (YNH, ONH) and within-subject factors of ITD condition (L500, L250, Zero, R250, R500) and hemisphere (Left, Right). Analyses were completed based on amplitudes averaged across peak components (P1, N1, P2) and separately for each peak component. The results of those analyses are reported in Table 3.

To better appreciate the overall differences in response magnitude between hemispheres for each ITD condition and each group, absolute values of the amplitudes for primary peak components (P1, N1, P2) were averaged and plotted by ITD, hemisphere and group, as shown in Figure 6. Consistent with the contralateral bias in binaural processing, young adults (left panel) showed greater right hemisphere activity for left leading ITDs (i.e., L500, L250) and greater left hemisphere activity for right leading ITDs (R500, R25), albeit somewhat smaller hemispheric bias for right leading stimuli. Older adults showed similar patterns, but smaller hemispheric differences across all ITDs. As reported in Table 3, statistically significant results were observed for the main effect of ITD [$F(4,712) = 7.826$, $p < 0.001$], as well as significant interactions between ITD and age [$F(4,712) = 3.522$, $p = 0.011$] and notably, between ITD and hemisphere [$F(4,712) = 15.854$, $p < 0.001$]. *Post-hoc* pairwise comparisons with Bonferroni correction showed that significant hemispheric differences were present in the YNH group for ITD conditions of L500 ($p < 0.05$) and L250 ($p < 0.01$) but only for L500 ($p < 0.05$) in the ONH group (see Figure 6). Further, significant differences in the ONH group were present in the right hemisphere only between the ITD conditions of L500 and L250 ($p < 0.01$) and between L500 and R250 ($p < 0.05$). Although not illustrated in a graphic format but reported in Table 3, analyses completed for each peak component showed significant three-way interactions between ITD, age and hemisphere for both P1 [$F(4,232) = 4.071$, $p = 0.003$] and N1 [$F(4,232) = 2.902$, $p = 0.030$] components. These results indicate a relatively complex relationship regarding how ITD cues are processed between hemispheres, over time (i.e., latency-based peak components) across age groups.

To further examine these complexities, we evaluated contralateral bias with the laterality index (LI) measure. The absolute values of peak component amplitudes in left and right hemisphere ROIs were averaged and used to compute the LI for each participant and each ITD condition. Figure 7 shows the mean LI results for YNH (left panel) and ONH (right panel) groups for each of the ITD conditions. Consistent with the contralateral bias model, left-leading ITDs produced greater lateralization toward the right hemisphere, whereas right-leading ITDs produced greater lateralization toward the left hemisphere. Statistical analyses based on a mixed model repeated measures ANOVA are reported in Table 4. The results showed that ITD had a significant effect on laterality [$F(4,232) = 4.843$, $p = 0.002$], but no significant differences were observed across age groups and no significant interactions between ITD and age were measured. Although both age groups

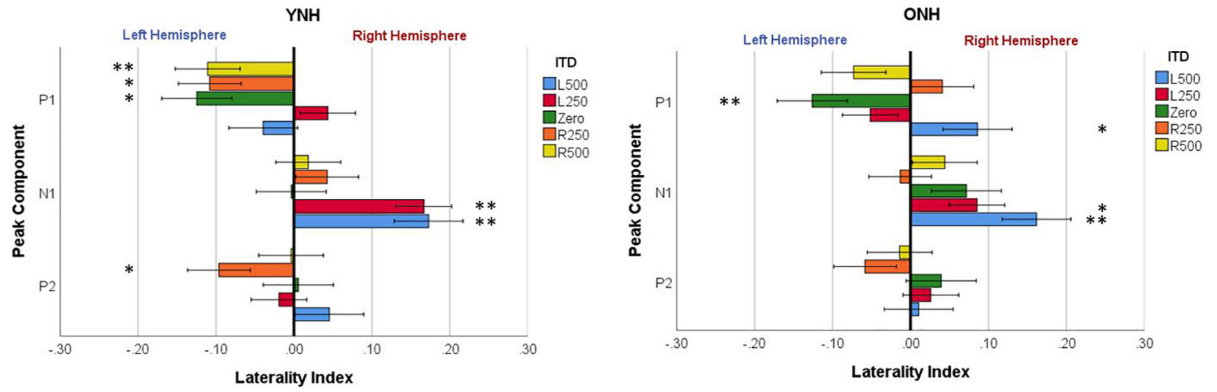



FIGURE 8
Mean laterality index for each interaural time difference (ITD) condition plotted relative to latency of peak components (P1, N1, P2) for YNH **(left panel)** and ONH **(right panel)** groups. Significant effect of peak component on laterality plotted relative to ITD condition is indicated for the following *p*-values: $^{*}p < 0.05$, $^{**}p < 0.01$.

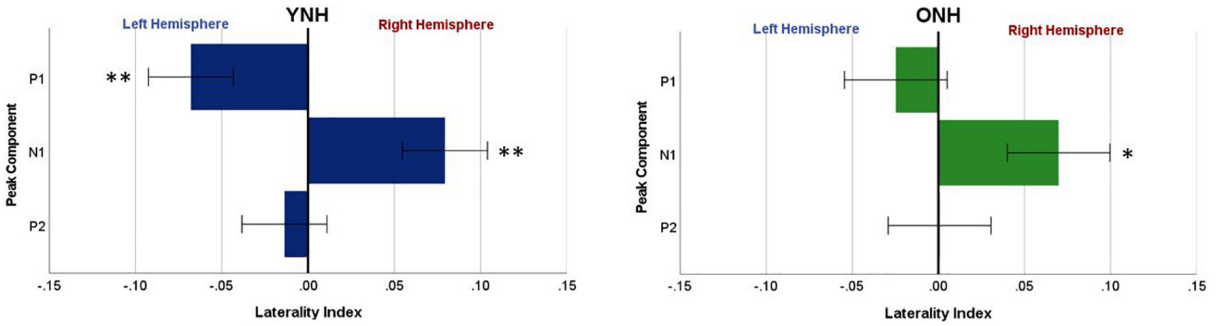



FIGURE 9
Mean laterality index, averaged across interaural time difference (ITD), plotted relative to latency of peak components (P1, N1, P2) for YNH **(left panel)** and ONH **(right panel)** groups. Significant laterality (relative to zero) for given components for each group separately is indicated for the following *p*-values: $^{*}p < 0.05$, $^{**}p < 0.01$.

demonstrated similar patterns across ITD, a one sample *t*-test was used to determine the extent to which each ITD condition produced significant asymmetry relative to zero. As indicated by the asterisks in Figure 7, the YNH group had statistically significant right lateralized activity for left-leading ITDs of L500 and L250 ($^{*}p < 0.05$), and significant left lateralized activity for one right-leading ITD of R250 ($^{**}p < 0.01$). The ONH group, on the other hand, only produced significant right hemisphere laterality for one left-leading L500 condition ($^{**}p < 0.01$). These results are consistent with reduced hemispheric asymmetry in the older listeners during ITD processing.

To evaluate further the potential dynamic nature of lateralization over time, LI was quantified for each ITD condition in relation to the temporal sequence of CAEP peak components (P1, N1, P2). As indicated above, a one-sample *t*-test was used to determine which ITD conditions produced asymmetry relative zero for P1, N1, and P2 for each group. Figure 8 shows that some ITD conditions were lateralized differentially based on timing of the peak component for both YNH (left panel) and ONH (right panel) groups. For the YNH group, P1 was significantly left lateralized most notably for right-leading ITDs (R500, L250, Zero), whereas N1 was right lateralized for left-leading ITDs (L500, L250) and P2 shifted back to the left lateralization for one right-leading condition (R250). For the ONH group, significant laterality was observed in four ITD conditions across P1 and N1 peaks but was less orderly than the dynamic shifts observed for the YNH group. When averaged across ITD conditions, Figure 9 illustrates more directly how hemispheric laterality varied by timing of peak components. One-sample *t*-tests revealed that both groups had a similar dynamic pattern of left to right lateralization for P1 and N1, respectively, but the ONH group had reduced and non-significant P1 lateralization as compared to the YNH group. The statistical results of the repeated measures ANOVA evaluating laterality for each peak component for effects of ITD and age group are reported in Table 4. The results demonstrate a significant effect of ITD for P1 and N1 (as shown in Figure 8) but no significant effect of age group or interaction between age

and ITD for any peak. Thus, based on laterality index measures, hemispheric dynamics during binaural temporal processing are influenced not only by the ITD stimulus condition but also by the time interval of the evoked response. Notably, the relationship between these factors is diminished in older relative to younger adults.

## 4. Discussion and conclusion

The overall goal of this project was to better understand the impact of advancing age on binaural processing and specifically on the encoding of binaural cues that support such processing. Here, we investigated the effects of advancing age on neural encoding of low-frequency ITD cues and the commonly observed contralateral bias in cortical processing, as indexed by both CAEP and BIC measures. The study design also allowed for assessment of the HAROLD hypothesis of reduced hemispheric asymmetry (or contralateral bias) as it might apply to auditory sensory processing in older adults during ITD processing.

### 4.1. Age-related changes in cortical processing of ITDs

Binaural processing, as evaluated for a range of different measures, is often degraded with advancing age (e.g., Dubno et al., 2008; Eddins and Hall, 2010; Ozmeral et al., 2016; Eddins and Eddins, 2018; Eddins et al., 2018; Gallun and Best, 2020). Although such degradation likely impacts everyday activities such as speech understanding in noisy backgrounds, the nature of such age-related processing changes and their underlying mechanisms are not fully understood or characterized (Gallun and Best, 2020). The present study was designed specifically to investigate the impact of aging on processing of low-frequency dominant ITD cues and their cortical representation in both younger and older normal-hearing listeners. Based on cortical responses to passively presented static ITDs, the results clearly demonstrated that grand average sensor-level CAEP responses were significantly larger for older than younger normal-hearing listeners across ITD conditions (see **Figure 3** and **Table 1**). The responses for both groups, however, were not systematically altered by ITD. Additionally, when evaluating responses by peak component (P1, N1, P2), there were no systematic differences based on age or ITD condition (see **Table 1**).

The overall enhanced responses with age to static ITD stimuli are consistent with a down-regulation of inhibitory processing, as suggested by previous animal studies (e.g., Willott et al., 1997; Caspary et al., 2008) as well as human evoked potential studies of binaural processing (e.g., Eddins et al., 2018). This is in contrast to alternative age-related changes thought to result from reduced temporal synchrony (or increased temporal jitter), which would lead to smaller response amplitudes in older versus younger adults (e.g., Pichora-Fuller et al., 2007; Ozmeral et al., 2016). Amplitude reductions in the CAEP in older versus younger adults were observed in previous studies in response to dynamic changes in consecutive ITDs (Ozmeral et al., 2016). These changes were attributed to reduced temporal synchrony in the older group. Likewise, age-related reductions in CAEP amplitudes in older versus younger adults also have been observed during selective attention to spatial changes in sound location using comparable low-frequency ITD stimulus conditions (Ozmeral et al., 2021). Thus, the impact of age on whether CAEP amplitudes are enhanced or reduced appears to be more related to the nature of the stimulus presentation (static versus dynamic) and task demands (passive versus attention) rather than the attributes of the binaural stimulus *per se* (i.e., ITD).

To determine if other neural measures of binaural processing might shed more light on potential age-related differences in underlying function, the BIC was derived from sensor-level CAEP measures for each ITD condition. Like the CAEP analyses, the BIC results (when averaged across peak components) also demonstrated significant amplitude differences between age groups, where older listeners revealed significantly larger BIC responses across all ITD conditions (see **Figure 4** and **Table 2**), yet ITD itself did not produce differences in amplitude for either group nor was there a significant interaction between ITD and age group. When BIC responses were analyzed independent by peak component (P1, N1, P2), the main effect of age was equally robust and significant, whereas ITD did not produce differences in amplitude nor were there significant interactions between ITD and age group for any of the peak components. To our knowledge, no previous studies of CAEP-based BIC measures with changes in ITD have been reported. Studies investigating ABR- and MLR-based BIC measures over a range ITDs, however, have shown mixed results in terms of how BIC responses change with ITD. Some studies have reported decreased BIC amplitudes and increased latencies with increasing ITD (McPherson and Starr, 1995; Riedel and Kollmeier, 2006), while a more recent normative study of ABR BIC showed no significant change in amplitude with ITD and substantial variability across a group of 40 young to middle-age normal-hearing participants (Sammeth et al., 2020). Differences in results between the present CAEP-based BIC responses and those from ABR- and MLR-based BIC measures are not surprising given the differences in the location of anatomical generators and additional contributors along the auditory pathway. Cortical measures in the present study likely have their origin in the brainstem, but they may reflect a decrease in temporal precision of the onset response to ITDs due to additional synaptic connections between the brainstem and cortex. The reduced temporal precision in CAEP compared to ABR or MLR BIC measures may then lead to smaller amplitude differences between ITD conditions. Nonetheless, both sensor-level CAEP and BIC analyses in the present study clearly

demonstrate that advancing age leads to enhanced cortical responses to low-frequency ITD stimuli when presented in a passive, static mode with no systematic response variation across ITD conditions.

## 4.2. Age-related changes in hemispheric asymmetry during ITD processing

An important focus of this study was to test the hypothesis that advancing age in adults with normal hearing based on pure-tone thresholds leads to measurable changes in the expected contralateral bias in hemispheric processing of ITD cues often observed in studies of binaural processing (von Kriegstein et al., 2008; Salminen et al., 2009; Briley et al., 2013; Gutschalk and Steinmann, 2015; Stecker et al., 2015; McLaughlin et al., 2016; Ozmeral et al., 2016). Based on source-localized responses quantified from ROIs encompassing the primary auditory cortex in each hemisphere, the results showed that indeed ITD cues did elicit contralateral bias with greater response magnitudes in right hemisphere for left-leading ITDs and in left hemisphere for right-leading ITDs (see **Figures 5**, **6**). These results are consistent with previous studies that have assessed both interaural timing and level difference encoding in the cortex using different neuroimaging methodology (Stecker et al., 2015; McLaughlin et al., 2016). Statistical analyses of data from the present study indicate that not only was there a main effect of ITD on CAEP source response magnitudes, but significant interactions also were revealed between ITD and hemisphere as well as between ITD and age group (see **Table 3**). These results demonstrate that ITD processing varies across hemisphere and that cortical processing of ITDs is differentially impacted by age. Notably, this novel outcome shows that advancing age leads to reduced hemispheric differences in source magnitudes when processing ITD cues (see **Figure 6** and **Table 3**).

The most robust characterization of changes in hemispheric asymmetry was revealed with the laterality index analyses, as illustrated in **Figures 7–9**. First, both younger and older adults showed contralateral bias for some ITD conditions, as would be predicted from previous studies. Importantly, however, **Figure 7** shows that younger listeners had significant laterality for both left- (L500, L250) and right-leading (R250) ITDs, whereas older listeners only showed significant laterality for one left-leading condition (L500). This novel demonstration of age effects on the laterality of cortical processing of ITDs is consistent with the HAROLD model such that older adults have reduced hemispheric asymmetry during binaural processing of ITD cues. Given that these stimuli were presented in a passive listening modality, it is less likely that the observed reduction in asymmetry results from compensatory mechanisms but instead may be linked to dedifferentiation in cortical processing of ITDs. Further studies are warranted to confirm the mechanistic source(s) of this age-related change. In addition, although the average hearing thresholds for both groups were within about 5 dB HL of one another at the frequency region of interest (500 to 750 Hz), there were differences between groups for higher frequency thresholds (≥4,000 Hz). Thus, the minimal differences in hearing sensitivity at 500 Hz and the influence of slightly poorer hearing two octaves above (i.e., ≥4,000 Hz) in the older group cannot be ruled out as a contributing factor to the observed age-related differences. Future work with clinically-significant hearing loss in the frequency region of interest can more definitively establish the impacts of typical age-related hearing loss.

Another novel outcome from this investigation, also related to hemispheric asymmetry, was revealed when examining the laterality index by each peak component of the CAEP source-localized response. When evaluating laterality by both ITD condition and peak component (see **Figure 8**), the LI results clearly demonstrated that hemispheric laterality shifts dynamically over time relative to the latency of the peak component of the evoked response. That is, the early response corresponding to P1 (~40–70 ms post-stimulus onset) was significantly lateralized toward the left hemisphere and was driven largely by right-leading stimuli (e.g., R500, R250), whereas laterality during N1 (~80–130 ms post-stimulus onset) was lateralized toward the right hemisphere and driven primarily by left-leading ITDs (L500, L250). P2 (~160–240 ms), on the other hand, showed significant lateralization for only one ITD condition (R250) and one group (YNH). These results support the possibility that contralateral bias in hemispheric processing of ITD cues may be hierarchically processed such that right hemifield stimuli are processed earlier than left hemifield stimuli. Furthermore, as illustrated in **Figure 9**, both age groups showed similar laterality patterns as a function of peak component latencies, albeit older listeners showed reduced and non-significant laterality corresponding to P1. Taken together, the overall laterality results provide robust evidence of an age-related reduction in hemispheric asymmetry during ITD processing that is dynamically influenced over the time frame of the cortical evoked response. The relevance of such a binaural temporal processing scheme within and across hemispheres warrants further exploration and evaluation.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by University of South Florida Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

AE wrote the initial draft of the manuscript with editorial contributions from EO and DE. All authors contributed equally to the experimental design, implementation, data collection, analysis, and interpretation.

## Funding

This work was supported in part by NIH NIDCD F32 DC013724 (EO) and NIH NIA P01 AG009524 (DE and AE).

## Acknowledgments

We are grateful for the valuable contributions of Alicia Durkin and Ionna Tagarelli Christiansen during the data collection for this project.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

Benichoux, V., Ferber, A., Hunt, S., Hughes, E., and Tollin, D. (2018). Across species "natural ablation" reveals the brainstem source of a noninvasive biomarker of binaural hearing. *J. Neurosci.* 38, 8563–8573. doi: 10.1523/JNEUROSCI.1211-18.2018

Brand, A., Behrend, O., Marquardt, T., McAlpine, D., and Grothe, B. (2002). Precise inhibition is essential for microsecond interaural time difference coding. *Nature* 417, 543–547. doi: 10.1038/417543a

Briley, P. M., Kitterick, P. T., and Summerfield, A. Q. (2013). Evidence for opponent process analysis of sound source location in humans. *J. Assoc. Res. Otolaryngol.* 14, 83–101. doi: 10.1007/s10162-012-0356-x

Brown, A. D., Anbuhl, K. L., Gilmer, J. I., and Tollin, D. J. (2019). Between-ear sound frequency disparity modulates a brain stem biomarker of binaural hearing. *J. Neurophysiol.* 122, 1110–1122. doi: 10.1152/jn.00057.2019

Cabeza, R. (2002). Hemispheric asymmetry reduction in older adults: The HAROLD model. *Psychol. Aging* 17:85. doi: 10.1037/0882-7974.17.1.85

Cabeza, R., Grady, C. L., Nyberg, L., McIntosh, A. R., Tulving, E., Kapur, S., et al. (1997). Age-related differences in neural activity during memory encoding and retrieval: A positron emission tomography study. *J. Neurosci.* 17, 391–400. doi: 10.1523/JNEUROSCI.17-01-00391.1997

Caspary, D. M., Ling, L., Turner, J. G., and Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *J. Exp. Biol.* 211, 1781–1791. doi: 10.1242/jeb.013581

Dale, A. M., Liu, A. K., Fischl, B. R., Buckner, R. L., Belliveau, J. W., Lewine, J. D., et al. (2000). Dynamic statistical parametric mapping: Combining fMRI and MEG for high-resolution imaging of cortical activity. *Neuron* 26, 55–67. doi: 10.1016/S0896-6273(00)81138-1

Destrieux, C., Fischl, B., Dale, A., and Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* 53, 1–15. doi: 10.1016/j.neuroimage.2010.06.010

Dobie, R. A., and Berlin, C. I. (1979). Binaural interaction in brainstem-evoked responses. *Arch. Otolaryngol.* 105, 391–398. doi: 10.1001/archotol.1979.00790190017004

Dubno, J. R., Ahlstrom, J. B., and Horwitz, A. R. (2008). Binaural advantage for younger and older adults with normal hearing. *J. Speech Lang Hear Res.* 51, 539–556. doi: 10.1044/1092-4388(2008/039)

Dykstra, A. R., Burchard, D., Starzynski, C., Riedel, H., Rupp, A., and Gutschalk, A. (2016). Lateralization and binaural interaction of middle-latency and late-brainstem components of the auditory evoked response. *J. Assoc. Res. Otolaryngol.* 17, 357–370. doi: 10.1007/s10162-016-0572-x

Eddins, A. C., and Eddins, D. A. (2018). Cortical correlates of binaural temporal processing deficits in older adults. *Ear Hear.* 39:594. doi: 10.1097/AUD.0000000000000518

Eddins, A. C., Ozmeral, E. J., and Eddins, D. A. (2018). How aging impacts the encoding of binaural cues and the perception of auditory space. *Hear. Res.* 369, 79–89. doi: 10.1016/j.heares.2018.05.001

Eddins, D. A., and Hall, J. W. (2010). "Binaural processing and auditory asymmetries," in *The aging auditory system*, eds S. Gordon-Salant, D. R. Frisina, A. N. Popper, and R. R. Fay (New York, NY: Springer), 135–165. doi: 10.1007/978-1-4419-0993-0_6

Festini, S. B., Zahodne, L., and Reuter-Lorenz, P. A. (2018). "Theoretical perspectives on age differences in brain activation: HAROLD, PASA, CRUNCH–how do they STAC up?," in *Oxford research encyclopedia of psychology*, (Oxford: Oxford University Press). doi: 10.1093/acrefore/9780190236557.013.400

Fowler, C. G., and Horn, J. H. (2012). Frequency dependence of binaural interaction in the auditory brainstem and middle latency responses. *Am. J. Audiol.* 21, 190–198. doi: 10.1044/1059-0889(2012/12-0006)

Gallun, F. J., and Best, V. (2020). "Age-related changes in segregation of sound sources," in *Aging and hearing*, eds K. S. Helfer, E. L. Bartlett, A. N. Popper, and R. R. Fay (Cham: Springer), 143–171. doi: 10.1007/978-3-030-49367-7_7

Gaumond, R. P., and Psaltikidou, M. (1991). Models for the generation of the binaural difference response. *J. Acoust. Soc. Am.* 89, 454–456. doi: 10.1121/1.400482

Goksoy, C., Demirtas, S., Yagcioglu, S., and Ungan, P. (2005). Interaural delay-dependent changes in the binaural interaction component of the guinea pig brainstem responses. *Brain Res.* 1054, 183–191. doi: 10.1016/j.brainres.2005.06.083

Gordon, K. A., Salloum, C., Toor, G. S., van Hoesel, R., and Papsin, B. C. (2012). Binaural interactions develop in the auditory brainstem of children who are deaf: Effects of place and level of bilateral electrical stimulation. *J. Neurosci.* 32, 4212–4223. doi: 10.1523/JNEUROSCI.5741-11.2012

Gramfort, A., Papadopoulo, T., Olivi, E., and Clerc, M. (2010). OpenMEEG: Opensource software for quasistatic bioelectromagnetics. *Biomed. Eng. Online* 9, 1–20. doi: 10.1186/1475-925X-9-45

Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika* 24, 95–112. doi: 10.1007/BF02289823

Gutschalk, A., and Steinmann, I. (2015). Stimulus dependence of contralateral dominance in human auditory cortex. *Hum. Brain Mapp.* 36, 883–896. doi: 10.1002/hbm.22673

Hancock, K. E., Noel, V., Ryugo, D. K., and Delgutte, B. (2010). Neural coding of interaural time differences with bilateral cochlear implants: Effects of congenital deafness. *J. Neurosci.* 30, 14068–14079. doi: 10.1523/JNEUROSCI.3213-10.2010

Haywood, N. R., Undurraga, J. A., Marquardt, T., and McAlpine, D. (2015). A comparison of two objective measures of binaural processing: The interaural phase modulation following response and the binaural interaction component. *Trends Hear.* 19:2331216515619039. doi: 10.1177/2331216515619039

He, S., Brown, C. J., and Abbas, P. J. (2010). Effects of stimulation level and electrode pairing on the binaural interaction component of the electrically evoked auditory brain stem response. *Ear Hear.* 31:457. doi: 10.1097/AUD.0b013e3181d5d9bf

Henkin, Y., Yaar-Soffer, Y., Givon, L., and Hildesheimer, M. (2015). Hearing with two ears: Evidence for cortical binaural interaction during auditory processing. *J. Am. Acad. Audiol.* 26, 384–392. doi: 10.3766/jaaa.26.4.6

Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., and Evans, A. C. (1998). Enhancement of MR images using registration for signal averaging. *J. Comput. Assist. Tomogr.* 22, 324–333. doi: 10.1097/00004728-199803000-00032

Hu, H., and Dietz, M. (2015). Comparison of interaural electrode pairing methods for bilateral cochlear implants. *Trends Hear.* 19:2331216515617143. doi: 10.1177/2331216515617143

Hu, H., Kollmeier, B., and Dietz, M. (2016). "Suitability of the binaural interaction component for interaural electrode pairing of bilateral cochlear implants," in *Physiology, psychoacoustics and cognition in normal and impaired hearing*, eds P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, and C. Lanting (Cham: Springer), 57–64. doi: 10.1007/978-3-319-25474-6_7

Jeffress, L. A. (1948). A place theory of sound localization. *J. Comp. Physiol. Psychol.* 41:35. doi: 10.1037/h0061495

Konishi, M. (2003). Coding of auditory space. *Ann. Rev. Neurosci.* 26, 31–55. doi: 10.1146/annurev.neuro.26.041002.131123

Kybic, J., Clerc, M., Abboud, T., Faugeras, O., Keriven, R., and Papadopoulo, T. (2005). A common formalism for the integral formulations of the forward EEG problem. *IEEE Trans. Med. Imaging* 24, 12–28. doi: 10.1109/TMI.2004.837363

Laumen, G., Ferber, A. T., Klump, G. M., and Tollin, D. J. (2016). The physiological basis and clinical use of the binaural interaction component of the auditory brainstem response. *Ear Hear.* 37:e276. doi: 10.1097/AUD.0000000000000301

Lavoie, B. A., Hine, J. E., and Thornton, R. D. (2008). The choice of distracting task can affect the quality of auditory evoked potentials recorded for clinical assessment. *Int. J. Audiol.* 47, 439–444. doi: 10.1080/14992020802033109

Learmonth, G., Benwell, C. S., Thut, G., and Harvey, M. (2017). Age-related reduction of hemispheric lateralisation for spatial attention: An EEG study. *Neuroimage* 153, 139–151. doi: 10.1016/j.neuroimage.2017.03.050

Magezi, D. A., and Krumbholz, K. (2010). Evidence for opponent-channel coding of interaural time differences in human auditory cortex. *J. Neurophysiol.* 104, 1997–2007. doi: 10.1152/jn.00424.2009

McAlpine, D. (2005). Creating a sense of auditory space. *J. Physiol.* 566, 21–28. doi: 10.1113/jphysiol.2005.083113

McLaughlin, S. A., Higgins, N. C., and Stecker, G. C. (2016). Tuning to binaural cues in human auditory cortex. *J. Assoc. Res. Otolaryngol.* 17, 37–53. doi: 10.1007/s10162-015-0546-4

McPherson, D. L., and Starr, A. (1993). Binaural interaction in auditory evoked potentials: Brainstem, middle-and long-latency components. *Hear. Res.* 66, 91–98. doi: 10.1016/0378-5955(93)90263-Z

McPherson, D. L., and Starr, A. (1995). Auditory time-intensity cues in the binaural interaction component of the auditory evoked potentials. *Hear. Res.* 89, 162–171. doi: 10.1016/0378-5955(95)00134-1

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x

Ozmeral, E. J., Eddins, D. A., and Eddins, A. C. (2016). Reduced temporal processing in older, normal-hearing listeners evident from electrophysiological responses to shifts in interaural time difference. *J. Neurophysiol.* 116, 2720–2729. doi: 10.1152/jn.00560.2016

Ozmeral, E. J., Eddins, D. A., and Eddins, A. C. (2019). Electrophysiological responses to lateral shifts are not consistent with opponent-channel processing of interaural level differences. *J. Neurophysiol.* 122, 737–748. doi: 10.1152/jn.00090.2019

Ozmeral, E. J., Eddins, D. A., and Eddins, A. C. (2021). Selective auditory attention modulates cortical responses to sound location change in younger and older adults. *J. Neurophysiol.* 126, 803–815. doi: 10.1152/jn.00609.2020

Pettigrew, C. M., Murdoch, B. E., Ponton, C. W., Kei, J., Chenery, H. J., and Alku, P. (2004). Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. *J. Am. Acad. Audiol.* 15, 469–485. doi: 10.3766/jaaa.15.7.2

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., and Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hear. Res.* 223, 114–121. doi: 10.1016/j.heares.2006.10.009

Riedel, H., and Kollmeier, B. (2002). Comparison of binaural auditory brainstem responses and the binaural difference potential evoked by chirps and clicks. *Hear. Res.* 169, 85–96. doi: 10.1016/S0378-5955(02)00342-8

Riedel, H., and Kollmeier, B. (2006). Interaural delay-dependent changes in the binaural difference potential of the human auditory brain stem response. *Hear. Res.* 218, 5–19. doi: 10.1016/j.heares.2006.03.018

Salminen, N. H., May, P. J., Alku, P., and Tiitinen, H. (2009). A population rate code of auditory space in the human cortex. *PLoS One* 4:e7600. doi: 10.1371/journal.pone.0007600

Salminen, N. H., Takanen, M., Santala, O., Lamminsalo, J., Altoe, A., and Pulkki, V. (2015). Integrated processing of spatial cues in human auditory cortex. *Hear. Res.* 327, 143–152. doi: 10.1016/j.heares.2015.06.006

Salminen, N. H., Tiitinen, H., Yrttiaho, S., and May, P. J. (2010). The neural code for interaural time difference in human auditory cortex. *J. Acoust. Soc. Am.* 127, EL60–EL65. doi: 10.1121/1.3290744

Sammeth, C. A., Greene, N. T., Brown, A. D., and Tollin, D. J. (2020). Normative study of the binaural interaction component of the human auditory brainstem response as a function of interaural time differences. *Ear Hear.* 42:629. doi: 10.1097/AUD.0000000000000964

Skrandies, W. (1990). Global field power and topographic similarity. *Brain Topogr.* 3, 137–141. doi: 10.1007/BF01128870

Smith, Z. M., and Delgutte, B. (2007). Using evoked potentials to match interaural electrode pairs with bilateral cochlear implants. *J. Assoc. Res. Otolaryngol.* 8, 134–151. doi: 10.1007/s10162-006-0069-0

Stecker, G. C., McLaughlin, S. A., and Higgins, N. C. (2015). Monaural and binaural contributions to interaural-level-difference sensitivity in human auditory cortex. *Neuroimage* 120, 456–466. doi: 10.1016/j.neuroimage.2015.07.007

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: A user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716

Ungan, P., and Yagcioglu, S. (2002). Origin of the binaural interaction component in wave P4 of the short-latency auditory evoked potentials in the cat: Evaluation of serial depth recordings from the brainstem. *Hear. Res.* 167, 81–101. doi: 10.1016/S0378-5955(02)00351-9

Ungan, P., Yağcioğlu, S., and Özmen, B. (1997). Interaural delay-dependent changes in the binaural difference potential in cat auditory brainstem response: Implications about the origin of the binaural interaction component. *Hear. Res.* 106, 66–82. doi: 10.1016/S0378-5955(97)00003-8

Van Yper, L. N., Vermeire, K., De Vel, E. F., Battmer, R. D., and Dhooge, I. J. (2015). Binaural interaction in the auditory brainstem response: A normative study. *Clin. Neurophysiol.* 126, 772–779. doi: 10.1016/j.clinph.2014.07.032

von Kriegstein, K., Griffiths, T. D., Thompson, S. K., and McAlpine, D. (2008). Responses to interaural time delay in human cortex. *J. Neurophysiol.* 100, 2712–2718. doi: 10.1152/jn.90210.2008

Willott, J. F., Milbrandt, J. C., Bross, L. S., and Caspary, D. M. (1997). Glycine immunoreactivity and receptor binding in the cochlear nucleus of C57BL/6J and CBA/CaJ mice: Effects of cochlear impairment and aging. *J. Comp. Neurol.* 385, 405–414. doi: 10.1002/(SICI)1096-9861(19970901)385:3<405::AID-CNE5>3.0.CO;2-7

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 11 January 2023
DOI 10.3389/fnins.2022.1057605



OPEN ACCESS

EDITED BY
Lina Reiss,
Oregon Health and Science University,
United States

REVIEWED BY
Inyong Choi,
The University of Iowa, United States
Andrew Dimitrijevic,
University of Toronto, Canada

*CORRESPONDENCE
Waldo Nogueira
✉ nogueiravazquez.waldo@mh-hannover.de

SPECIALTY SECTION
This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Neuroscience

RECEIVED 29 September 2022
ACCEPTED 20 December 2022
PUBLISHED 11 January 2023

CITATION
Dolhopiatenko H and Nogueira W
(2023) Selective attention decoding in
bimodal cochlear implant users.
*Front. Neurosci.* 16:1057605.
doi: 10.3389/fnins.2022.1057605



# Selective attention decoding in bimodal cochlear implant users

Hanna Dolhopiatenko and Waldo Nogueira*

Department of Otolaryngology, Hannover Medical School and Cluster of Excellence Hearing4all, Hanover, Germany

The growing group of cochlear implant (CI) users includes subjects with preserved acoustic hearing on the opposite side to the CI. The use of both listening sides results in improved speech perception in comparison to listening with one side alone. However, large variability in the measured benefit is observed. It is possible that this variability is associated with the integration of speech across electric and acoustic stimulation modalities. However, there is a lack of established methods to assess speech integration between electric and acoustic stimulation and consequently to adequately program the devices. Moreover, existing methods do not provide information about the underlying physiological mechanisms of this integration or are based on simple stimuli that are difficult to relate to speech integration. Electroencephalography (EEG) to continuous speech is promising as an objective measure of speech perception, however, its application in CIs is challenging because it is influenced by the electrical artifact introduced by these devices. For this reason, the main goal of this work is to investigate a possible electrophysiological measure of speech integration between electric and acoustic stimulation in bimodal CI users. For this purpose, a selective attention decoding paradigm has been designed and validated in bimodal CI users. The current study included behavioral and electrophysiological measures. The behavioral measure consisted of a speech understanding test, where subjects repeated words to a target speaker in the presence of a competing voice listening with the CI side (CIS) only, with the acoustic side (AS) only or with both listening sides (CIS+AS). Electrophysiological measures included cortical auditory evoked potentials (CAEPs) and selective attention decoding through EEG. CAEPs were recorded to broadband stimuli to confirm the feasibility to record cortical responses with CIS only, AS only, and CIS+AS listening modes. In the selective attention decoding paradigm a co-located target and a competing speech stream were presented to the subjects using the three listening modes (CIS only, AS only, and CIS+AS). The main hypothesis of the current study is that selective attention can be decoded in CI users despite the presence of CI electrical artifact. If selective attention decoding improves combining electric and acoustic stimulation with respect to electric stimulation alone, the hypothesis can be confirmed. No significant difference in behavioral speech understanding performance when listening with CIS+AS and AS only was found, mainly due to the ceiling effect observed with these two listening modes. The main finding of the current study is the possibility to decode selective attention in CI users even if continuous artifact is present. Moreover, an amplitude reduction of the forward transfer response function (TRF) of selective attention decoding

was observed when listening with CIS+AS compared to AS only. Further studies to validate selective attention decoding as an electrophysiological measure of electric acoustic speech integration are required.

KEYWORDS

cochlear implant, selective attention, electric acoustic stimulation, electrophysiological measures, central integration, bimodal hearing, bimodal stimulation, electroencephalography

## 1. Introduction

The growing group of cochlear implant (CI) users includes subjects with preserved acoustic hearing on the opposite side to the CI. The combination of electric and contralateral acoustic stimulation, also referred to as bimodal stimulation, usually results in an improvement in sound localization (Ching et al., 2004; Potts et al., 2009; Arndt et al., 2011; Firszt et al., 2012; Prejban et al., 2018; Galvin et al., 2019), music perception (Kong et al., 2005; Ching et al., 2007), tinnitus suppression (Van de Heyning et al., 2008; Galvin et al., 2019) and quality of life (Galvin et al., 2019) compared to monaural listening. Moreover, subjects with bimodal stimulation can integrate electric and acoustic information to improve their speech understanding (Ching et al., 2004; Kong et al., 2005; Dorman et al., 2008; Potts et al., 2009; Vermeire and Van de Heyning, 2009; Yoon et al., 2015; Devocht et al., 2017). However, the observed benefits present high variability across subjects (Ching et al., 2007; Crew et al., 2015) and some subjects even experience worsened speech performance with bimodal stimulation (Litovsky et al., 2006; Mok et al., 2006; Galvin et al., 2019). This variability in speech outcomes with bimodal listening may be associated with the effectiveness of the speech integration between electric and acoustic stimulation. Some previous works suggested that this integration has a central origin (Yang and Zeng, 2013; Reiss et al., 2014; Fowler et al., 2016; Balkenhol et al., 2020). However, the integration mechanisms and its impact on bimodal benefit requires investigation.

Different mechanisms might contribute into electric acoustic integration of speech: the integration of complementary speech information conveyed through electric and acoustic stimulation, the integration of similar speech information conveyed electrically and acoustically or the combination of the two mechanisms. Reiss et al. (2014) showed that bimodal CI users obtain abnormal spectral integration, which might lead to speech perception interference between the electric and the acoustic stimulation sides. To solve this, they suggested to reduce overlap in frequency information transmitted through electric and acoustic stimulation (Reiss et al., 2012a,b). Fowler et al. (2016) also investigated the reduction of frequency overlap in bimodal CI users and observed that subjects with better residual hearing (<60 dB HL at 250 and 500 Hz) might benefit when low frequency information is removed on the CI side. In contrast, Fu et al. (2017) showed that bimodal perception is not significantly impacted when changing the CI input low-cutoff frequency, claiming that bimodal CI users do not benefit from the mismatch correction. However, that study was conducted using a vocoder to simulate bimodal hearing in normal hearing subjects. The study of Kong and Braida (2011) assumed that bimodal listeners do not integrate available cues from both listening sides but rather rely on the cues processed by the dominant stimulation. However, Yoon et al. (2015) demonstrated that bimodal benefit does not depend on the performance of the dominant acoustic side alone but can be predicted by the difference between performances of the two stimulation modalities. Therefore, authors concluded that the bimodal benefit is a result of the integration between electric and acoustic stimulation.

The benefit of electric acoustic stimulation in bimodal CI users is usually measured behaviorally using clinical speech performance tests. These tests suffer from test-retest variability, cannot be applied to people with missing behavioral response and do not provide insights about the underlying physiological mechanisms related to electric acoustic integration. The understanding of these physiological mechanisms may provide novel approaches to program the CI and consequently improve speech perception in bimodal CI users. EEG is promising as an objective measure of speech integration for bimodal listening, however, its application is challenging because it is influenced by the CI electrical artifact.

Nowadays, there is a growing interest in the use of cortical auditory evoked potentials (CAEPs) as an objective measure of sound perception in NH listeners (Martin et al., 2007; Stapells, 2009; Papesh et al., 2015) and in CI users (Pelizzone et al., 1987; Ponton et al., 1996; Firszt et al., 2002; Maurer et al., 2002; Sharma et al., 2002). It has been shown that CAEPs provide information about binaural interaction at central level in NH listeners by analyzing the deviation of binaural responses from the sum of monaural responses (i.e. binaural interaction component (BIC) analysis) (McPherson and Starr, 1993; Jancke et al., 2002; Henkin et al., 2015). CAEPs were also measured in people with asymmetric hearing loss, revealing that the sound at cortical

level is processed similarly for acoustic alone and electric alone stimulation (Sasaki et al., 2009; Balkenhol et al., 2020; Wedekind et al., 2020, 2021). However, the amount of studies investigating electric acoustic integration at cortical level in bimodal CI users is limited. The current study investigates the possibility to record CAEPs when listening with the CI side (CIS) alone, the acoustic side (AS) alone and both sides simultaneously (CIS+AS).

One of the main disadvantages of CAEPs is that they require the use of relatively short and simple stimuli. Therefore, the relation between CAEPs and speech understanding is not easy to establish. Another alternative EEG measure, which recently has gained significant interest as an objective measure is neural tracking of the envelope of an attended speech source (Ding and Simon, 2012; Giraud and Poeppel, 2012; Power et al., 2012; Mirkovic et al., 2015; OSullivan et al., 2015). The paradigm in which in addition to the attended speaker also an ignored speaker is introduced, is called selective attention decoding. Two linear approaches exist for selective attention decoding, the forward and the backward models. Both approaches are based on least mean square error minimization between audio features and neural signals. Most previous studies performed stimulus-response mapping in the forward direction, i.e. using forward models to investigate how the system generates or encodes information (Haufe et al., 2014). By applying the forward model, the temporal response function (TRF), which describes the relationship between speech and neural recordings, is obtained. The morphology of the TRF resembles the classical N1P2 complex of the late evoked potentials (Lalor et al., 2009; Crosse et al., 2016). Analysis of the N1P2 TRF complex might provide different information than the N1P2 complex of CAEPs due to the utilization of more ecological speech stimuli, as selective attention is decoded using continuous speech streams. In order to investigate how speech features are decoded from the neural representation, one can apply the backward model (Mesgarani et al., 2009; Ding and Simon, 2012; Pasley et al., 2012; Mirkovic et al., 2015; OSullivan et al., 2015; Crosse et al., 2016). By using the backward model, the speech stimulus is reconstructed from the neural activity recordings. The backward model explores the accuracy of decoding by analyzing speech features of reconstructed and original speech stimuli.

Recently, the possibility to predict speech intelligibility from selective attention decoding has been shown in NH listeners (Keitel et al., 2018; Vanthornhout et al., 2018; Dimitrijevic et al., 2019; Etard and Reichenbach, 2019; Lesenfants et al., 2019), in hearing impaired listeners with hearing aids (Petersen et al., 2017) and in bilateral (Paul et al., 2020) and monaural CI users (Nogueira and Dolhopiatenko, 2022). However, the application of such objective measures in CI users is still challenging because of the CI electrical artifact leaking into the EEG recordings (Hofmann and Wouters, 2010; Somers et al., 2010; Deprez et al., 2017). Artifact rejection techniques such as independent component analysis (ICA) can suppress artifacts in EEG, however, the full removal of the CI electrical artifact can not be ensured. Nevertheless, some previous works showed that it is still feasible to decode selective attention in CI users (Nogueira et al., 2019a,b; Aldag et al., 2022). In this regard, these previous works showed that the maximum differentiation between the attended and the ignored speaker occurs at 200–400 ms after stimulus onset showing a minimization effect of the CI electrical artifact at this time interval. However, as selective attention is recorded to the continuous speech, the impact of the CI electrical artifact cannot be fully discarded and more evidences that selective attention decoding is possible in CI users are necessary. Therefore, the main goal of the current study is to confirm the feasibility to decode selective attention in bimodal CI users, which will provide further evidence on the use of continuous EEG recordings to speech stimuli in this population, despite the presence of the CI electrical artifact. This work investigates selective attention decoding in bimodal CI users when listening with CIS only, AS only and both sides together (CIS+AS). It is hypothesized that combined electric and acoustic stimulation results in improved selective attention decoding with respect to listening with CIS alone. If this hypothesis is confirmed, it can be concluded that it is possible to decode selective attention in CI users, as the additional neural activity provided by the acoustic stimulation is used to improve the decoding, even if CI artifact is present. Moreover, the confirmation of the main hypothesis might open the possibility to further investigate selective attention decoding as a measure of speech integration between electric and acoustic stimulation in bimodal CI users using continuous speech which is a natural and ecologically valid signal.

To find a descriptive link between speech understanding and selective attention decoding, the current study also included a behavioral measure. The behavioral measure consisted of a speech understanding performance test to a target speaker in the presence of a competing talker. Speech material was presented using the three listening modes (CIS only, AS only, and CIS+AS). The second part of the study included recording of EEG. The possibility to record cortical responses to short stimuli with all three listening modes was demonstrated through CAEPs. Afterwards, selective attention decoding, which is a novel approach when applied to bimodal CI users, was measured. In the selective attention paradigm, a target and a competing talker were presented to the subjects using the three listening modes (CIS only, AS only, CIS+AS). The main goal of this study is to investigate the feasibility to decode selective attention in CI users despite the presence of CI electrical artifact. Furthermore, first attempts to investigate the potential of selective attention decoding as a speech integration measure between electric and acoustic stimulation in bimodal CI users were conducted.

# 2. Materials and methods

## 2.1. Participants

Ten subjects participated in the study (mean age: 57.7). All participants were implanted with an Oticon Medical CI and had a device experience of 6–48 months. Demographics of the participants are shown in Table 1. Prior to the experiment, all participants provided written informed consent and the study was carried out in accordance with the Declaration of Helsinki principles, approved by the Ethics Committee of the Hannover Medical School.

A pure tone audiogram on the non-implanted ear in unaided condition was measured *via* a calibrated audiometry system (CAS AD2117, Audio-DATA, Duvensee, Germany). The audiograms for the study participants are presented in Figure 1.

## 2.2. Behavioral paradigm

The German Hochmair-Schulz-Moser sentence test (HSM test) (Hochmair-Desoyer et al., 1997) was used to assess speech understanding behaviorally. Each list consists of 20 semantically structured sentences, uttered by a male or by a female talker. Two sentence lists were presented to the acoustic side only (AS only), to the CI side only (CIS only) and to both sides simultaneously (CIS+AS). Target and interference speech streams were co-located and presented at 0 dB signal-to-interference ratio (SIR) between the target (male/female) and the interference (female/male) speech stream. Subjects were instructed to attend to the target talker and to repeat all words after each sentence. The speech stream to be attended was randomized within subjects and is indicated in Table 1. The attended speech stream was kept the same through the whole experiment. The speech understanding performance score was calculated in percentage of correct recalled words per listening mode. Speech material was presented to the CI side *via* the Oticon Bluetooth Streamer and to the acoustic side *via* inner-ear phones (E-A-RTONE Gold 3A, 3M, St. Paul, Minneapolis). For subjects wearing a hearing aid on the contralateral side to the CI, speech material presented to the AS was preprocessed using a digital hearing aid implemented in a PC. The hearing aid was based on the half-gain rule amplification according to the measured audiogram (Lybarger, 1963). This hearing aid implementation has been successfully used in previous studies in our group (Krüger et al., 2022). Stimulus presentation

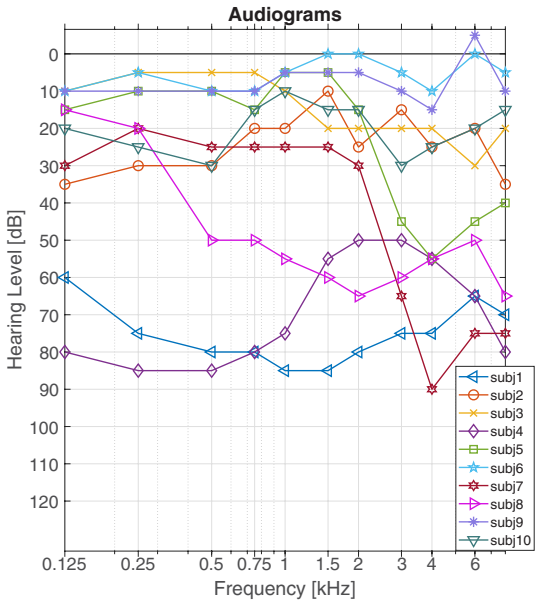



FIGURE 1
Audiograms of study participants.

TABLE 1 Demographics of participants.

| ID | Sex | Age | Hearing aid | Implanted side | Duration of deafness (y) | CI Experience (y) | Stream to be attended |
|---|---|---|---|---|---|---|---|
| 1 | M | 63 | Yes | Left | 55 | 4 | Male |
| 2 | F | 62 | No | Right | 3 | 4 | Female |
| 3 | F | 54 | No | Left | 42 | 4 | Male |
| 4 | M | 62 | Yes | Left | 13 | 1.5 | Male |
| 5 | M | 68 | No | Left | 4 | 2 | Female |
| 6 | M | 47 | no | Left | 1 | 3 | Female |
| 7 | M | 73 | No | Right | 12 | 4 | Male |
| 8 | M | 57 | Yes | Right | 19 | 0.5 | Female |
| 9 | M | 30 | No | Left | 0.07 | 3.4 | Male |
| 10 | F | 61 | No | Right | 15 | 1.5 | Female |

was controlled by the Presentation Software (Neurobehavioral Systems, Inc., Berkeley, CA, United States; version 20.1). The presentation level for the CIS and AS was set to a most comfortable level, using a seven point loudness rating-scale (from 1 to 7: from extremely soft to extremely loud; 4 - most comfortable level).

## 2.3. Electrophysiological paradigm

The electrophysiological part of the experiment consisted of EEG recordings. The recording was conducted in an electromagnetically and acoustically shielded booth. High-density continuous EEG was recorded using a SynAmps RT System with 64 electrodes mounted in a customized, infracerebral electrode cap (Compumedics Neuroscan, Australia). The reference electrode was placed on the nose tip; two additional electrodes were placed on the mastoids. Impedances were controlled and maintained below 15 kΩ. Electrodes with high impedance were excluded from further analysis. Each subject was instructed to sit relaxed, avoid any movements, and to keep their eyes open in order to minimize physiological artifacts. All material was presented *via* Bluetooth Streaming to the CIS and *via* inner-ear phones to the AS. For AS in CI users with a hearing aid on the contralateral side, all presented material was processed with the same digital hearing aid and adjusted in loudness to their MCL exactly in the same manner as in the behavioral paradigm described in Section 2.2.

### 2.3.1. Cortical auditory evoked potentials

#### 2.3.1.1. Stimuli

To maximize responses, a broadband noise of 50 ms duration was used as a stimulus. To ensure time synchronization between both listening sides, the delay between electric and acoustic stimulation needs to be considered. The technical CI delay was measured using an oscilloscope and a research implant unit. The stimulus was presented through the Bluetooth Streamer to the CI processor, and the delay between audio start and the start of electrical stimulation resulted in 30 ms. On the acoustic side however, the delay between stimulus onset and the auditory nerve response is frequency dependent and variable across subjects (Elberling et al., 2007). It was estimated from the literature that the average delay for a stimulus to travel from the outer ear to the auditory nerve is 7 ms (Elberling et al., 2007), therefore, the stimulus for AS started 23 ms after the onset of the stimulus on the CIS. Note that delay compensation was conducted at group level and not adjusted individually for each subject. The stimuli were presented with an inter-stimulus interval of 1 s. In total, 100 trials were recorded. The EEG data was recorded with a sampling rate of 20 kHz.

#### 2.3.1.2. Processing

Recorded EEG data was processed through the EEGLAB MATLAB toolbox (Delorme and Makeig, 2004). ICA based on second-order blind identification (SOBI) was applied to the recorded data to remove physiological and CI electrical artifacts (Kaur and Singh, 2015). Data recorded with CIS only and with CIS+AS listening mode was concatenated prior to SOBI in order to ensure equal portion of the removed CI electrical artifact in both listening modes. The topology and the signal in the time and spectral domain were visually analyzed for each component. On average 1.8 (std:±0.64) components for each subject were removed from the data. EEG with the suppressed artifacts was afterwards epoched in the time interval ranging from −200 to 1,000 ms. The CAEPs were obtained from the vertex electrode (Cz). The signal was filtered between 1 and 15 Hz and re-referenced to the mean of the two mastoid electrodes.

### 2.3.2. Selective attention decoding

#### 2.3.2.1. Stimuli

For the selective attention paradigm HSM sentence lists with a male and a female talker at 0 dB SIR were used. The speech stream to be attended was kept the same as in the behavioral part of the experiment (Table 1). For each listening condition (CIS only, AS only, and CIS+AS) 8 lists were presented, resulting in approximately 6 min of stimulation per listening mode. To extend the training dataset for the decoder, additional speech material consisting of two audio story books were used. The story books included two German narrations ("A drama in the air" by Jules Verne, narrated by a male speaker and "Two brothers" by the Grimm brothers, narrated by a female speaker) at 0 dB SIR. In total, 36 min of story (12 min per listening mode) were presented. To ensure the continuous engagement of the CI user when listening to the corresponding speaker, questions to the context of the presented speech material were asked every 2 min. The presented speech material was randomized across listening conditions to avoid the influence of the material. EEG data was recorded with a sampling rate of 1,000 Hz.

#### 2.3.2.2. Processing

EEG data was processed offline in MATLAB (MATLAB, 2018) and the EEGLAB toolbox (Delorme and Makeig, 2004). SOBI artifact rejection was applied to the EEG data to suppress physiological and CI electrical artifact. The location of the CI and the signal in the time and spectral domain of each component were analyzed. On average, 3.5 (std: ±1.08) components were removed from the data. Afterwards, the EEG data was split into the trials corresponding with the duration of each sentence list and 1 min segments of the story. Next, the digital signal was band-pass filtered for

frequencies 2–8 Hz and downsampled to 64 Hz. The envelopes of the original attended and unattended speech streams were extracted through the Hilbert transform. The envelopes were filtered with a low-pass filter having cut-off frequency of 8 Hz and downsampled to 64 Hz. Selective attention was analyzed using the forward and the backward model approaches (Crosse et al., 2016). By applying the forward model, the TRF was obtained. By using the backward model the speech stimulus was reconstructed from the neural activity recordings. The correlation coefficient between the original envelope of the attended audio and the reconstructed envelope (attended correlation coefficient $\rho_A$) as well as the correlation coefficient between the original envelope of the unattended audio and the reconstructed envelope (unattended correlation coefficient $\rho_U$) were calculated. Selective attention decoding was analyzed in terms of $\rho_A$ and the difference between $\rho_A$ and $\rho_U$ ($\rho_{Diff}$). Both, forward and backward models were applied across time lag. The time lag performs a time shift of the EEG signal that reproduces the physiological delay between the audio presentation and its processing up to the cortex (OSullivan et al., 2015). In total 38 lags spanning the interval from 16 to 608 ms were used. The lag window, over which reconstruction was conducted, was set to $\triangle$ = 16 ms. The regularization parameter $\lambda$ was set to 100 to maximize the peak amplitudes of the TRF for the forward model and to 0.01 to maximize the difference between the attended and unattended correlation coefficients for the backward model. Further details on the analysis of TRFs and correlation coefficients across $\lambda$ can be found in the Supplementary material. For more details on the reconstruction procedure see Nogueira et al. (2019a,b). A classical leave-one-out cross-validation approach was used to train and test the decoder. HSM lists and the story were used to train the decoder. Only HSM sentences were used for testing, resulting in 8 folds for cross-validation (corresponding to the amount of lists) with each listening mode.

## 2.4. Statistical analysis

Statistical analysis was carried out with the SPSS software (version 26, IBM). The effect of listening mode on the investigated parameters was explored through a repeated measures analysis of variance (ANOVA). Pairwise comparisons between listening modes were conducted with *post-hoc* analysis based on the *t*-test for each pair of observations. To avoid type I error for multiple comparisons, Bonferroni correction was applied. For non-normally distributed data, the non-parametric Friedman test followed by a *post-hoc* Wilcoxon signed-rank test for pairwise comparisons was applied to the data.

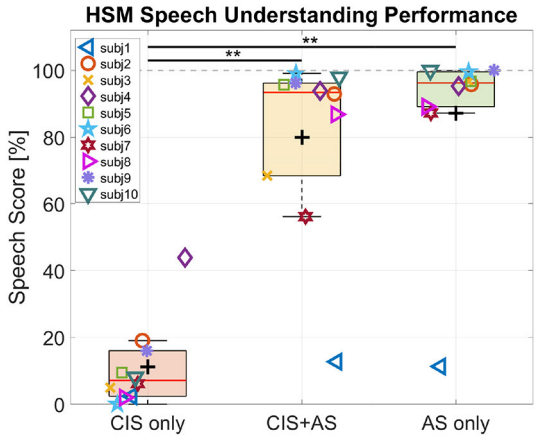



FIGURE 2
Behavioral speech understanding performance for three listening modes: CIS only, CIS+AS, and AS only. Asterisk indicates significant difference between a pair of conditions revealed through the Wilcoxon signed-rank test.

# 3. Results

## 3.1. Behavioral paradigm

Figure 2 shows the individual speech understanding performance for each listening mode.

On average, the highest score was observed with AS only listening mode (87.18%), followed by CIS+AS listening mode (80.01%). The lowest score was obtained with CIS only (11.21%) and can be explained by the high difficulty of the task for the participants when listening with CIS alone. A Friedman test revealed a significant effect of listening mode on speech performance scores [$\chi^2_{(2)} = 18.200; p < 0.001$]. *Post-hoc* analysis with the Wilcoxon signed-rank test with a Bonferroni correction resulted in a significant effect for the pairs CIS+AS—CIS only ($p = 0.005$) and AS only—CIS only ($p = 0.005$). No significant difference between CIS+AS and AS only was observed.

## 3.2. Cortical auditory evoked potentials

Figure 3 presents the averaged CAEPs across subjects after SOBI artifact rejection for the three listening modes (CIS only, AS only, and CIS+AS). Subject 1 was excluded from the analysis due to the low quality of the recorded signal.

In general, it was possible to distinguish the cortical response with all three listening modes. The peak-to-peak amplitude of the N1P2 was estimated for each subject individually. An ANOVA analysis revealed a significant effect of listening mode

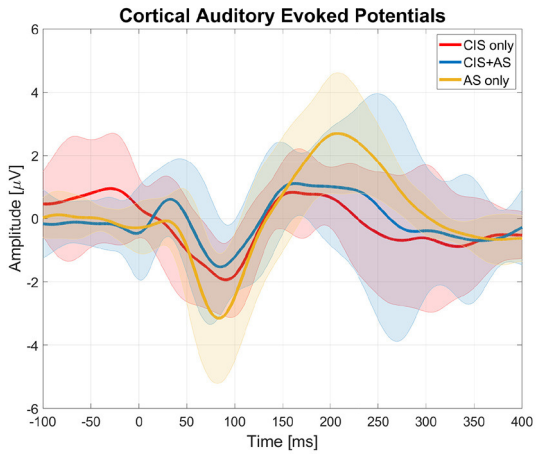



FIGURE 3
Cortical auditory evoked potentials (CAEPs) obtained from the central electrode (Cz) with three listening modes (CIS only, AS only, and CIS+AS) averaged across subjects. The thick lines represent the mean values across subjects and the shaded areas represent the standard deviation across subjects.

on the N1P2 amplitude [$F_{(2,16)} = 6.544$; $p = 0.008$]. A *post-hoc* $t$-test with Bonferroni correction revealed a significantly higher N1P2 amplitude with AS only than with CIS only listening mode ($p = 0.014$). No significant differences between CAEPs recorded with CIS only and CIS+AS listening modes were found.

## 3.3. Selective attention decoding

### 3.3.1. Temporal response function

Figure 4A presents the mean TRF across subjects, where the TRF represents the decoder weights of the forward model approach. The TRF were analyzed comparing the first negative (N1) and second positive (P2) peaks for each listening mode and listener. The analysis revealed highest N1P2 peak-to-peak amplitude for the AS only, followed by the CIS+AS and the lowest amplitude for the CIS only listening mode. Moreover, weights of the TRF at the N1 and P2 peaks were estimated for each subject and presented in the form of topographical maps per each listening mode (Figure 4B). The weight distribution is similar across all listening modes, however, the activation power with the CIS only listening mode is visibly weaker.

From Figure 4A, it can be observed that the latencies of the TRF peaks for the CIS+AS and the AS only listening modes are similar, while the latency for CIS only is delayed. Moreover, the amplitude of the N1P2 peak of the TRF was compared to the amplitude of the N1P2 peaks obtained from CAEPs presented in Section 3.2. A significant correlation between the N1P2 peak-to-peak amplitude from TRFs and CAEPs was observed for CIS only and AS only listening modes (CIS only: $r = 0.715, p = 0.031$; AS only: $r = 0.793, p = 0.011$) (Figure 5). For the CIS+AS listening mode, no significant correlation between the N1P2 amplitude derived from CAEPs and TRFs was found. This may be explained by the temporal delay correction between electric and acoustic stimulation implemented in CAEP measurements, which was not applied during the selective attention decoding experiment.

### 3.3.2. Selective attention correlation coefficients

Figure 6 presents the $\rho_A$ and $\rho_U$ coefficients across lags obtained from selective attention decoding using the backward model after SOBI artifact rejection. Note, that lag $\triangle$ is used to time shift the EEG, modeling the physiological delay required for a sound to travel along the auditory pathway up to the cortex.

The correlation coefficients across lags for the AS only condition present a morphology consistent with the morphology reported in NH listeners (Nogueira et al., 2019a). In NH listeners, the typical morphology of the $\rho_A$ curve presents two peaks at around 100 and 250 ms associated with different stages of neural processing. The correlation coefficients obtained with the CIS only and CIS+AS listening modes at early lags were higher than with AS only indicating a contribution of the CI electrical artifact. SOBI artifact rejection suppressed part of this artifact, however, full removal could not be achieved. In order to minimize the effect of the CI artifact, a later lag interval was chosen for further analysis. Based on previous works (Nogueira et al., 2019a,b), the chosen lag interval spanned the time between 208 and 304 ms, which also corresponds to the second peak of the $\rho_A$ curve for the AS only condition (Figure 6). At that chosen lag interval, a $t$-test comparing the $\rho_A$ and the $\rho_U$ coefficients for each listening mode revealed a significant difference for the CIS+AS ($p < 0.001$) and for the AS only ($p < 0.001$) listening modes, but not for the CIS only mode ($p = 0.405$) due to the small differences between the $\rho_A$ and the $\rho_U$. This result confirms the possibility to decode selective attention in CI users despite the presence of CI electrical artifact.

Furthermore, we focus our analysis only on the difference between the attended and the unattended correlation coefficients ($\rho_{Diff}$), which reduces the impact of the CI artifact (Nogueira and Dolhopiatenko, 2022). Figure 7 shows the $\rho_{Diff}$ at the lag interval 208–304 ms for each listening mode. The ANOVA test revealed a significant effect of listening mode on the $\rho_{Diff}$ $F_{(2,18)} = 23.640; p < 0.001$. The *post-hoc* pairwise $t$-test comparison showed a significant difference for the pairs CIS+AS—CIS only ($p = 0.003$) and AS only—CIS only ($p = 0.011$). Note that the behavioral speech understanding scores were also significantly different for the same pairs of comparisons.

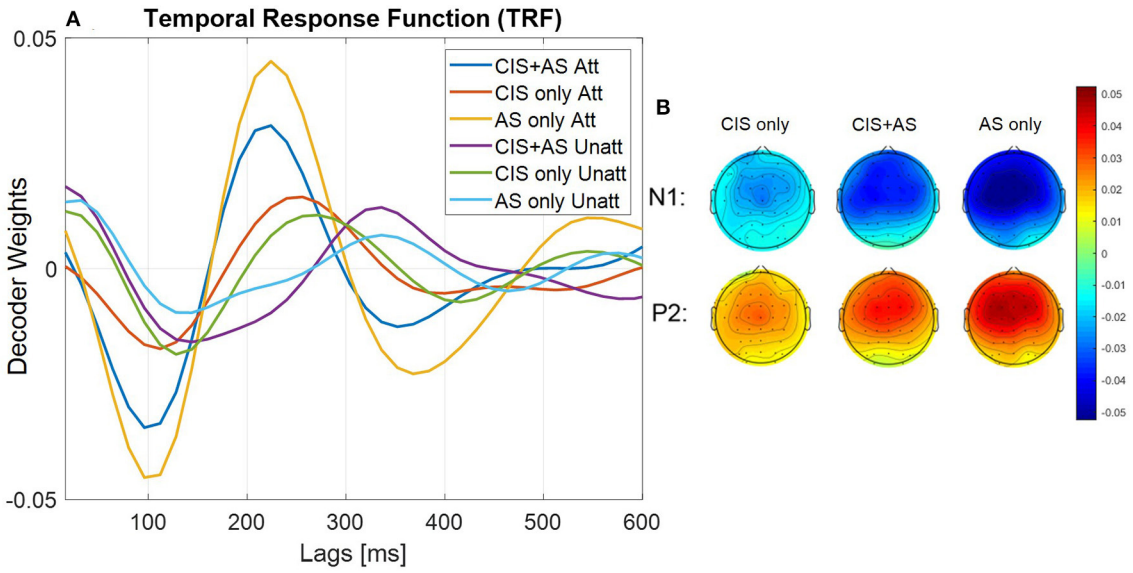



FIGURE 4
**(A)** Forward transfer response function (TRF) averaged across subjects. Attended TRF and unattended TRF are estimated using attended and unattended decoder respectively; **(B)** Topographical maps show TRF weights across subjects at first and second peaks of attended curve for each listening mode.

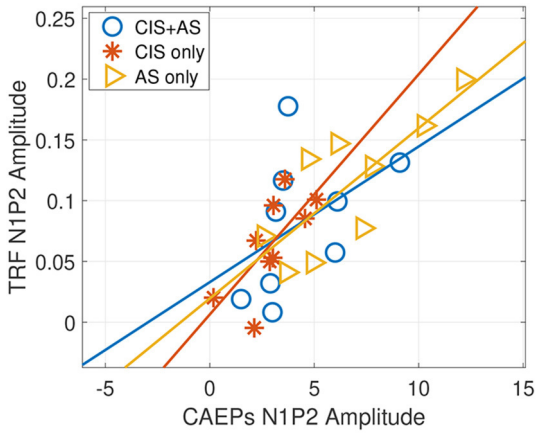



FIGURE 5
Pearson correlation between N1P2 peak amplitudes of the temporal response function (TRF) and N1P2 peak amplitudes of cortical auditory evoked potentials (CAEPs) with three listening modes: CIS only, CIS+AS, and AS only.

Pearson correlation between the behavioral speech score and the $\rho_{Diff}$ revealed a significant correlation only between $\rho_{Diff}$ and the speech score obtained with the CIS only listening mode ($r = 0.712, p = 0.021$) (Figure 8). A lack of significance in correlation between speech understanding performance and the selective attention correlation coefficients for AS only and CI+AS listening modes can be explained by the ceiling effect observed in the behavioral speech understanding performance.

## 4. Discussion

The main goal of this work was to investigate a possible electrophysiological measure of speech integration between electric and acoustic stimulation in bimodal CI users. An electrophysiological paradigm based on CAEPs to short stimuli showed the feasibility to record cortical responses with CIS only, AS only and CIS+AS listening modes. As an electrophysiological measure of speech integration, decoding of selective attention was proposed and validated in bimodal CI users. The results of the study confirmed that it is possible to decode selective attention in CI users despite the presence of CI electrical artifact in the EEG. Moreover, this work investigated how selective attention decoding is related to behavioral speech understanding performance. No bimodal benefit in speech understanding with respect to listening with the better ear was found, mainly due to the ceiling effects observed when listening with the CIS+AS and the AS only listening modes.

### 4.1. Speech understanding performance

Based on previous studies (e.g., Ching et al., 2004; Kong et al., 2005; Dorman et al., 2008; Potts et al., 2009; Vermeire and Van de Heyning, 2009; Yoon et al., 2015; Devocht et al.,

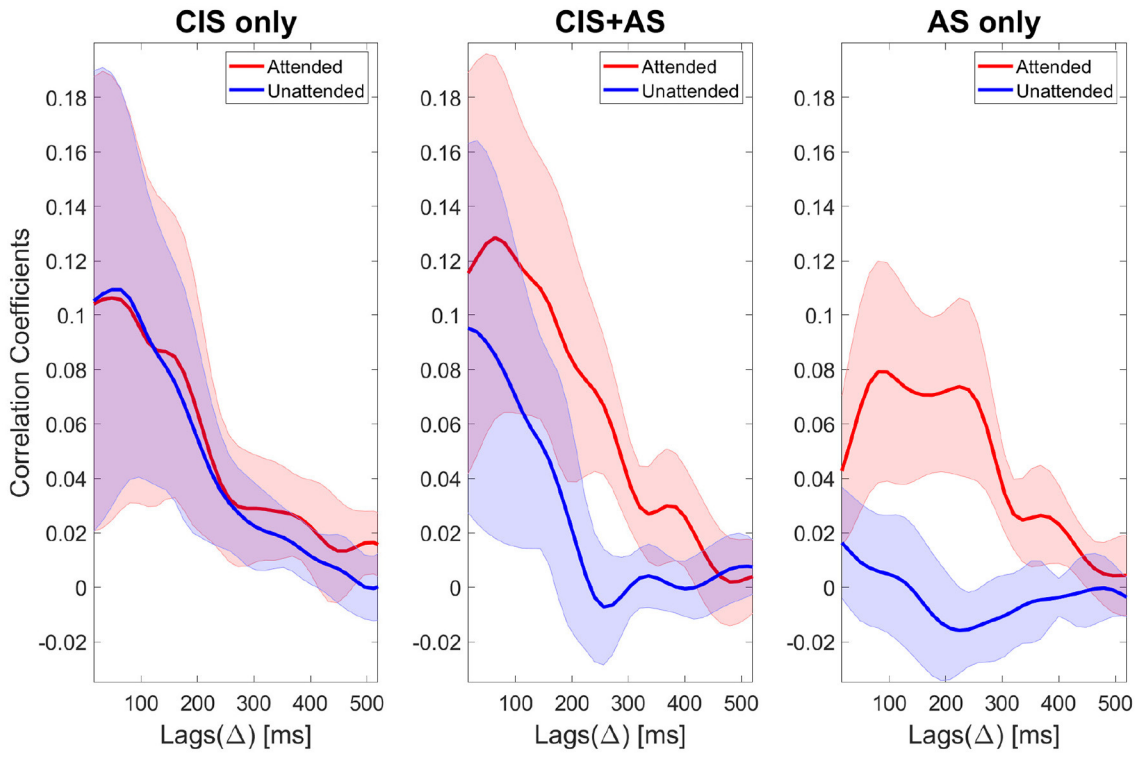



FIGURE 6
Attended correlation coefficients (red color) and unattended correlation coefficients (blue color) of selective attention decoding for three listening modes: CIS only **(left)**, CIS+AS **(center)**, and AS only **(right)**. Correlation coefficients are denoted as corr coeff and calculated across lags (Δ). The thick lines represent the mean values and the shaded areas represent the standard deviation across subjects.

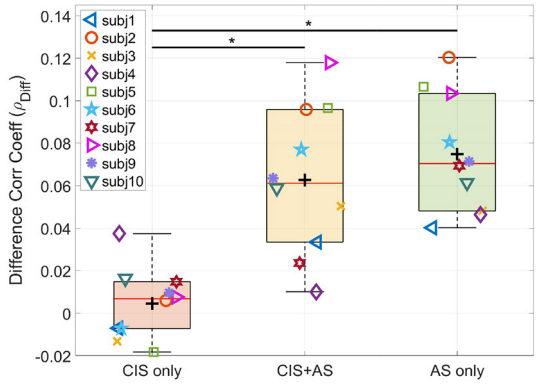



FIGURE 7
Difference between attended and unattended correlation coefficients ($\rho_{Diff}$) of selective attention decoding at the 208–304 ms lag interval. Asterisk indicates significance between pair of observations revealed by the *t*-test.

2017), this work assumed a benefit in speech understanding for bimodal CI users when listening with CIS+AS listening mode in comparison to listening with the CIS only or with the AS only mode. Nevertheless, the interpretation of the reported results in the literature about the benefit of electric acoustic stimulation in bimodal CI users depends on the reference listening mode used to report the bimodal benefit and the inclusion criteria of the subjects participating in these studies. For instance, in agreement with the results of the current study, some previous studies have shown a benefit of bimodal listening compared to CIS only listening mode, but not compared to AS only listening mode (Mok et al., 2006; Devocht et al., 2017). In contrast, the study of Potts et al. (2009) observed a bimodal benefit compared to CIS only and to AS only, however, the authors of the study recruited candidates for bilateral CI implantation, i.e. with poor residual hearing.

Nevertheless, because of a ceiling effect observed in the speech scores with the CIS+AS and AS only listening modes, it was not possible to demonstrate a possible bimodal benefit compared to the best performing ear for some of the study participants. Despite this, two subjects obtained lower speech scores with the CIS+AS than with the AS only listening mode. Demographical data for these two subjects (Table 1) was analyzed and revealed long duration of deafness for subject 3, which can explain the reduction in speech understanding when listening with both sides compared to the better ear

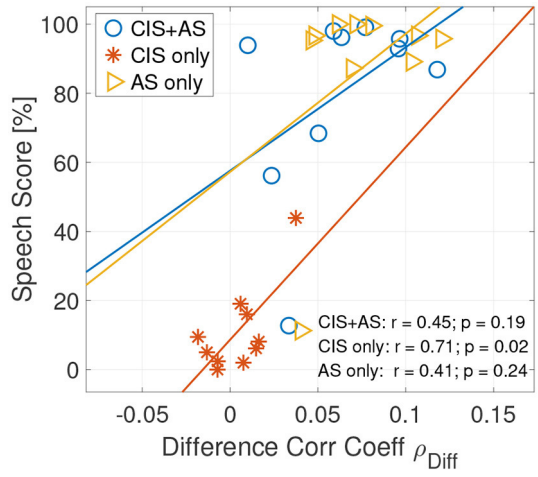



FIGURE 8
Correlation plot between the speech understanding scores and the attended correlation coefficients $\rho_A$ **(left)** and between the speech understanding scores and the difference between the attended and the unattended correlation coefficients $\rho_{Diff}$ **(right)** of selective attention decoding at the 208–304 ms lag interval.

(Cohen and Svirsky, 2019). Moreover, a significant negative correlation between duration of deafness and speech scores with the CIS+AS listening mode was observed across all subjects ($r = -0.846, p = 0.004$). Subject 7 presented shorter duration deafness but reported not using the CI frequently in daily life, which probably explains the reduction in performance observed with the CIS+AS listening mode for this subject. To confirm the benefit of using the CI in daily life, subjects were additionally asked to answer questions regarding their listening experience in different acoustic situations. Results of the questionnaire are presented in the Supplementary material. Interestingly, all participants reported a benefit of using the CI in daily life. A number of previous studies showed benefit of bimodal hearing for CI users through improved quality of life (Galvin et al., 2019), drop in self-reported listening effort (Devocht et al., 2017) or even tinnitus suppression (Van de Heyning et al., 2008). Therefore, bimodal hearing provides benefits to CI users but these could not be measured through the speech understanding task proposed in this study.

Another possible explanation for the interference effect observed in two bimodal subjects when listening with CIS+AS is the reduced integration between electric and acoustic stimulation in our group of subjects. According to Yoon et al. (2015), bimodal benefit is greater in subjects that obtain similar performance with the CIS alone and the AS alone. Subjects recruited in the current study had normal or close to normal hearing with the AS only and relatively poor performance with the CIS only, due to the long duration of deafness prior to implantation. Therefore, this might have led to reduced integration between electric and acoustic stimulation. Moreover, Reiss et al. (2014) and Fowler et al. (2016) suggested that bimodal CI users can better integrate mismatched rather than matched spectral information across listening sides. The participants of the current study had a good residual hearing causing broad frequency range overlap across ears. As a result, this abnormal broad spectral integration may have lead to speech perception interference when listening with the CIS+AS compared to AS alone.

On the other hand, the interpretation of the results about bimodal benefit reported in the literature depends on the utilized materials and tests. The present work investigated the benefit of electric acoustic stimulation on speech understanding with a co-located target and interferer presented at the same level. The same speech material was presented in both ears, which does not allow to measure some binaural effects such as spatial release from masking or binaural squelch. Moreover, the interferer consisted of a speech signal which reduces speech understanding in CI users compared to the utilization of non-intelligible maskers, such as stationary or babble noise (Dieudonné and Francart, 2020). Studies using a similar paradigm as the one used in the current study observed no speech understanding improvement in bimodal CI users compared to the better ear performance (Vermeire and Van de Heyning, 2009; Galvin et al., 2019; Dieudonné and Francart, 2020). In the current study, we decided to use the same material for the behavioral and for the selective attention decoding paradigm such that the results of both experiments could be compared to each other. The selective attention paradigm, which has been extensively validated in our previous works in CI users, is based on a target and an interferer speech streams presented at the same level to reduce the effect of the CI artifact.

## 4.2. Cortical auditory evoked potentials

It was possible to measure CAEPs and to distinguish the N1P2 complex with all three listening modes for 9 out of 10 participants. Subject 1 was excluded from the analysis due to the low quality of the EEG signal caused by high impedances of the EEG electrodes. The peak-to-peak N1P2 latencies and amplitudes were in the range of 85–130 ms and 3–7 $\mu$V, respectively. These results are in agreement with the results reported in NH listeners (Martin et al., 2007; Stapells, 2009; Papesh et al., 2015) and CI users (Pelizzone et al., 1987; Ponton et al., 1996; Maurer et al., 2002; Sharma et al., 2002).

The highest N1P2 amplitude of 6.7 $\mu$V was obtained for the AS only condition. No significant difference between the N1P2 measured with the CIS+AS and the CIS only listening modes was observed. Previous studies have shown greater N1P2 responses with bilateral stimulation compared to monaural stimulation in NH listeners. The mentioned study claimed that the greater response evoked by the bilateral stimulus compared to the

monaural stimulus can be explained by binaural integration or fusion of stimuli across both ears (Butler et al., 1969). In the current study higher N1P2 amplitudes for the CIS+AS listening mode compared to the CIS only or AS only listening modes were expected due to possible synergetic integration of electric and acoustic stimulation (Ching et al., 2001; Kong et al., 2005; Kong and Carlyon, 2007). However, the responses for the CIS+AS and the CIS only listening modes were not significantly different.

One possible explanation for the reduced bimodal response is the time processing difference or the lack of synchronization between the two listening sides. The time delay for the CIS is caused by the CI sound processor and the implant. The delay for an acoustic stimulus to reach the auditory nerve comprises ear canal, middle ear and the basilar membrane traveling wave delays. In this work, the processing delay for the CIS was measured through an oscilloscope and the delay for the AS was estimated from the literature. While the delay for the CIS is device dependent and has negligible variability across CI users with Oticon Medical CI, the time delay on the AS is less obvious, it depends on the individual anatomy and physiology of the ear, and it is frequency dependent due to the tonotopic organization of the cochlea. If to compare the estimated traveling wave delays provided by different authors utilizing different measurement techniques, a high variability across studies can be observed (Elberling et al., 2007). In this work, the delay between the two listening sides was not individually compensated, which may have caused reduced electric acoustic integration and consequently reduced bimodal CAEP responses. One possible solution for an individual delay compensation in bimodal CI users, is to correct the delay based on the wave V of the auditory brainstem response (ABR) as proposed by Zirn et al. (2015). However, the implementation of this procedure in the current work would have dramatically increased the time required to conduct the experiment. Nevertheless, the individual temporal synchronization between the two listening sides using ABRs has to be considered for future work.

## 4.3. Selective attention decoding

The topographical analysis of the decoder weight distribution for the forward model approach revealed weaker activation when listening with the CIS only than with the CIS+AS or the AS only listening modes. This weaker activation may be related to the difficulties experienced by the bimodal CI users to concentrate on the desired speech stream using the CIS only listening mode. The TRF morphology across lags for all three listening modes is consistent with previous reported results in NH listeners (Crosse et al., 2016) and CI users (Paul et al., 2020). For TRFs the highest N1P2 peak-to-peak amplitude was obtained when listening with the AS only, followed by the CIS+AS and the CIS only listening modes. The amplitude reduction of the TRF curve for CIS+AS listening mode may be explained by reduced integration between electric and acoustic stimulation or interference caused by the CI when listening with CIS+AS. Such an interference effect was observed at least in two subjects in the speech understanding performance test. Unfortunately, it was not possible to establish a relation between TRF amplitude and behavioral speech understanding performance in the current study due to ceiling effects observed in the speech understanding test with the AS only and CIS+AS listening modes. Moreover, a delay between TRF peaks for the AS only and the CIS only listening modes was observed. Therefore, the reduction of the TRF amplitude for the CIS+AS listening mode compared to the AS only listening mode can be also explained by the lack of the temporal synchronization between electric and acoustic stimulation.

As the TRF curve resembles the N1P2 complex of CAEPs, the individual N1P2 amplitudes from TRFs were compared to the N1P2 amplitudes of the CAEP responses.A significant correlation between both measures was observed for CIS only and AS only listening modes. For the CIS+AS listening mode no significant correlation was observed, possibly because a delay compensation between both sides was applied in the CAEP measurements but not in the selective attention paradigm. In the future, the impact of interaural delay on selective attention decoding should be further investigated.

The correlation coefficients of backward selective attention decoding with the CIS+AS and the CIS only listening modes obtained high values for the first lags, probably because of the contribution of residual CI artifact. As the target and interference were presented at 0 dB SIR, the contribution of the CI artifact is almost equal for both the attended and the unattended speech envelopes. Therefore, in absence of neural activity, the correlation coefficients to the attended and unattended envelopes should be almost identical, as demonstrated by an artifact model in our previous study (Nogueira et al., 2019a). In the current study, only a small difference between the attended and unattended correlation coefficients in the CIS only condition was observed. This is not surprising, taking into account the poor behavioral speech understanding performance obtained by the study participants when listening with the CIS only mode. Meanwhile, when listening with the CIS+AS mode, a significant difference between the attended and unattended correlation coefficients was observed. This result confirms the possibility to decode selective attention in CI users despite the presence of residual CI electrical artifact leaking into the EEG. Two peaks were observed at 100 and 220 ms. Coming back to the CI artifact model mentioned before, high correlation coefficients at early lags up to 80 ms followed by a decay ending at around 150 ms have been observed in our previous study (Nogueira et al., 2019a). Therefore, the first peak might be contaminated by the CI artifact, but the second peak might be less contaminated by the artifact. For this reason, this second peak could potentially be a valid parameter to compare selective attention decoding between different listening modes. The time occurrence of the

second peak also corresponds to the late locus of attention reported by Power et al. (2012). Therefore, the lag interval of 208–304 ms was chosen for further analysis. Moreover, we focus our analysis on the difference between the attended and the unattended correlation coefficients, which further reduces the effect of the CI artifact as shown in previous studies (Paul et al., 2020; Nogueira and Dolhopiatenko, 2022). Besides that, the analysis of the difference between attended and unattended correlation coefficients might reduce the impact of some individual factors, such as artifact, skin thickness or electrode impedance. The comparison of the difference correlation coefficient revealed higher values for the CIS+AS and AS only listening modes compared to the CIS only listening mode, which is consistent with the speech understanding behavioral results. The correlation between selective attention decoding coefficients and behavioral data was significant only for the CIS only listening mode. A lack of significance for the CIS+AS and AS only modes can be explained by the ceiling effect in the speech understanding scores observed in these conditions. For this reason, the results of the current study cannot conclude whether selective attention decoding can be used as an electric acoustic speech integration measure. An extension of the dataset including bimodal CI users with less residual hearing or the use of different speech understanding performance tests to avoid ceiling or floor effects need to be considered for future work.

## 5. Conclusion

This work demonstrates that it is possible to decode selective attention in bimodal CI users. This result provides more evidence on the use of continuous EEG recordings to speech stimuli in CI users despite the presence of continuous electric artifact. The analysis of CAEPs and TRFs from selective attention decoding demonstrated an amplitude reduction when listening with CIS+AS relative to listening with AS only. The outcomes of this study may pave the way toward novel speech integration measures for bimodal CI users using EEG to continuous stimuli. However, further validation of these measurements are required.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee of the Hannover Medical School. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

HD implemented the algorithm and conducted the data collection and analysis. HD and WN contributed to the experimental setup and wrote the manuscript. WN conceived the study and contributed to data analysis and interpretation. All authors contributed to the article and approved the submitted version.

## Funding

This work is a part of the project BiMoFuse: Binaural Fusion between Electric and Acoustic Stimulation in Bimodal CI Subjects (ID: 20-1588, PI: WN) funded by William Demant Foundation. This work was also supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) cluster of excellence EXC 2177/1. Part of these results is part of the project that received funding from the European Research Council (ERC) under the European Union's Horizon-ERC programme (Grant agreement READIHEAR No. 101044753. PI: WN).

## Acknowledgments

The authors would like to thank Manuel Segovia-Martinez, Aswin Wijetillake, and colleagues from Oticon Medical for their contribution in this work. The authors would also like to thank all subjects who participated in the study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1057605/full#supplementary-material

# References


Aldag, N., Büchner,1, A., Lenarz, T., and Nogueira, W. (2022). Towards decoding selective attention through cochlear implant electrodes as sensors in subjects with contralateral acoustic hearing. *J. Neural Eng.* 19, 2772–2785. doi: 10.1088/1741-2552/ac4de6

Arndt, S., Aschendorff, A., Laszig, R., Beck, R., Schild, C., Kroeger, S., et al. (2011). Comparison of pseudobinaural hearing to real binaural hearing rehabilitation after cochlear implantation in patients with unilateral deafness and tinnitus. *Otol. Neurotol.* 32, 39–47. doi: 10.1097/MAO.0b013e3181fcf271

Balkenhol, T., Wallhäusser-Franke, E., Rotter, N., and Servais, J. (2020). Cochlear implant and hearing aid: objective measures of binaural benefit. *Front. Neurosci.* 14, 586119. doi: 10.3389/fnins.2020.586119

Butler, R. A., Keidel, W. D., and Spreng, M. (1969). An investigation of the human cortical evoked potential under conditions of monaural and binaural stimulation. *Acta Otolaryngol.* 68, 317–326. doi: 10.3109/00016486909121570

Ching, T., Incerti, P., and Hill, M. (2004). Binaural benefits for adults who use hearing aids and cochlear implants in opposite ears. *Ear. Hear.* 25, 9–21. doi: 10.1097/01.AUD.0000111261.84611.C8

Ching, T., van Wanrooy, E., and Dillon, H. (2007). Binauralbimodal fitting or bilateral implantation for managing severe to profound deafness: a review. *Trends Amplif.* 11, 161–192. doi: 10.1177/1084713807304357

Ching, T. Y. C., Psarros, C., Hill, M., Dillon, H., and Incerti, P. (2001). Should children who use cochlear implants wear hearing aids in the opposite ear? *Ear. Hear.* 22, 365–380. doi: 10.1097/00003446-200110000-00002

Cohen, S., and Svirsky, M. (2019). Duration of unilateral auditory deprivation is associated with reduced speech perception after cochlear implantation: a single-sided deafness study. *Cochlear Implants Int.* 20, 51–56. doi: 10.1080/14670100.2018.1550469

Crew, J., Galvin, J., r., Landsberger, D., and Fu, Q. (2015). Contributions of electric and acoustic hearing to bimodal speech and music perception. *PLoS ONE* 10, 0120279. doi: 10.1371/journal.pone.0120279

Crosse, M. J., Di Liberto, G. M., Bednar, A., and Lalor, E. C. (2016). The multivariate temporal response function (mtrf) toolbox: a matlab toolbox for relating neural signals to continuous stimuli. *Front. Hum. Neurosci.* 10, 604. doi: 10.3389/fnhum.2016.00604

Delorme, A., and Makeig, S. (2004). Eeglab: an open-source toolbox for analysis of single-trial eeg dynamics. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

Deprez, H., Gransier, R., Hofmann, M., Wieringen, A., Wouters, J., and Moonen, M. (2017). Characterization of cochlear implant artifacts in electrically evoked auditory steady-state responses. *Biomed. Signal Process. Control* 31, 127–138. doi: 10.1016/j.bspc.2016.07.013

Devocht, E., Janssen, A., Chalupper, J., Stokroos, R., and George, E. (2017). The benefits of bimodal aiding on extended dimensions of speech perception: intelligibility, listening effort, and sound quality. *Trends Hear.* 21, 1–20. doi: 10.1177/2331216517727900

Dieudonn,é, B., and Francart, T. (2020). Speech understanding with bimodal stimulation is determined by monaural signal to noise ratios: No binaural cue processing involved. *Ear. Hear.* 41, 1158–1171. doi: 10.1097/AUD.0000000000000834

Dimitrijevic, A., Smith, M., Kadis, D., and et al. (2019). Neural indices of listening effort in noisy environments. *Sci. Rep.* 9, 11278. doi: 10.1038/s41598-019-47643-1

Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109

Dorman, M., Gifford, R., Spahr, A., and McKarns, S. (2008). The benefits of combining acoustic and electric stimulation for the recognition of speech, voice and melodies. *Audiol. Neurootol.* 13, 105–112. doi: 10.1159/000111782

Elberling, C., Don, M., Cebulla, M., and Stürzebecher, E. (2007). Auditory steady-state responses to chirp stimuli based on cochlear traveling wave delay. *J. Acoust. Soc. Am.* 122, 2772–2785. doi: 10.1121/1.2783985

Etard, O., and Reichenbach, T. (2019). Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *J. Neurosci.* 39, 1828–1818. doi: 10.1523/JNEUROSCI.1828-18.2019

Firszt, J., Chambers, R., Kraus, N., and Reeder, R. (2002). Neurophysiology of cochlear implant users i: effects of stimulus current level and electrode site on the electrical abr, mlr, and n1-p2 response. *Ear. Hear.* 23, 502–515. doi: 10.1097/00003446-200212000-00002

Firszt, J., Holden, L., Reeder, R., Cowdrey, L., and King, S. (2012). Cochlear implantation in adults with asymmetric hearing loss. *Ear. Hear.* 32, 521–533. doi: 10.1097/AUD.0b013e31824b9dfc

Fowler, J. R., Eggleston, J. L., Reavis, K. M., McMillan, G. P., and Reiss, L. A. (2016). Effects of removing low-frequency electric information on speech perception with bimodal hearing. *J. Speech Lang. Hear. Res.* 59, 99–109. doi: 10.1044/2015_JSLHR-H-15-0247

Fu, Q., Galvin, J., and Wang, X. (2017). Integration of acoustic and electric hearing is better in the same ear than across ears. *Sci. Rep.* 7, 2500. doi: 10.1038/s41598-017-12298-3

Galvin, J. J., Fu, Q.-J., Wilkinson, E., Mills, D., Hagan, S., Lupo, J. E., et al. (2019). Benefits of cochlear implantation for single-sided deafness: data from the house clinic-university of southern california-university of california,. *Ear. Hear.* 40, 766–781. doi: 10.1097/AUD.0000000000000671

Giraud, A.-L., and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* 15, 511–517. doi: 10.1038/nn.3063

Haufe, S., Meinecke, F., Görgen, K., Dähne, S., Haynes, J.-D., Blankertz, B., et al. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage* 87, 96–110. doi: 10.1016/j.neuroimage.2013.10.067

Henkin, Y., Yaar-Soffer, Y., Givon, L., and Hildesheimer, M. (2015). Hearing with two ears: evidence for cortical binaural interaction during auditory processing. *J. Am. Acad Audiol.* 26, 384–392. doi: 10.3766/jaaa.26.4.6

Hochmair-Desoyer, I., Schulz, E., Moser, L., and Schmidt, M. (1997). The hsm sentence test as a tool for evaluating the speech understanding in noise of cochlear implant users. *Am. J. Otol.* 18, 83.

Hofmann, M., and Wouters, J. (2010). Electrically evoked auditory steady state responses in cochlear implant users. *J. Assoc. Res. Otolaryngol.* 11, 267–282. doi: 10.1007/s10162-009-0201-z

Jancke, L., Wiistenberg, T., Schulze, K., and Heinze, H. (2002). Asymmetric hemodynamic responses of the human auditory cortex to monaural and binaural stimulation. *Hear. Res.* 170, 166–178. doi: 10.1016/S0378-5955(02)00488-4

Kaur, C., and Singh, P. (2015). "EEG artifact suppression based on SOBI based ICA using wavelet thresholding," in *IEEE 2015 2nd International Conference on Recent Advances in Engineering Computational Sciences* (Chandigarh: IEEE), 1–4. doi: 10.1109/RAECS.2015.7453319

Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking inauditory and motor cortex reflects distinct linguistic features. *PLos Biol.* 16, e2004473. doi: 10.1371/journal.pbio.2004473

Kong, Y., and Braida, L. (2011). Cross-frequency integration for consonant and vowel identification in bimodal hearing. *J. Rehabil. Res. Dev.* 54, 959–980. doi: 10.1044/1092-4388(2010/10-0197)

Kong, Y., and Carlyon, R. (2007). Improved speech recognition in noise in simulated binaurally combined acoustic and electric stimulation. *J. Acoust. Soc. Am.* 121, 3717–3727. doi: 10.1121/1.2717408

Kong, Y., Stickney, G., and Zeng, F.-G. (2005). Speech and melody recognition in binaurally combined acoustic and electric hearing. *J. Acoust. Soc. Am.* 3, 1351–1361. doi: 10.1121/1.1857526

Krüger, B., Büchner, A., and Nogueira, W. (2022). Phantom stimulation for cochlear implant users with residual low-frequency hearing. *Ear. Hear.* 43, 631–645. doi: 10.1097/AUD.0000000000001121

Lalor, E. C., Power, A. J., Reilly, R. B., and Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Neurophysiol* 102, 349–359. doi: 10.1152/jn.90896.2008

Lesenfants, D., Vanthornhout, J., Verschueren, E., Decruy, L., and Francart, T. (2019). Predicting individual speech intelligibility from the neural tracking of acoustic- and phonetic-level speech representations. *Hear. Res.* 380, 1–9. doi: 10.1016/j.heares.2019.05.006

Litovsky, R., Johnstone, P., and Godar, S. (2006). Benefits of bilateral cochlear implants and/or hearing aids in children. *Int. J. Audiol.* 45, 78–91. doi: 10.1080/14992020600782956

Lybarger, S. F. (1963). *Simplified Fitting System for Hearing Aid*. Canonsburg, PA: Radioear Corp.

Martin, B., Tremblay, K., and Stapells, S. (2007). "Principles and applications of cortical auditory evoked potentials," in *Auditory Evoked Potentials. Basic Principles*

*and Clinical Application* (Philadelphia, PA: Lippincott Williams and Wilkins), 482–507.

MATLAB (2018). *9.7.0.1190202 (R2019b)*. Natick, MA: The MathWorks Inc.

Maurer, J., Collet, L., Pelster, H., Truy, E., and Gallégo, S. (2002). Auditory late cortical response and speech recognition in digisonic cochlear implant users. *Laryngoscope* 112, 2220–2224. doi: 10.1097/00005537-200212000-00017

McPherson, D., and Starr, A. (1993). Binaural interaction in auditory evoked potentials: brainstem, middle- and long-latency components. *Hear. Res.* 66, 91–98. doi: 10.1016/0378-5955(93)90263-Z

Mesgarani, N., David, S. V., Fritz, J. B., and Shamma, S. A. (2009). Influence of context and behavior on stimulus reconstruction from neural activity in primary auditory cortex. *Neurophysiol* 102, 3329–3339. doi: 10.1152/jn.91128.2008

Mirkovic, B., Debener, S., Jaeger, M., and De Vos, M. (2015). Decoding the attended speech stream with multi-channel EEG: implications for online, daily-life applications. *J. Neural Eng.* 12, 046007. doi: 10.1088/1741-2560/12/4/046007

Mok, M., Grayden, D., Dowell, R., and Lawrence, D. (2006). Speech perception for adults who use hearing aids in conjunction with cochlear implants in opposite ears. *J. Rehabil. Res. Dev*. 49, 338–351. doi: 10.1044/1092-4388(2006/027)

Nogueira, W., Cosatti, G., Schierholz, I., Egger, M., Mirkovic, B., and Buchner, A. (2019a). Towards decoding selective attention from single-trial eeg data in cochlear implant users. *IEEE Trans. Biomed. Eng.* 67, 38–49. doi: 10.1109/ICASSP40776.2020.9054021

Nogueira, W., and Dolhopiatenko, H. (2022). Predicting speech intelligibility from a selective attention decoding paradigm in cochlear implant users. *J. Neural Eng.* 19, 5991. doi: 10.1088/1741-2552/ac599f

Nogueira, W., Dolhopiatenko, H., Schierholz, I., Büchner, A., Mirkovic, B., Bleichner, M. G., et al. (2019b). Decoding selective attention in normal hearing listeners and bilateral cochlear implant users with concealed ear eeg. *Front. Neurosci.* 13, 720. doi: 10.3389/fnins.2019.00720

OSullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2015). Attentional selection in a cocktail party environment can be decoded from single-trial eeg. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Papesh, M., Billings, C., and Baltzell, L. (2015). Background noise can enhance cortical auditory evoked potentials under certain conditions. *Clin. Neurophysiol.* 126, 31319–31330. doi: 10.1016/j.clinph.2014.10.017

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Bio.l* 10, e1001251. doi: 10.1371/journal.pbio.1001251

Paul, B., Uzelac, M., and Chan, E. (2020). Poor early cortical differentiation of speech predicts perceptual difficulties of severely hearing impaired listeners in multitalker environments. *Sci. Rep.* 10, 6141. doi: 10.1038/s41598-020-63103-7

Pelizzone, M., Hari, R., Mäkel,ä, J., Kaukoranta, E., and Montandon, P. (1987). Cortical activity evoked by a multichannel cochlear prosthesis. *Acta Otolaryngol.* 103, 632–636.

Petersen, E. B., Wöstmann, M., Obleser, J., and Lunner, T. (2017). Neural tracking of attended versus ignored speech is differentially affected by hearing loss. *J. Neurophysiol.* 117, 18–27. doi: 10.1152/jn.00527.2016

Ponton, C., Don, M., Eggermont, J., Waring, M., and Masuda, A. (1996). Maturation of human cortical auditory function: differences between normal-hearing children and children with cochlear implants. *Ear. Hear.* 17, 430–437. doi: 10.1097/00003446-199610000-00009

Potts, L., Skinner, M., Litovsky, R., Strube, M., and Kuk, F. (2009). Recognition and localization of speech by adult cochlear implant recipients wearing a digital hearing aid in the nonimplanted ear (bimodal hearing). *J. Am. Acad. Audiol.* 20, 353–373. doi: 10.3766/jaaa.20.6.4

Power, A. J., Foxe, J. J., Forde, E.-J., Reilly, R. B., and Lalor, E. C. (2012). At what time is the cocktail party? a late locus of selective attention to natural speech. *Eur. J. Neurosci.* 35, 1497–1503. doi: 10.1111/j.1460-9568.2012.08060.x

Prejban, D., Hamzavi, J., Arnoldner, C., Liepins, R., Honeder, C., Kaider, A., et al. (2018). Single sided deaf cochlear implant users in the difficult listening situation: speech perception and subjective benefit. *Otol. Neurotol.* 39, e803-e809. doi: 10.1097/MAO.0000000000001963

Reiss, L., Ito, R., Eggleston, J., and Wozny, D. (2014). Abnormal binaural spectral integration in cochlear implant users. *J. Assoc. Res. Otolaryngol.* 15, 235–248. doi: 10.1007/s10162-013-0434-8

Reiss, L., Perreau, A., and CW., T. (2012a). Effects of lower frequency-to-electrode allocations on speech and pitch perception with the hybrid short-electrode cochlear implant. *Audiol. Neurotol.* 17, 357–372. doi: 10.1159/000341165

Reiss, L., Turner, C., Karsten, S., Ito, R., Perreau, A., McMenomey, S., et al. (2012b). *Electrode Pitch Patterns in Hybrid and Long-Electrode Cochlear Implant Users: Changes Over Time and Long-Term Data*. San Diego, CA: Midwinter Research Meeting of the Association for Research in Otolaryngology.

Sasaki, T., Yamamoto, K., Iwaki, T., and Kubo, T. (2009). Assessing binaural/bimodal advantages using auditory event-related potentials in subjects with cochlear implants. *Auris Nasus Larynx* 36, 541–546. doi: 10.1016/j.anl.2008.12.001

Sharma, A., Dorman, M., and Spahr, A. (2002). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear. Hear.* 23, 532–539. doi: 10.1097/00003446-200212000-00004

Somers, B., Francart, T., and Bertrand, A. (2010). A generic EEG artifact removal algorithm based on the multi-channel wiener filter. *J. Neural Eng.* 15, 36007. doi: 10.1088/1741-2552/aaac92

Stapells, D. R. (2009). "Cortical event-related potentials to auditory stimuli," in *Handbook of Clinical Audiology Edition* (Baltimore, MD: Lippingcott; Williams Wilkins).

Van de Heyning, P., Vermeire, K., Diebl, M., Nopp, P., Anderson, I., and De Ridder, D. (2008). Incapacitating unilateral tinnitus in single-sided deafness treated by cochlear implantation. *An. Otorhinolaryngol.* 117, 645–652. doi: 10.1177/000348940811700903

Vanthornhout, J., Decruy, L., Wouters, J., Simon, J. Z., and Francart, T. (2018). Speech intelligibility predicted from neural entrainment of the speech envelope. *J. Assoc. Res. Otolaryngol.* 19, 181–191. doi: 10.1007/s10162-018-0654-z

Vermeire, K., and Van de Heyning, P. (2009). Binaural hearing after cochlear implantation in subjects with unilateral sensorineural deafness and tinnitus. *Audiol. Neurootol.* 14, 163–171. doi: 10.1159/000171478

Wedekind, A., Rajan, G., Van Dun, B., and Távora-Vieira, D. (2020). Restoration of cortical symmetry and binaural function: cortical auditory evoked responses in adult cochlear implant users with single sided deafness. *PLoS ONE* 15, e0227371. doi: 10.1371/journal.pone.0227371

Wedekind, A., Távora-Vieira, D., Nguyen, A. T., Marinovic, W., and Rajan, G. P. (2021). Cochlear implants in single-sided deaf recipients: near normal higher-order processing. *Clin. Neurophysiol.* 132, 449–456. doi: 10.1016/j.clinph.2020.11.038

Yang, H., and Zeng, F. (2013). Reduced acoustic and electric integration in concurrent-vowel recognition. *Sci. Rep.* 3, 1419. doi: 10.1038/srep01419

Yoon, Y., Shin, Y., Gho, J., and Fu, Q. (2015). Bimodal benefit depends on the performance difference between a cochlear implant and a hearing aid. *Cochlear Implants Int.* 16, 159–167. doi: 10.1179/1754762814Y.0000000101

Zirn, S., Arndt, S., Aschendorff, A., and Wesarg, T. (2015). Interaural stimulation timing in single sided deaf cochlear implant users. *Hear. Res.* 328, 148–156. doi: 10.1016/j.heares.2015.08.010

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 17 January 2023
DOI 10.3389/fnins.2022.1080398



OPEN ACCESS

EDITED BY
Lina Reiss,
Oregon Health and Science University, United States

REVIEWED BY
Erol J. Ozmeral,
University of South Florida, United States
Marzieh Amiri,
Ahvaz Jundishapur University of Medical Sciences, Iran

*CORRESPONDENCE
Kavassery Venkateswaran Nisha
✉ nishakv@aiishmysore.in

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 26 October 2022
ACCEPTED 20 December 2022
PUBLISHED 17 January 2023

CITATION
Nisha KV, Uppunda AK and Kumar RT (2023) Spatial rehabilitation using virtual auditory space training paradigm in individuals with sensorineural hearing impairment. *Front. Neurosci.* 16:1080398. doi: 10.3389/fnins.2022.1080398



# Spatial rehabilitation using virtual auditory space training paradigm in individuals with sensorineural hearing impairment

Kavassery Venkateswaran Nisha*, Ajith Kumar Uppunda and Rakesh Trinesh Kumar

Department of Audiology, All India Institute of Speech and Hearing (AIISH), Mysore, India

**Purpose:** The present study aimed to quantify the effects of spatial training using virtual sources on a battery of spatial acuity measures in listeners with sensorineural hearing impairment (SNHI).

**Methods:** An intervention-based time-series comparison design involving 82 participants divided into three groups was adopted. Group I ($n$ = 27, SNHI-spatially trained) and group II ($n$ = 25, SNHI-untrained) consisted of SNHI listeners, while group III ($n$ = 30) had listeners with normal hearing (NH). The study was conducted in three phases. In the pre-training phase, all the participants underwent a comprehensive assessment of their spatial processing abilities using a battery of tests including spatial acuity in free-field and closed-field scenarios, tests for binaural processing abilities (interaural time threshold [ITD] and level difference threshold [ILD]), and subjective ratings. While spatial acuity in the free field was assessed using a loudspeaker-based localization test, the closed-field source identification test was performed using virtual stimuli delivered through headphones. The ITD and ILD thresholds were obtained using a MATLAB psychoacoustic toolbox, while the participant ratings on the spatial subsection of speech, spatial, and qualities questionnaire in Kannada were used for the subjective ratings. Group I listeners underwent virtual auditory spatial training (VAST), following pre-evaluation assessments. All tests were re-administered on the group I listeners halfway through training (mid-training evaluation phase) and after training completion (post-training evaluation phase), whereas group II underwent these tests without any training at the same time intervals.

**Results and discussion:** Statistical analysis showed the main effect of groups in all tests at the pre-training evaluation phase, with *post hoc* comparisons that revealed group equivalency in spatial performance of both SNHI groups (groups I and II). The effect of VAST in group I was evident on all the tests, with the localization test showing the highest predictive power for capturing VAST-related changes on Fischer discriminant analysis (FDA). In contrast, group II demonstrated no changes in spatial acuity across timelines of measurements. FDA revealed increased errors in the categorization of NH as SNHI-trained at post-training evaluation compared to pre-training evaluation, as the spatial performance of the latter improved with VAST in the post-training phase.

**Conclusion:** The study demonstrated positive outcomes of spatial training using VAST in listeners with SNHI. The utility of this training program can be extended to other clinical population with spatial auditory processing deficits such as auditory neuropathy spectrum disorder, cochlear implants, central auditory processing disorders etc.

KEYWORDS

**virtual auditory space training, localization, virtual acoustics, binaural cue processing, interaural threshold differences, perceptual ratings, spatial hearing, spatial processing**

## Introduction

Deficits in spatial hearing secondary to hearing loss have a direct bearing on day-to-day communication in listening environments (Abel et al., 2000), such as listening in noise (Kidd et al., 2005) and reverberation (Takahashi, 2009). The impact of hearing loss in listeners with sensorineural hearing impairment (SNHI) on their ability to use auditory spatial cues is readily observable on most psychoacoustical measures. Investigations on binaural processing reported poorer values in time [interaural time difference (ITD)] (Hawkins and Wightman, 1980; Kinkel et al., 1991), intensity (interaural level difference [ILD]) (Kinkel et al., 1991; Gabriel et al., 1992; Spencer et al., 2016), phase (interaural phase difference [IPD]) (Lacher-Fougère and Demany, 2005; Neher et al., 2011), and interaural cross-correlation (Gabriel et al., 1992; Spencer et al., 2016) in listeners with SNHI compared to their normal-hearing counterparts. In natural environments, changes in ITDs are usually accompanied by corresponding changes in IPDs and ILDs; however, precise control of the acoustic parameters in laboratory conditions powers the investigators to manipulate either of the cues alone or in combination, thus aiding in understanding the role of each cue (ITD, ILD, and IPD) in spatial processing.

The spatial deficits seen in SNHI listeners on the psychoacoustical measures affect their spatial performance in free-field (Best et al., 2010; van den Bogaert et al., 2011; Kuk et al., 2013; Brimijoin and Akeroyd, 2016) and closed-field scenarios (Chung et al., 2008; van Esch et al., 2013; Brimijoin and Akeroyd, 2014; Brungart et al., 2017). Apart from showing deficits in psychoacoustical measures, SNHI listeners also experience perceptual difficulties in everyday listening (Noble and Gatehouse, 2006). Spatial perception is paramount for comfortable listening in daily environments (Risoud et al., 2018), and deficits in spatial hearing places SNHI listeners at a disadvantage on a variety of tasks, including spatial navigation, speech understanding, communication in adverse listening environments (Best et al., 2010), and lowered self-confidence in their social interactions. Undoubtedly, the poor ability to localize sound accurately is a common source of frustration for SNHI listeners (Subramaniam et al., 2005). In addition, the increased spatial disability in SNHI listeners was associated with other avoidance behaviors such as the desire to escape from situations in which sounds were confusing and caused nervousness (Noble et al., 1995).

Akeroyd and Whitmer (2016) reviewed 12 studies on the localization of real sources in different quadrants (right, left, front, and back) of the acoustic field to calculate root mean square (RMS) localization error. They reported that localization/RMS error was 5° higher in SNHI listeners compared to normal hearing (NH) listeners in right-left hemifields. When considering directional acuity in the front–back dimension, mean front–back confusion rates for NH listeners were found to range from 0.1 to 5%, while those with SNHI ranged between 10 and 26% (Best et al., 2010; van den Bogaert et al., 2011). Studies exploring spatial acuity in SNHI listeners in closed-field environments using virtual sources report high front-back confusion in both SNHI and NH listeners (relative to confusions in free field), with the former exhibiting greater errors than the latter (Chung et al., 2008; van Esch et al., 2013; Brimijoin and Akeroyd, 2014; Brungart et al., 2017). Furthermore, on subjective ratings, listeners with SNHI are known to experience more serious localization difficulties that increase with the degree of HI (Noble et al., 1997; Glyde et al., 2013). Deficits in a number of peripheral and central processes including reduced audibility (Brimijoin and Akeroyd, 2016), impaired frequency selectivity (Strelcyk and Dau, 2009), poor temporal resolution, altered filtered shapes (Baker and Rosen, 2002; Bernstein and Oxenham, 2006), and increased spectral and temporal masking (Le Goff et al., 2013) can be conceived as factors for impaired spatial processing in SNHI listeners.

Although literature highlights the impact of compromised spatial acuity in human communication in listeners with SNHI, remediation programs aimed at resolving spatial deficits are surprisingly few. Some of the notable strides in enhancing spatial acuity have used hearing aids (HAs) with novel algorithms for auditory spatial coding (Drennan et al., 2005), gain settings (Keidser et al., 2011), bilateral synchronization (Johnson et al., 2017), and listening configurations (van den Bogaert et al., 2006;

Neher et al., 2009). Although theoretically enhancing temporal, spectral, and intensity cues aiding directional perception should be possible using novel spatial processing algorithms in HAs, studies involving the localization ability of HA users have shown poor-to-mixed results (refer to reviews—Denk et al., 2019; Zheng et al., 2022). Reports suggest slight improvements in localization ability after a period of acclimatization in SNHI listeners (Noble and Byrne, 1990; Drennan et al., 2005) however, contrary evidence on decreased localization performance after HAs usage is also available abundantly (Byrne and Noble, 1998; van den Bogaert et al., 2011; Akeroyd, 2014). A number of factors such as the design of HA, degree of loss, cognition, age of the participant, testing condition (aided or unaided, noise or quiet, open or closed earmolds), and test material (speech, noises with various bandwidths, center frequencies, spectral slopes, and real-world sounds such as telephone ring) could have influenced these results. Akeroyd and Whitmer (2016) reviewed 36 studies that compared directional acuity of unaided and bilaterally aided hearing-impaired (mostly mild to severe sloping and older hearing-impaired adults) listeners and reported only a slight 1° difference between the aided and unaided scores (the RMS error: RMS values were 12° for unaided listening and 13° for aided listening). Although within-subject variability was seen between aided and unaided conditions with the differences reaching statistical significance in a few reports (Keidser et al., 2009; van den Bogaert et al., 2011), contrary evidence is also available (Drennan et al., 2005; Best et al., 2010; Brungart et al., 2014). Only four of 36 reviewed studies (11.11%) showed a benefit of aiding of at least 1°, whereas more than 20 studies (55.55%) showed a deficit of aiding of at least 1° and nine (=25%) reports showed a deficit of 3° or more. Taken together with the multiplicity of differences between studies (HA styles, test materials, and conditions) in Akeroyd and Whitmer's (2016) review, the results can be treated to be more or less representative of the effect of aiding on spatial perception. The use of HAs (as seen in studies earlier) can negatively affect spatial perception as they reduce the HRTF cues and also distort ITD and ILD cues.

Alternatively, minimal improvements in spatial performance are reported when spatial training programs use loudspeakers in free field (Tyler et al., 2010; Kuk et al., 2014) or interaural difference training (Wright and Fitzgerald, 2001; Rowan and Lutman, 2006; Spierer et al., 2007; Zhang and Wright, 2007; Ortiz and Wright, 2009) under closed field (headphones), although these improvements were not clinically or statistically significant. The clinical utility of training-related remedial programs in auditory spatial perception is limited by a number of factors, such as those related to study design (small sample sizes, heterogeneous outcome measures, inconsistent use of control groups, and limits of generalization) as well as those related to technical aspects such as length of the training programs and the cost–benefit ratio.

All the research efforts in documenting the effects of spatial deficits can be productive only if promising intervention strategies are devised. In addition to the minimal spatial acuity improvements reported in SNHI listeners consequent to training, a lot of issues related to study design and technical aspects question their utility in day-to-day practice. It is unknown whether everyday localization accuracy can be facilitated by training and whether improvements identified in a laboratory setting can be sustained and generalized.

The present study is intervention-based research that investigated the effect of virtual auditory space training (VAST) (Nisha and Kumar, 2017, 2018) on spatial acuity in listeners with SNHI. VAST is a novel paradigm that relies on auralization techniques to synthesize spatial percepts called virtual acoustic stimuli, which cause an illusionary effect of natural sound-field localization within the head (King et al., 2001). The virtual stimuli are constructed by superposing the target stimuli with the non-individualized HRTFs (refer to methods for stimulus generation), that enriches the stimuli with important spatial cues such as the ITDs, ILDs, spectral, and HRTFs. Thus, generated virtual stimuli are played back within the head using headphones in a systematically graded order of spatial difficulty (refer to methods for the hierarchy of stimulus presentation) and the listener is trained to achieve mastery in each level under self-supervision (refer to methods for detailed training procedure). The use of the VAST paradigm is proven to be as effective as free-field spatial training using loudspeakers in fine-tuning spatial skills of NH listeners (Nisha and Kumar, 2022) and has promising implications on cortical re-organization (Nisha and Kumar, 2019a). In addition, the use of virtual stimuli identical to those used in the VAST paradigm has been effective in documenting the effects of SNHI on cortical processing (Nisha and Kumar, 2019a,b), the effects of maturational and aging-related changes across life-span (Nisha et al., 2023), and the musical training effects (Nisha et al., 2022). The present study aimed to validate the efficacy of the VAST paradigm in resolving spatial perception deficits in SNHI listeners. Specifically, the objectives of the study were to document the spatial processing abilities in SNHI and compare the same with NH listeners using spatial acuity measures in the free field and closed field, binaural cue processing, and subjective ratings. The study additionally aimed to compare the pre-, mid-, and post-training performance of SNHI listeners on the above psychoacoustic spatial measures at different timelines as a function of VAST.

# Methods

## Participants

A total of 82 participants in the age range of 35–55 years were recruited for the present study, and they were divided into three groups. Groups I and II consisted of listeners

with mild to moderate (Katz, 2015) flat SNHI (Pittman and Stelmachowicz, 2003). While the group I ($n = 27$, 17 males, 10 females, $M_{age} = 43.8 \pm 10.44$ years SD) consisted of individuals with SNHI who underwent spatial training (VAST), group II consisted of individuals with SNHI who did not receive spatial training ($n = 25$, 12 males, 13 females, $M_{age} = 43.7 \pm 5.92$ years SD). In addition, group III ($n = 30$, 17 males, 13 females, $M_{age} = 38.55 \pm 3.25$ years SD) had participants with normal peripheral hearing sensitivity. The sample size considered in the study was calculated using G*Power version 3.1.9.4 (Faul et al., 2007) based on the effect size reported by Kuk et al. (2014) on spatial training in SNHI listeners. Although the sample size calculated was only 8 participants in each group for an effect size of 1.39, a higher sample size of at least 25 was considered for each group in the present study. In contrast to Kuk et al.'s (2014) study, where the SNHI participants who underwent spatial training also used HAs, the current study was performed on SNHI listeners who did not use HAs. The induction of SNHI listeners who had no previous exposure of HAs helped us to avoid potential limitations of HAs on auditory spatial processing (refer to the "Introduction" Section). The sample size used by us (group I = 27, group II = 25, and group III = 30) was, therefore, deemed appropriate for measuring changes due to VAST.

All the participants were subjected to a detailed case history to rule out any external and middle ear pathology. Pure tone audiometry was conducted on all participants where both the air conduction (250–8,000 Hz) and bone conduction thresholds (250–4,000 Hz) were obtained at the octave frequencies using the modified Hughson and Westlake procedure given by Carhart and Jerger (1959). The air conduction and bone conduction thresholds of these participants were tested using a Piano inventis audiometer (Inventis, Padova, Italy) by routing stimulus through Telephonics TDH 39 earphones (Telephonics, Farmingdale, NY, USA) and B71 bone vibrator (RadioEar, Kimmetrics, Smithsburg, MD, USA). Based on four frequency pure tone AC thresholds (pure tone average (PTA) obtained as an average of thresholds at 500, 1,000, 2,000, and 4,000 Hz), participants with mild to moderate SNHI formed groups I and II. The SNHI was further confirmed by the absence of otoacoustic emissions and acoustic reflexes (both ipsilateral and contralateral reflex) recorded using standard recording protocol in Otodynamics ILO V6 DP Echoport (Otodynamics Ltd., Hatfield, Herts, UK) and Inventis Clarinet (Inventis Inc., Padova, Italy) instruments, respectively. Among these two groups, the former received formal spatial training using VAST and the latter served as the control group that did not receive any spatial training. Participants with NH sensitivity, i.e., PTA <25 dB HL were selected for group III. Figure 1 shows the thresholds for (Figure 1A) right and (Figure 2B) left ears across the groups. The group differences in the thresholds were verified using one-way ANOVA [right ear: $F_{(2,79)} = 171.31$, $p < 0.001$; left ear: $F_{(2,79)} = 121.82$, $p < 0.001$], followed by Bonferroni comparisons which showed that the thresholds of the two SNHI groups were similar (right ear: $p = 0.75$; left ear: $p = 0.35$) and significantly higher (right and left ears: $p < 0.001$) than the NH group.

Participants with any neurological and cognitive deficits were excluded from the study. Mini-Mental State Examination (MMSE) (Folstein et al., 1983) translated into Indian English (Milman et al., 2018) was administered to rule out any pathological cognitive decline (all participants scored more than 24 points). Group equivalency in cognition across groups was also cross-checked statistically from the results of one-way ANOVA, which showed no main effect of the group [$F_{(2,79)} = 1.09$, $p = 0.34$] on MMSE scores. In addition, the influence of musical training and aptitude in participants of the study was ruled out based on informal interviews (for any formal history of musical training) and cutoff scores ($\leq$18) on mini-Profile of Music Perception Skills (mini-PROMS, Zentner and Strauss, 2017). The results of one-way ANOVA for mini-PROM verified that there were no differences in musical aptitude across the three groups [$F_{(2,79)} = 1.19$, $p = 0.31$].

Written consent was obtained from all the participants, which confirmed their willingness to participate in the research study. The study was approved by the institutional core committee on research and adhered to the institutional ethical guidelines of bio-behavioral research involving human subjects (Venkatesan, 2009) under the reference number SH/CDN/ARF-AUD-4/2018-19.

## Research design

A mixed design (which includes both between-subjects design and within-subjects design) (Schiavetti and Metz, 2006) based on the intervention-control model was used. Standard group comparison (SNHI vs. NH) was adopted to study the spatial processing differences between the groups (between-subjects design), while a time-series (pre-, mid-, and post-training) design was used to evaluate the effect of VAST on the auditory spatial performance of SNHI listeners (within the subject design).

## Procedure

The study was conducted in three phases, i.e., pre-training, training, and post-training phase.

### Phase I: Pre-training evaluation phase

In the pre-training phase, the spatial acuity of all listeners was assessed using a test battery comprising three psychoacoustical measures and one subjective measure as discussed in the following sections. The stimulus presentation

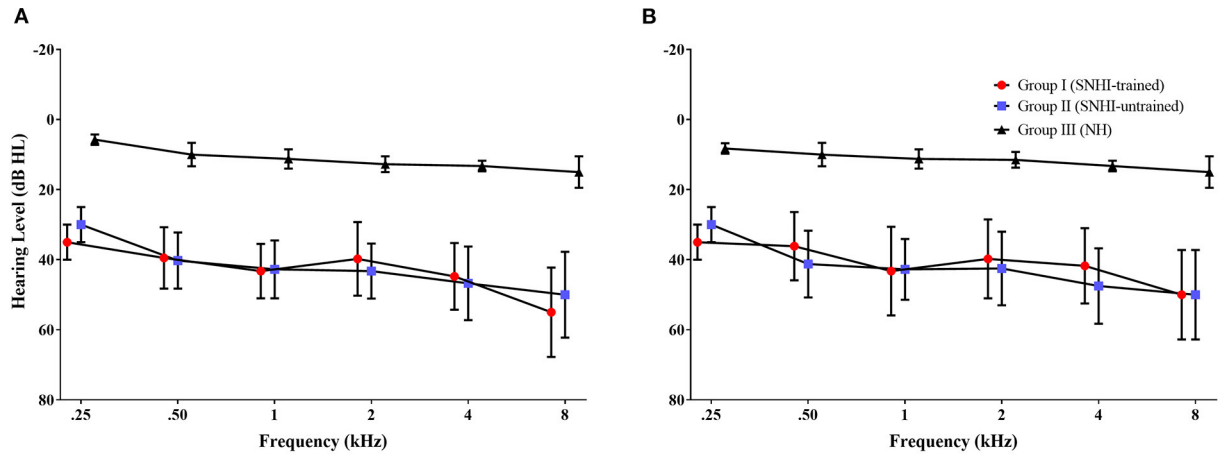



FIGURE 1
The mean hearing thresholds of participants in each group across audiometric frequency range for **(A)** right ear and **(B)** left ear. The error bar represents ± 1 SD.

level was maintained constant at 80 dB SPL for all the psychoacoustical tests employed in this study.

### Test of spatial acuity in free field (localization test)

White noise bursts of 250 ms (inclusive of 5 ms rise and fall time) were generated using the AUX viewer software (mono, 32 bits, 44,100 sampling rate), calibrated to the output level of 80 dB SPL (using sound level meter; Bruel and Kjaer 2270, Naerem, Denmark), and used as stimuli for the test of spatial acuity in free field (localization test). The stimuli were loaded on a personal computer and were assigned to 18 audio tracks using the Cubase software (Steinberg Media Technologies GmbH, Hamburg). One of the 18 tracks delivered the stimuli to the corresponding loudspeakers *via* Aurora mixer (Lynx Studio Technology Inc., California, USA). The 18 loudspeakers (Genelec 8020B BI-amplified monitoring system, Finland) were placed in a concentric circle with a spacing of 20° from each other covering a complete spatial field spanning 360° azimuth, as shown in Figure 2A. The 18 loudspeaker array with 20° separation used in the current study is in accordance with the localization setups used for spatial studies on hearing-impaired listeners (Lorenzi et al., 1999; Drennan et al., 2005; Keidser et al., 2009).

The participant was asked to sit comfortably on a chair placed at a distance of 1 m in a center of the loudspeaker array in the localization chamber (semi-anechoic). After the delivery of the stimulus, the participant was asked to judge the location of the loudspeaker that emitted the sound and respond by writing the number corresponding to the loudspeaker. The test started with a pilot trial where 10 random presentations of the stimuli were given to make the participant familiarize to the task. In the testing phase, the stimuli were presented five times to each of the 18 speakers (in random order based on the track sequence in Cubase), and the participant was asked to respond by writing the speaker's number on a response sheet. Once the response to a particular trial was completed, the participant was asked to indicate its completion with a thumbs-up sign to the experimenter. The inter-trial interval for stimulus presentation depended on the response time of the participant. The next stimulus was presented only after the participant registered his response to the previous stimuli. The test was terminated after a total of 90 presentations (18 loudspeaker locations × 5 repetitions) and was completed in ∼15 min.

The responses were then entered in the user interface built for spatial error analysis in paradigm experimenter builder software (Perception Research Systems, 2007). A program written in a Python script running in the background of the interface recorded the target and the response location. The overall localization errors (RMS error) were computed in accordance with the formula given by Rakerd and Hartmann (1986). RMS error represents the root mean square of actual response deviation (in ° azimuth) from the target location. This was done for all the participants across the three groups.

### Test of spatial acuity in the closed field (virtual auditory space identification test)

VASI comprised of presentation of acoustic stimulus under the headphone at the target azimuths, which mimicked the free-field environment presentation. Although the stimuli were presented through the circumaural headphones (Sennheiser HD 280 PRO, Wedemark, Germany), the use of appropriate equalization techniques provided good azimuth replication that was comparable with the spatial hearing performance of individuals with NH in free-field environments (Pralong and Carlile, 1996; Zhong and Xie, 2013). Virtual percepts in the VASI test were created by convolving 250 ms white noise bursts

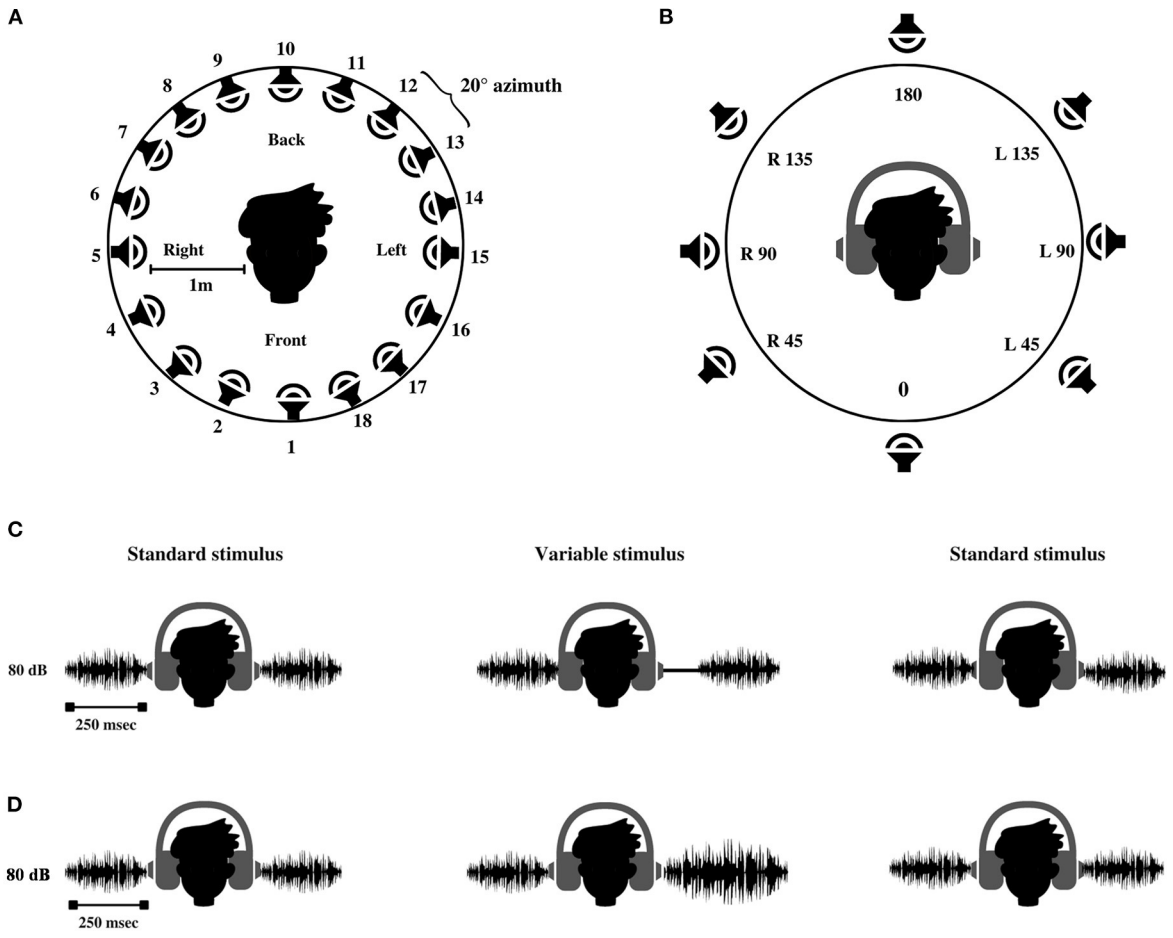



FIGURE 2
Schematic representation of **(A)** loudspeakers setup used for testing spatial acuity in free field. **(B)** Interface used for stimulus and response acquisition in virtual auditory space identification (VASI) test. The alphanumerical code represents the location of spatial percept created within the head. 0°-At the midline front, 180°-At the midline back, R45–45° azimuth toward the right ear; R90–90° azimuth toward the right ear, R135–135° azimuth toward the right ear; L45–45° azimuth toward the left ear, L90–90° azimuth toward the left ear, and L135–135° azimuth toward the left ear. **(C)** Interaural time difference test and **(D)** interaural level difference test.

with a non-individualized HRTF obtained from the sound lab (Slab3d) database. The sound lab (Slab 3D) version 6.7.3 (Spatial Auditory Displays lab, 2012) was used to control the generation of the virtual percepts in eight target locations: Midline front: 0° azimuth, midline back: 180° azimuth, and 45°, 90°, and 135° azimuth to the right and left. All the stimuli had constant elevation (0° azimuth) and distance (1 m). The virtual stimuli were synthesized to be identical to the free-field stimuli in terms of overall level (80 dB SPL) and duration (250 ms). The generated stimuli were calibrated to a level of 80 dB SPL using the sound level meter (Bruel and Kjaer 2270, Naerem, Denmark). The synthesized stimuli were loaded into the paradigm software, which controlled stimulus delivery and response acquisition using a graphical user interface, as shown in Figure 2B.

To allow for familiarization to VASI stimuli, the participants were encouraged to use practice runs. A dummy head (Figure 2B) was displayed on the monitor screen during these runs. The participant was instructed to use the mouse to click on each virtual location (no more than five trials per stimulus), and the corresponding virtual sound was emitted. In the testing phase, the eight stimuli were presented ten times at each location in random order. The order of presentation is randomized using a designated function in the stimulus characteristics window of the paradigm software. The participant was asked to attend to the virtual stimuli and click the mouse pointer on the position of the dummy head (Figure 2B), corresponding to the perceived location in the head. No feedback was given during the familiarization and testing phase. The test involved a presentation of a total of 80 virtual stimuli (eight loudspeaker

locations × ten repetitions) to all the participants and was completed in ∼15 min. After the completion of the experiment, the data corresponding to the target and the response locations stored in the output (excel) file was derived. The data comprising the VASI accuracy scores for each virtual location and overall VASI score (aggregate score of all eight virtual locations) were computed using a confusion matrix script (Gnanateja, 2014) in MATLAB version 2021b (The MathWorks Inc., Natick, MA, USA).

### Binaural processing (ITD and ILD thresholds)

The binaural processing abilities of the participants in the study were assessed using ITD and ILD thresholds. The test of ITD and ILD involved the presentation of two identical signals to both ears, with one ear receiving the signal slightly earlier or at a higher intensity relative to the other. The lowest intensity level at which a person reports the difference in intensity between the two ears is considered the ILD threshold. In the ITD test, the smallest time delay that a person can identify is considered a threshold for ILD. The difference in the time of arrival or the intensity between the two ears created lateralization of the stimuli toward one ear, which is to be detected by the participant.

The binaural abilities in the current study were assessed using a psychoacoustic toolbox (Soranzo and Grassi, 2014) implemented in MATLAB version 2021b (The MathWorks Inc., Natick, MA, USA). The ITD and ILD thresholds were measured in the three-interval forced-choice method. A two-down one-up staircase procedure was followed, which converged at 70.7% psychometric function (Levitt, 1971). Among the three stimuli in one trial, two were standard stimuli, and one was the variable stimulus. White noise bursts (250 ms, stereo, 16 bit, 44,100 sampling frequency, 80 dB SPL) similar in terms of binaural intensity and time of arrival, producing a midline sensation, were designated as standard stimuli. The variable stimuli were similar to the standard stimuli, except that it produced lateralization to the right ear due to inherent delay (introduced in the left ear) and increased intensity in one channel (increased intensity in one ear in the ILD test). The participants were instructed to compare the intensity of the signal between the two ears and report the interval in which the sound was lateralized to the right ear. The variable stimuli either led (ITD) or were heard louder (ILD) in the right ear, as shown in Figures 2C, D.

The time delay and intensity given in variable intervals are changed adaptively based on the response of the participant. The starting level of the delay for variable stimulus in the ITD test was 30 ms, which decreased by half when the participant recorded correct judgments in two successive trials (or) doubled when the participant made an incorrect judgment. For the test of ILD, starting level of variable stimuli was 20 dB higher in the right ear. The level of signal changed successively by a step size of 2 dB as the test progressed. The level was reduced by 2 dB when the participant recorded two successive correct responses and increased by 2 dB on registering an incorrect response. The test was terminated after 10 reversals, and the last four reversals were averaged to calculate the ITD and ILD thresholds. The ITD and ILD thresholds were tabulated and subjected to statistical analyses.

## Subjective ratings (Spatial sub-section of Spatial, Qualities, and Hearing Questionnaire in Kannada)

Participants rated the perceptual difficulties in spatial orientation using the Spatial sub-section of SSQ (Gatehouse and Noble, 2004), translated to Kannada (SSQ-K) (Shetty et al., 2019). This list contained 17 items that are administered on an 11-point rating scale, where 0 represents the minimal ability and 10 represents the complete ability to locate the sound source accurately.

## Phase II: Spatial training and mid-training evaluation phase

Participants in group I (SNHI-trained) underwent virtual acoustic space (VAST) using a hierarchy of graded VAS stimuli. The spatial training was performed using VAS stimuli as it facilitated definite simulation of spatial (azimuth) location within the head (Wenzel et al., 1993; Hartmann and Wittenberg, 1996). The use of VAS stimuli for spatial training allowed us to have systematic control on varying the levels of spatial perception difficulty, which were introduced using important source lateralization cues, namely, the length/duration of the signal and the number of locations. In addition, VAST enhanced the practicality of the spatial training paradigm as its implementation required only minimal equipment, which was easily portable, and the participants could undergo spatial training at home as well.

The VAST paradigm used in the current study was adapted from Kuk et al. (2014). The VAST paradigm was proven to be effective in resolving front-back confusion in NH listeners (Nisha and Kumar, 2017, 2022), and thus its application in SNHI listeners was conceptualized in the present study. The complexity of VAS stimuli varied adaptively in terms of their durations (stages: 2,000, 1,000, 500, and 300 ms) and the number of locations (levels 4, 6, and 8), as shown in Figure 3. Irving and Moore (2011) showed that a longer stimulus is easier to localize and easier to train than a shorter stimulus. It is also easier to judge the spatial locations of virtual sources that are distant apart than those that are closely spaced (Carlile et al., 2016). Training commenced from the easiest level ($S_1L_1$: 2,000 ms, four locations) and accurate judgments were counted. The VASI accuracy at each level was calculated, based on which they progressed to the next difficult level based on 70% VASI accuracy

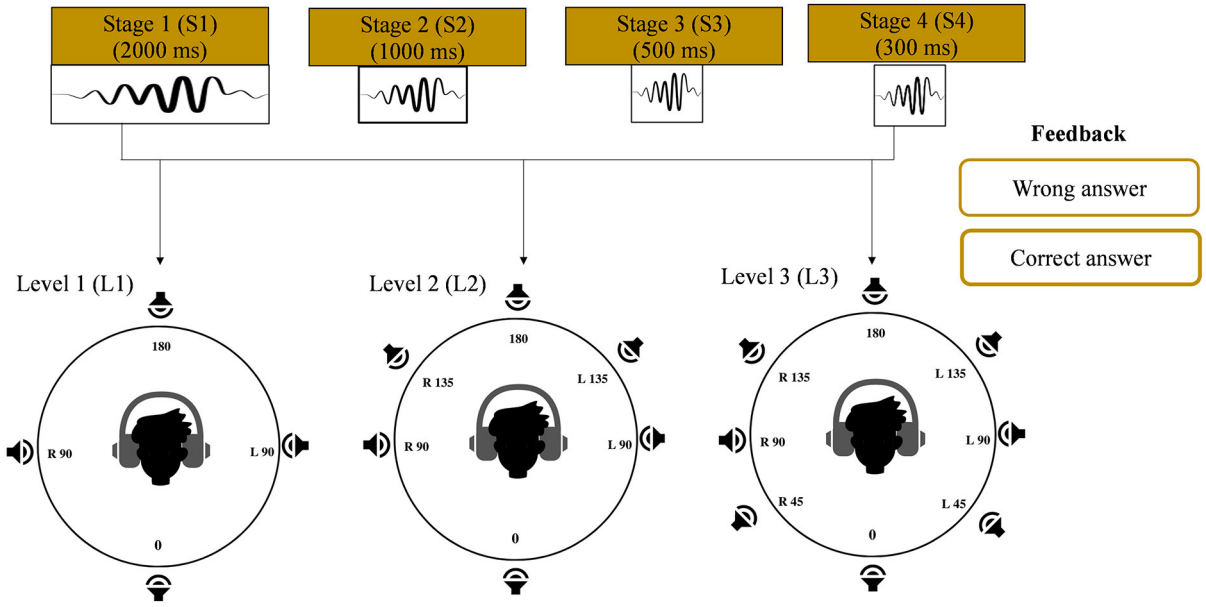



FIGURE 3
Hierarchy of stimulus (duration and number of virtual locations parameters) presentation in the training phase of the study. Progression from left to right represents easy to difficult conditions. Feedback given during the training is also shown.

criteria. The most challenging level ($S_4L_3$) was stimuli in stage 4 with a duration of 300 ms and eight locations.

### VAST paradigm

The user interface for VAST was built using the Paradigm experimental builder software. Two interfaces were separately designed for familiarization and training modules. In each of these interfaces, a display of a dummy head with four, six, and eight locations corresponding to levels 1, 2, and 3, respectively, is configured, as shown in Figure 3. The training was carried out in two modules. Module I was a familiarization task, while module II was a training task. In the familiarization module, the participants were encouraged to get familiar with the VAS stimuli (not more than 10 trials per stimulus) using a practice run. In this run, a graphical interface consisting of a dummy head with varying VAS locations (4 in level 1, 6 in level 2, and 8 in level 3) was displayed on the monitor. The participants were asked to click the mouse pointer on the position of the dummy head, and the corresponding VAS sound file was played. In the training module, the stimuli were played randomly and the participants were instructed to cautiously attend to the stimuli and click the mouse pointer on the position of the dummy head (Figure 3) corresponding to the perceived location in the head. Each VAS stimulus was randomly presented 7 times, thus making a total of 28 (4 locations × 7 repetitions) presentations for level 1, 42 (6 locations × 7 repetitions) presentations for level 2, and 56 (eight locations × seven repetitions) presentations for level 3 in each stage (stage 1: 2,000 ms; stage 2: 1,000 ms; stage 3: 500 ms; and stage 4: 300 ms). Once the participant registered his response through the mouse click, corrective feedback on the response was given. The correct responses were acknowledged, while the incorrect responses were compared with the correct location. 70% criterion was set up to progress from each level. The total VAST paradigm was completed in 5–8 sessions (30 min each), depending on the rate of learning and mastery obtained by the participant.

Halfway through training, the spatial skills of group I participants were re-assessed using all the tests conducted in the pre-training phase. This assessment provided an opportunity to understand the time course/pattern of spatial learning in group I participants who underwent VAST. All the tests were also conducted in group II participants (SNHI-untrained), which constituted the second evaluation (at a similar timeline as the post-training evaluation in group I) in them. Following this evaluation, the participants in group I completed the remaining stages of training.

## Phase III: Post-training evaluation phase

Immediately after the completion of training (0–5 days post-training), the spatial test battery was re-administered to the group I participants. The spatial tests were also administered on group II participants (SNHI-untrained), which constituted the third evaluation (at a similar timeline as the post-training evaluation in group I) in them.

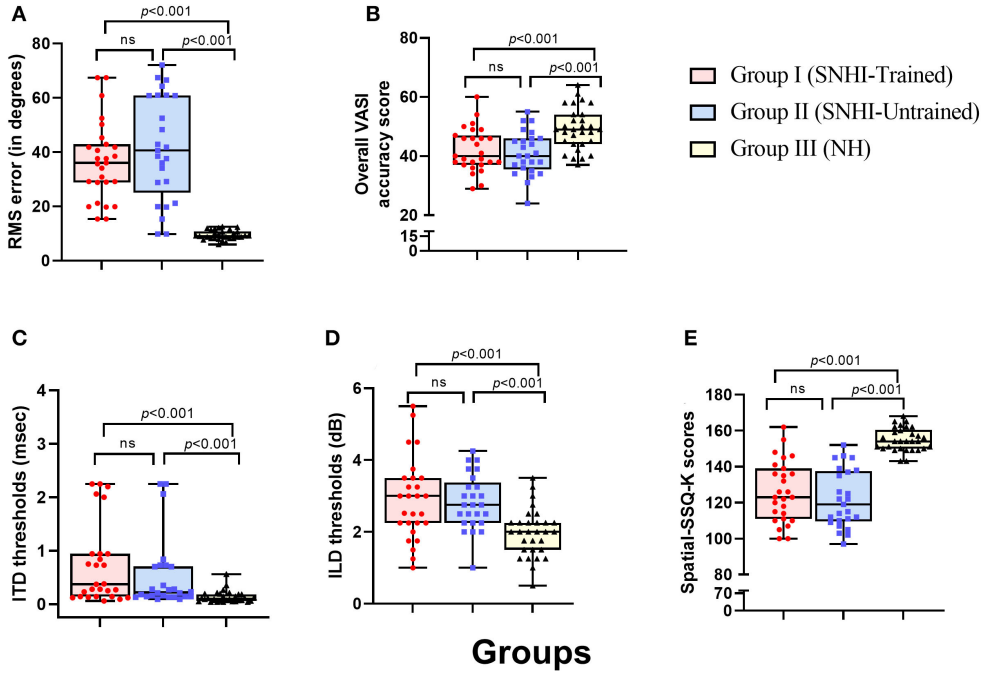



FIGURE 4
Comparison of spatial acuity across groups in **(A)** test of spatial acuity in free field, **(B)** test of spatial acuity in closed-field, **(C)** interaural time difference thresholds, **(D)** interaural level difference thresholds, and **(E)** subjective ratings on spatial sub-section of SSQ-K. The horizontal line in each box plot at center represents the median, while the "+" sign indicates the mean for each test. The box area corresponds to interquartile range, while the error bar indicates interquartile deviation. The results of Bonferonni comparisons are also indicated.

## Statistical analysis

The data obtained from all the tests were subjected to statistical analyses using the IBM Statistical Package for the Social Sciences (SPSS) version 25 software (IBM Corp., Armonk, NY, USA). Shapiro–Wilk test of normality was employed to check if the data follow the normal or non-normal distribution. For the data that followed a normal distribution, multivariate analysis of variance (MANOVA) and follow-up pairwise comparisons using independent *t*-tests (with Bonferroni's correction) were carried out for each measure of spatial acuity. These tests were conducted to compare the performance of SNHI (groups I and II) and NH (group III) participants. However, the Mann–Whitney *U*-test was performed for data that did not adhere to normal distribution. For analyzing the effect of training across evaluation phases in group I and II participants, within-subject tests of repeated measure ANOVA, followed by Bonferroni's test or Friedman test and then by Dunn–Bonferroni's test were carried out for normal and non-normal data distributions, respectively.

In addition, to explore the impact of VAST on various behavioral measures of spatial acuity in group I (SNHI-trained) and to compare the same with the spatial performance in group II (SNHI-untrained) and NH participants (group III), Fischer's discriminant analysis (FDA) was done. A default mathematical operation (Di = a + b1 x 1 + b2 x 2 +…+ bnxn; Di = predicted discriminant score; a = a constant, x = predictor; and b = discriminant coefficient) run in SPSS version 25 for group categorization was employed in the study. The main purpose of discriminant function (DF) analysis in this study was for group segregation and identification of the optimal spatial measure (RMS error, VASI scores, ITD and ILD thresholds, and spatial subsection of SSQ-K scores) that best predicts VAST-related changes. The FDA was performed for each measurement phase separately (pre-, mid-, and post-evaluation), while the error in classification at each phase is also reported.

# Results

## Comparison of the spatial performance of listeners with sensorineural hearing impairment and normal hearing sensitivity

The descriptive statistics showing the median and mean along with the interquartile range for all the measures of spatial acuity measures used in the study; localization (RMS) error scores, overall VASI scores, ITD and ILD thresholds, and perceptual SSQ ratings across the three groups (group

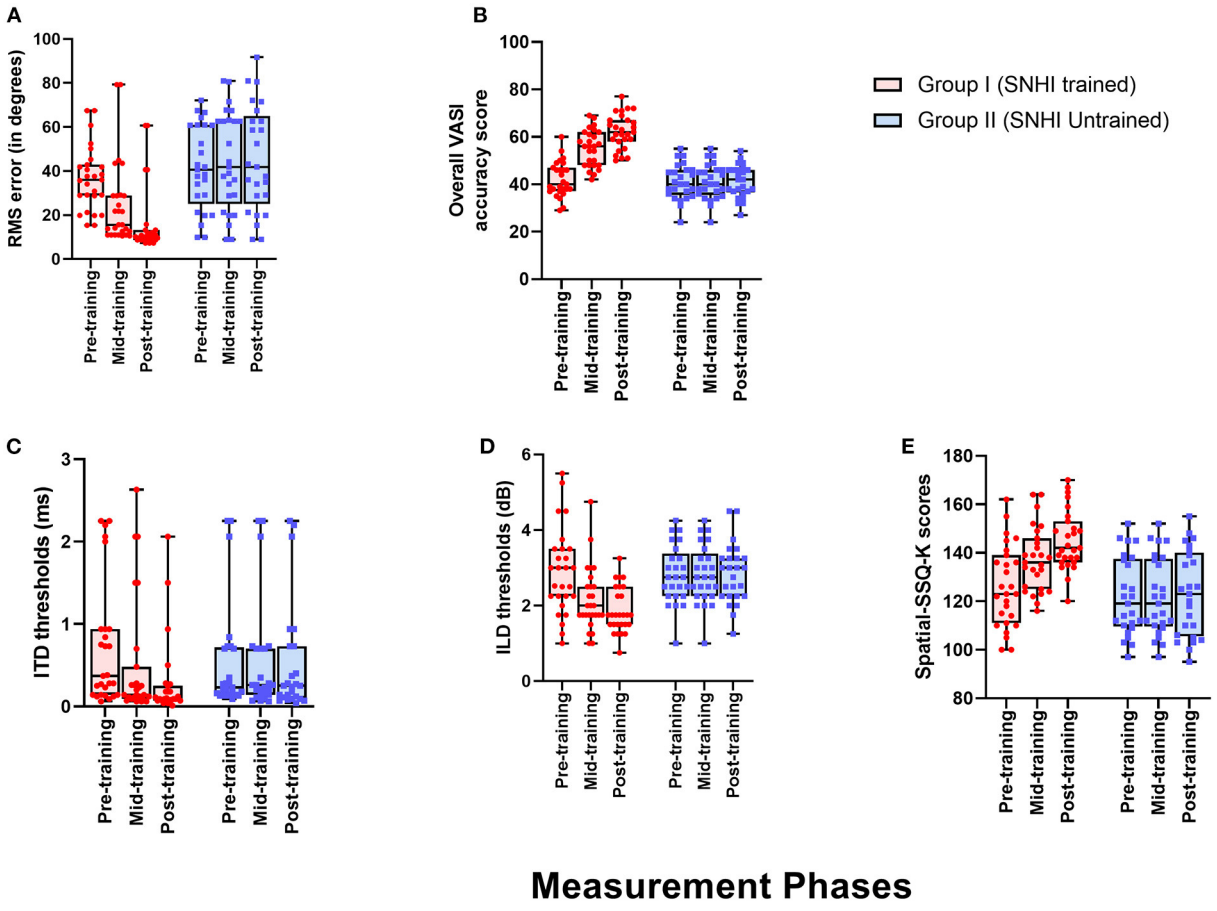



FIGURE 5
Comparison of spatial acuity of group I (SNHI-trained) and group II (SNHI-untrained) participants as a function of VAST in **(A)** test of spatial acuity in free field, **(B)** test of spatial acuity in closed-field, **(C)** interaural time difference thresholds, **(D)** interaural level difference thresholds, and **(E)** subjective ratings. The horizontal line in each box plot at center represents the median, while the "+" sign indicates the mean for each test. The box area corresponds to interquartile range, while the error bar indicates interquartile deviation. The Bonferroni comparisons between pre-, mid-, and post-evaluations were given for only group I participants, as similar comparisons for group II yielded no difference between evaluations.

I: SNHI-trained; group II: SNHI-untrained; and group III: NH listeners) obtained at pre-training evaluation revealed that SNHI participants (groups I and II) demonstrated spatial acuity deficits compared to NH, as reflected in Figure 4. Results of the MANOVA test showed the main effect of the group for the localization test [$F_{(2,79)} = 46.78$, $p < 0.001$, $\eta_p^2 = 0.54$], VASI test [$F_{(2,79)} = 10.51$, $p < 0.001$, $\eta_p^2 = 0.21$], ILD [$F_{(2,79)} = 9.93$, $p < 0.001$, $\eta_p^2 = 0.20$], and spatial sub-section of SSQ-K [$F_{(2,79)} = 49.11$, $p < 0.001$, $\eta_p^2 = 0.55$]. The *post hoc* comparisons using the Bonferroni test indicated that RMS errors in the free-field localization test of group I (36.63°± 14.37 SD) were similar ($p > 0.05$) to group II (40.96°± 17.25 SD), although both groups I and II registered significantly higher ($p < 0.001$) RMS errors than NH listeners (9.41°± 1.73). A similar trend was also observed in VASI, ILD, and spatial subsection of SSQ-K tests with group equivalency of both SNHI groups (groups I and II), who demonstrated significantly poorer ($p < 0.001$) spatial acuity scores (i.e., groups I and II had lower VASI and SSQ scores and higher ILD thresholds) compared to the NH group. The non-parametric Kruskal–Wallis test also revealed a significant main effect of the group for ITD [$H$ (2) = 27.25, $p \leq 0.001$, $\eta^2(H) = 0.27$]. Upon *post hoc* analyses using Dunn–Bonferroni pairwise comparisons, participants with SNHI (groups I and II) were shown to have significantly higher ($p < 0.001$) ITD thresholds compared to individuals with NH sensitivity indicative of binaural temporal cue processing deficits in them. Although SNHI listeners registered poorer spatial acuity than NH listeners on all spatial acuity tests, the performance between the former two groups was similar, indicating their group equivalency.

## Effect of VAST on spatial acuity measures

To evaluate the effect of VAST on the spatial acuity of SNHI listeners, the pre-, mid-, and post-training scores of group I along with two evaluations (at time-intervals equivalent of pre- and post-training evaluations) of group II were compared, with the median and mean along with the interquartile range is reflected in Figures 5A–E.

The spatial performance of group I (SNHI-trained) participants who underwent VAST improved as a function of training in all the measures, as reflected in the improvement of median/mean scores and reduction in variability with progression in spatial training, as shown in Figure 5. No noticeable change was seen in both median/mean in the group II (SNHI-untrained) participants, who did not undergo any formal training. The statistical significance of such differences explored using repeated measure ANOVA (3 evaluation phases) or Friedman test for each measure separately, along with their corresponding effect sizes in groups I and II, is shown in Table 1. The significant main effect of the evaluation phase was observed for all the spatial acuity measures in group I (SNHI-trained), while no main effect of the evaluation phase ($p > 0.05$) was seen for SNHI-untrained (Table 1), who showed no observable changes in their spatial acuity across evaluations.

The group I participants who underwent VAST showed improved spatial acuity not only on the overall VASI, but the spatial training benefits were also evident on the location-wise scores, as shown in Figure 6. The SNHI-trained group demonstrated improved VASI scores at the mid- and post-training evaluations, relative to the pre-training evaluation phase, while no changes in virtual location perception were seen in the SNHI-untrained group. Scores obtained by NH listeners are also depicted for comparison purposes. The improvement of virtual location identification seen in the SNHI-trained group was statistically significant on the Friedman test (along with Dunn–Bonferroni *post hoc*), while no main effect of the evaluation phase was seen in the SNHI-untrained group, as shown in Table 2. On closer visual inspection of Figure 6, VASI scores at each location in the SNHI-trained group not only improved across evaluation phases but also outperformed NH, in the post-training phase at all virtual locations.

## Identifying the optimal test for measuring VAST-related changes in spatial processing in SNHI

The discriminant functional analysis generated two DFs that effectively categorized VAST-related changes in spatial acuity, combinedly for all the tests considered in the study. While $DF_1$ was statistically the most robust function ($p < 0.001$) for the group segregation based on spatial processing abilities, DF2 was not significant. The extent of variability explained by $DF_1$ and $DF_2$ across measurement phases is shown in Table 3. The variability explained by $DF_2$ was relatively less, ranging between 1.50 (pre-training) and 29.5 (post-training).

Table 4 shows the relative contribution (weights) of each test in group membership (SNHI-trained, SNHI-untrained, and NH) on $DF_1$ across the three evaluation phases (pre-, mid-, and post-training). The coefficient with large absolute values corresponds to RMS error (free-field localization test measure for spatial acuity) for all the evaluation phases, indicative of the higher predictive power of this metric for group categorization.

Based on the weights (Table 4), the canonical $DF_1$ obtained in the study for each evaluation phase is summarized below:

Pre-training $DF_1$: (0.66 × RMS error) + (0.11 × ITD thresholds) + (0.19 × ILD thresholds)–(0.12 × overall VASI)–(0.57 × SSQ-K).

Mid-training $DF_1$: (0.56 × RMS error) + (0.05 × ITD thresholds) + (0.36 × ILD thresholds)–(0.06 × overall VASI)–(0.67 × SSQ-K).

Post-training $DF_1$: (0.54 × RMS error) + (0.07 × ITD thresholds) + (0.27 × ILD thresholds)–(0.43 × overall VASI)–(0.39 × SSQ-K).

The analyses of the DF1 function across evaluation phases (as summarized by weightages in equations earlier and Table 4) identified RMS error as the most sensitive metric that can capture spatial perception benefits derived through VAST in SNHI listeners. The combined group plot obtained using the results of FDA was plotted using $DF_1$ on abscissa and $DF_2$ on the ordinate, and a cluster of classification values of spatial performance across tests for different groups is shown in Figure 7. The cluster of classification values for all the groups was calculated by multiplying the standardized canonical DF coefficient by the test results of each individual on the five associated spatial measures and summing these products. Thus, calculated mean and individual scores for each group (group centroids) on the two DFs are shown in Figure 7. On visual inspection of the combined group plot, it can be seen that DF1 helped in the effective categorization of the differences in auditory spatial performance between the groups at all three phases of evaluation.

The combined group (Figure 7) plot also depicted the emergence of two distinct clusters of spatial performance on $DF_1$. In the pre-training evaluation phase (Figure 7A), while the symbols corresponding to SNHI groups (SNHI-trained: red circles; SNHI-untrained: blue squares) were concentrated on the right side of the $DF_1$, the symbols circles (black triangles) denoting the NH group emerged as a distinct cluster on the left side of the function. The marked disparity seen in the distribution of the two clusters (two SNHI groups relative to the NH group) at the pre-training phase (Figure 7A) became less apparent at the mid-training phase (Figure 7B), wherein a gradual shift in spatial acuity performance of SNHI-trained

TABLE 1 Results of Friedman test for main effect of evaluation phase and follow-up adjusted Bonferroni's pairwise comparisons for SNHI-trained and SNHI-untrained groups in each test of spatial processing.

| | Group I (SNHI trained) | | | | Group II (SNHI untrained) |
|---|---|---|---|---|---|
| Tests | Main effect of evaluation phase | Bonferroni comparisons | | | Main effect of evaluation phase |
| | | Pre-training | Mid-training | Post-training | |
| RMS Error | $\chi^2_{(2)} = 48.67, p < 0.001$, Kendall's W = 0.90 | $p = .01$ | $p < .001$ | $p < .001$ | $F_{(2,48)} = 1.90, p = 0.31$ |
| VASI | $F_{(2,52)} = 105.02, p < 0.001, \eta^2_p = 0.80$ | $p < .001$ | $p < .001$ | $p < .001$ | $F_{(2,48)} = 0.03, p = 0.97$ |
| ITD | $\chi^2_{(2)} = 38.34, p < 0.001$, Kendall's W = 0.70 | $p = .01$ | $p = .01$ | $p < .001$ | $\chi^2_{(2)} = 3.58, p = 0.17$ |
| ILD | $\chi^2_{(2)} = 38.00, p < 0.001$, Kendall's W = 0.70 | $p = .01$ | $p = .02$ | $p < .001$ | $F_{(2,48)} = 0.01, p = 0.99$ |
| SSQ | $F_{(2,52)} = 89.31, p < 0.001, \eta^2_p = 0.78$ | $p < .001$ | $p < .001$ | $p < .001$ | $F_{(2,48)} = 0.61, p = 0.55$ |

(red circles) due to VAST is materialized (seen as a shift in predicted scores toward NH). At the mid-evaluation phase (Figure 7B), the distribution of clusters involving red circles (SNHI-trained) moved slightly toward the right (relative to the concentration of red symbols in the pre-training condition) on the $DF_1$. This cluster movement toward the NH group (black triangles) indicated the initial realization of benefits derived from VAST on behavioral spatial acuity measures. It is also important to note that the SNHI-untrained group did not show any visible movement in the mid-training phase, as opposed to the SNHI-trained indicative of no spatial performance changes in them. The separation of the clusters (SNHI-trained and NH) became nearly extinct separation on $DF_1$ at the post-training phase (Figure 7C). At the post-training evaluation phase (Figure 7C), the cluster distribution of SNHI-trained (red circles) advanced further to the right of $DF_1$, causing the superimposition of red and black symbols. This camouflaging of the cluster distributions on the $DF_1$ in the post-training evaluation phase signals the materialization of positive outcomes of VAST in group I (SNHI-trained) participants.

Furthermore, the errors in the prediction of group membership based on classification results of DF analysis also revealed that differentiating the groups based on $DF_1$ caused no confusion in classifying NH into the same group (predicted membership of NH was similar to original NH) at pre-training condition, as shown in Table 5. This delineation was not readily apparent between the SNHI-trained and SNHI-untrained groups, with a consistent overlap occurring between them. Only 55.66% of SNHI-trained and 60.00% of SNHI-untrained were correctly classified, accounting for the error of 44.34 and 40%, respectively. This finding showed that at pre-training conditions, the spatial acuity of the trained and untrained groups was similar, making the group membership prediction difficult for the algorithm. At the mid-training phase, the categorization error of the SNHI-trained decreased to 29.63%, while that of SNHI-untrained dropped to 20.0%, indicating the change in auditory spatial performance of the SNHI-trained group consequent to the VAST paradigm, which in turn successfully segregated the spatial performance of the former group from the latter. Complementary to the same, the accuracy of group prediction also increased to 70.37% and 80.0% for the SNHI-trained and SNHI-untrained groups, respectively. This trend in decreased group classification errors and increased prediction accuracy was persistent in the post-training phase as well, with chances of error in grouping reduced to 22.20 and 16% for SNHI-trained and SNHI-untrained groups. However, the errors of classifying SNHI-trained as NH increased in the post-training evaluation, with 6 NH being wrongly classified as SNHI-trained (as opposed to only 2 NH misclassified as SNHI-trained in mid-evaluation and none in pre-training), indicative of spatial performance in few participants in the SNHI-trained group nearing spatial abilities of NH participants.

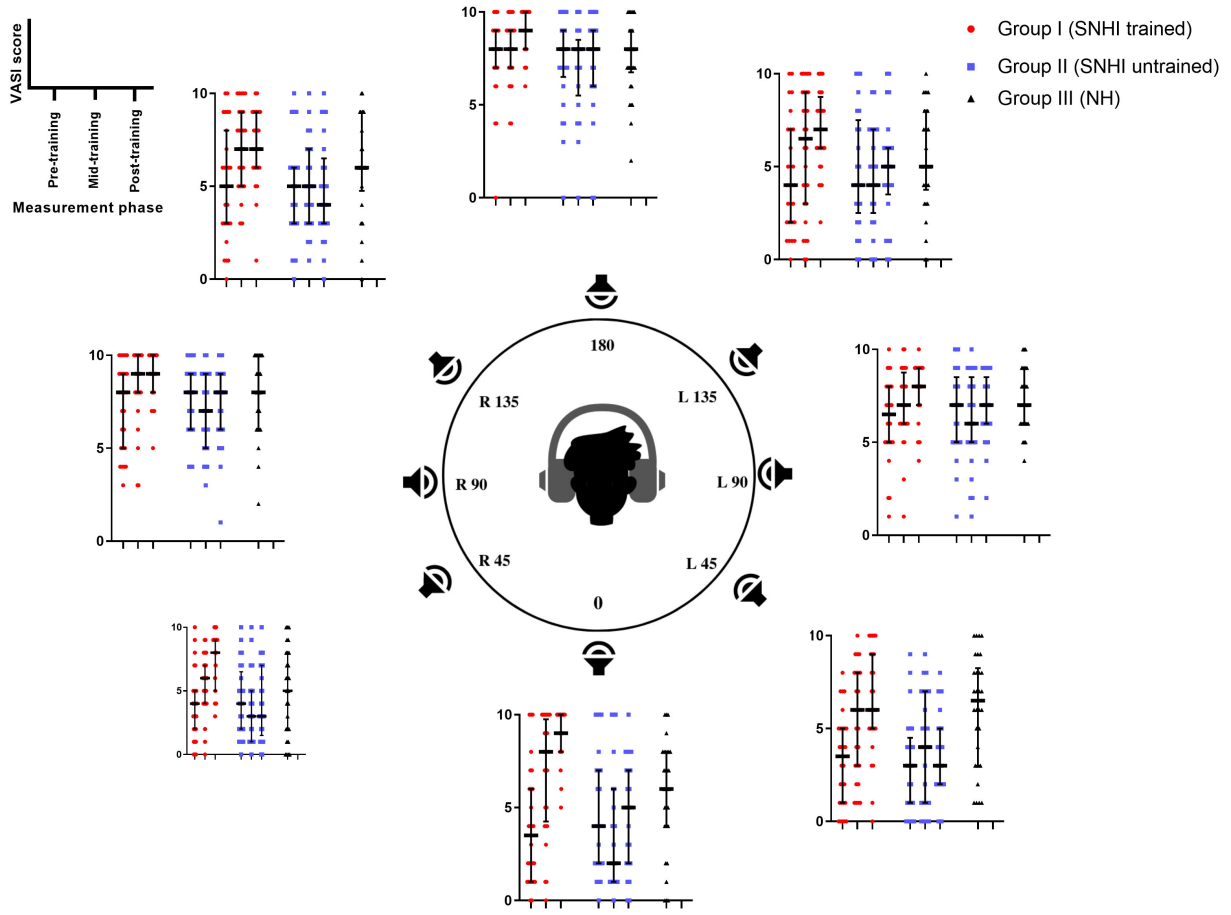



FIGURE 6
Comparison of VASI scores of group I (SNHI-trained) and group II (SNHI-untrained) participants as a function of VAST across in virtual location. The inner dummy head panel represents eight VAS locations used in the study, while the outer panels denote VASI scores of each participant corresponding to VAS location mentioned in the inner panel at three evaluation phases (pre-, mid-, and post-training). The horizontal line in each box plot at center represents the median, while the "+" sign indicates the mean for each test. The box area corresponds to interquartile range, while the error bar indicates interquartile deviation. The VASI of group III (NH listeners) is given only for comparison purpose. Maximum score per location is 10.

## Discussion

In an intervention-based research design, the present study investigated the application of VAST in resolving auditory spatial deficits in SNHI listeners, apart from comparing the same to NH listeners. The findings of MANOVA and Kruskal–Wallis *H*-test indicated a significant main effect of group, with the consequent *post hoc* tests revealing higher spatial resolution skills in NH listeners, in terms of both precision (lower localization errors and lower ITD and ILD thresholds) and accuracy (higher VASI and spatial sub-section of SSQ-K scores) compared to both SNHI groups (SNHI-trained and SNHI-untrained). In contrast, the participants with SNHI who were either in VAST-trained or VAST-untrained groups demonstrated similarity in their spatial performance (equally poorer skills), suggestive of group equivalency prior to spatial training. The spatial acuity deficits seen in SNHI listeners (SNHI-trained and SNHI-untrained) in the present study are in consensus with literature accounts on different spatial measures such as localization (Häusler et al., 1983; van den Bogaert et al., 2006), lateralization (Kubo et al., 1998; Spierer et al., 2007), and binaural cue processing (Koehnke et al., 1995; Smith-Olinde et al., 1998; Spencer et al., 2016). Deficits in a number of peripheral processes, such as impaired frequency selectivity (Turner et al., 1999; Bernstein and Oxenham, 2006; Hopkins and Moore, 2011), temporal resolution (Arehart, 1998; Bianchi et al., 2016), and altered auditory filter shapes (Glasberg and Moore, 1986; Dubno and Dirks, 1989) in individuals with SNHI can be conceived as factors that account for these group differences.

The results of the within-subject analysis in the SNHI-trained group showed significant improvements in all spatial tests at mid-evaluation and further refinement of the same

TABLE 2 Results of Friedman test for main effect of evaluation phase and follow-up adjusted Bonferroni's pairwise comparisons for SNHI-trained and SNHI-untrained groups at each virtual location in VASI test.

| | Group I (SNHI trained) | | | | Group II (SNHI untrained) |
|---|---|---|---|---|---|
| Virtual location | Main effect of evaluation phase | Bonferroni comparisons | | | Main effect of evaluation phase |
| | | Pre-training | Mid-training | Post-training | |
| R45 | $\chi^2_{(2)} = 24.74, p < 0.001$, Kendall's W = 0.46 | $p = .12$ | $p = .02$ | $p < .001$ | $\chi^2_{(2)} = 0.27, p = 0.87$ |
| R90 | $\chi^2_{(2)} = 6.71, p = 0.04$, Kendall's W = 0.12 | $p = .52$ | $p = 1.00$ | $p = .05$ | $\chi^2_{(2)} = 1.16, p = 0.45$ |
| R135 | $\chi^2_{(2)} = 6.15, p = 0.05$, Kendall's W = 0.11 | $p = .26$ | $p = 1.00$ | $p = .05$ | $\chi^2_{(2)} = 0.02, p = 0.98$ |
| 180 | $\chi^2_{(2)} = 9.78, p = 0.01$, Kendall's W = 0.18 | $p = 1.00$ | $p = .06$ | $p = .03$ | $\chi^2_{(2)} = 0.53, p = 0.77$ |
| L135 | $\chi^2_{(2)} = 17.43, p < 0.001$, Kendall's W = 0.32 | $p = .03$ | $p = .74$ | $p = .001$ | $\chi^2_{(2)} = 0.51, p = 0.77$ |
| L90 | $\chi^2_{(2)} = 6.81, p = 0.04$, Kendall's W = 0.31 | $p = .54$ | $p = 1.00$ | $p = .05$ | $\chi^2_{(2)} = 4.08, p = 0.13$ |
| L45 | $\chi^2_{(2)} = 9.78, p = 0.01$, Kendall's W = 0.18 | $p = .02$ | $p = .83$ | $p < .001$ | $\chi^2_{(2)} = 2.02, p = 0.36$ |
| 0 | $\chi^2_{(2)} = 30.02, p < 0.001$, Kendall's W = 0.18 | $p = .15$ | $p = .01$ | $p < .001$ | $\chi^2_{(2)} = 0.53, p = 0.77$ |

The italic text correspond to significance value 'p' or test statistic, as used conventionally in reporting.

TABLE 3 Eigen values, Wilk's lambda (λ), and percentage of variance for the standardized discriminant functions ($DF_1$ and $DF_2$) in pre-, mid-, and post-training evaluation phases.

| Evaluation phase | Discriminant function | Eigen value | % of variance | Wilk's lambda (λ) | Chi-square test |
|---|---|---|---|---|---|
| Pre-training | 1 | 2.45 | 98.50 | 0.28 | $\chi^2_{(10)} = 98.36, p < 0.001$ |
| | 2 | 0.04 | 1.50 | 0.96 | $\chi^2_{(4)} = 2.91, p = 0.57$ |
| Mid-training | 1 | 2.12 | 79.00 | 0.21 | $\chi^2_{(10)} = 121.87, p < 0.001$ |
| | 2 | 0.56 | 21.00 | 0.64 | $\chi^2_{(4)} = 34.32, p = 0.06$ |
| Post-training | 1 | 2.37 | 70.50 | 0.15 | $\chi^2_{(10)} = 146.67, p < 0.001$ |
| | 2 | 0.99 | 29.50 | 0.50 | $\chi^2_{(4)} = 53.10, p = 0.14$ |

The italic text correspond to significance value 'p' or test statistic, as used conventionally in reporting.

at post-training evaluation (Table 2), indicative of positive outcomes of VAST. The present study demonstrated that the effect of the VAST paradigm was not restricted to just the trained stimuli but was generalizable to other tasks such as localization of real sources, thresholds of ILD and ITD, and perceptual ratings (Figures 5A–E). While the improvements in VASI score reflect stimulus-specific learning, i.e., learning the task for which one is trained, improvements on other spatial tasks signal perceptual learning (Wright and Fitzgerald, 2001). Thus, the paradigm used in the study seems to validate the process of supervised learning in which related perceptual networks calibrate each other in a goal-directed way (Knudsen, 1984), as recorded in several reports on perceptual learning (Zahorik et al., 2006; Wright et al., 2010). The generalization effect reported in the current study is also supported by the observation of Ortiz and Wright (2009) who found that

ITD/ILD training effects were generalized to temporal acuity (GAP detection) skills (not implicitly trained) apart from improving ITD/ILD thresholds. The success derived from the VAST is further strengthened by findings in group II participants, who continued to demonstrate spatial deficits at the mid- and post-training evaluation. This benefit derived from VAST can be explained by the modified auditory adaptation–feedback model of spatial processing proposed by Mendonça (2014).

Auditory adaptation–feedback model in its original form was constructed to explain auditory spatial adaptability to altered signals. Drawing parallels to the current spatial training paradigm, i.e., VAST, altered signals in the original model are equated to distorted (due to consequence of SNHI) direction-dependent inputs (binaural and spectral cues). These cues, which lack precise binaural/spectral information, are further combined in the peripheral auditory system to estimate the virtual source position. Owing to degraded inputs at the peripheral auditory system in SNHI listeners, an incorrect space percept of the virtual source is formed. This incorrect virtual auditory space percept is then followed by feedback.

TABLE 4 Contribution (weights) of auditory spatial measures for group membership on discriminant function 1 ($DF_1$).

| Predictor variable/tests | Pre-training | Mid-training | Post-training |
|---|---|---|---|
| RMS error (free-field test) | 0.66 | 0.56 | 0.54 |
| Overall VASI scores | −0.12 | −0.06 | −0.43 |
| ITD thresholds | 0.11 | 0.05 | 0.07 |
| ILD thresholds | 0.19 | 0.36 | 0.27 |
| Spatial-subsection of SSQ-K scores | −0.57 | −0.67 | −0.39 |

As approached in VAST, the feedback given was the corrective response feedback of virtual sound location. The correct responses are acknowledged, while the incorrect responses are compared and contrasted with the correct item. Tyler et al. (2010) also reported success when participants were allowed to compare loudspeaker locations. In VAST, the feedback is compared to the original virtual auditory space percept. If no differences are found (i.e., in the case of correct response), the original virtual sound percept is strengthened. If the feedback is substantially different from the percept, then a new cue combination rule (set of ITD, ILD, and spectral cues) is created. On multiple repetitions of the feedback, the new cue combination gets further strengthened, and a new spatial percept is created. A forward–backward loop between the perceptual mechanisms involved in original virtual space perception and feedback is created. We postulate that spatial learning occurs precisely from this loop. The application of feedback in perceptual learning is advocated in learning theory, which supports the idea that best learning occurs when listeners are motivated (Amitay et al., 2010). When the listeners sense that the task is impossible to succeed, or when they feel the limited challenge in the task, their motivation for learning diminishes. However, changing the stimulus difficulty (length and number of virtual locations) adaptively in VAST ensured that the motivation level of listeners during training was maintained high.

The benefits derived from the spatial learning loop can be influenced by other non-acoustic perceptual factors (Andéol et al., 2015), including its ability to maintain the motivation level of the listener (Amitay et al., 2010). In accordance with the same, the VAST paradigm started with an easier stimulus (longer duration), and difficulty was adaptively adjusted based on the listeners' performance level. Furthermore, as mentioned earlier, visual and auditory

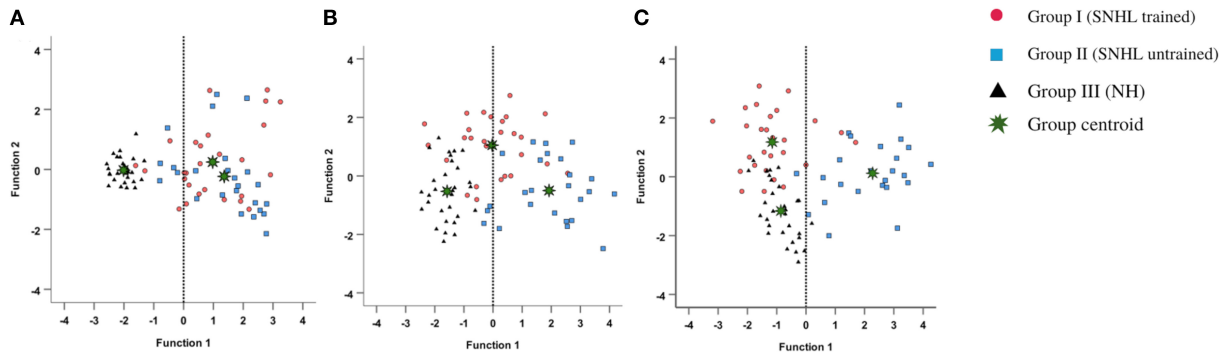



FIGURE 7
Grouping participants based on canonical discriminant scores derived for behavioral measures of spatial acuity at **(A)** pre-, **(B)** mid-, and **(C)** post-training evaluation phases. The red circles, blue squares, and black triangles correspond to discriminant scores of group I (SNHI-trained), group II (SNHI-untrained), and group III (NH) participants, respectively. The categorization of the groups is evident on the significant function, i.e., DF1. Function 2 is merely used for plotting the abscissa values, and group classification is not elaborately discussed as this function was not significant.

TABLE 5 Classification results for groups in pre-, mid-, and post-training evaluation phases.

| Evaluation phase | Groups | Original count | | | Percentage classified (%) | | |
|---|---|---|---|---|---|---|---|
| | | Group I (SNHI-trained) | Group II (SNHI-untrained) | Group III (NH) | Group I (SNHI-trained) | Group II (SNHI-untrained) | Group III (NH) |
| Pre-training (73.22% of original grouped cases correctly classified) | Group I (SNHI-trained) | 15 | 8 | 0 | 55.66 | 32.00 | 0.00 |
| | Group II (SNHI-untrained) | 10 | 15 | 0 | 37.00 | 60.00 | 0.00 |
| | Group III (NH) | 2 | 2 | 30 | 7.44 | 8.00 | 100.00 |
| | Total | 27 | 25 | 30 | 100.00 | 100.00 | 100.00 |
| Mid-training (81.77% of original grouped cases correctly classified) | Group I (SNHI-trained) | 19 | 2 | 2 | 70.37 | 8.00 | 6.67 |
| | Group II (SNHI-untrained) | 2 | 20 | 0 | 7.41 | 80.00 | 0.00 |
| | Group III (NH) | 6 | 3 | 28 | 22.22 | 12.00 | 93.33 |
| | Total | 27 | 25 | 30 | 100.0 | 100.0 | 100.00 |
| Post-training (80.55% of original grouped cases correctly classified) | Group I (SNHI-trained) | 21 | 1 | 6 | 77.78 | 4.0 | 20.00 |
| | Group II (SNHI-untrained) | 2 | 21 | 0 | 7.41 | 84.0 | 0.00 |
| | Group III (NH) | 4 | 3 | 24 | 14.81 | 12.0 | 80.00 |
| | Total | 27 | 25 | 30 | 100.0 | 100.0 | 100.00 |



10.3389/fnins.2022.1080398

feedback was provided during the training (listen and compare). This was done to encourage self-correction and help participants know which directions required more attention. This feedback along with a structured hierarchy of stimulus presentation forms the core strength of VAST and serves as a primary basis for the realization of benefits derived from it.

The spatial abilities of listeners with SNHI who underwent VAST (group I) were also compared to NH listeners (group III) to measure the extent of spatial learning in them using FDA. Discriminant analyses showed that the RMS error measure of the free-field localization test had the highest predictive power for group categorization in all the evaluation phases (Table 4), suggestive of its robust sensitivity to VAST-related benefits. Specifically, the benefit derived due to VAST in reducing localization error in the current study (36.02° in pre-training to 10.11° in post-training evaluation, Figure 5A) is relatively higher than the improvements seen in other psychoacoustic (overall VASI, Figure 5B, and binaural processing scores, Figures 5C, D) and subjective (spatial-sub-section of SSQ, Figure 5E) measures used in the study. The improvement seen in localization abilities secondary to VAST reported in the present study is similar to the earlier reports in the literature by Tyler et al. (2010), who demonstrated that the pre-training localization score consisted of an average RMS error of 24°, while the post-training score was 17° RMS error. Further combined group plot (Figure 7) of FDA showed distinct clusters of SNHI listeners (SNHI-trained and SNHI-untrained) from NH listeners in the pre-training evaluation phase, although there was considerable overlapping of the former two groups (SNHI-trained and SNHI-untrained) with a relatively higher prediction error between these groups (Table 5). In the mid-training followed by post-training, the error in classification decreased between the SNHI groups (SNHI-trained and SNHI-untrained), showing the improved scores as a function of VAST in the trained group, making it easier for group membership prediction. Complimentary to the earlier, there was a rightward movement of the trained-SNHI centroid toward the NH group (Figure 7), which further could be attributed to the improved spatial performance in the trained-SNHI group, wherein spatial performance of a few individuals of this target group overlapped with the spatial skills of NH listeners leading to misjudgment of SNHI-trained as NH (which was not otherwise visible in pre-training evaluation phase).

## Conclusion

The findings from the current investigation highlighted the efficacy of VAST as an intervention tool for remediating spatial deficits in SNHI listeners. The success derived from VAST has promising implications for rehabilitative audiologists, as the tool has good clinical applicability. Although the VAST paradigm was done under laboratory conditions, it can also be adopted for spatial training at home as it requires only minimal equipment (laptop, paradigm player software, and headphones with good frequency response) at the user's end. The feasibility and applicability of the VAST paradigm with equipment already available at home make this protocol even more practical for implementation. The spatial training paradigm can also be extended to other clinical populations with spatial difficulties, such as individuals with central auditory processing disorder (CAPD), spatial processing disorders, and auditory neuropathy spectrum disorder (ANSD) after gathering research evidence. Future studies in this field should focus on the endurance of the learned capabilities over time, generalization limits, and the role of other cognitive factors in assessing the effects of VAST.

## Data availability statement

The data analyzed in this study will be made available by authors on request, for educational purposes. Requests should be directed to the corresponding author.

## Ethics statement

The studies involving human participants were reviewed and approved by AIISH Ethical Committee for biobehavioral research. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

KN: conceptualization of the study, formulation of the methods, data collection, data pruning and statistical analysis, and the original draft preparation. AU: conceptualization of the study, formulation of the method, supervision, and the original draft preparation. RK: writing and visual depiction of data. All authors contributed to the article and approved the submitted version.

## Funding

The study was a part of All India Institute of Speech and Hearing (AIISH) research funds (Project No. SH/CDN/ARF-AUD-4/2018-19), which is a research initiative at the institutional level.

## Acknowledgments

The authors thank the Director and Head of the Department, Audiology (All India Institute of Speech and Hearing, Mysuru, affiliated with the University of Mysuru) for permitting us to carry out the study. The authors would like to thank all the participants of the study for their consent and cooperation during the data collection.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Abel, S. M., Giguáre, C., Consoli, A., and Papsin, B. C. (2000). The effect of aging on horizontal plane sound localization. *J. Acoust. Soc. Am.* 108, 743–752. doi: 10.1121/1.429607

Akeroyd, M. A. (2014). An overview of the major phenomena of the localization of sound sources by normal-hearing, hearing-impaired, and aided listeners. *Trends Hear.* 18, 1–7. doi: 10.1177/2331216514560442

Akeroyd, M. A., and Whitmer, W. M. (2016). "Spatial hearing and hearing aids," in *Hearing Aids, Springer Handbook of Auditory Research*, ed G. Popelka (Cham: Springer International Publishing), 253–289. doi: 10.1007/978-3-319-33036-5_7

Amitay, S., Halliday, L., Taylor, J., Sohoglu, E., and Moore, D. R. (2010). Motivation and intelligence drive auditory perceptual learning. *PLoS ONE* 5, e9816. doi: 10.1371/journal.pone.0009816

Andéol, G., Savel, S., and Guillaume, A. (2015). Perceptual factors contribute more than acoustical factors to sound localization abilities with virtual sources. *Front. Neurosci.* 8, 451. doi: 10.3389/fnins.2014.00451

Arehart, K. H. (1998). Effects of harmonic content on complex-tone fundamental-frequency discrimination in hearing-impaired listeners. *J. Acoust. Soc. Am.* 95, 3574. doi: 10.1121/1.409975

Baker, R. J., and Rosen, S. (2002). Auditory filter nonlinearity in mild/moderate hearing impairment. *J. Acoust. Soc. Am.* 111, 1330–1339. doi: 10.1121/1.1448516

Bernstein, J. G. W., and Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination : sensorineural hearing loss. *J. Acoust. Soc. Am.* 120, 3929–3945. doi: 10.1121/1.2372452

Best, V., Kalluri, S., McLachlan, S., Valentine, S., Edwards, B., and Carlile, S. (2010). A comparison of CIC and BTE hearing aids for three-dimensional localization of speech. *Int. J. Audiol.* 49, 723–732. doi: 10.3109/14992027.2010.484827

Bianchi, F., Fereczkowski, M., Zaar, J., Santurette, S., and Dau, T. (2016). Complex-Tone pitch discrimination in listeners with sensorineural hearing loss. *Trends Hear.* 20, 233121651665579. doi: 10.1177/2331216516655793

Brimijoin, W. O., and Akeroyd, M. A. (2014). The moving minimum audible angle is smaller during self motion than during source motion. *Front. Neurosci.* 8, 273. doi: 10.3389/fnins.2014.00273

Brimijoin, W. O., and Akeroyd, M. A. (2016). The effects of hearing impairment, age, and hearing aids on the use of self motion for determining front/back location. *J. Am. Acad. Audiol.* 27, 588–600. doi: 10.3766/jaaa.15101

Brungart, D. S., Cohen, J., Cord, M., Zion, D., and Kalluri, S. (2014). Assessment of auditory spatial awareness in complex listening environments. *J. Acoust. Soc. Am.* 136, 1808–1820. doi: 10.1121/1.4893932

Brungart, D. S., Cohen, J. I., Zion, D., and Griffin, R. (2017). The localization of non-individualized virtual sounds by hearing impaired listeners. *J. Acoust. Soc. Am.* 141, 2870–2881. doi: 10.1121/1.4979462

Byrne, D., and Noble, W. (1998). Optimizing sound localization with hearing aids. *Trends Amplif.* 3, 51–73. doi: 10.1177/108471389800300202

Carhart, R., and Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *J. Speech Hear. Disord.* 24, 330–345. doi: 10.1044/jshd.2404.330

Carlile, S., Fox, A., Orchard-Mills, E., Leung, J., and Alais, D. (2016). Six degrees of auditory spatial separation. *J. Assoc. Res. Otolaryngol.* 17, 209–221. doi: 10.1007/s10162-016-0560-1

Chung, K., Neuman, A. C., and Higgins, M. (2008). Effects of in-the-ear microphone directionality on sound direction identification. *J. Acoust. Soc. Am.* 123, 2264–2275. doi: 10.1121/1.2883744

Denk, F., Ewert, S. D., and Kollmeier, B. (2019). On the limitations of sound localization with hearing devices. *J. Acoust. Soc. Am.* 146, 1732–1744. doi: 10.1121/1.5126521

Drennan, W. R., Gatehouse, S., Howell, P., Van Tasell, D., and Lund, S. (2005). Localization and speech-identification ability of hearing-impaired listeners using phase-preserving amplification. *Ear Hear.* 26, 461–472. doi: 10.1097/01.aud.0000179690.30137.21

Dubno, J. R., and Dirks, D. D. (1989). Auditory filter characteristics and consonant recognition for hearing-impaired listeners. *J. Acoust. Soc. Am.* 85, 1666–1675. doi: 10.1121/1.397955

Faul, F., Erdfelder, E., Lang, A., and Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 175–191. doi: 10.3758/BF03193146

Folstein, M. F., Robins, L. N., and Helzer, J. E. (1983). The mini-mental state examination. *Arch. Gen. Psychiatry* 40, 812. doi: 10.1001/archpsyc.1983.01790060110016

Gabriel, K. J., Koehnke, J., Colburn, H. S., and Gabriel, J. (1992). Frequency dependence of binaural performance in listeners with impaired binaural hearing. *J. Acoust. Soc. Am.* 91, 336–347. doi: 10.1121/1.402776

Gatehouse, S., and Noble, W. (2004). The speech, spatial and qualities of hearing scale (SSQ). *Int. J. Audiol.* 43, 85–99. doi: 10.1080/14992020400050014

Glasberg, B. R., and Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *J. Acoust. Soc. Am.* 79, 1986–1033. doi: 10.1121/1.393374

Glyde, H., Cameron, S., Dillon, H., Hickson, L., and Seeto, M. (2013). The effects of hearing impairment and aging on spatial processing. *Ear Hear.* 34, 15–28. doi: 10.1097/AUD.0b013e3182617f94

Gnanateja, N. (2014). *Consonant Confusion Matrix*. Available online at: https://in.mathworks.com/matlabcentral/fileexchange/46461-consonant-confusion-matrix (accessed December 7, 2022).

Hartmann, W. M., and Wittenberg, A. (1996). On the externalization of sound images. *J. Acoust. Soc. Am.* 99, 3678–3688. doi: 10.1121/1.414965

Häusler, R., Colburn, S., and Marr, E. (1983). Sound localization in subjects with impaired hearing: spatial-discrimination and interaural-discrimination tests. *Acta Otolaryngol.* 96, 1–62. doi: 10.3109/00016488309105590

Hawkins, D. B., and Wightman, F. L. (1980). Interaural time discrimination ability of listeners with sensorineural hearing loss. *Audiology* 19, 495–507. doi: 10.3109/00206098009070081

Hopkins, K., and Moore, B. C. J. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *J. Acoust. Soc. Am.* 130, 334–349. doi: 10.1121/1.3585848

Irving, S., and Moore, D. R. (2011). Training sound localization in normal hearing listeners with and without a unilateral ear plug. *Hear. Res.* 280, 100–108. doi: 10.1016/j.heares.2011.04.020

Johnson, J. A., Xu, J., and Cox, R. M. (2017). Impact of hearing aid technology on outcomes in daily life III : localization. *Ear Hear.* 38, 746–759. doi: 10.1097/AUD.0000000000000473

Katz, J. (2015). *Handbook of Clinical Audiology*. 7th Edn. Philadelphia, PA: Wolters Kluwer

Keidser, G., Convery, E., and Hamacher, V. (2011). The effect of gain mismatch on horizontal localization performance: *Hear. J.* 64, 26. doi: 10.1097/01.HJ.0000394541.95207.c7

Keidser, G., O'Brien, A., Hain, J.-U., McLelland, M., and Yeend, I. (2009). The effect of frequency-dependent microphone directionality on horizontal localization performance in hearing-aid users. *Int. J. Audiol.* 48, 789–803. doi: 10.3109/14992020903036357

Kidd, G., Arbogast, T. L., Mason, C. R., and Gallun, F. J. (2005). The advantage of knowing where to listen. *J. Acoust. Soc. Am.* 118, 3804–3815. doi: 10.1121/1.2109187

King, A. J., Schnupp, J. W. H., and Doubell, T. P. (2001). The shape of ears to come: dynamic coding of auditory space. *Trends Cogn. Sci.* 5, 261–270. doi: 10.1016/S1364-6613(00)01660-0

Kinkel, M., Kollmeier, B., and Holube, I. (1991). Binaurales hören bei normalhörenden und schwerhörigen I : meßmethode und meßergebnisse (binaural hearing in normal-hearing andimpaired-hearing listeners. I. Measurement methods andresults). *Audiol. Akust.* 30, 192–201.

Knudsen, E. I. (1984). The role of auditory experience in the development and maintenance of sound localization. *Trends Neurosci.* 7, 326–330. doi: 10.1016/S0166-2236(84)80081-8

Koehnke, J., Culotta, C. P., Hawley, M. L., and Colburn, H. S. (1995). Effects of reference Interaural time and intensity differences on binaural performance in listeners with normal and impaired hearing. *Ear Hear.* 16, 331–353. doi: 10.1097/00003446-199508000-00001

Kubo, T., Sakashita, T., Kusuki, M., Kyunai, K., Ueno, K., Hikawa, C., et al. (1998). Sound lateralization and speech discrimination in patients with sensorineural hearing loss. *Acta Otolaryngol.* 538, 63–69. doi: 10.1080/00016489850182765

Kuk, F., Keenan, D. M., Lau, C., Crose, B., and Schumacher, J. (2014). Evaluation of a localization training program for hearing impaired listeners. *Ear Hear.* 35, 652–666. doi: 10.1097/AUD.0000000000000067

Kuk, F., Korhonen, P., Lau, C., Keenan, D., and Norgaard, M. (2013). Evaluation of a pinna compensation algorithm for sound localization and speech perception in noise. *Am. J. Audiol.* 22, 84–93. doi: 10.1044/1059-0889(2012/12-0043)

Lacher-Fougère, S., and Demany, L. (2005). Consequences of cochlear damage for the detection of interaural phase differences. *J. Acoust. Soc. Am.* 118, 2519–2526. doi: 10.1121/1.2032747

Le Goff, N., Buchholz, J. M., and Dau, T. (2013). "Modeling horizontal localization of complex sounds in the impaired and aided impaired auditory system," in *The Technology of Binaural Listening*, eds J. Blauert, and J. Blauert (NewYork, NY: Springer), 121–144. doi: 10.1007/978-3-642-37762-4_5

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *J. Acoust. Soc. Am.* 49, 467–477. doi: 10.1121/1.1912375

Lorenzi, C., Gatehouse, S., and Lever, C. (1999). Sound localization in noise in hearing-impaired listeners. *J. Acoust. Soc. Am.* 105, 3454–3463. doi: 10.1121/1.424672

Mendonça, C. (2014). A review on auditory space adaptations to altered head-related cues. *Front. Neurosci.* 8, 219. doi: 10.3389/fnins.2014.00219

Milman, L. H., Faroqi-Shah, Y., Corcoran, C. D., and Damele, D. M. (2018). Interpreting mini-mental state examination performance in highly proficient bilingual Spanish–English and Asian Indian–English speakers: demographic adjustments, item analyses, and supplemental measures. *J. Speech, Lang. Hear. Res.* 61, 847–856. doi: 10.1044/2017_JSLHR-L-17-0021

Neher, T., Behrens, T., Carlile, S., Jin, C., Kragelund, L., Petersen, A. S., et al. (2009). Benefit from spatial separation of multiple talkers in bilateral hearing-aid users: effects of hearing loss, age, and cognition. *Int. J. Audiol.* 48, 758–774. doi: 10.3109/14992020903079332

Neher, T., Laugesen, S., Søgaard Jensen, N., and Kragelund, L. (2011). Can basic auditory and cognitive measures predict hearing-impaired listeners' localization and spatial speech recognition abilities? *J. Acoust. Soc. Am.* 130, 1542–1558. doi: 10.1121/1.3608122

Nisha, K. V., Durai, R., and Konadath, S. (2022). Musical training and its association with age-related changes in binaural, temporal, and spatial processing. *Am. J. Audiol.* 31, 669–683. doi: 10.1044/2022_AJA-21-00227

Nisha, K. V., and Kumar, A. U. (2017). Virtual auditory space training-induced changes of auditory spatial processing in listeners with normal hearing. *J. Int. Adv. Otol.* 13, 118–127. doi: 10.5152/iao.2017.3477

Nisha, K. V., and Kumar, A. U. (2018). *Effects of Training Regime on Behavioural and Electrophysiological Correlates of Auditory Spatil Processing in Individuals With Sensorineural Hearing Impairement*. Thesis, published at University of Mysore, India. Available online at: http://192.168.100.26:8080/xmlui/handle/123456789/70 (accessed March 01, 2022).

Nisha, K. V., and Kumar, A. U. (2019a). Pre-attentive neural signatures of auditory spatial processing in listeners with normal hearing and sensorineural hearing impairment: a comparative study. *Am. J. Audiol.* 28, 437–449. doi: 10.1044/2018_AJA-IND50-18-0099

Nisha, K. V., and Kumar, A. U. (2019b). Cortical reorganization following auditory spatial training in listeners with sensorineural hearing impairment: a high-density electroencephalography study. *J. Acoust. Soc. Am.* 145, 1906. doi: 10.1121/1.5101915

Nisha, K. V., and Kumar, A. U. (2022). Effects of spatial training paradigms on auditory spatial refinement in normal-hearing listeners: a comparative study. *J. Audiol. Otol.* 26, 113–121. doi: 10.7874/jao.2021.00451

Nisha, K. V., Kumar, A. U., and Konadath, S. (2023). Effects of maturation and chronological aging on auditory spatial processing: a cross-sectional study across lifespan. *Am. J. Audiol.* doi: 10.1044/2022_AJA-22-00113

Noble, W., and Byrne, D. (1990). A comparison of different binaural hearing aid systems for sound localization in the horizontal and vertical planes. *Br. J. Audiol.* 24, 335–346. doi: 10.3109/03005369009076574

Noble, W., Byrne, D., and Ter-Horst, K. (1997). Auditory localization, detection of spatial separateness, and speech hearing in noise by hearing impaired listeners. *J. Acoust. Soc. Am.* 102, 2343–2352. doi: 10.1121/1.419618

Noble, W., and Gatehouse, S. (2006). Effects of bilateral versus unilateral hearing aid fitting on abilities measured by the speech, spatial, and qualities of hearing scale (SSQ). *Int. J. Audiol.* 45, 172–181. doi: 10.1080/14992020500376933

Noble, W., Ter-Horst, K., and Byrne, D. (1995). Disabilities and handicaps associated with impaired auditory localization. *J. Am. Acad. Audiol.* 6, 129–140.

Ortiz, J. A., and Wright, B. A. (2009). Contributions of procedure and stimulus learning to early, rapid perceptual improvements. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 188–194. doi: 10.1037/a0013161

Perception Research Systems. 2007. *Paradigm Stimulus Presentation*. Available online at: http://www.paradigmexperiments.com (accessed March 1, 2021).

Pittman, A. L., and Stelmachowicz, P. G. (2003). Hearing loss in children and adults: audiometric configuration, asymmetry, and progression. *Ear Hear.* 24, 198–205. doi: 10.1097/01.AUD.0000069226.22983.80

Pralong, D., and Carlile, S. (1996). "Generation and validation of virtual auditory space," in *Virtual Auditory Space: Generation and Applications*, ed S. Carlile (Berlin: Springer), 109–151. doi: 10.1007/978-3-662-22594-3_4

Rakerd, B., and Hartmann, W. M. (1986). Localization of sound in rooms. III: onset and duration effects. *J. Acoust. Soc. Am.* 80, 1695–1706. doi: 10.1121/1.394282

Risoud, M., Hanson, J. N., Gauvrit, F., Renard, C., Lemesre, P. E., Bonne, N. X., et al. (2018). Sound source localization. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* 135, 259–264. doi: 10.1016/j.anorl.2018.04.009

Rowan, D., and Lutman, M. E. (2006). Learning to discriminate interaural time differences: an exploratory study with amplitude-modulated stimuli. *Int. J. Audiol.* 45, 513–520. doi: 10.1080/14992020600801434

Schiavetti, N., and Metz, D. E. (2006). *Evaluating Research in Communicative Disorders*. Boston, MA: Allyn & Bacon

Shetty, H. N., Palaniappan, V., Chambayil, S. S., and Syeda, A. (2019). Assessment of localization ability – a subjective tool in Kannada version. *J. Indian Speech Lang. Hear. Assoc.* 33, 1–7. doi: 10.4103/jisha.JISHA_2_18

Smith-Olinde, L., Koehnke, J., and Besing, J. (1998). Effects of sensorineural hearing loss on interaural discrimination and virtual localization. *J. Acoust. Soc. Am.* 103, 2084–2099. doi: 10.1121/1.421355

Soranzo, A., and Grassi, M. (2014). PSYCHOACOUSTICS: a comprehensive MATLAB toolbox for auditory testing. *Front. Psychol.* 5, 712. doi: 10.3389/fpsyg.2014.00712

Spatial Auditory Displays Lab (2012). *Sound Lab (SLAB 3d).* Available online at: http://slab3d.sonisphere.com/ (accessed July 19, 2022); http://human-systems.arc.nasa.gov/SLAB/ (accessed December 9, 2022).

Spencer, N. J., Hawley, M. L., and Colburn, H. S. (2016). Relating interaural difference sensitivities for several parameters measured in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 140, 1783–1799. doi: 10.1121/1.4962444

Spierer, L., Tardif, E., Sperdin, H., Murray, M. M., and Clarke, S. (2007). Learning-induced plasticity in auditory spatial representations revealed by electrical neuroimaging. *J. Neurosci.* 27, 5474–5483. doi: 10.1523/JNEUROSCI.0764-07.2007

Strelcyk, O., and Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *J. Acoust. Soc. Am.* 125, 3328–3345. doi: 10.1121/1.3097469

Subramaniam, K., Eikelboom, R. H., Eager, K. M., and Atlas, M. D. (2005). Unilateral profound hearing loss and the effect on quality of life after cerebellopontine angle surgery. *Otolaryngol. Neck Surg.* 133, 339–346. doi: 10.1016/j.otohns.2005.05.017

Takahashi, T. (2009). A novel view of hearing in reverberation. *Neuron* 62, 6–7. doi: 10.1016/j.neuron.2009.04.004

Turner, C. W., Chi, S., and Flock, S. (1999). Limiting spectral resolution in sensorineural hearing loss. *J. Speech, Lang. Hear. Res.* 42, 773–784. doi: 10.1044/jslhr.4204.773

Tyler, R. S., Witt, S. A., Dunn, C. C., and Wang, W. (2010). Initial development of a spatially separated speech-in-noise and localization training program. *J. Am. Acad. Audiol.* 21, 390–403. doi: 10.3766/jaaa.21.6.4

van den Bogaert, T., Carette, E., and Wouters, J. (2011). Sound source localization using hearing aids with microphones placed behind-the-ear, in-the-canal, and in-the-pinna. *Int. J. Audiol.* 50, 164–176. doi: 10.3109/14992027.2010.537376

van den Bogaert, T., Klasen, T. J., Moonen, M., Van Deun, L., and Wouters, J. (2006). Horizontal localization with bilateral hearing aids: without is better than with. *J. Acoust. Soc. Am.* 119, 515. doi: 10.1121/1.2139653

van Esch, T. E. M., Kollmeier, B., Vormann, M., Lyzenga, J., Houtgast, T., Hällgren, M., et al. (2013). Evaluation of the preliminary auditory profile test battery in an international multi-centre study. *Int. J. Audiol.* 52, 305–321. doi: 10.3109/14992027.2012.759665

Venkatesan, S. (2009). *Ethical Guidelines for Bio Behavioral Research Involving Human Subjects*. All India Institute of Speech and Hearing, Mysore, India.

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.* 94, 111–123. doi: 10.1121/1.407089

Wright, B. A., and Fitzgerald, M. B. (2001). Different patterns of human discrimination learning for two interaural cues to sound-source location. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12307–12312. doi: 10.1073/pnas.211220498

Wright, B. A., Wilson, R. M., and Sabin, A. T. (2010). Generalization lags behind learning on an auditory perceptual task. *J. Neurosci.* 30, 11635–11639. doi: 10.1523/JNEUROSCI.1441-10.2010

Zahorik, B. P., Sundareswaran, V., Wang, K., and Tam, C. (2006). Perceptual recalibration in human sound localization: learning to remediate front-back reversals. *J. Acoust. Soc. Am.* 120, 343–359. doi: 10.1121/1.2208429

Zentner, M., and Strauss, H. (2017). Assessing musical ability quickly and objectively: development and validation of the Short-PROMS and the mini-PROMS. *Ann. N.Y. Acad. Sci.* 1400, 33–45. doi: 10.1111/nyas.13410

Zhang, Y., and Wright, B. A. (2007). Similar patterns of learning and performance variability for human discrimination of interaural time differences at high and low frequencies. *J. Acoust. Soc. Am.* 121, 2207–2216. doi: 10.1121/1.2434758

Zheng, Y., Swanson, J., Koehnke, J., and Guan, J. (2022). Sound localization of listeners with normal hearing, impaired hearing, hearing aids, bone-anchored hearing instruments, and cochlear implants: a review. *Am. J. Audiol.* 31, 819–834. doi: 10.1044/2022_AJA-22-00006

Zhong, X., and Xie, B. (2013). "Binaural reproduction through headphones," in *Head-Related Transfer Function And Virtual Auditory Display,* ed B. Xie (Florida: InTech), 265–282.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 26 January 2023
DOI 10.3389/fnins.2023.1000079



OPEN ACCESS

EDITED BY
Yi Zhou,
Arizona State University, United States

REVIEWED BY
Ying Xu,
Western Sydney University, Australia
John C. Middlebrooks,
University of California, Irvine, United States
Anuprasad Sreenivasan,
Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), India
Thomas Biberger,
University of Oldenburg, Germany

*CORRESPONDENCE
Samuel S. Smith
✉ samuel_smith@meei.harvard.edu
Michael A. Akeroyd
✉ Michael.Akeroyd@nottingham.ac.uk

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 21 July 2022
ACCEPTED 02 January 2023
PUBLISHED 26 January 2023

CITATION
Smith SS, Sollini J and Akeroyd MA (2023) Inferring the basis of binaural detection with a modified autoencoder. *Front. Neurosci.* 17:1000079. doi: 10.3389/fnins.2023.1000079



# Inferring the basis of binaural detection with a modified autoencoder

Samuel S. Smith *, Joseph Sollini and Michael A. Akeroyd *

Hearing Sciences, Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham, United Kingdom

The binaural system utilizes interaural timing cues to improve the detection of auditory signals presented in noise. In humans, the binaural mechanisms underlying this phenomenon cannot be directly measured and hence remain contentious. As an alternative, we trained modified autoencoder networks to mimic human-like behavior in a binaural detection task. The autoencoder architecture emphasizes interpretability and, hence, we "opened it up" to see if it could infer latent mechanisms underlying binaural detection. We found that the optimal networks automatically developed artificial neurons with sensitivity to timing cues and with dynamics consistent with a cross-correlation mechanism. These computations were similar to neural dynamics reported in animal models. That these computations emerged to account for human hearing attests to their generality as a solution for binaural signal detection. This study examines the utility of explanatory-driven neural network models and how they may be used to infer mechanisms of audition.

KEYWORDS

**binaural (two-ear) hearing effect, hearing, cross-correlation (CC), signal detection algorithm, representational learning**

## 1. Introduction

In everyday listening, it is commonplace for a sound of interest to be masked by simultaneous background sounds such as noises. If a target sound is in a different direction to a noise then they will arrive at different times to each of the ears. The auditory system takes advantage of this difference to improve the target's detectability. In the laboratory, the prototypical method to quantify this improvement is to compare detection thresholds when (1) the signal has a different interaural time difference (ITD) to the noise, versus when (2) the signal and noise have the same ITD (**Figure 1**). The amount by which the former threshold is reduced in comparison to the latter is called the "binaural masking level difference" (BMLD). The value of the BMLD depends systematically on how the ITDs differ (Durlach, 1972; Durlach and Colburn, 1978) and can be as large as 15 dB at low frequencies (Hirsh, 1948; Hirsh and Burgeat, 1958). Yet, it is an open question as to what the neural mechanisms underlying human binaural detection are.

For example, midbrain and cortical recordings in non-human species lend support to a cross-correlation mechanism comparing auditory signals across the ears (Palmer and Shackleton, 2002; Lane and Delgutte, 2005; Gilbert et al., 2015). In contrast, human behavior appears to be equally well, if not better, described by a noise-cancelation scheme (Durlach, 1963; Breebaart et al., 2001a; Culling, 2007). Computational models have been built demonstrating

that the cross-correlation framework and the noise-cancelation framework are both empirically feasible (Durlach, 1972; Colburn, 1977). Discrepancies between frameworks have not been resolved with human imaging data (Sasaki et al., 2005; Wack et al., 2012, 2014; Fowler, 2017), for which resolution and response variability are key limitations. As the neural activity in brain regions underlying binaural detection cannot be directly recorded in humans, we considered alternative methods of scrutiny from the field of machine learning.

The human-like "behavior" achievable with deep neural networks, combined with their unpremeditated network of computations, have seen them advocated as a new generation of model organisms (Scholte, 2018). These models can effectively approximate any mathematical function (Hornik et al., 1989), are resource efficient, relatively easy to record from and perturb activity in, and are not limited by species-specific ecology. In principle, if a network can be built that corresponds with human behavior, then knowing how that network works might give insight into the underlying human mechanisms. Yet, to date, the inner workings of neural networks configured to handle binaural audition have received limited consideration (Adavanne et al., 2018; Vecchiotti et al., 2019; Francl and McDermott, 2022), and almost exclusively in the context of binaural localization rather than detection. One potential stumbling block when interrogating the inner workings of neural network analogs is their black-box nature. However, network architectures that put mechanistic interpretability at the forefront (such as modified autoencoders that have shown promise in the field of physics; Higgins et al., 2017; Iten et al., 2018) could help overcome this.

Here, we trained neural network models to imitate the phenomena of binaural signal detection under human-like behavioral constraints, then interrogated their inner workings to discover *how* they operated. In three stages of work, we first sought validation of our methodology. Second, we developed networks that operated on waveforms to predict binaural detection performance. Third, we explored how the waveform-based networks operated, examining how they internally represented information. We discovered that not only did networks learn to make predictions similar to human behavior, but representations were found to have striking similarities with a cross-correlation mechanism similar to animal models (McAlpine et al., 1996; Lane and Delgutte, 2005; Asadollahi et al., 2010; Gilbert et al., 2015). Our key insight–that these computations emerged to account for human hearing–attests to their generality as a solution for binaural signal detection and illustrates the benefits of machine learning methods.

# 2. Results

## 2.1. Proof-of-principle: Inferring a latent binaural variable

Our goal was to use neural network models as a tool to infer computations underlying binaural detection in humans. Such an approach has proven successful in the field of physics (Iten et al., 2018). For example, in the case of predicting the movement of a pendulum, networks have correctly inferred an influential role of variables such as spring constant and damping factor. First, to demonstrate the feasibility of this methodology in the context of binaural hearing, we trained a network on a reduced example. We wanted to verify that, in the process of predicting the dynamics of a fully defined system, the network would infer the same latent variable as within said system. Accordingly, we trained networks to mimic a system of equations derived under the "equalization-cancelation" (EC) framework (Durlach, 1972, part IV.B; see Eq. 1 in Section "Materials and methods"), which is effective at reproducing the key phenomena of the detection of a pure tone signal masked by a broadband noise (Durlach, 1963; Klein and Hartmann, 1981; Breebaart et al., 2001a; Hartmann and McMillon, 2001; Culling, 2007; Wan et al., 2010). The framework proposes that the interaural configuration of the masking noise is "equalized" (=applying an internal time delay to the waveform from one ear to compensate for, or equalize for, the external temporal disparity compared to the waveform from the other ear) and "canceled" (=subtracting the equalized waveforms from one another), resulting in a more detectable signal. These EC operations give rise to a latent representation that can be captured by the variable $\varphi$ (**Figure 1B**, left, see Eq. 1 in Section "Materials and methods" for details). In the EC framework, this variable is used to predict the consequent improvement in signal detection from binaural processing over monaural processing, i.e., BMLDs. In particular, we were interested as to whether a neural network would automatically infer the latent variable $\varphi$ in the process of predicting BMLDs as described under the EC system of equations.

We trained a neural network, with a modified autoencoder architecture, to predict the *binaural* improvement in signal detection (i.e., BMLDs) based on four parameters describing the *monaural* arrival times of a 500 Hz signal and broadband noise at each ear. The input/output training data were drawn from EC equations fit to human psychophysics (**Figure 1B**). Following training, we tested the network on parametric combinations of BMLDs for which it had not been trained and discovered that its root-mean-square (RMS) error was just 0.075 dB. We took this as evidence that the network was able to successfully generalize its performance. The network correctly predicted larger BMLDs when the signal had a non-zero ITD and the masking noise did not, and vice versa (**Figure 1C** and **Supplementary Figure 1A**). Interrogating the computations latent within the network provided insight into how it operated. Because the network utilized a modified autoencoder architecture, its inputs were "encoded" into a simpler representation, the latent representation, by passing information through a bottleneck at the center of the network (**Figure 1B**, right). When we looked at the bottleneck node's activation values (its numerical readout), we saw that its activation almost exactly matched the latent variable in the EC framework, $\varphi$ (**Figure 1D**; Pearson's $R = 0.9994$, $p < 0.001$), even though the model was never directly informed of that variable.

In summary, within this fully defined system, the network was able to infer the appropriate latent variable in accounting for BMLD dynamics and therefore reinforced our premise.

## 2.2. Modified autoencoder accounted for binaural detection psychophysics

In our first stage, we provided the network with four parameters quantifying a signal in a noise, whereas in reality the human auditory system would be presented with *waveforms* of a signal combined with masking noise. How these waveforms are processed as to confer a

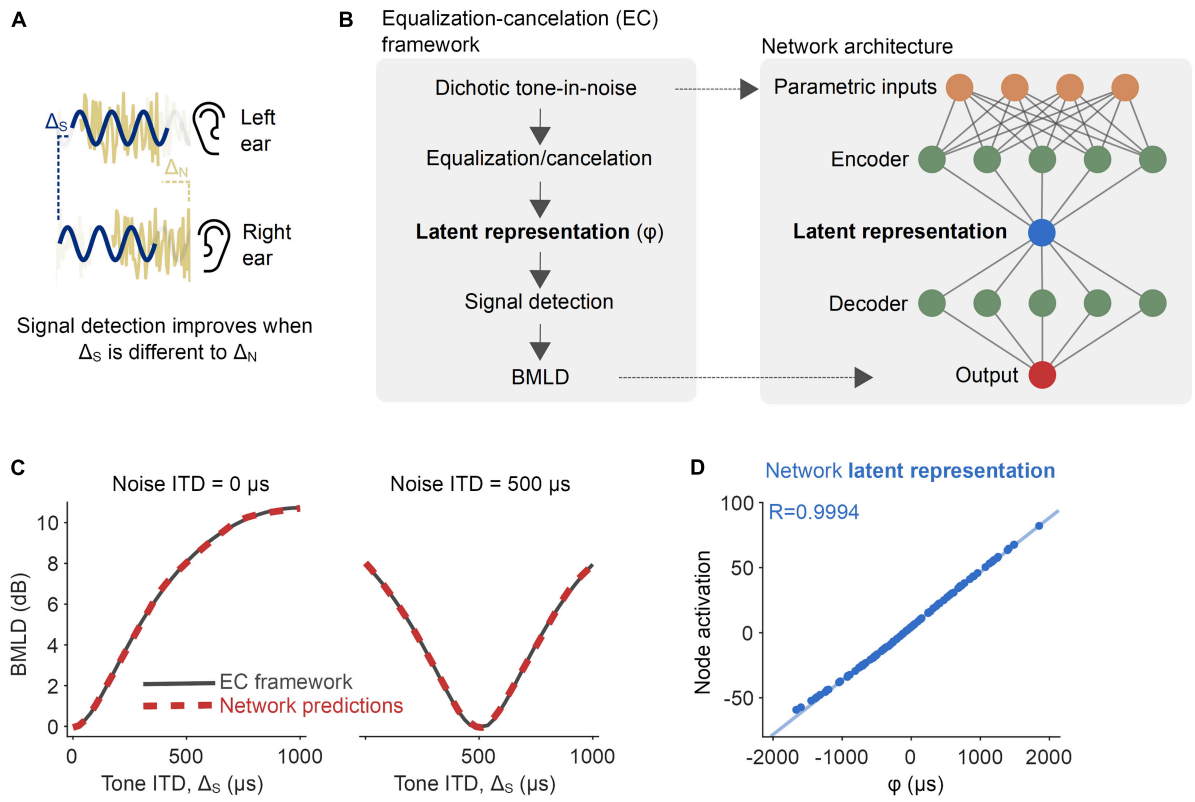



FIGURE 1
Proof-of-principle: Inferring a latent binaural variable. **(A)** The detection of a signal (sine wave denoted in navy blue) is improved if its interaural disparity is different from that of the noise (noise waveform denoted in yellow). **(B)** A neural network was trained to predict binaural masking level differences (BMLDs), as described by the equalization-cancelation (EC) framework (left). The network had a modified autoencoder architecture, in which the central layer acted as an information bottleneck. **(C)** BMLDs were numerically calculated by the EC framework (black) and estimated by the trained neural network (red-dashed), for a 500 Hz pure tone signal and noise at varying interaural time differences (ITDs). **(D)** A node central within the network had activation values entirely consistent with the latent variable as formally defined by the EC framework (φ, the signal's post-EC ITD).

binaural advantage is an open question, nor does the EC framework make any explicit proposal about how equalization parameters would be derived from said waveforms (Durlach, 1972; Wan et al., 2010). Additionally, humans display a graded psychometric performance as signal level is varied, from an inability to full detection, for which detection thresholds only offer a single-value snapshot at one chosen performance level.

Accordingly, in the second part of our work, we advanced our network/training paradigm to incorporate these aspects of binaural detection. Namely, input into the networks were vectors describing waveforms simulated as arriving at the left and right "ears" (see the top of the schematic in Figure 2B). Further, networks were constrained to predict detection rates to which a graded psychometric function could be fit (see Figure 3A). We also generalized the training data to represent signals coming from random azimuthal locations in the frontal horizontal plane, restricting the range of incorporated ITDs to within an approximate human physiological range (±655 μs; Figure 2A). To generate BMLD estimates, we retained the set of equations used in Section "2.1. Proof-of-principle: Inferring a latent binaural variable" (which were fed parameters from which waveforms were constructed), as they represent good fits to human binaural psychophysics (Durlach, 1972) and augment the availability of training data. To account for the increased complexity, the autoencoder was modified to have two layers of nodes at the "encoder" and "decoder" stages and allowed for multiple (10) nodes in the central layer of the network (Figure 2B). We ran 60 separate networks, each trained on the same data, but with varying constraints as to how independently each central node represented information. This was determined by a parameter β that specified whether the emphasis was given to the predictive accuracy of the network or the interpretation and simplicity of its latent representations. This was specified within the network's cost function, a function that specifies to what end a network should be optimized during training (see Eq. 4 in Section "5.2. Modified autoencoder network"). Based on the form of the cost function, we see that a higher value of β prioritizes the interpretation and simplicity of latent representations over predictive accuracy. Interestingly, we found that networks with a non-zero, but intermediate, value of β best accounted for a held-out set of data (Figure 2C), showing that some constraints on information encoding were better than none.

The optimal network had a root mean square error of 2.5% for the validation dataset (these networks predict detection rate, hence why the unit is % and not dB). We found this network was able to closely replicate the psychometric functions for the improvement in signal detection as the presented tone increased in level amongst a 60 dB SPL broadband noise (Figure 3A). From these data, we were able to regress functions from which to derive detection thresholds (defined as a performance level *d'* of 1) and, in turn, calculate BMLDs. We found that the network's BMLDs increased as the difference between tone ITD and noise ITD increased (Figure 3B

10.3389/fnins.2023.1000079

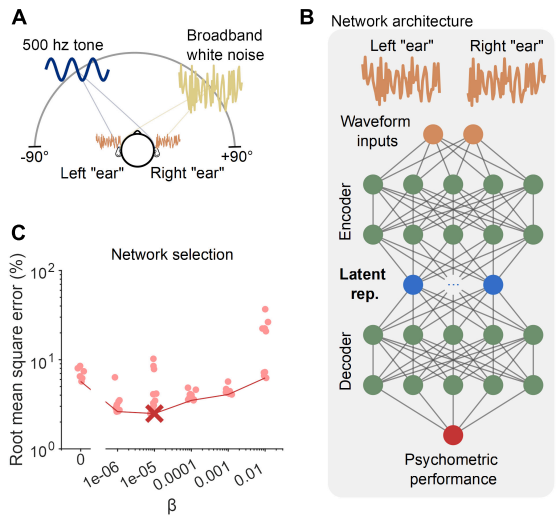



FIGURE 2
Network training and configuration. **(A)** Data from a simulated frontal field binaural detection task were used to train neural networks to detect a 500 Hz pure tone (sine wave denoted in navy blue) in broadband noise (yellow noise waveform). Locations of the tone and noise were chosen at random on each trial and were equally likely to come from each azimuthal location. **(B)** The modified autoencoder network received left/right "ear" waveforms as inputs, and had five hidden layers, with the central layer containing 10 nodes–constrained by the parameter β in their information transmission. **(C)** Error for 60 networks (10 for each value of β, see Section "Materials and methods") tested on a held-out validation dataset. The red circles indicate the errors for the 60 networks. The red cross marks the optimally performing network, and the red line bounds the networks with minimum error for each value of β.

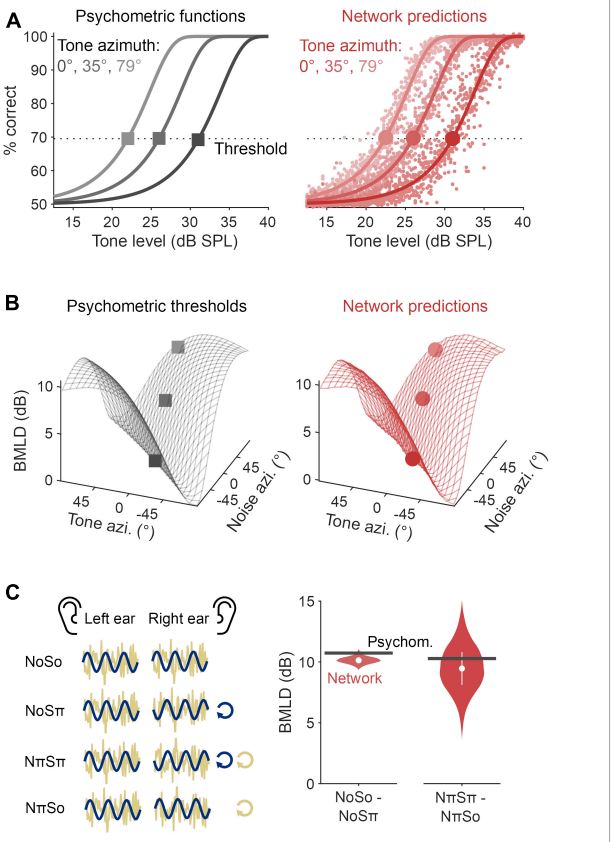



FIGURE 3
Modified autoencoder accounted for binaural detection psychophysics. **(A)** Psychometric functions quantifying tone detection as a function of tone level masked by a 60 dB SPL Gaussian noise (left, black). These functions are drawn for tones presented from three azimuths, relative to a noise presented directly in front. The optimal neural network model was able to approximate these psychometric functions (red, right), from which detection thresholds (corresponding to a d-prime of 1) and binaural masking level differences (BMLDs) could be calculated. **(B)** Psychophysical estimates (left, black) of human BMLDs for a 500 Hz tone presented in noise, each with interaural time differences (ITDs) mapped from differing azimuths. Alongside are the optimal network's predictions (right, red). Markers representing thresholds as defined in panel **(A)** are overlaid. **(C)** (Left panel) A schematic of the laboratory stimuli configurations denoted as NoSo, NoSπ, NπSπ, and NπSo. (Right panel) BMLDs were derived for experimental stimulus configurations: NoSo/NoSπ, NπSπ/NπSo (π, for a 500 Hz signal, is beyond the range of ITDs used during training).

and Supplementary Figure 1B). For example, in diotic noise (noise ITD = 0) with a tone placed at the far left, detection thresholds were significantly enhanced by 9 dB (two-sided unpaired $t$-test, $p < 0.001$), matching human BMLD behavior (Durlach and Colburn, 1978).

To allow a comparative assessment of the neural network models and previously published work, we also presented networks with stimulus configurations typically employed in the laboratory to study binaural detection. These include tones and noise in popular laboratory configurations, either in-phase or completely out-of-phase across the ears. In the literature, these stimuli are denoted as NoSo, NoSπ, NπSπ, and NπSo, where N refers to the noise, S the pure tone signal, with the subscripts denoting interaural phase difference (IPD) in radians (see Figure 3C, left panel). Importantly, none of these stimuli were used in training, nor can most occur in everyday listening. These stimuli have ITDs that are frequency dependent and can be greater than the range permitted by head width. For example, a 500 Hz pure tone with an IPD of π corresponds to an ITD of 1,000 μs, whereas the typical value for the largest ITD due to a head is 655 μs (Woodworth et al., 1954). As our networks were trained on ITDs within the head's range, this meant networks had no prior exposure to this magnitude of ITD and so it was unclear how they would function over this range. We found that when the noise signal had zero IPD, the BMLD for the corresponding homophasic (NoSo) and antiphasic (NoSπ) tone conditions was 10.1 dB, an effect that was statistically significant (two-sided unpaired $t$-test, $p < 0.001$). Comparatively, when instead the noise signal was interaurally out-of-phase, the predicted BMLD for the corresponding homophasic (NπSπ) and antiphasic (NπSo) stimuli was 9.5 dB, and again significant (two-sided unpaired $t$-test, $p < 0.001$). These BMLDs are similar to those typically measured in laboratory research (Durlach and Colburn, 1978) and with estimates from the psychophysical equations (10.7 and 10.3 dB, respectively; Figure 3C).

## 2.3. Latent representations imitate neural signature of population-level cortical activity

In the third stage, we investigated *how* the model achieved this behavior. To do this we, first, looked at the network's latent representations and considered them relative to known binaural phenomena. Prior animal neural data have shown that the stimulus conditions depicted in Figure 3C (NoSo/NoSπ and NπSπ/NπSo)

hint at a unique signature of binaural detection processing (Gilbert et al., 2015). In guinea pig cortical recordings, population spike counts dropped amongst a No signal as a 500 Hz tone went from So to Sπ (**Figure 4A**). Conversely, amongst an Nπ signal, as a pure tone transitioned from Sπ to So, population spike counts *increased*. The neural dynamics contrast, yet in both conditions binaural detection thresholds improved. We would not expect such opposing dynamics under an EC framework–a signal and a noise that are interaurally out-of-phase with one another should consistently give rise to a less "canceled" signal representation than if they were in phase with one another. Instead, this neural signature is more in line with the dynamics expected under a cross-correlation framework (demonstrated in Gilbert et al., 2015; see Section "5.9. Binaural cross-correlation algorithm" for more details on binaural cross-correlation).

We, therefore, examined the latent representation of the NoSo/NoSπ and NπSπ/NπSo stimulus conditions within the central layer of our network. To do this we needed to determine which nodes in this layer were operational, in the sense that they had non-trivial output values. We found that this was true of six nodes, whereas the remaining four had adapted to produce negligible outputs to comply with constraints on information transmission (**Figure 4B**). We found that the operational nodes exhibited opposing dynamics in response to the two pairs of homophasic/antiphasic stimuli (**Figure 4C**), although the directionality of these opposing dynamics varied across the six nodes (we believe that this is a consequence of the nodes being able to take any real number, and hence this directionality can be ignored). On average, the change in activation for NoSo/NoSπ was opposite to NπSπ/NπSo for all six operational nodes (a $2^{-6} = 0.016$ chance). Although mean differences were significant (two-sided unpaired *t*-tests for tone-level of 35 dB SPL, $p < 0.001$ for all), trial-to-trial values were noisy and overlapping [two-sample K-S test between NoSo/NoSπ and NπSπ/NπSo conditions, for a tone level of 35 dB SPL, D ranged from 0.076 ($n_4$) to 0.43 ($n_2$), $p < 0.001$ for all (see Section "5.10. Statistical analysis")], to be expected given the input waveforms were dominated by Gaussian noise. Some of this variance was due to the partial representation of non-binaural stimulus properties (e.g., monaural tone phase) that had not been adequately disregarded early in the network. Some of this variance could be accounted for based on the activity of other central nodes (**Supplementary Figure 2A**). With such co-variation accounted for, we saw a further enhanced contrast for the NoSo/NoSπ and NπSπ/NπSo stimulus conditions, markedly at threshold levels (**Supplementary Figures 2B, C**).

In summary, given that the network predicted similar magnitudes of BMLDs for NoSo/NoSπ and NπSπ/NπSo, *and* broadly captured opposing dynamics for these stimulus conditions, we conclude that the network imitated this key signature of binaural detection.

## 2.4. Encoder network dynamics matched those of a cross-correlator

Finally, in order to further understand the encoder network that lies between the waveform inputs and the latent representations described in the network's central layer, we examined ITD tuning. To determine this, we computed noise delay functions in nodes within the encoding network (**Figure 5A**), i.e., their activation values in response to noises presented with varying ITDs. Tuning was quantified by regressing a Gabor function onto the noise-delay function (Lane and Delgutte, 2005), i.e., the combination of a cosine windowed by a Gaussian (overlaid in **Figure 5A**). For nodes in the encoder's first layer, we observed significant ITD tuning in 63 out of 100 nodes (**Figure 5B**). By the encoder's second layer, significant ITD tuning had emerged in all 100 nodes. Estimates of each node's best ITD (i.e., the ITD that gives the maximum activation) were derived from Gabor fits (to account for nodes that were cyclical in their noise delay responses, the best ITD was attributed to the most central tuning peak). In both the first and second layers of the encoder network, we observed a wide distribution of best ITDs, both within the simulated head range, and beyond it (**Figure 5C**).

Importantly, one framework that is both commensurate with the earlier results (Section "2.3. Latent representations imitate neural signature of population-level cortical activity") and found in animal models is that of a binaural cross-correlator mechanism (McAlpine et al., 1996; Lane and Delgutte, 2005; Asadollahi et al., 2010; Gilbert et al., 2015). The concept is predicated on the existence of coincidence detectors that encode temporally offset signals. To deduce whether our network had automatically learned to operate like a cross-correlator, we measured nodal activations in responses to the laboratory tone-in-noise conditions: NoSo, NoSπ, NπSπ, and NπSo (**Figure 5D**). When a signal was presented amongst an in-phase noise (No), responses were largest for nodes with best ITDs near 0 μs and decreased as best ITDs were increasingly non-zero. Conversely, amongst an out-of-phase noise (Nπ), responses were lowest for nodes with best ITDs near 0 μs and increased as best ITDs deviated away from this. The effects of the tone phase on node dynamics were more subtle, although these dynamics were also in accordance with a node's tuning properties. Nodes tuned to smaller ITDs responded most to in-phase tones (So) and least to out-of-phase tones (Sπ), and vice-versa for nodes tuned to larger ITDs. These dynamics are consistent with a cross-correlation model.

Computationally, a binaural cross-product can be calculated by summing the point-by-point product of two temporally offset signals. Comparative outputs from a simple binaural cross-correlation algorithm (namely for signals in noise passed through narrow-band filters centered at 500 Hz) are shown in **Figure 5E**. We saw a significant correlation between the network and the cross-correlation calculation (with local averaging: Pearson's $r = 0.91$, $p \ll 0.001$; without: Pearson's $r = 0.36$, $p \ll 0.001$). When looking across all 60 of the networks that we trained, we found that the more a network made predictions that matched the psychophysical data, the more similar its encoder network was to a cross-correlator (**Figure 5F**).

# 3. Discussion

Binaural detection of a signal masked by noise is a well-standardized laboratory measurement that underpins important theories of auditory processing. However, the underlying mechanisms involved remain uncertain. Here, we used machine learning methods to infer potential mechanisms underlying human-like binaural detection. We found that our neural networks were able to successfully utilize interaural discrepancies across dichotic signal-in-noise waveforms to predict human-like binaural detection behavior. Notably, similarities with animal neural dynamics and a binaural cross-correlator were emergent within the network. We emphasize that these dynamics were not hard-coded into the

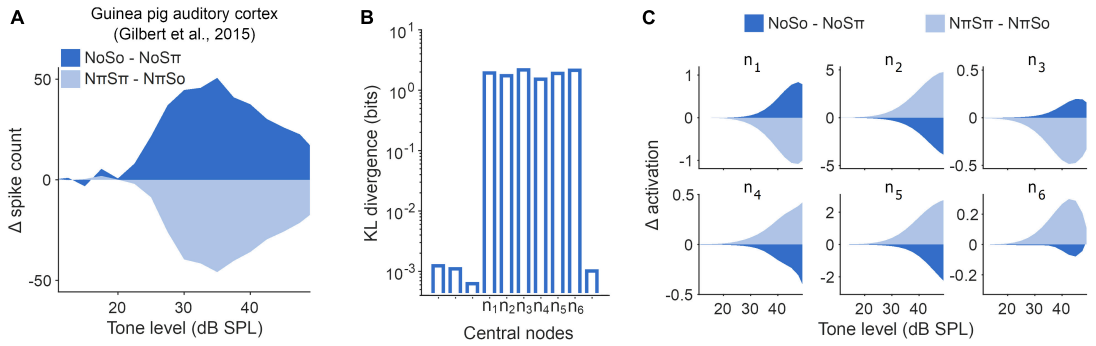



FIGURE 4
Latent representations imitated signature of population-level cortical activity. **(A)** Change in population masked rate-level functions recorded from guinea pig auditory cortex (Gilbert et al., 2015) in response to changes in experimental binaural stimuli NoSo/NoSπ (dark blue) and NπSπ/NπSo (light blue). **(B)** Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951) between each individual node and a unit Gaussian. Unlabeled nodes along the x-axis were deemed to be suppressed during training. **(C)** Rate-level functions for the operational central nodes in the optimal network, comparable to panel **(A)**.

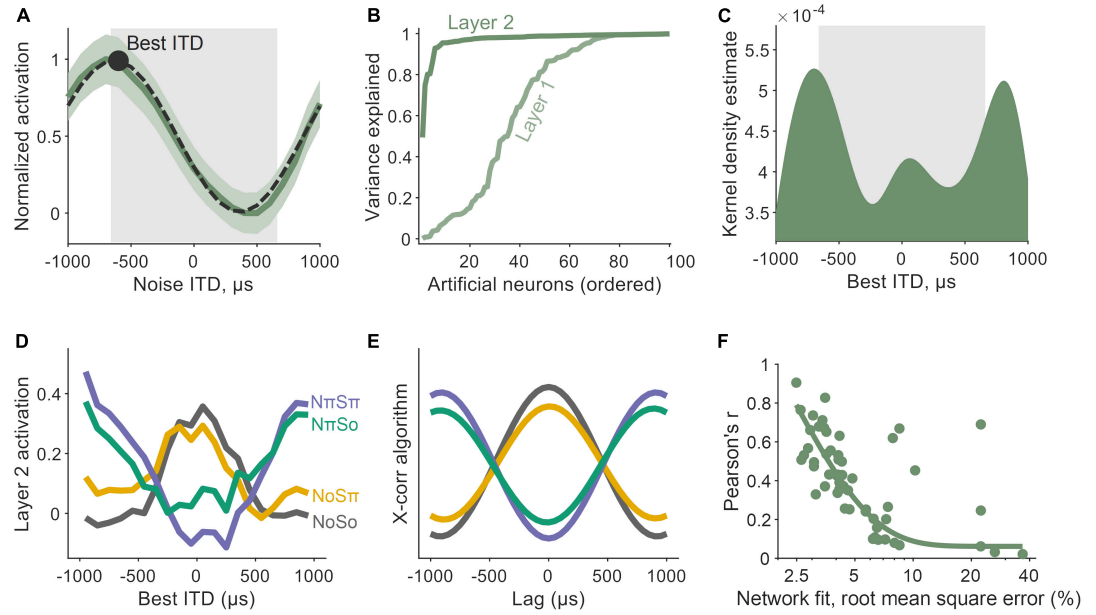



FIGURE 5
Encoder network dynamics matched those of a cross-correlator. **(A)** Interaural time difference (ITD) tuning emerged as a property of nodes within the early encoder layer of the network. The activation values of an example node are shown to vary as a function of noise ITD (dark green). Tuning was characterized by Gabor functions (black, dashed) with peaks defined as a node's best ITD (black circle). The gray box underlays represent the ITD-limit for our training simulation. **(B)** The proportion of variance explained ($R^2$) by Gabor fits, although high in Layer 1 (light green) of the encoder, was widespread by Layer 2 (darker green). **(C)** Best ITD distribution for nodes in Layer 2, characterized by a kernel density estimate (bandwidth of 200 μs). Again, the gray box underlay represents the ITD-limit for our training simulation. **(D)** Activation values of Layer 2 nodes for binaural detection stimuli: NoSo, NoSπ, NπSπ, NπSo (color-coded). Smoothed with a 600 μs moving average window. **(E)** The profiles in 5D were similar to a simple cross-correlation (X-corr) algorithm. **(F)** The better a network predicted psychophysical data (x-axis), the more similar its encoder network to a cross-correlator (y-axis).

network, they were learned, and highlight their importance in the context of signal detection, not just the more commonly referenced function of sound localization (Joris and Yin, 2007). These findings promote the understanding of how neural network models operate as an effective tool for investigating the basis of binaural processing.

## 3.1. The basis of binaural detection

In our study, we utilized a set of equations originally derived under the assumptions of the EC framework (Durlach, 1972), treating them as accurate numerical fits to human binaural psychophysical data (see Section "5.1. Binaural detection rates and thresholds"). This is the case, and was in part the motivation, for the experimental parameters investigated in this study [i.e., the EC framework fits well to human psychophysics for a 500 Hz tone and ITDs, but not for ILDs (Wan et al., 2010)]. Yet, our findings overall support a different process for interaural detection, namely, cross-correlation. The distinction is important because Domnitz and Colburn (1976) provided statistical evidence that, under certain assumptions, models based on temporal or phase differences (as the EC framework is) provide similar predictions of tone-in-noise detection to interaural

correlation-based models. They concluded that comparing binaural detection predictions made by both classes of models is insufficient to disentangle underlying mechanisms. To circumvent this, we inverted the conventional forward-approach to modeling, and instead reverse engineered our models. We discovered that our models developed a cross-correlation mechanism to reproduce psychophysical data. We also observed that central nodes broadly reproduced opposing dynamics to NoSo/NoSπ and NπSπ/NπSo, consistent with population neural activity in animal models. In contrast, one would expect that an EC-like noise cancelation scheme would operate similarly for both NoSo/NoSπ and NπSπ/NπSo stimulus conditions, and hence would not exhibit these opposing dynamics. Further, we found that additional mechanisms that utilize *a priori* knowledge of the masker, as have been proposed for some EC models (Hawley et al., 2004), are not required in order to account for binaural detection behavior. Taken together, one interpretation of our findings is that, in its analytical form, the EC framework captures the "computational goal" of the system (Marr and Poggio, 1976), enacted *via* means of a binaural cross-correlator. An alternative interpretation is that, although binaural cross-correlation produced a sufficient decision variable for the detection of simple stimuli, for more complex tasks and stimuli (e.g., speech recognition), binaural cross-correlation could instead be used to derive optimal delay parameters within a hybrid EC framework (Culling, 2020).

Despite the occurrence of the earlier mentioned network dynamics, the model exhibited flaws including dynamics that were less tangible. For example, we observed instances in which central nodes partially represented seemingly irrelevant stimulus properties, e.g., monaural phase. As opposed to the encoder network filtering out these stimulus properties, the network appeared to separately represent this co-variation and account for it at a later stage. This is possibly a consequence of the modified autoencoder architecture's preference for capturing separate latent variables in separate nodes (Higgins et al., 2017; Iten et al., 2018), potentially augmented by an over-resourced "decoder" network. In addition to these divergent dynamics, for some extreme stimulus configurations, we observed some slight discrepancies in predicted and ground truth detection thresholds, although we stress that relative differences (i.e., BMLDs) were accurately predicted. We trained our models on stimuli with ITDs limited by a typical head size (i.e., ±655 μs). However, there is evidence that natural sound statistics can incorporate ITDs beyond this limit (Młynarski and Jost, 2014). Training networks on such distributions may improve the predictive performance for extreme stimulus configurations.

We note that we have modeled only a fraction of the BMLD conditions that have been experimentally tested (see Breebaart et al., 2001a,b,c; Bernstein and Trahiotis, 2017). It will be of interest to learn how far a model like ours can further generalize to other parametric laboratory stimuli. Potential tests range from confirming more standard results such as the effect of the interaural correlation of the noise (Robinson and Jeffress, 1963; van der Heijden and Trahiotis, 1997; Bernstein and Trahiotis, 2020) to exploring results that apparently require extensions such as longer delay lines (van der Heijden and Trahiotis, 1999, but see Encke and Dietz, 2022 and Eurich et al., 2022, for an opposing interpretation). Given that our model is essentially a "stationary signal" model, at minimum an extended set of training stimuli would likely be necessary for detecting dynamically changing signals, such as those demonstrating "binaural sluggishness" (Kollmeier and Gilkey, 1990).

### 3.2. Neural network analogs of auditory processing

Understanding of binaural detection in humans has been mired due to ambiguity regarding whether animal neurophysiology data satisfactorily accounts for human psychophysics. Whilst not a substitution for "ground-truth" neurophysiology, treating deep neural networks as a model organism (Scholte, 2018) appears to be a promising approach to bridging together neural and behavioral data. Recent neural network studies have described correlates with broad organizational principles in the auditory system (Kell et al., 2018; Koumura et al., 2019; Khatami and Escabí, 2020) and asked questions of "why" a neural system operates in a particular way. Here, we focused on the question of "how" a system operates, for the well characterized binaural phenomena of improved detection of a 500-Hz tone in noise. Despite the notable computational similarities between our trained networks and neural observations, comparisons between neural network models and neural biology come accompanied by an asterisk. The network was not constructed with the goal of accurately mimicking neuronal biophysics or hierarchical complexity, but instead a trade-off was made in which a modified autoencoder architecture (Iten et al., 2018) was applied to facilitate interpretation and optimization. In future work, the limits of this network architecture could be further examined and improved by considering how predicted BMLDs are influenced by spectral and temporal qualities of the masker and target signals (Breebaart et al., 2001b,c). Further scaling of this modeling approach, for example, to examine interaural level differences or across-frequency integration, would also likely be insightful. However, any impact on interpretability should be weighed (even in this, arguably simplified, context the network dynamics were non-trivial), and such models are first contingent on the generation of suitably large psychophysical datasets.

## 4. Conclusion

In conclusion, our results newly demonstrate that neural network models, utilizing a modified autoencoder architecture, can discover key computations underlying binaural hearing. Latent activity within the model corroborates observations made in animal physiology and speaks to their generality as a solution to binaural detection. The work demonstrates the potential for machine learning methods to help bridge the gap between neurophysiology and psychophysics.

## 5. Materials and methods

### 5.1. Binaural detection rates and thresholds

The framework of equalization and cancelation (Durlach, 1972) has human psychophysical support, accurate in predicting binaural masking level difference (BMLD) data (Durlach, 1972), binaural pitch phenomena (Durlach, 1972; Klein and Hartmann, 1981; Hartmann and McMillon, 2001), and underpinning other models of binaural hearing (Breebaart et al., 2001a). Although psychophysical predictions made under this framework do not extend to individual

differences, they are sufficient to consider presumed commonalities across individuals. Numerical predictions of BMLDs in decibels were calculated from phenomenological equations derived from this framework (Durlach, 1972; Wan et al., 2010):

$$\mathrm{BMLD}\left(\tau_S, \tau_N\right) = 10\log_{10}\max\left\{\frac{k-\cos\left(\omega_0\varphi\right)}{k-\gamma\left(\tau_N-\tau_0\right)}, 1\right\} \quad (1)$$

where $\tau_S$ and $\tau_N$ are the interaural time lags of the signal and noise, $\omega_0$ is the angular frequency of the pure tone signal, $k = \left(1+\sigma_\varepsilon^2\right)e^{\omega_0^2\sigma_\delta^2}$ where $\sigma_\varepsilon^2$ and $\sigma_\delta^2$ are jitter (internal noise) parameters with the values proposed by Durlach ($\sigma_\varepsilon^2 = 0.25$ and $\sigma_\delta^2 = 105$ μs), $\gamma$ is the normalized envelope of the autocorrelation of the narrow-band noise output of a triangular-gain filter centered at the target tone frequency, and $\tau_0$ is an optimal time equalization parameter. The parameter $\varphi = \tau_S - \tau_N$ represents the difference in interaural time of the tone and noise signals (in Section "2.1. Proof-of-principle: Inferring a latent binaural variable" we examined whether a neural network could discover this parameter). The values of the other parameters were chosen according to Durlach's (1972) original formulation in which the model was fit to human data.

Psychometric functions were derived from BMLDs calculated in Eq. 1 (Egan et al., 1969), with detection thresholds defined as equivalent to a *d'* of 1 in a yes-no experiment (Green and Swets, 1966):

$$\text{detection rate} = 100\Phi\left(\frac{m_0 10^{0.1(\mathrm{BMLD}+a-23)}}{2}\right) \quad (2)$$

where BMLD is from Eq. 1, $a$ is pure tone pressure amplitude in decibels, and $\Phi$ is the cumulative normal distribution. Assuming a nominal diotic detection threshold of 31 dB SPL, we can solve for $m_0$:

$$m_0 = \frac{2\Phi^{-1}\left(0.69\right)}{10^{0.8}} \approx 0.16 \quad (3)$$

where $a$ is 31 dB SPL, BMLD is 0 dB, and detection rate is 69 %.

## 5.2. Modified autoencoder network

We ran autoencoder-based, three-part neural network models (Higgins et al., 2017; Iten et al., 2018).The three parts are the encoder, central, and decoder layers. Networks took input values that were passed through exponential linear unit (ELU) layer(s), referred to as the "encoder" portion of the network. This was followed by a single central layer with Gaussian node(s) (≥the number of parameters varied in the generation of training stimuli) with minimal uncorrelated representations, constrained by a parameter β which balances network regularization versus network interpretation. This was followed by further ELU layer(s), referred to as the "decoder" portion of the network. All layers were fully connected and feed forward. The Adam optimization algorithm (Kingma and Ba, 2014) was used to minimize the cost function:

$$C_\beta\left(\hat{x}, x, \sigma, \mu\right) = ||\hat{x}-x||_2^2 - \frac{\beta}{2}\sum_i \log\left(\sigma_i^2\right) - \mu_i^2 - \sigma_i^2 \quad (4)$$

where $\hat{x}$ and $x$ are predicted and ground truth outputs, respectively (subscript 2 is the L2 norm, superscript 2 is squaring), $\sigma$ and $\mu$ are the standard deviation and mean of Gaussian nodes, respectively, and the $i$ subscripts reference separate central nodes. Architecture meta-parameters were influenced by those described in Iten et al. (2018). Network weights and biases were randomly initialized. The number of training instances employed in each iterative update of network parameters (i.e., the batch size) was set to 256. The learning rate (training hyperparameter) was set to $5 \times 10^{-4}$ for 1,000 epochs (i.e., total passes of the entire training dataset).

## 5.3. Parameter-based network

The first "proof-of-principle" network (see Section "2.1. Proof-of-principle: Inferring a latent binaural variable") took four parametric inputs representing the arrival times of each of a 500 Hz pure tone and broadband noise at each ear. The network was trained to predict BMLDs as specified in Eq. 1. The network was trained and validated (95%/5% split, respectively) on 100 000 instances of monaural tone and noise arrival times, each randomly drawn from between 0 and 2,000 μs. The encoder and decoder portions each had one layer with 100 ELU nodes. The central layer had two nodes (one was suppressed during training) with β set to $10^{-5}$.

## 5.4. Waveform-based networks

In our second model (Sections "2.2. Modified autoencoder accounted for binaural detection psychophysics" to "2.4. Encoder network dynamics matched those of a cross-correlator") we trained networks using waveforms of a signal combined with masking noise. In this way, and in contrast to the "proof-of-principle" network, individual stimulus characteristics were not initially known by the system. These networks took 800 input values, representative of simulated left ear and right ear waveforms, each of 400 samples as simulated from a pure tone and noise mapped to different angles in the azimuth. Networks were trained to predict the corresponding detection rates, as specified in Eq. 2. Training/validation (95%/5% split, respectively) was performed with 1,000,000 instances of a random phase tone in randomly generated white noise. Pure tones had 10 periods, completing one period per 40 samples. Pure tones were treated as 500 Hz for generating estimates in Eq. 1. Pure tones were set to levels between 0 and 50 dB SPL. Pure tones were masked by randomly distributed broadband noise (50–5,000 Hz, limited by 6th order Butterworth bandpass filter) with an overall level of 60 dB SPL. The tone and noise were gated simultaneously. Tones and noises were simulated with ITDs mapped from two independent angles in the azimuth between −90° (far left) and 90° (far right). ITDs were derived from Woodworth's equation (Woodworth et al., 1954), assuming a head radius of 0.0875 m. Based upon this formula and waveform sampling, the azimuth had an effective resolution between 5.6° and 10.3°, depending upon the region within it.

The encoder and decoder portions of the network each had two 100-neuron ELU layers. The central layer of networks had 10 Gaussian nodes. As the optimization of artificial neural networks was non-deterministic, and we wished to derive a network representative of a global minimum, ten networks were trained for each value of β, namely, 0, $10^{-6}$, $10^{-5}$, $10^{-4}$, $10^{-3}$, and $10^{-2}$ giving 60 in total. The network with the least root mean square error between predicted detection rates and ground truth for the validation dataset was selected for further analysis. Central nodes were considered operational if the Kullback–Leibler divergence (Kullback and Leibler, 1951) between their individual responses and a unit Gaussian was larger than 0.1 bits.

## 5.5. An example network calculation

We illustrate the computations from input-to-output in the waveform-based networks described in Section "5.4. Waveform-based networks" and schematized in **Figure 2B**. First, a weighted sum is performed on the input vector (representative of the left/right ear waveforms) and this is passed through a non-linear function. Formally, $f(x) = \mathrm{a}(\mathrm{W}^T x)$, where $x$ is the input vector, $\mathrm{W}^T$ is a vector of trainable weights (incorporating a bias term), and a() is the non-linear "activation" function defined as:

$$\mathrm{a}(z) = \begin{cases} z, & z \geq 0 \\ e^z - 1, & z < 0 \end{cases}. \quad (5)$$

This computation gives us the "activation value" for one artificial neuron (also referred to as a node). This process is repeated 100 times, once for each of the 100 neurons in the layer–where each neuron has its own unique set of weights. We effectively have a multivariate function between the network inputs and the first layer of neurons. This transform is then repeated where the outputs of the first layer of neurons become the inputs to the next. Ultimately, we end up with 100 activation values corresponding to the number of neurons in the final layer of the "encoder." Separate weighted sums of these 100 values are computed to represent mean and standard deviation parameters describing ten latent Gaussian distributions. These parameters form the basis of the information bottleneck of the autoencoder. These parameters are used to generate 10 randomly sampled values, $\mu_i + \sigma_i \epsilon$, where $\mu_i$ and $\sigma_i$ are the mean and standard deviation parameters defining the $i$-th latent Gaussian distribution, and $\epsilon \sim \mathrm{N}(0, 1)$ a random normally distributed number. These randomly drawn values are then used as inputs to the "decoder" network. The computations of the "decoder" mimic the "encoder," but with separately defined weights, and with one final weighted sum output–the predicted binaural detection performance. For more thorough details on the modified autoencoder architecture, please see Iten et al. (2018).

## 5.6. Network predictions

Binaural masking level difference (BMLD) predictions were generated by averaging outputs for 10 repeats of a given stimulus configuration (i.e., stimulus ITDs would be fixed whilst other parameters were randomized 10 times). For the waveform-based networks, BMLDs had to be derived based on detection rates. To determine detection thresholds, the mean of 10 detection rates for tone levels, set between 0 and 50 dB SPL in 2.5 dB SPL steps, were regressed with a psychometric curve (Eq. 2; **Figure 3A**). BMLDs were predicted for (i) random phase tones amongst randomly generated broadband noise with ITDs each mapped from fixed azimuthal locations spaced between ±90° (corresponding to the effective resolution, namely, 0°, ±5.61°, ±11.27°, ±16.97°, ±22.76°, ±28.67°, ±34.73°, ±41.01°, ±47.56°, ±54.45°, ±61.80°, ±69.77°, ±78.60°, and ±88.71°), and (ii) random phase tones amongst randomly generated broadband noise each either in- or out-of-phase (i.e., NoSo, NoSπ, NπSπ, and NπSo).

## 5.7. Artificial neural representations

Artificial neuron activation values (=a node's numerical expression) were measured in response to the stimuli configurations in Section "5.6. Network predictions." Activation values were also measured as a function of ITD for broadband noise only (50–5,000 Hz, 60 dB SPL). ITDs ranged from −2,000 to 2,000 μs in steps of 100 μs. For the parametric-based network, central layer activation values were measured in response to 100 random stimulus generations. For the waveform-based networks, activations were measured in response to 5,000 random stimulus generations.

## 5.8. ITD tuning

Interaural time difference (ITD) tuning was quantified by fitting a Gabor function (Lane and Delgutte, 2005) to noise delay responses. The parametric expression for a Gabor function is:

$$G = Ae^{-(ITD-bITD)^2/2s^2} \cos\left(2\pi F\left(ITD - bITD\right)\right) + C \quad (6)$$

in which we characterized a node's best ITD as the parameter *bITD*, $F$ is the tuning curve frequency, $A$ is a scaling factor (constrained to be positive), $C$ is a constant offset, and $s$ is a decay constant. These parameters were fit with the non-linear least squares algorithm curve_fit in SciPy (Virtanen et al., 2020). An $F$-test was used to assess whether a Gabor function was a significantly better fit to noise delay responses than a linear function of ITD.

## 5.9. Binaural cross-correlation algorithm

For comparative purposes, we ran a standard psychophysical model of binaural cross-correlation (Akeroyd, 2017). It produced an output approximating an ensemble of neurons rather than individual spike trains. The stimuli NoSo, NoSπ, NπSπ, and NπSo were generated for a 35 dB SPL tone and a 60 dB SPL randomly distributed broadband noise (the algorithm utilized computer representations of dB SPL). Stimuli were sampled at 20 kHz and were 1 s in duration. Signals were passed through gammatone filters centered at 500 Hz and passed through a non-linear model of neural transduction (Meddis et al., 1990). The outputs were then delayed relative to one another, and the cross-products were calculated and summated.

## 5.10. Statistical analysis

We performed Student's two-tailed $t$-tests (assuming unequal variance) to assess differences between BMLDs. Pearson product-moment correlation was calculated between the average responses of nodes to NoSo, NoSπ, NπSπ, and NπSo, and the delay matched outputs of a binaural cross-correlation algorithm (see Section "5.9. Binaural cross-correlation algorithm"). Correlations were calculated with, and without, local averaging (within 600 μs). Student's two-tailed $t$-tests (assuming unequal variance) and two-sample Kolmogorov–Smirnov tests were performed to compare changes in central node activation values for homophasic/antiphasic stimuli pairs. The D statistic of the Kolmogorov–Smirnov test is the absolute maximum distance between the cumulative distribution functions of the two samples. The $p$-value returned by the Kolmogorov–Smirnov test is the probability that the null hypothesis, that two samples were drawn from the same distribution, is rejected. For the outlined statistical analyses, the criterion for significance was set to $p = 0.05$. Violin plots were used to capture data probability density in **Figure 3C** and **Supplementary Figure 2**. The lightly shaded underlay

in Figure 5A shows standard errors. In Figure 5F an exponential curve was robustly fit with the least absolute residual method.

## Data availability statement

The main materials presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://github.com/Hearing-Sciences/BinauralDetection_DNN.

## Author contributions

SS and MA: conceptualization. SS: methodology, investigation, and writing – original draft. SS, JS, and MA: interpretation and writing – review and editing. MA: funding acquisition, resources, and supervision. All authors contributed to the article and approved the submitted version.

## Funding

SS and MA were supported by the Medical Research Council (grant number: MR/S002898/1). JS was funded by a Nottingham Research Fellowship from the University of Nottingham.

## Acknowledgments

We are grateful for access to the University of Nottingham's Augusta High Performance Computer service. We thank Alan Palmer for assistance in sharing previously published physiology data and to Wilten Nicola for his valuable comments during the writing phase.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2023.1000079/full#supplementary-material

SUPPLEMENTARY FIGURE 1
**(A)** The prediction error between the equalization-cancelation (EC) framework and network predictions in Figure 1C. **(B)** The prediction error between the EC framework and network predictions in Figure 3B.

SUPPLEMENTARY FIGURE 2
**(A)** Some central nodes orthogonally represented stimulus-properties. For example, $n_3$ sinusoidally varied in activation value as a function of monaural tone phase. Shown for NoSo with tone level at 20 dB SPL. **(B)** Near threshold (tone level of 20 dB SPL), the distribution of values when comparing the change in $n_1$ activation between NoSo/NoSπ (dark blue, left) and NπSπ/NπSo (light blue, right) are considerably overlapping. Two-sample KS test statistic, D, is 0.11, $p < 0.001$. **(C)** When the co-variate captured by $n_3$ is controlled for (e.g., looking at when $n_3 < 0$, i.e., monaural tone phase between 0 and π) the distinction between the conditions is clearer. Two-sample KS test statistic, D, is 0.8, $p < 0.001$.

## References

Adavanne, S., Politis, A., Nikunen, J., and Virtanen, T. (2018). Sound event localization and detection of overlapping sources using convolutional recurrent neural networks. *IEEE J. Sel. Top. Signal Process* 13, 34–48. doi: 10.1109/JSTSP.2018.2885636

Akeroyd, M. (2017). *A binaural cross-correlogram toolbox for MATLAB*. Farmington, CT: University of Connecticut Health Center.

Asadollahi, A., Endler, F., Nelken, I., and Wagner, H. (2010). Neural correlates of binaural masking level difference in the inferior colliculus of the barn owl (*Tyto alba*). *Eur. J. Neurosci.* 32, 606–618. doi: 10.1111/j.1460-9568.2010.07313.x

Bernstein, L. R., and Trahiotis, C. (2017). An interaural-correlation-based approach that accounts for a wide variety of binaural detection data. *J. Acoust. Soc. Am.* 141, 1150–1160. doi: 10.1121/1.4976098

Bernstein, L. R., and Trahiotis, C. (2020). Binaural detection as a joint function of masker bandwidth, masker interaural correlation, and interaural time delay: Empirical data and modeling. *J. Acoust. Soc. Am.* 148, 3481–3488. doi: 10.1121/10.0002869

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001a). Binaural processing model based on contralateral inhibition. I. Model structure. *J. Acoust. Soc. Am.* 110, 1074–1088. doi: 10.1121/1.1383297

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001b). Binaural processing model based on contralateral inhibition. II. Dependence on spectral parameters. *J. Acoust. Soc. Am.* 110, 1089–1104. doi: 10.1121/1.1383298

Breebaart, J., van de Par, S., and Kohlrausch, A. (2001c). Binaural processing model based on contralateral inhibition. III. Dependence on temporal parameters. *J. Acoust. Soc. Am.* 110, 1105–1117. doi: 10.1121/1.1383299

Colburn, H. S. (1977). Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. *Cit. J. Acoust. Soc. Am.* 61:525. doi: 10.1121/1.381294

Culling, J. (2020). "Equalization-cancellation revisited," in *Proceedings of the forum acusticum*, Madrid, 1913–1917.

Culling, J. F. (2007). Evidence specifically favoring the equalization-cancellation theory of binaural unmasking. *J. Acoust. Soc. Am.* 122:2803. doi: 10.1121/1.2785035

Domnitz, R. H., and Colburn, H. S. (1976). Analysis of binaural detection models for dependence on interaural target parameters. *J. Acoust. Soc. Am.* 59, 598–601. doi: 10.1121/1.380904

Durlach, N. I. (1963). Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am.* 35, 1206–1218. doi: 10.1121/1.1918675

Durlach, N. I. (1972). "Binaural signal detection: Equalization and cancella- tion theory," in *Foundations of modern auditory theory*, ed. J. V. Tobias (New York, NY: Academic Press), 369–462.

Durlach, N. I., and Colburn, H. S. (1978). "Binaural phenomena," in *Handbook of perception, Vol IV, Hearing*, eds E. C. Carterette and M. P. Friedman (New York, NY: Academic), 365–447. doi: 10.1016/B978-0-12-161904-6.50017-8

Egan, J. P., Lindner, W. A., and McFadden, D. (1969). Masking-level differences and the form of the psychometric function. *Percept. Psychophys.* 6, 209–215. doi: 10.3758/BF03207019

Encke, J., and Dietz, M. (2022). A hemispheric two-channel code accounts for binaural unmasking in humans. *arXiv* [Preprint]. arXiv2111.04637. doi: 10.1038/s42003-022-04098-x

Eurich, B., Encke, J., Ewert, S. D., and Dietz, M. (2022). Lower interaural coherence in off-signal bands impairs binaural detection. *J. Acoust. Soc. Am.* 151, 3927–3936. doi: 10.1121/10.0011673

Fowler, C. G. (2017). Electrophysiological evidence for the sources of the masking level difference. *J. Speech Lang. Hear. Res.* 60, 2364–2374. doi: 10.1044/2017_JSLHR-H-16-0251

Francl, A., and McDermott, J. H. (2022). Deep neural network models of sound localization reveal how perception is adapted to real-world environments. *Nat. Hum. Behav. 2022* 61, 111–133. doi: 10.1038/s41562-021-01244-z

Gilbert, H. J., Shackleton, T. M., Krumbholz, K., and Palmer, A. R. (2015). The neural substrate for binaural masking level differences in the auditory cortex. *J. Neurosci.* 35, 209–220. doi: 10.1523/JNEUROSCI.1131-14.2015

Green, D. M., and Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York, NY: Wiley.

Hartmann, W. M., and McMillon, C. D. (2001). Binaural coherence edge pitch. *J. Acoust. Soc. Am.* 109, 294–305. doi: 10.1121/1.1331680

Hawley, M. L., Litovsky, R. Y., and Culling, J. F. (2004). The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer. *J. Acoust. Soc. Am.* 115, 833–843. doi: 10.1121/1.1639908

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., et al. (2017). "B-VAE: Learning basic visual concepts with a constrained variational framework," in *Proceedings of the 5th international conference on learning representations, ICLR 2017 - conference track proceedings*, Toulon.

Hirsh, I. J. (1948). The influence of interaural phase on interaural summation and inhibition. *J. Acoust. Soc. Am.* 20, 536–544. doi: 10.1121/1.1906407

Hirsh, I. J., and Burgeat, M. (1958). Binaural effects in remote masking. *J. Acoust. Soc. Am.* 30, 827–832. doi: 10.1121/1.1909781

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Netw.* 2, 359–366. doi: 10.1016/0893-6080(89)90020-8

Iten, R., Metger, T., Wilming, H., del Rio, L. L., and Renner, R. (2018). Discovering physical concepts with neural networks. *Phys. Rev. Lett.* 124:010508. doi: 10.1103/PhysRevLett.124.010508

Joris, P., and Yin, T. C. T. (2007). A matter of time: Internal delays in binaural processing. *Trends Neurosci.* 30, 70–78. doi: 10.1016/j.tins.2006.12.004

Kell, A. J. E., Yamins, D. L. K., Shook, E. N., Norman-Haignere, S. V., and McDermott, J. H. (2018). A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron* 98, 630.e–644.e. doi: 10.1016/j.neuron.2018.03.044

Khatami, F., and Escabí, M. A. (2020). Spiking network optimized for word recognition in noise predicts auditory system hierarchy. *PLoS Comput. Biol.* 16:e1007558. doi: 10.1371/journal.pcbi.1007558

Kingma, D. P., and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv* [Preprint]. arXiv1412.6980.

Klein, M. A., and Hartmann, W. M. (1981). Binaural edge pitch. *J. Acoust. Soc. Am.* 70, 51–61. doi: 10.1121/1.386581

Kollmeier, B., and Gilkey, R. H. (1990). Binaural forward and backward masking: Evidence for sluggishness in binaural detection. *J. Acoust. Soc. Am.* 87, 1709–1719. doi: 10.1121/1.399419

Koumura, T., Terashima, H., and Furukawa, S. (2019). Cascaded tuning to amplitude modulation for natural sound recognition. *J. Neurosci.* 39, 5517–5533. doi: 10.1523/JNEUROSCI.2914-18.2019

Kullback, S., and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.* 22, 79–86. doi: 10.1214/aoms/1177729694

Lane, C. C., and Delgutte, B. (2005). Neural correlates and mechanisms of spatial release from masking: Single-unit and population responses in the inferior colliculus. *J. Neurophysiol.* 94, 1180–1198. doi: 10.1152/jn.01112.2004

Marr, D., and Poggio, T. (1976). *From understanding computation to understanding neural circuitry*. Cambridge: Massachusetts Institute of Technology.

McAlpine, D., Jiang, D., and Palmer, A. R. (1996). Binaural masking level differences in the inferior colliculus of the guinea pig. *J. Acoust. Soc. Am.* 100, 490–503. doi: 10.1121/1.415862

Meddis, R., Hewitt, M. J., and Shackleton, T. M. (1990). Implementation details of a computation model of the inner hair-cell/auditory-nerve synapse. *J. Acoust. Soc. Am.* 87, 1813–1816. doi: 10.1121/1.399379

Młynarski, W., and Jost, J. J. (2014). Statistics of natural binaural sounds. *PLoS One* 9:e108968. doi: 10.1371/journal.pone.0108968

Palmer, A. R., and Shackleton, T. M. (2002). The physiological basis of the binaural masking level difference. *Acta Acust. United Acust.* 88, 312–319.

Robinson, D. E., and Jeffress, L. A. (1963). Effect of varying the interaural noise correlation on the detectability of tonal signals. *J. Acoust. Soc. Am.* 35, 1947–1952.

Sasaki, T., Kawase, T., Nakasato, N., Kanno, A., Ogura, M., Tominaga, T., et al. (2005). Neuromagnetic evaluation of binaural unmasking. *Neuroimage* 25, 684–689. doi: 10.1016/j.neuroimage.2004.11.030

Scholte, H. S. (2018). Fantastic DNimals and where to find them. *Neuroimage* 180, 112–113. doi: 10.1016/j.neuroimage.2017.12.077

van der Heijden, M., and Trahiotis, C. (1997). A new way to account for binaural detection as a function of interaural noise correlation. *J. Acoust. Soc. Am.* 101, 1019–1022. doi: 10.1121/1.418026

van der Heijden, M., and Trahiotis, C. (1999). Masking with interaurally delayed stimuli: The use of "internal" delays in binaural detection. *J. Acoust. Soc. Am.* 105, 388–399. doi: 10.1121/1.424628

Vecchiotti, P., Ma, N., Squartini, S., and Brown, G. J. (2019). "End-to-end binaural sound localisation from the raw waveform," in *Proceedings of the ICASSP, IEEE international conference on acoustics, speech and signal processing - proceedings*, (Manhattan, MNY: Institute of Electrical and Electronics Engineers Inc), 451–455. doi: 10.1109/icassp.2019.8683732

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* 17, 261–272.

Wack, D. S., Cox, J. L., Schirda, C. V., Magnano, C. R., Sussman, J. E., Henderson, D., et al. (2012). Functional anatomy of the masking level difference, an fMRI study. *PLoS One* 7:e41263. doi: 10.1371/journal.pone.0041263

Wack, D. S., Polak, P., Furuyama, J., and Burkard, R. F. (2014). Masking level differences – a diffusion tensor imaging and functional mri study. *PLoS One* 9:e88466. doi: 10.1371/journal.pone.0088466

Wan, R., Durlach, N. I., and Colburn, H. S. (2010). Application of an extended equalization-cancellation model to speech intelligibility with spatially distributed maskers. *J. Acoust. Soc. Am.* 128, 3678–3690. doi: 10.1121/1.3502458

Woodworth, R., Barber, B., and Schlosberg, H. (1954). *Experimental psychology*. New York, NY: Rinehart and Winston.

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 01 February 2023
DOI 10.3389/fnins.2023.1067937



OPEN ACCESS

EDITED BY
Huiming Zhang,
University of Windsor, Canada

REVIEWED BY
A. John Van Opstal,
Radboud University, Netherlands
Michael Pecka,
Ludwig Maximilian University of Munich, Germany

*CORRESPONDENCE
Ana Sanchez Jimenez
✉ ana.sanchezjimenez@dpag.ox.ac.uk
Andrew J. King
✉ andrew.king@dpag.ox.ac.uk
Fernando R. Nodal
✉ fernando.nodal@dpag.ox.ac.uk

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 12 October 2022
ACCEPTED 10 January 2023
PUBLISHED 01 February 2023

CITATION
Sanchez Jimenez A, Willard KJ, Bajo VM, King AJ and Nodal FR (2023) Persistence and generalization of adaptive changes in auditory localization behavior following unilateral conductive hearing loss. *Front. Neurosci.* 17:1067937. doi: 10.3389/fnins.2023.1067937



# Persistence and generalization of adaptive changes in auditory localization behavior following unilateral conductive hearing loss

Ana Sanchez Jimenez*, Katherine J. Willard, Victoria M. Bajo, Andrew J. King* and Fernando R. Nodal*

Department of Physiology, Anatomy and Genetics, University of Oxford, Oxford, United Kingdom

**Introduction:** Sound localization relies on the neural processing of binaural and monaural spatial cues generated by the physical properties of the head and body. Hearing loss in one ear compromises binaural computations, impairing the ability to localize sounds in the horizontal plane. With appropriate training, adult individuals can adapt to this binaural imbalance and largely recover their localization accuracy. However, it remains unclear how long this learning is retained or whether it generalizes to other stimuli.

**Methods:** We trained ferrets to localize broadband noise bursts in quiet conditions and measured their initial head orienting responses and approach-to-target behavior. To evaluate the persistence of auditory spatial learning, we tested the sound localization performance of the animals over repeated periods of monaural earplugging that were interleaved with short or long periods of normal binaural hearing. To explore learning generalization to other stimulus types, we measured the localization accuracy before and after adaptation using different bandwidth stimuli presented against constant or amplitude-modulated background noise.

**Results:** Retention of learning resulted in a smaller initial deficit when the same ear was occluded on subsequent occasions. Each time, the animals' performance recovered with training to near pre-plug levels of localization accuracy. By contrast, switching the earplug to the contralateral ear resulted in less adaptation, indicating that the capacity to learn a new strategy for localizing sound is more limited if the animals have previously adapted to conductive hearing loss in the opposite ear. Moreover, the degree of adaptation to the training stimulus for individual animals was significantly correlated with the extent to which learning extended to untrained octave band target sounds presented in silence and to broadband targets presented in background noise, suggesting that adaptation and generalization go hand in hand.

**Conclusions:** Together, these findings provide further evidence for plasticity in the weighting of monaural and binaural cues during adaptation to unilateral conductive hearing loss, and show that the training-dependent recovery in spatial hearing can generalize to more naturalistic listening conditions, so long as the target sounds provide sufficient spatial information.

KEYWORDS

**perceptual learning, training, plasticity, adaptation, spatial hearing, binaural, monaural spectral cues, ferret**

## Introduction

Sensory experience plays a vital role in calibrating neural circuits in the brain, so that they can be optimized to the prevailing sensory conditions. The experience-dependent plasticity of the maturing brain enables the processing that takes place within those circuits to adjust to growth-related changes in sensory inputs and to be matched to the statistics of the environment during sensitive periods of development (Keating and King, 2013; Kumpik and King, 2019). Plasticity in later life allows perceptual skills to improve with practice and affords a capacity to compensate for altered inputs associated with sensory disorders (Irvine, 2018; Kumpik and King, 2019).

In the auditory domain, the recovery of sound localization accuracy in mammals experiencing a unilateral conductive hearing loss has been widely used as a model for training-dependent plasticity (Keating and King, 2015). Sound localization is achieved by the computation of binaural and monaural spatial cues that result from the way sounds interact with the head and external ears (Grothe et al., 2010). Interaural time differences (ITDs) and interaural level differences (ILDs) provide the primary basis for localizing sounds in the horizontal plane, and the relationship between these binaural cues and directions in space is altered by occluding one ear, resulting in much less accurate localization judgments. Previous work has shown that adult humans and ferrets can be trained to adapt to this monaural hearing perturbation, thereby substantially recovering their localization accuracy (Kacelnik et al., 2006; Kumpik et al., 2010; Keating et al., 2016). This entails plasticity in the way monaural and binaural cues are used to localize sound (Keating and King, 2015), which is dependent on the functional integrity of auditory cortical circuits (Bajo et al., 2010, 2019; Nodal et al., 2010).

Adaptation to monaural hearing loss can be achieved by reweighting the different localization cues—with greater reliance placed on the unchanged monaural spectral cues available at the open ear—or through compensatory changes in binaural cue sensitivity (Kacelnik et al., 2006; Kumpik et al., 2010; Keating et al., 2016; Zonooz and Van Opstal, 2019). In addition, if a fixed sound level is used, the localization responses of adult humans wearing an earplug in one ear become more dependent during training on spatially ambiguous monaural head-shadow cues (Zonooz and Van Opstal, 2019). Evidence for cue reweighting is provided by the absence of an aftereffect when the earplug is removed, i.e., post-plug localization performance is indistinguishable from the normal-hearing pre-plug condition (Kacelnik et al., 2006; Kumpik et al., 2010; Keating et al., 2016; Bajo et al., 2019; Zonooz and Van Opstal, 2019). Cue reweighting provides a particularly effective strategy for adapting to asymmetric hearing loss because this helps to maintain accurate sound localization under different hearing conditions.

Spatial cue reweighting is therefore thought to be context specific, disappearing when normal binaural inputs are again experienced following earplug removal, and may arise from changes in the relative reliability of different localization cues (Van Wanrooij and Van Opstal, 2007; Keating and King, 2015). However, the cortex-dependent learning induced by monaural occlusion appears to leave a memory trace that can be retrieved when the same ear is re-plugged (Bajo et al., 2019). It is not known how long this memory trace is retained or whether the previously learned strategy for adapting to unilateral conductive hearing loss is ear specific. This is likely to be important in clinical conditions, such as otitis media with effusion (Hogan et al., 1997), where recurrent hearing loss is often experienced.

The extent to which perceptual training generalizes to untrained stimuli determines its therapeutic value. Previous studies have assessed adaptation to unilateral conductive hearing loss by presenting target sounds against a silent background in order to maximize learning. However, more naturalistic listening conditions can change abruptly as we navigate our daily lives, as do the target sounds we need to attend to. Therefore, determining whether auditory spatial learning generalizes beyond the training stimulus and to more complex listening environments is crucial for designing effective training protocols as part of rehabilitation strategies for hearing-impaired individuals.

In this study, we address both the persistence and generalization of auditory adaptive learning. Our results show that the cue reweighting that occurs when adult ferrets are first trained with abnormal auditory spatial cues takes place more readily when subsequent periods of monaural occlusion are experienced, indicating that the effects of learning persist for at least several months. However, adaptation to conductive hearing loss in one ear appears to reduce the capacity of the auditory system to compensate for occlusion of the contralateral ear. We also show that the training-dependent recovery in localization accuracy does generalize to other acoustic conditions and that the degree of adaptation correlates with the generalization of learning.

## Materials and methods

### Animals

Eleven adult female pigmented ferrets (*Mustela putorius*), aged ~6 months at the start of the study and sourced from Marshal BioResources (United Kingdom), were used. All procedures were approved by the Committee on Animal Care and Ethical Review at the University of Oxford and licensed by the Home Office under the Animals (Scientific Procedures) Act (1986).

### Hardware and training procedures

Ferrets were trained by positive reinforcement on an approach-to-target sound localization task, using water as the reward. During the testing periods, the animals were housed in enriched cages in groups of 2 or 3 individuals with *ad libitum* access to dry food and controlled access to water. They were usually tested twice daily, with each session typically lasting 20–30 min, and received water as a reward contingent on their performance on the sound localization task. The animals' body weight and water intake during behavioral testing were monitored daily to ensure that a welfare threshold of 15% weight drop was not reached, which would otherwise have resulted in the temporary suspension of testing. If required, extra water in the form of pureed food was provided at the end of each day to meet the estimated daily need of 60 ml/kg (based on daily measurements made on ferrets with free access to water in the animal facility at the University of Oxford).

Behavioral testing was performed in a circular arena (Ø 140 cm) located inside a soundproof chamber (**Figure 1A**). The arena was equipped with 12 loudspeakers (FRS 8, Visaton, Crewe,

United Kingdom) positioned at 30° intervals around its periphery in the horizontal plane. These loudspeakers were used to present target sounds for testing the ferrets' sound localization accuracy, and an additional overhead speaker was located at the center of the chamber to provide background noise when required. Auditory stimuli were produced using MATLAB (MathWorks, United States) and presented using an RX8 multi I/O processor and two SA-8 power amplifiers (Tucker-Davies Technologies (TDT), Alachua, FL, United States). The output of each loudspeaker was digitally matched and flattened across sound frequencies.

Ferrets were trained to stand on a platform at the center of the arena and lick a waterspout facing the 0° position for a variable time, 300–500 ms, which triggered the presentation of a target sound stimulus from one of the 12 peripheral loudspeakers. An infrared beam at the back of the platform and a proximity infrared detector at the central waterspout were used to ensure the animal was positioned correctly when the stimulus was presented. The animal then had to signal its location by approaching the loudspeaker to obtain a water reward from its associated waterspout, which was equipped with an infrared detector and located just below the loudspeaker at the same azimuth. Only correct responses were rewarded (typically 150–250 μl of water depending on the animal and matched across all reward spouts). An incorrect response was followed by a "correction trial," whereby the same sound was presented from the same location as in the previous unsuccessful trial. As in our previous work, up to two correction trials were allowed. If both correction trials elicited incorrect responses, an "easy trial" was triggered, in which the same sound was presented continuously until a response was made. Data from correction and easy trials were not included in the analysis.

We also obtained a more absolute measure of localization accuracy by tracking the ferret's head movement using a reflective strip attached to the midline of the animal's head. A high-speed camera located above the central platform captured the reflectance of the head-strip for 1 s from the onset of the target sound. The image acquisition rate was 60 frames per second (FPS) for the first part of the study (Imaging Source DMK21BF04) and 500 FPS for the second part of the study, following an upgrade to a high-speed camera (DMK 37AUX287).

## Experimental design

All animals were trained and tested on the approach-to-target sound localization task in a quiet background (hereafter referred to as "silence") for 1–2 months until they showed a stable level of performance before the ear-plugging experiment started. Sound localization accuracy was measured using single bursts of broadband noise (low-pass cut-off frequency 30 kHz) in blocks of constant stimulus duration (2000, 1000, 500, 200, 100, and 40 ms). Within each block, we pseudorandomly varied both the stimulus location across the 12 loudspeakers and its level over a 56–84 dB SPL range in steps of 7 dB.

A reversible unilateral conductive hearing loss was induced by inserting an earplug (E-A-R Classic 3M) into the external auditory meatus and securing this in place with a silicone mold (Otoform KC, Dreve Otoplastik GmbH, Unna, Germany) placed in the concha of the external ear. Earplug insertion and removal were performed under sedation (medetomidine hydrochloride 0.1 mg/kg i.m., Domitor Orion Pharma, Reading, United Kingdom). An otoscopic examination was performed and a tympanogram (Kamplex KLT25 Audiometer, P.C. Werth, London, United Kingdom) was obtained both before insertion and after removal of an earplug to check the health status of the external and middle ear. No abnormalities were detected. Sedation was reversed with atipamezole hydrochloride (0.5 mg/kg s.c., Antisedan, Orion Pharma, Reading, United Kingdom). The acoustical effects of these earplugs have been characterized in previous studies in our laboratory in ferrets (Moore et al., 1989; Keating et al., 2013) and humans (Kumpik et al., 2010; Keating et al., 2016). We did not determine whether their properties changed during each period of monaural occlusion as this would have involved sedating the animals to remove and reposition the earplugs.

As in our previous work (Kacelnik et al., 2006; Bajo et al., 2010, 2019), we trained ferrets to adapt to an earplug using 1,000-ms long broadband noise bursts as the target stimuli on a background of silence. Individual ferrets wore an earplug in one ear for typically ≥10 days, until their performance stabilized over three consecutive days at a mean score of ≥70% correct across all 12 loudspeaker locations. We then removed the earplug and measured their localization accuracy again. We next examined either the persistence of adaptive learning by re-plugging the same ear and then the contralateral ear, or the generalization of learning using different target sounds and background noise combinations.

## Persistence of learning

To explore the persistence of the adaptive learning with one ear occluded, one group of 4 ferrets was subjected to several periods of unilateral conductive hearing loss interspersed within variable periods of normal binaural hearing.

Once their localization performance under normal hearing conditions had been measured using broadband noise bursts as target sounds in silence, a temporary unilateral conductive hearing loss was induced by inserting an earplug in the right ear (**Figure 1B**). Target sound levels and locations were pseudorandomized as described above. This first period of right ear occlusion (right plug 1) was followed by 7 days of normal hearing and then by another period of right ear occlusion (right plug 2) [data from right plug 1 and right plug 2 were reported in Bajo et al. (2019) as part of the control group in that study]. After a 6-month-long break in which the ferrets experienced normal hearing conditions, localization performance was measured during two further periods of right ear occlusion (right plug 3 and right plug 4). Finally, two periods of left ear occlusion were conducted (left plug 1 and left plug 2).

## Generalization of learning

A second group of 7 ferrets was used to explore whether the adaptive learning induced by training animals wearing an earplug to localize broadband noise target sounds on a silent background extended to other stimuli and acoustic conditions that were not available during training.

The sound localization performance of the animals was measured using different stimuli before, during and after occlusion of one ear. As before, adaptation to the earplug was achieved by training the animals to localize 1,000-ms broadband noise burst targets presented in silence, after which the animals were tested using different stimuli and background sounds with the earplug still in place. The target

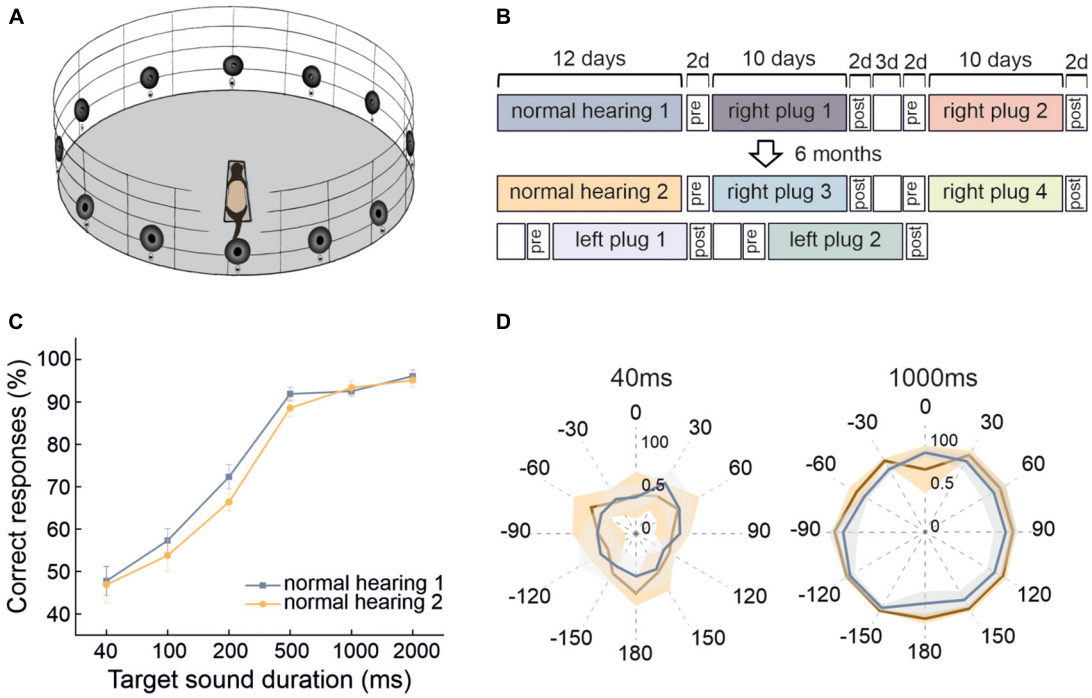



FIGURE 1
Stability of ferret sound localization performance over time when normal spatial cues are available. **(A)** Schematic of the circular testing arena with 12 loudspeakers and reward ports located at 30° intervals around the perimeter and a central platform on which the animal stands facing position 0° in order to trigger each stimulus presentation. **(B)** Timeline of behavioral testing to explore the persistence of adaptive learning over multiple periods of monaural occlusion interspersed by periods of normal binaural hearing. These animals were first trained to adapt with the right ear plugged and then with the left ear plugged. **(C)** Localization performance (percentage of correct trials averaged across all 12 target locations and sound levels) at different sound durations for two normal hearing periods that were separated by extensive localization training with one ear occluded. Error bars show the standard error of the mean. **(D)** Spatial distribution of the percentage correct scores from both normal hearing periods for the 40- and 1,000-ms broadband noise bursts. Lines represent the means and shaded areas the standard error of these scores across animals.

sound was either broadband (BB) noise (low-pass cutoff at 30 kHz) or one-octave wide narrowband noise (NB) centered on 16 kHz. The duration of the target sound was again 1,000 ms, and its level and location were pseudorandomly varied as before. These stimuli were presented either in silence (s) or in the presence of continuous background noise presented from an overhead loudspeaker. The background noise consisted of a broadband stimulus (low-pass cut-off at 30 kHz), whose envelope was either unmodulated (um) or amplitude modulated (am) at 5 Hz. The combination of these targets and backgrounds produced five different testing conditions: BBs, NBs, BBum, NBum, and NBam. The combination of broadband target sounds presented in an amplitude-modulated background (BBam) was discarded as a testing condition, because no difference was observed in localization accuracy between modulated and unmodulated backgrounds with this target sound in normal hearing conditions. Broadband target sounds were therefore only presented in silence (BBs) or with an unmodulated background (BBum).

## Role of dynamic and spectral cues

To explore the extent to which dynamic spatial cues that might be provided by the movement of the head during stimulus presentation and the spectral cues available at the non-occluded ear contribute to adaptation to unilateral conductive hearing loss, we included additional sessions during daily sound localization training with an earplug in place. Ferrets ($n$ = 4) were trained to localize 1,000-ms broadband noise bursts in silence in the same fashion as described in previous sections. Every 2 days, we interspersed broadband and narrowband sounds (1/6-octave bandwidth centered at 15 kHz frequency to restrict the cues to ILDs) (Keating et al., 2015) at two different durations (200 and 1,000 ms), with each of these 4 stimulus types being presented with equal probability. The level of the broadband noise bursts was pseudorandomized as described previously, whereas the level of the narrowband stimuli was kept constant at 84 dB SPL, except for 10% of the trials in which it was 56 dB SPL. Incorrect responses to the broadband target sounds were followed by correction and easy trials, whereas no correction or easy trials were provided after an incorrect response on the narrowband trials. This was done to ensure that the animals did not learn a different strategy when localizing these stimuli that might have interfered with the normal adaptation process with broadband targets. However, all correct responses were rewarded to maintain the motivation of the animals.

## Data analysis

### Head orienting responses

The angular position of the head-strip in each video frame was used to produce head movement traces. Trials in which <10% of consecutive frames contained measurable data were rejected. The start of a head turn was defined as a movement exceeding a threshold speed of 50 deg/s in the same direction (initial direction) over at least

three consecutive frames. The timing of the last frame before a change in this initial direction was taken as the end of the head turn. The final head bearing was defined as the mean head angle calculated over the three frames before the end of the head movement. For the learning generalization experiment, traces were down-sampled to 50 FPS to make the data analysis comparable to the learning persistence experiment (**Figure 1B**), in which the sample rate was 60 FPS.

### Statistical analysis

The rate of adaptation for each of the periods of monaural occlusion was computed by fitting regression lines to the percentage of correct responses across training day, with the shaded areas in the figures representing 95% confident intervals. The probability of making a correct response was compared across animals, hearing conditions (unilateral earplug or no earplug), and stimulus characteristics using a linear mixed model with a Bernoulli-distributed data, probit link function. Final head bearings were compared using repeated measures ANOVA. Data normality was checked using Q-Q plots and applying the Shapiro–Wilk test. All statistical tests were performed using RStudio (RStudio: Integrated Development for R. RStudio, Inc., Boston, MA, United States) or SPSS (IBM, SPSS Inc., Armonk, NY, United States).

## Results

### Effects of sound localization training on task performance

Over the course of collecting behavioral data on the sound-localization task, we observed little change in the performance of the ferrets under normal hearing conditions (i.e., during the testing runs carried out without an earplug) (**Figure 1**). The accuracy of their responses on the approach-to-target task improved as the duration of the broadband target sound was increased (from 47.8 ± 3.4% correct for 40 ms to 96.1 ± 1.5% correct for 2,000 ms) (GLMM, $p < 0.0001$) (**Figures 1C, D**). Despite extensive testing, which included several periods of monaural occlusion (**Figure 1B**), their performance across different stimulus durations remained very stable when normal binaural inputs were available, other than a small reduction in percentage correct scores at intermediate target durations (100–500 ms) (GLMM, $p = 0.001$) (**Figure 1C**). The lack of improvement in localization accuracy over time for broadband stimuli providing access to the full range of spatial cues suggests that no perceptual learning had taken place. Furthermore, the consistent performance across different stimulus locations (**Figure 1D**) gave no indication of any long-lasting change in the way these cues were processed and integrated under normal hearing conditions, despite the ferrets experiencing several intervening periods during which they learned to adapt their behavior to conductive hearing loss in one ear.

### Behavioral adaptation to unilateral conductive hearing loss

The approach-to-target responses of one cohort of ferrets ($n = 7$) are shown in **Figure 2A**. They localized 1,000-ms broadband noise busts with a consistently high level of accuracy when normal binaural and monaural spatial cues were available, with their performance declining from ∼95% in the last pre-plug session to ∼30% correct when one of their ears (the left in this case) was first occluded. The ferrets initially mislocalized almost every stimulus presented in the frontal region of the hemifield ipsilateral to the earplug (see polar plots at the top of this figure). Over the first few days of monaural occlusion, a gradual improvement in performance occurred at all other locations, including those on the side of the open ear, which was then followed by a recovery in localization accuracy at the frontal ipsilateral locations. This pattern of adaptation was remarkably consistent across animals.

We also measured the accuracy of the ferrets' sound-evoked head orienting movements, as these are thought to involve different neural circuits from those required for the approach-to-target responses (Thompson and Masterton, 1978; Lomber et al., 2001; Nodal et al., 2012; Isa et al., 2021; **Figures 2B, C**). Data from one example ferret (F2005) show that when the earplug was inserted, the initial head turns made following sound presentation were clearly biased toward the side of the open ear, independent of the location of the target (**Figure 2B**, first 3 plug days). Adaptation of the head orienting movement then occurred with training on the sound-localization task, which was manifest as a gradual recovery in orienting responses toward the side of the plugged ear for stimulus locations on that side (**Figure 2B**, last 3 days). It should be noted, however, that the degree of adaptation varied between animals and was generally not complete, as shown by the undershoot in the responses made in that hemifield by ferret F2005 on the last 3 plugging days (**Figure 2C**). This mirrors the asymmetry in the approach-to-target behavior at the corresponding stage (**Figure 2A**). As with the approach-to-target responses, the metrics of the head turns measured on the days after earplug removal resembled the pre-plug data (**Figure 2B**).

Adaptation of the head orienting responses was quantified for each ferret as the area between the average final head bearing across all animals prior to earplug insertion and the final head bearing measured for individual animals on the side of the occluded ear in the last 3 days of training (**Figure 2C**). This was then compared to the adaptation index derived from the approach-to-target behavior, which was computed as the ratio between the percentage correct responses in the last 3 days of monaural occlusion and the 3 days prior to earplug insertion. The comparison of these two measures for individual animals revealed that they were highly correlated (slope significantly different from 0; $R^2 = 0.94$, $p < 0.001$) (**Figure 2D**), with the ferrets that achieved the highest percentage correct scores also showing greatest adaptation of the head orienting responses, as reflected by a smaller difference between pre-plug and adapted final head bearing.

### Equivalent adaptation to occlusion of each ear

Before investigating the effects of further exposure to unilateral conductive hearing loss in either the same or the contralateral ear, it was first necessary to show that ferret localization behavior adapts equally well to occlusion of either the left or the right ear. This is illustrated in **Figure 3** for the sound-evoked head orienting responses recorded from two groups of ferrets, one of which was plugged in the left ear (the animals used in **Figure 2**) and another in the right ear. In the normal hearing (pre-plug) sessions, the final head bearing of the ferrets varied systematically with the target location out to

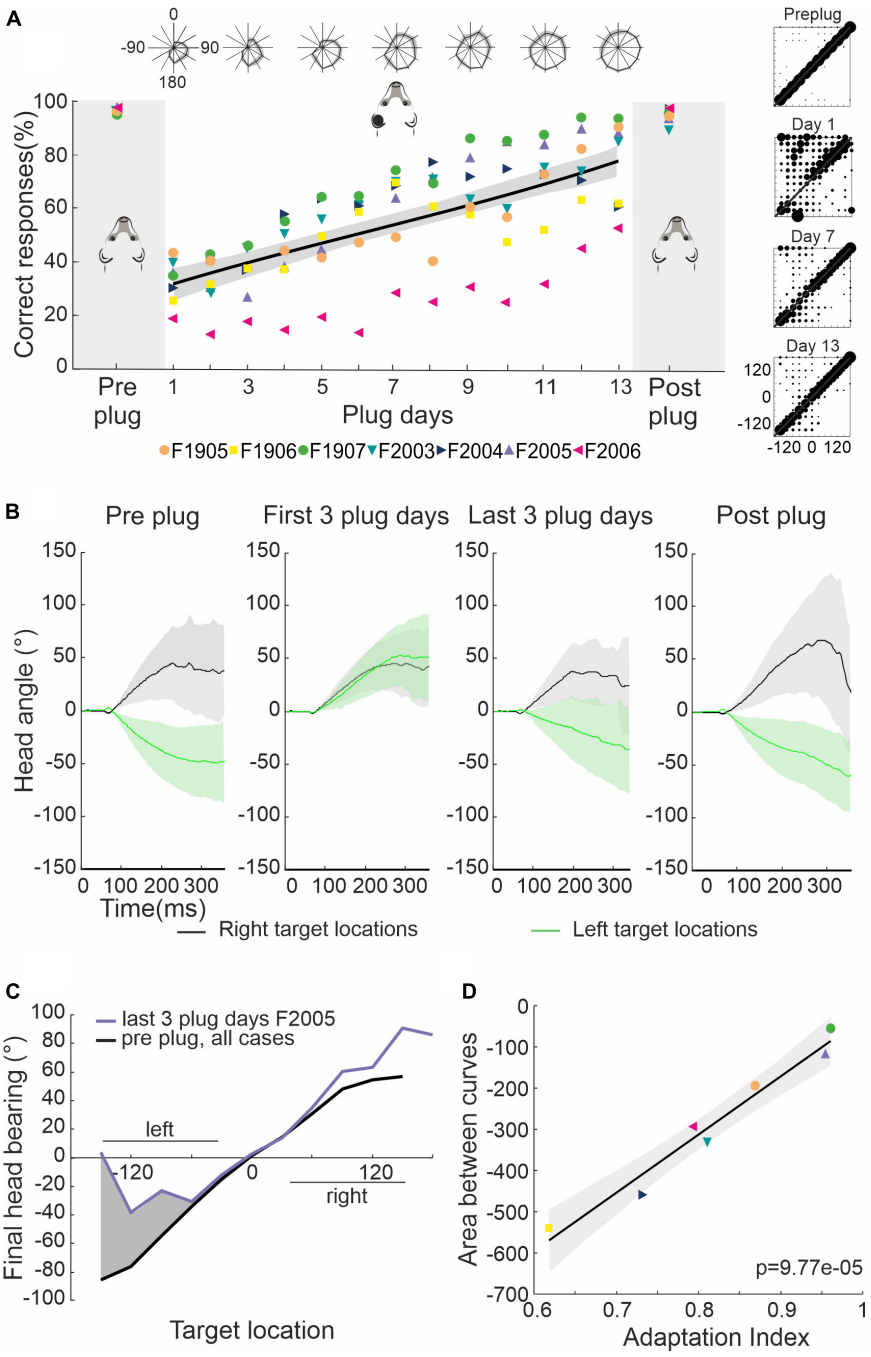



FIGURE 2
Adaptive changes in approach-to-target and head orienting responses over the course of monaural occlusion. **(A)** Percentage correct scores for the approach-to-target behavior. Different symbols indicate individual ferrets. Their mean scores over 3 days prior to earplug insertion and 3 days after earplug removal are highlighted by the gray regions on either side of the scores obtained for each day of training with the left ear occluded, which for these animals lasted 13–16 days (days 1–13 only are shown). The black line and shaded area represent the linear regression and the 95% confident intervals for the data obtained with the earplug in place. The polar plots at the top show the mean ± S.E. score across animals at different target locations for every 2 days of training. The stimulus-response plots on the right show the distribution of response locations selected (*y*-axis) as a function of target location (*x*-axis) before and at three different time points after earplug insertion. **(B)** Mean ± S.E. changes in head angle for left (green) and right (black) target locations for ferret F2005 (measured at 500 FPS and low-pass filtered at 50 Hz for plotting). Positive values correspond to rightward turns and negative values correspond to leftward turns of the head. Plots from left to right correspond to the 3 days before earplug insertion, the first 3 days with the ear plugged, the last 3 days with the ear plugged, and the 3 days following earplug removal. Occluding the left ear disrupted head orienting responses made to locations on the left, which progressively recovered as approach-to-target performance improved with training. **(C)** Mean final head bearing versus target location before earplug insertion averaged across all animals (black) and for the last 3 days of earplug adaptation for the example animal F2005 (purple). The shaded region represents the area between the two curves on the side of the occluded ear, computed as an index of adaptation of the head orienting responses. **(D)** Relationship between the adaptation index (ratio of the percentage correct scores on the last 3 days of plugging and before earplug insertion), which reflects recovery with training of the approach-to-target behavior, and adaptation of the head orienting responses (the area between the head orientation curves, as in panel **C**) for all seven animals shown in panel **(A)**.

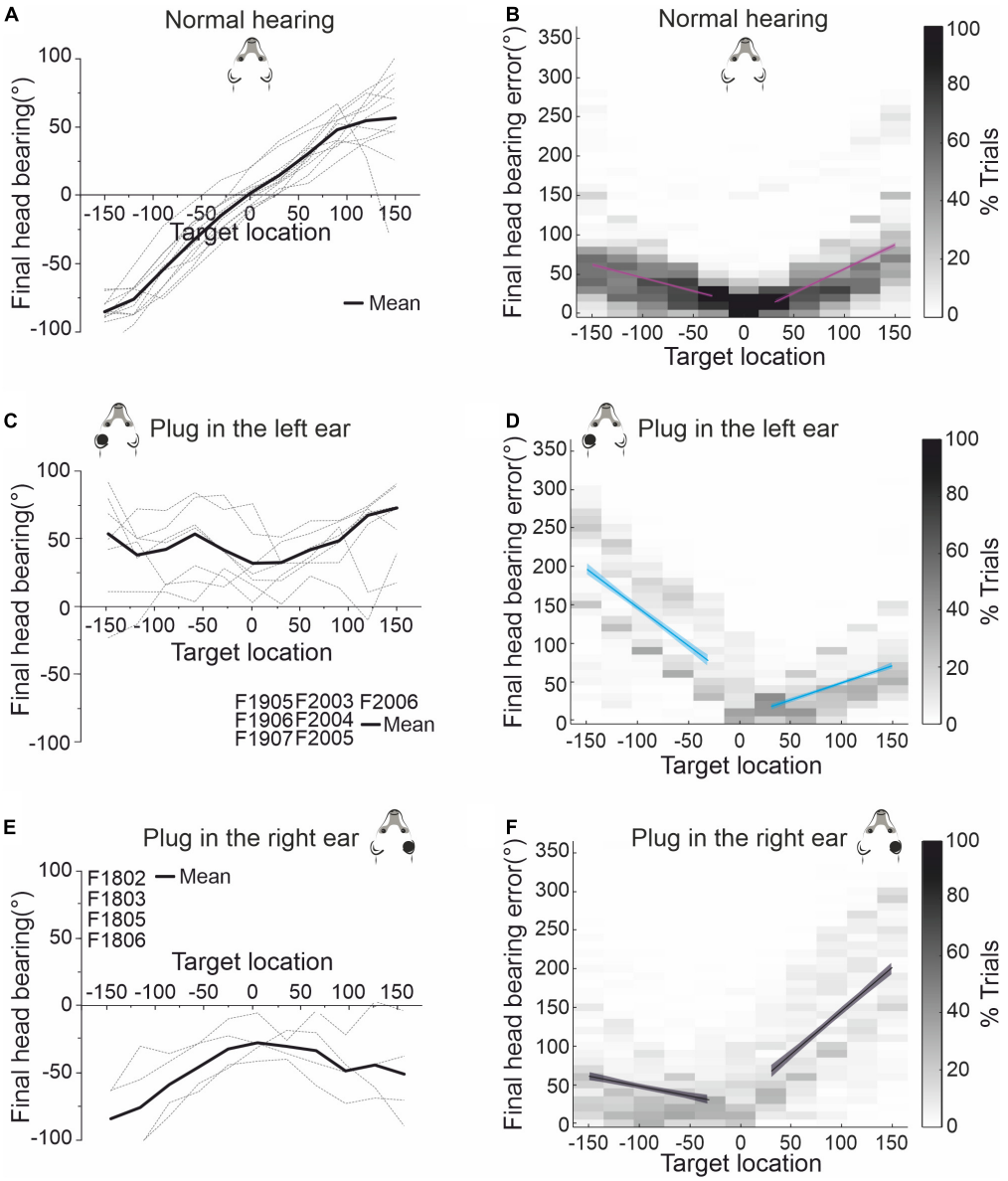



FIGURE 3
Effect of plugging either the left or the right ear on the accuracy of sound-evoked head-orienting responses. **(A)** Average final head bearing versus target location in normal hearing conditions for each of the animals (gray dashed lines) that participated in this study. Positive values correspond to targets on the right and rightward head movements, and negative values to targets on the left and leftward movements. The solid line represents the mean across animals. **(B)** Mean final head bearing errors versus target location in normal binaural hearing conditions for all animals. Linear regressions and their 95% confidence intervals are fitted independently for right and left target locations (magenta lines). **(C)** Mean final head bearing versus target location in the first 3 days with the left ear occluded. **(D)** Mean final head bearing errors versus target location in the first 3 days with the left ear occluded. Linear regressions and their 95% confidence intervals are fitted independently for right and left target locations (blue). **(E)** Mean final head bearing versus target location in the first 3 days with the right ear occluded. **(F)** Mean final head bearing errors versus target location in the first 3 days with the left ear occluded. Linear regressions and their 95% confidence intervals are fitted independently for right and left target locations (black/gray).

**TABLE 1** Parameters of the fitted regression lines for head bearing errors in Figure 3.

| | y = a + bx | $R^2$ | a (intercept) | b (slope) | $p_b$ | tStat (b) |
|---|---|---|---|---|---|---|
| Preplug | Left side | 0.17 | 11.96 | −0.333 | <0.0001 | $t_{(2049)} = -20.220$ |
| | Right side | 0.28 | −4.00 | 0.610 | <0.0001 | $t_{(1822)} = 26.674$ |
| Plug in the right ear | Left side | 0.05 | 22.69 | −0.258 | <0.0001 | $t_{(854)} = -7.102$ |
| | Right side | 0.34 | 33.55 | 1.124 | <0.0001 | $t_{(833)} = 20.829$ |
| Plug in the left ear | Left side | 0.34 | 47.17 | −1.000 | <0.0001 | $t_{(762)} = -19.805$ |
| | Right side | 0.26 | 3.86 | 0.449 | <0.0001 | $t_{(777)} = 16.784$ |

TABLE 2 Parameters of the fitted regression lines during earplug adaptation.

| y = a + bx | $R^2$ | a (intercept) | b (slope) | $p_b$ | tStat (b) |
|---|---|---|---|---|---|
| Right plug 1 | 0.633 | 10.366 | 6.377 | 5.06e−10 | $t_{(38)} = 8.2695$ |
| Right plug 2 | 0.198 | 54.064 | 2.939 | 0.00234 | $t_{(38)} = 3.2623$ |
| Right plug 3 | 0.337 | 37.354 | 3.775 | 5.07e−05 | $t_{(38)} = 4.5681$ |
| Right plug 4 | 0.122 | 51.785 | 2.395 | 0.0154 | $t_{(38)} = 2.5363$ |
| Left plug 1 | 0.474 | 20.61 | 3.541 | 5.43e−07 | $t_{(38)} = 6.0149$ |
| Left plug 2 | 0.442 | 33.301 | 3.524 | 1.71e−06 | $t_{(38)} = 5.6525$ |

±150° azimuth (Figure 3A). As previously reported (Nodal et al., 2008), we observed an increasing undershoot in the final head bearing for progressively more eccentric targets (Figure 3B). There was no difference in the accuracy of the responses to targets in the left and right hemifields or in the absolute slope of the regression lines fitted to these data (GLMM, $p = 0.055$; Table 1).

The head-orienting responses were affected in exactly the same way by plugging either the left ear in one group of ferrets (Figures 3C, D) or the right ear in the other group (Figures 3E, F). The final head bearing was initially biased toward the side of the open ear, i.e., to the right when the left ear was plugged (Figure 3C) and to the left when the right ear was plugged (Figure 3E). Although the size of the head orienting errors across locations remained unchanged on the side of the open ear, much larger errors were made on the contralateral side, as indicated by the steeper slope of the linear fits to these data (Figures 3D, F and Table 1). This asymmetry in performance between the left and right hemifields then declined with training, as illustrated by the adaptive changes in head-orienting behavior for one example animal (F2005) in Figure 2B.

## Adaptive learning is partially retained when spatial cues are altered again

Previous work has suggested that adaptation to the altered cues resulting from monaural occlusion leave a "memory trace," since a second period of ear-plugging has a smaller impact on localization accuracy (Kacelnik et al., 2006; Bajo et al., 2019). Here, we explored further how this spatial learning and its memory or retrieval are affected by multiple periods of monaural occlusion and their relative timing. To that aim, the interval between the training periods was varied, as was the ear that was plugged.

The ferrets in this group ($n = 4$) first experienced four periods of right ear occlusion, separated by short (7 days) or long (6 months) periods of normal binaural hearing. To determine whether they were then able to adapt to a different set of altered spatial cues, we subsequently measured their performance when the contralateral left ear was occluded instead (see Figure 1B for timeline). We modeled adaptation to an earplug by fitting a simple regression line, for which the independent term (a) is inversely proportional to the initial disruption caused by insertion of the plug and the slope (b) represents the rate of adaptation (Table 2). Plugging the right ear for the first time (right plug 1) resulted in the most dramatic loss of sound localization accuracy from 97.6 ± 1.2% correct to 28.9 ± 9.3%, as indicated by the lowest value of the intercept ($a = 10.366$). Sound localization accuracy recovered with twice-daily training and the highest rate of adaptation ($b = 6.4$) was observed during this first period of monaural occlusion (Figure 4A). On all three subsequent occasions that the same ear was plugged, a smaller initial drop in performance was observed (Figure 4 and see intercepts in Table 2). Although less adaptation was therefore required, the animals' performance improved at a slower rate (less steep slopes) than during the first adaptation period, each time reaching a mean score of 70–80% correct after 10 days of monaural occlusion. All adaptation slopes were, however, statistically significant from zero (Table 2), indicating that the ferrets were relearning to accommodate the altered spatial cues after each period of normal binaural experience.

Despite the similar pattern of adaptation in the second, third and fourth periods of right ear occlusion, differences were observed based on the interval between the plugging runs. The right plug 2 and right plug 4 runs were preceded by a short period (7 days) of normal binaural hearing following a prior period of monaural occlusion (right plug 1 and right plug 3, respectively) and had the smallest initial localization errors when the earplug was first inserted and lowest slope values (Figures 4A, B, D, E and Table 2). This, therefore, suggests a greater retention of the previously learned adaptation. By contrast, the third period of right ear occlusion (right plug 3) was preceded by 6 months of normal hearing, and the adaptation profile more closely resembled that for right plug 1, i.e., a larger initial drop in performance (lower intercept value) and faster rate of adaptation (higher slope) than in the right plug 2 and right plug 4 runs (Figures 4C, E and Table 2).

The distribution of the localization errors produced by monaural occlusion was not homogeneous across target locations and was always greater in the hemifield ipsilateral to the occluded right ear. This pattern was apparent at the start of all 4 ear-plugging runs (Figures 4F, G top row), and partly reflects mis-localization of sounds presented on the side of the earplug to the opposite side (right-to-left errors were initially twice as high as left-to-right errors). However, in contrast to the first and third periods of right ear occlusion, the difference between the mean response errors in the left and right hemifields for the right plug 2 and 4 runs was not significant (GLMM, $p = 0.126$, $p = 0.99$, respectively) (Figure 4F), providing further evidence that the amount of prior normal binaural experience determines how the brain responds behaviorally to each episode of unilateral conductive hearing loss.

The spatial pattern of errors also provides insights into the potential basis for the recovery of sound localization accuracy at different stages during the adaptation process. Localization errors on the side of the open left ear were consistently low at the start of the second, third and fourth periods of right ear occlusion (Figures 4F, G), which could reflect a rapid switch to the use of monaural cues provided by the open ear. A greater and more gradual improvement in performance was found on the side of the plugged ear (Figure 4G) and over this time the incidence of both left-right and front-back

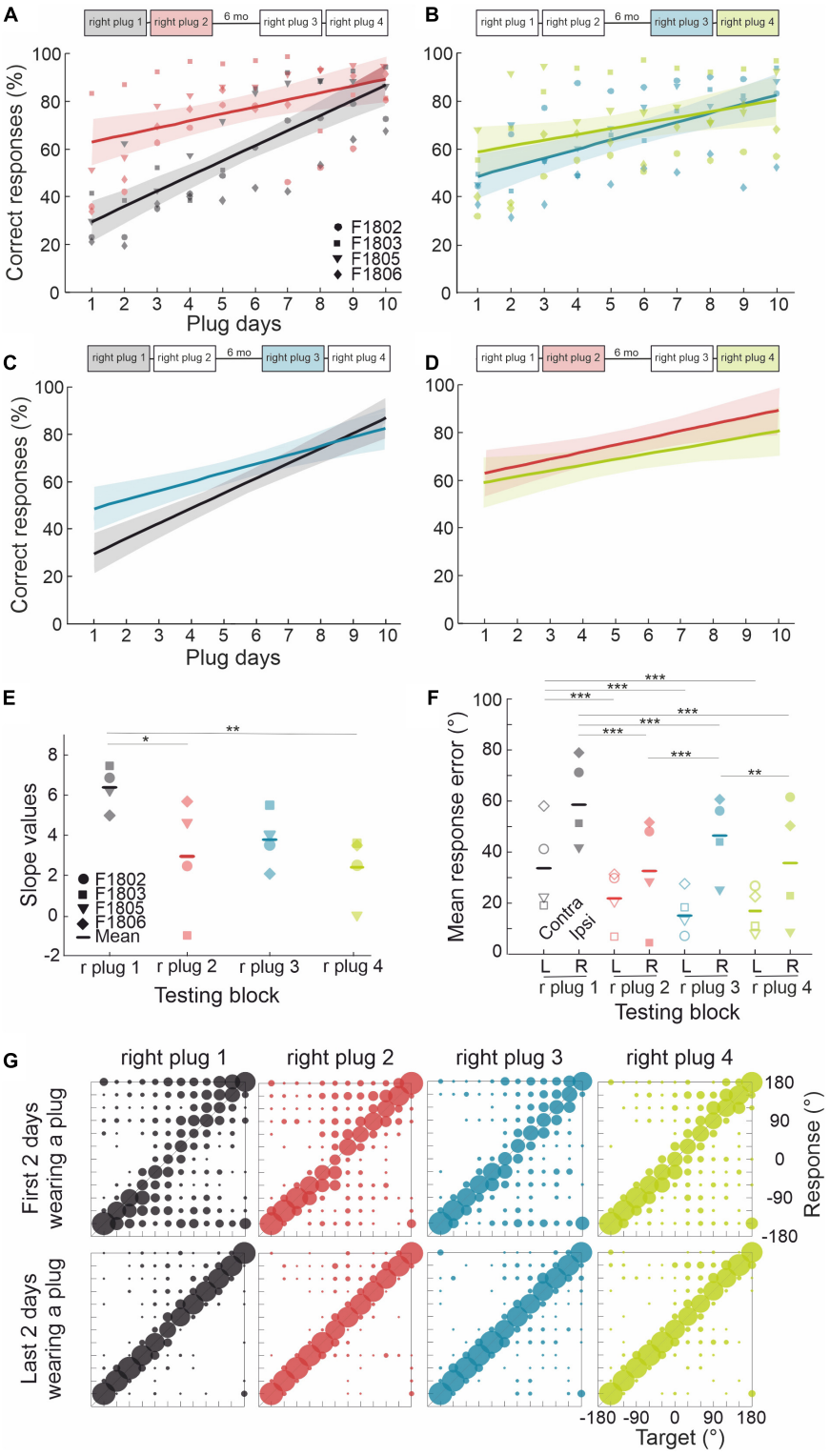



FIGURE 4
Effect of repeated occlusion of the same ear on adaptive learning. **(A)** Regression lines (shaded areas indicate 95% confidence intervals) and individual animal scores during the first (right plug 1, black line and gray symbols) and second (right plug 2, red line and symbols) periods of right ear occlusion, which were separated by 7 days of normal binaural hearing. **(B)** Regression lines and individual performance for the third (right plug 3, blue line and symbols) and fourth (right plug 4, green line and symbols) periods of right ear occlusion, which were also separated by 7 days of normal binaural hearing. A longer interval of 6 months separated the right plug 2 and right plug 3 runs. **(C)** Comparison of regression lines for the right plug 1 and right plug 3 runs, which were each preceded by long periods (~6 months) of normal binaural hearing. **(D)** Comparison of regression lines for the right plug 2 and right plug 4 runs, which were each preceded by 7 days of normal binaural hearing following the right plug 1 and right plug 3 runs, respectively. **(E)** Individual animal and mean regression slopes for all four periods of monaural occlusion. **(F)** Mean response error magnitude in the initial 2 days of each period of right ear occlusion (right plug 1 to 4). Errors are shown separately for ipsilateral (R, right) and contralateral (L, left) target locations. Midline locations (0° and 180°) were not included in this analysis. Asterisks represent statistical significance of paired comparisons (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$). **(G)** Response probabilities versus target location for the initial 2 days (top row) and the final 2 days (bottom row) of each of the four periods of right ear occlusion. For the numerical values and statistical analysis, refer to the main text.

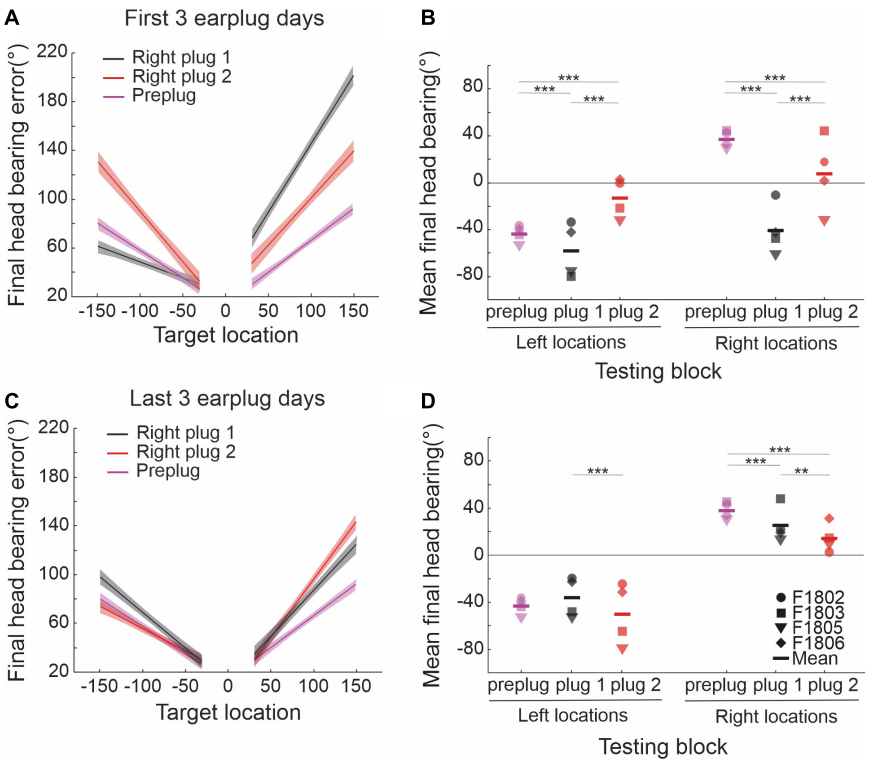



FIGURE 5
Persistence of adaptation in the sound-evoked head orienting responses. **(A)** Linear regression fits and their 95% confidence intervals for the mean final head bearing errors in the first 3 days of the right plug 1 and 2 runs (black and red lines, respectively). For comparison linear fits for preplug responses are shown in magenta. **(B)** Mean final head bearing shown separately for left and right target positions in the first 3 days of the right plug 1 and 2 runs. Midline locations (0° and 180°) were excluded. Note that the head orienting responses were no longer as biased toward the side of the open left ear at the start of the second period of right ear occlusion. **(C)** Linear regression fits and their 95% confidence intervals for the mean final head bearing errors in the pre-plug responses (magenta lines) and in the last 3 days of the right plug 1 and 2 runs (black and red lines, respectively). **(D)** Mean final head bearing shown separately for left and right target positions in the last 3 days of the right plug 1 and 2 runs. (**$p < 0.01$, ***$p < 0.001$).

TABLE 3 Parameters of the fitted regression lines for head bearing errors in Figure 5.

| | | y = a + bx | $R^2$ | a (intercept) | b (slope) | $p_b$ | tStat (b) |
|---|---|---|---|---|---|---|---|
| | Preplug | Left side | 0.14 | 12.84 | −0.449 | <0.0001 | $t_{(992)} = -12.946$ |
| | | Right side | 0.21 | 14.34 | 0.520 | <0.0001 | $t_{(977)} = 16.388$ |
| First 3 days | Right plug 1 | Left side | 0.05 | 22.69 | −0.258 | <0.0001 | $t_{(854)} = -7.102$ |
| | | Right side | 0.34 | 33.55 | 1.124 | <0.0001 | $t_{(833)} = 20.829$ |
| | Right plug 2 | Left side | 0.27 | 7.30 | −0.824 | <0.0001 | $t_{(494)} = -13.763$ |
| | | Right side | 0.24 | 23.19 | 0.779 | <0.0001 | $t_{(529)} = 13.058$ |
| Last 3 days | Right plug 1 | Left side | 0.22 | 10.38 | −0.587 | <0.0001 | $t_{(716)} = -14.299$ |
| | | Right side | 0.27 | 12.00 | 0.752 | <0.0001 | $t_{(641)} = 15.414$ |
| | Right plug 2 | Left side | 0.12 | 19.15 | −0.364 | <0.0001 | $t_{(849)} = -10.885$ |
| | | Right side | 0.35 | 1.21 | 0.950 | <0.0001 | $t_{(837)} = 21.099$ |

errors declined. This could be due to a more gradual remapping of the abnormal binaural cues to compensate for the effects of the earplug. Further support for this interpretation is provided by the observation that performance reaches ≥70% correct after a similar number of days irrespective of previous earplug experience and adaptation, the time elapsed between the periods of monaural occlusion, or the initial impact of plugging the ear on localization performance.

We have previously shown that the primary auditory cortex (A1) plays an essential role in adapting to an earplug in one ear, but is not required for retrieval of the memory trace that facilitates adaptation during a second period of monaural occlusion (Bajo et al., 2019). This raises the possibility that adaptive learning is consolidated in circuits that lie downstream from the cortex. The disruptive effects of auditory corticocollicular lesions on adaptation (Bajo et al., 2010) suggest that midbrain areas that are known to be more directly involved in the control of head orienting behavior may be important for the retrieval of learning. We therefore also examined the impact of a second period of monaural occlusion on the accuracy of the sound-evoked head movements.

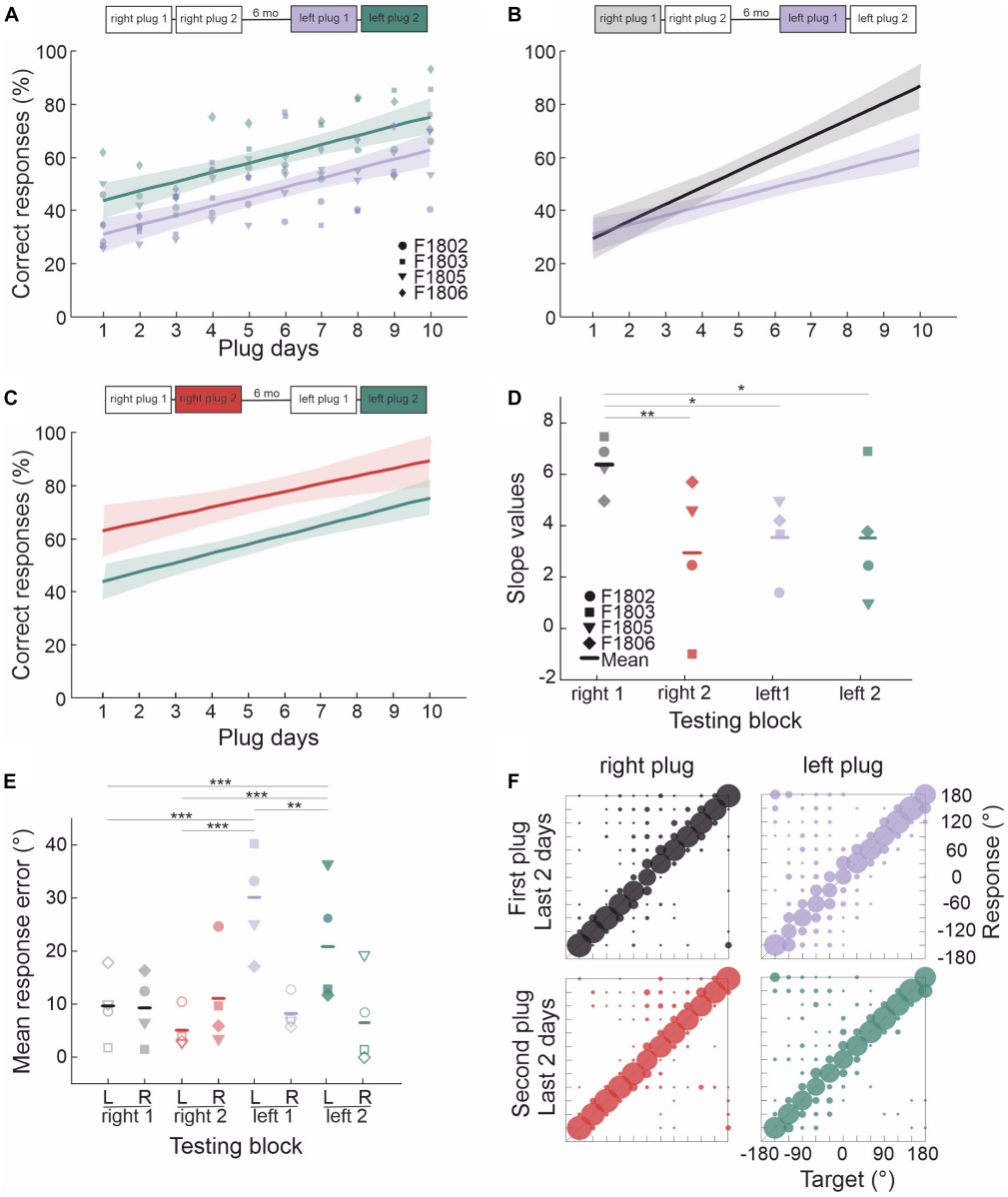



FIGURE 6
Effect of occluding the other ear following adaptation to unilateral conductive hearing loss. **(A)** Regression lines (shaded areas indicate 95% confidence intervals) and individual animal scores during the first (left plug 1, purple line and symbols) and second (left plug 2, green line and symbols) periods of left ear occlusion, which were separated by 7 days of normal binaural hearing. Unlike the right plug 1 and right plug 2 runs (Figure 4A) previously experienced by these animals, the slopes of the linear fits to these data were very similar. **(B)** Comparison of the regression lines for the right plug 1 (black/gray) and left plug 1 (purple) runs. Although the ferrets performed at the same level when each ear was first plugged, much less adaptation was seen when the hearing loss was experienced in the left ear following adaptation to right ear occlusion. **(C)** Comparison of the regression lines for the right plug 2 (red) and left plug 2 (green) runs. The slopes are very similar, but the difference in intercepts indicates less overall improvement in localization accuracy when the left ear was occluded. **(D)** Individual animal and mean regression slopes for the fits shown in panels **(A–C)**. **(E)** Mean response error magnitude in the final 2 days of each of these periods of monaural occlusion (right plug 1 and 2, left plug 1 and 2). Errors are shown separately for target locations ipsilateral (filled symbols) and contralateral (open symbols) to the side of the earplug. Midline locations (0° and 180°) were not included in this analysis. Asterisks represent statistical significance of paired comparisons (*$p < 0.05$, **$p < 0.01$, and ***$p < 0.001$). **(F)** Response probabilities versus target location for the final 2 days of the first right and left (top row) and the second right and left (bottom row) periods of monaural occlusion.

In contrast to the start of the first period of right ear occlusion (right plug 1), where a marked left-right asymmetry was observed in the size of the head orienting errors due to the animals consistently turning toward the side of the open ear, these responses were more symmetrically distributed to appropriate sides of the midline at the start of the second period of right ear occlusion (right plug 2) (Figures 5A, B and Table 3). In both cases, head orienting accuracy had partially recovered toward pre-plug values by the last 3 days of ear-plugging, with all the animals now responding to the appropriate side of space (Figures 5C, D). This pattern of adaptation in head orienting behavior across the two periods of right ear occlusion is consistent with that seen for the approach-to-target responses in showing that previous experience with altered spatial cues enables the ferrets to more readily adapt to a second period of unilateral conductive hearing loss in the same ear. The head orienting data therefore support the possibility that retrieval

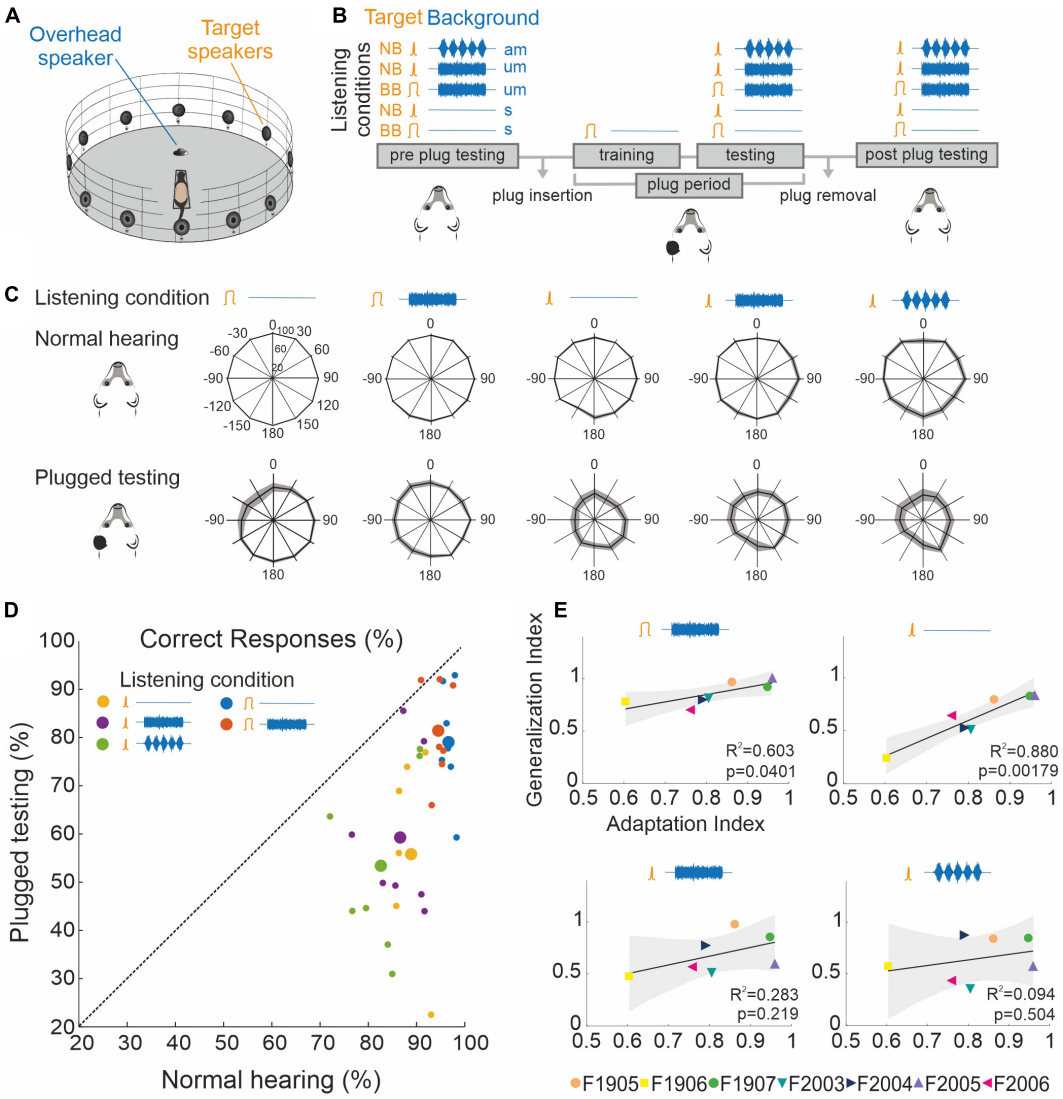



FIGURE 7
Generalization of training-dependent adaptation. **(A)** Schematic of the circular testing arena indicating the position of the peripheral target loudspeakers (as in Figure 1) and the loudspeaker located above the central platform used to generate background noise. **(B)** Timeline indicating the different combinations of target and background sounds used for behavioral training and testing. The two types of target sounds [broadband (BB) and narrowband (NB) noise] are depicted in orange and three possible backgrounds [silence (s), unmodulated BB noise (um), and BB noise amplitude modulated at 5 Hz (am)] are shown in blue. **(C)** Spatial distribution of mean (±S.E.) percent correct scores for each target and background combination under normal listening conditions (top row) and after adaptation to an earplug in the left ear (bottom row). **(D)** Mean percent correct scores for all target and background combinations under normal hearing conditions and after adaptation with one ear plugged. Large circles are the mean values and the small circles represent individual animals. Almost all values fall below the diagonal identity line, indicating more accurate responses with normal binaural inputs. Responses to BB targets were closer to the identity line and therefore showed less difference between the two hearing conditions. **(E)** Relationship between the adaptation index (the ratio of the percent correct scores for BB noise in silence in the final 3 days of adaptation to monaural occlusion and the preplug scores) and the generalization index [the ratio of the percent correct scores for each of the other stimulus types (shown at the top of the panels) measured with the earplug in place at the end of the adaptation run and the corresponding preplug scores].

of learning following a short period of normal hearing involves subcortical circuits.

## Adaptation to unilateral conductive hearing loss in the contralateral ear

We next examined whether ferrets that had learned to localize accurately following occlusion of one ear also exhibit similar adaptive plasticity when the direction of the binaural imbalance is reversed by plugging the other ear. The same group of ferrets that had experienced four periods of right ear occlusion now had their left ear plugged (Figure 6). This resulted in an initial drop in localization accuracy to ~30% correct, the same as that seen when the right ear was plugged for the first time (Figures 6A, B).

All the ferrets then exhibited a gradual improvement in performance with training, but they did so at a slower rate than in the equivalent first right ear plug run and recovered to a much lower level after 10 days of training with the left earplug in place (Figures 6B, D and Table 2). A further difference in localization accuracy between right plug 1 and left plug 1 was revealed by comparing the responses to sounds presented in the left and right hemifields. In the first left

TABLE 4 Generalized linear mixed model comparing different stimuli.

| | | | Normal hearing | | | | | After adaptation | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | BBs | BBum | NBs | NBum | NBam | BBs | BBum | NBs | NBum | NBam |
| Reference | Normal hearing | BBs | – | −0.171 | −0.610 | −0.728 | −0.915 | −1.080 | −0.924 | −1.740 | −1.613 | −1.799 |
| | | BBum | ** | – | −0.439 | −0.556 | −0.744 | −0.909 | −0.753 | −1.569 | −1.441 | −1.628 |
| | | NBs | *** | *** | – | −0.117 | −0.305 | −0.470 | −0.314 | −1.130 | −1.003 | −1.189 |
| | | NBum | *** | *** | ** | – | −0.188 | −0.352 | −0.196 | −1.012 | −0.885 | −1.071 |
| | | NBam | *** | *** | *** | *** | – | −0.165 | −0.009 | −0.825 | −0.698 | −0.884 |
| | After adaptation | BBs | *** | *** | *** | *** | *** | – | 0.156 | −0.660 | −0.533 | −0.719 |
| | | BBum | *** | *** | *** | *** | 0.816 | *** | – | −0.816 | −0.689 | −0.875 |
| | | NBs | *** | *** | *** | *** | *** | *** | *** | – | 0.127 | −0.059 |
| | | NBum | *** | *** | *** | *** | *** | *** | *** | *** | – | −0.186 |
| | | NBam | *** | *** | *** | *** | *** | *** | *** | 0.081 | *** | – |

Target: BB, broadband noise; NB, narrowband noise (one octave bandwidth, centered on 16 kHz). Background: s, silence; um, unmodulated noise; am, amplitude-modulated noise. $^{**}p < 0.01$ and $^{***}p < 0.001$; numbers are odds ratios.

TABLE 5 Sound localization accuracy for different stimuli.

| % Correct responses | | | | | |
|---|---|---|---|---|---|
| Normal hearing/After adaptation | | | | | |
| Ferret# | BBs | BBum | NBs | NBum | NBam |
| F1905 | 96.2/82.9 | 94.8/92.0 | 86.3/68.9 | 87.2/85.4 | 90.6/76.1 |
| F1906 | 98.2/59.2 | 95.2/74.4 | 93.0/22.5 | 91.6/44.0 | 76.6/44.0 |
| F1907 | 97.9/92.8 | 97.5/90.7 | 91.8/76.8 | 91.5/79.1 | 90.6/77.5 |
| F2003 | 96.4/77.7 | 94.6/78.0 | 91.0/47.5 | 90.9/47.5 | 84.9/31.0 |
| F2004 | 95.2/75.3 | 95.5/77.2 | 85.8/45.0 | 76.5/59.8 | 72.0/63.6 |
| F2005 | 95.4/91.6 | 91.0/91.8 | 88.0/73.9 | 83.0/49.8 | 79.5/44.6 |
| F2006 | 97.1/73.9 | 93.1/65.9 | 86.3/56.0 | 85.6/49.3 | 84.0/37.0 |

Target: BB, broadband noise; NB, narrowband noise (one octave bandwidth, centered on 16 kHz). Background: s, silence; um, unmodulated noise; am, amplitude-modulated noise.

earplug run, localization accuracy at the end of the adaptation period was significantly worse on the side of the plugged ear than on the side of the open ear (repeated measures ANOVA: effect of side, $p = 0.01$), whereas no difference between the left and right hemifields was found for the first right earplug run ($p = 0.139$) (Figures 6E, F). This is not because of any inherent difference in the capacity of the auditory system to adapt to conductive hearing loss in the two ears (Figure 3), nor is it likely to reflect any residual bias toward the side of the previously open ear since this declined as the ferrets adapted and disappeared when the earplug was removed. Rather, it appears to reflect the prior experience of the ferrets in repeatedly adapting to occlusion of the right ear.

A week after removing the earplug from the left ear, this ear was occluded for a second time (left plug 2). As with repeated plugging of the right ear, this resulted in a much smaller initial deficit than when the animals first received an earplug in the left ear (Figure 6A and Table 2), and their performance then improved with training at approximately the same rate as in left plug 1 (Figures 6A, D) and right plug 2 (Figures 6C, D). However, the animals localized less accurately at the end of the second period of left ear occlusion than after equivalent training with the right ear plugged (Figure 6C), due to their poorer performance on the left side, ipsilateral to the earplug, relative to the right side (repeated measures ANOVA, $p = 0.003$) (Figure 6E). These results therefore show that learning to localize sound with one ear occluded interferes with the capacity to subsequently adapt to conductive hearing loss in the other ear and that this effect persists over more than one period of monaural occlusion.

## Adaptation to unilateral conductive hearing loss generalizes beyond the training stimulus

To explore the extent to which the improvement in localization accuracy following adaptation to unilateral conductive hearing loss generalized beyond the standard broadband noise training paradigm, we tested the performance of a group of ferrets using a combination of different target sounds and acoustic backgrounds before and after they had been trained to adapt to an earplug. The target sounds were either broadband noise (BB), for which all acoustic spatial cues would have been available, or narrowband noise (NB) with a one-octave bandwidth centered on 16 kHz, selected to reduce the availability of binaural cues and to provide limited access to high-frequency spectral localization cues (Carlile, 1990; Keating et al., 2013). The target sounds were either presented in silence (s) or against a background of continuous broadband noise presented from an overhead speaker (Figure 7A), whose amplitude was either unmodulated (um) or modulated at 5 Hz (am) to add temporal structure to it (Figure 7B).

After measuring their localization accuracy with all combinations of targets and acoustic backgrounds (Figure 7C, Normal hearing), the animals were plugged in the left ear and trained as usual using 1,000-ms BB targets. Once they reached the adaptation criterion, the animals were retested on the above combination of target and acoustic backgrounds while still wearing the earplug (Figure 7C, Plugged testing).

Under normal hearing conditions and in a silent acoustic background, the ferrets localized the BB targets almost perfectly

TABLE 6 Adaptation and generalization indices.

| | Adaptation index | Generalization index | | | |
|---|---|---|---|---|---|
| Ferret# | BBs | BBum | NBs | NBum | NBam |
| F1905 | 0.86 | 0.97 | 0.80 | 0.98 | 0.84 |
| F1906 | 0.60 | 0.78 | 0.24 | 0.48 | 0.57 |
| F1907 | 0.95 | 0.93 | 0.84 | 0.86 | 0.85 |
| F2003 | 0.80 | 0.82 | 0.52 | 0.52 | 0.36 |
| F2004 | 0.79 | 0.81 | 0.52 | 0.78 | 0.88 |
| F2005 | 0.96 | 1.01 | 0.84 | 0.60 | 0.56 |
| F2006 | 0.76 | 0.71 | 0.65 | 0.57 | 0.44 |
| Mean ± SD | 0.82 ± 0.12 | 0.86 ± 0.11 | 0.63 ± 0.22 | 0.69 ± 0.19 | 0.64 ± 0.21 |

(96.4 ± 0.76%) and the NB target sounds slightly less well (88.9 ± 1.13%) (GLMM, $p < 0.001$, Table 4), particularly for posterior sound locations (Figure 7C, Normal hearing, Table 5). Reducing the bandwidth of the stimulus led to an increase in the incidence of both front-back errors (sounds presented in the frontal hemifield that were mislocalized into the posterior ipsilateral hemifield or vice versa; BB target, 0.94 ± 0.83%; NB target 1.99 ± 1.30%; paired sample *t*-test, df = 6, $p = 0.039$) and left-right errors (BB target, 0.23 ± 0.23%; NB target 0.81 ± 0.51%; paired sample *t*-test, df = 6, $p = 0.043$). Sound localization accuracy was also degraded to a small but significant extent in the presence of both types of background noise (GLMM, $p = 0.019$, Table 4).

Following training with one ear plugged, the ferrets showed the usual substantial recovery in their ability to localize BB noise bursts. Nevertheless, their sound localization accuracy across all stimuli was less accurate than that measured under normal hearing conditions prior to plugging one ear (Figures 7C, D and Table 4, GLMM, hearing condition, $p < 0.001$). This was particularly the case for the NB targets, with a greater difference in performance between BB and NB targets when one ear was occluded than prior to ear-plugging (GLMM, $p < 0.001$), as indicated by the greater distance to the identity line for the data from the NB targets in Figure 7D. The difference in localization performance with target bandwidth was also apparent from the incidence of front-back errors (BB target, 3.82 ± 2.92%; NB target 10.42 ± 6.22%; paired sample *t*-test, df = 6, $p = 0.002$) and left-right errors (BB target, 2.78 ± 2.43%; NB target 8.27 ± 6.19%; paired sample *t*-test, df = 6, $p = 0.011$). Following adaptation to monaural occlusion, the ferrets localized NB targets significantly worse on the side of the plugged ear (GLMM effect of target location, $p < 0.001$; odds ratio −0.385 relative to the best localized locations) (Figure 7C, Plugged testing).

Surprisingly, we found an improvement in performance in the presence of background noise for BB target sounds after the animals had adapted to the earplug (GLMM, $p < 0.001$, odds ratio = 0.156) (Figures 7C, D). This is also reflected by the greater value of the generalization index, the mean performance following adaptation relative to that obtained under normal hearing conditions (Figure 7E and Table 6). We observed a significant correlation between the degree of adaptation and generalization: ferrets that showed more complete recovery in their performance with training also localized BB targets against background noise more accurately, indicating that they were able to effectively segregate these sound sources despite the presence of a conductive hearing loss in one ear.

Greater inter-animal variability in performance was found with NB than BB target sounds (Figures 7C–E and Table 6). Ferrets that adapted more to the earplug when trained with BB sounds also localized NB targets in silence more accurately. Although they localized less accurately than with BB noise (Figures 7C, D), the significant correlation between the adaptation index and the generalization index (Figure 7E) suggests that sufficient spatial information that formed the basis for adapting to unilateral conductive hearing loss remained in these NB stimuli. The addition of background noise weakened this relationship, particularly when it was amplitude modulated, suggesting that background noise has a more disruptive effect on the transfer of learning when more limited spatial cues are provided by the target stimuli.

## Role of spectral cues and dynamic cues in adaptation to unilateral conductive hearing loss

The lack of an aftereffect following earplug removal is consistent with a primary adaptation strategy to unilateral conductive hearing loss in which the animals learn to rely more on the intact monaural spectral cues provided by the non-occluded ear than the normally dominant but now distorted binaural cues. However, adaptive changes in binaural cue sensitivity appear to contribute as well, as shown by the recovery in localization accuracy observed in humans when pure tones are interspersed during training with one ear occluded (Keating et al., 2016). Our finding that adaptation of sound localization in the horizontal plane is initially restricted to the side of the open ear, with improvements in performance on the plugged side occurring later during the period of monaural occlusion also supports the possibility that more than one process may underlie this training-dependent plasticity.

An additional factor that needs to be considered is that the relatively long training stimulus (1,000 ms noise bursts) may have allowed the animals to sample dynamic spatial cues as they were approaching the perceived sound source location. The very similar pattern of adaptation observed in the head orienting responses (Figure 2), which have a latency of ∼100–200 ms (Nodal et al., 2008, 2012), strongly suggests that adaptation relies primarily on the spatial information available at the onset of the target sound. Nevertheless, the extent of the adaptation has been shown to increase with the stimulus duration (Kacelnik et al., 2006), which could reflect greater integration time or multiple sampling of the available localization cues.

To gain a better understanding of the processes underlying training-dependent adaptation to unilateral conductive hearing loss, we interspersed sessions in which high-frequency narrowband sounds [1/6 octave bandwidth centered on 15 kHz, as in Keating et al. (2015)] were presented during the adaptation training with broadband noise. These sounds were designed to eliminate monaural spectral cues and ITDs, leaving only narrowband ILDs available. In addition, the stimuli in these trials were either 200 ms or 1,000 ms in duration to determine whether more adaptation occurred with the longer duration sounds.

As expected, the pre-plug scores showed that the ferrets localized high-frequency narrowband targets less accurately than the broadband sounds at stimulus durations of both 1,000 ms (Figure 8A) and 200 ms (Figure 8B). Including different stimulus

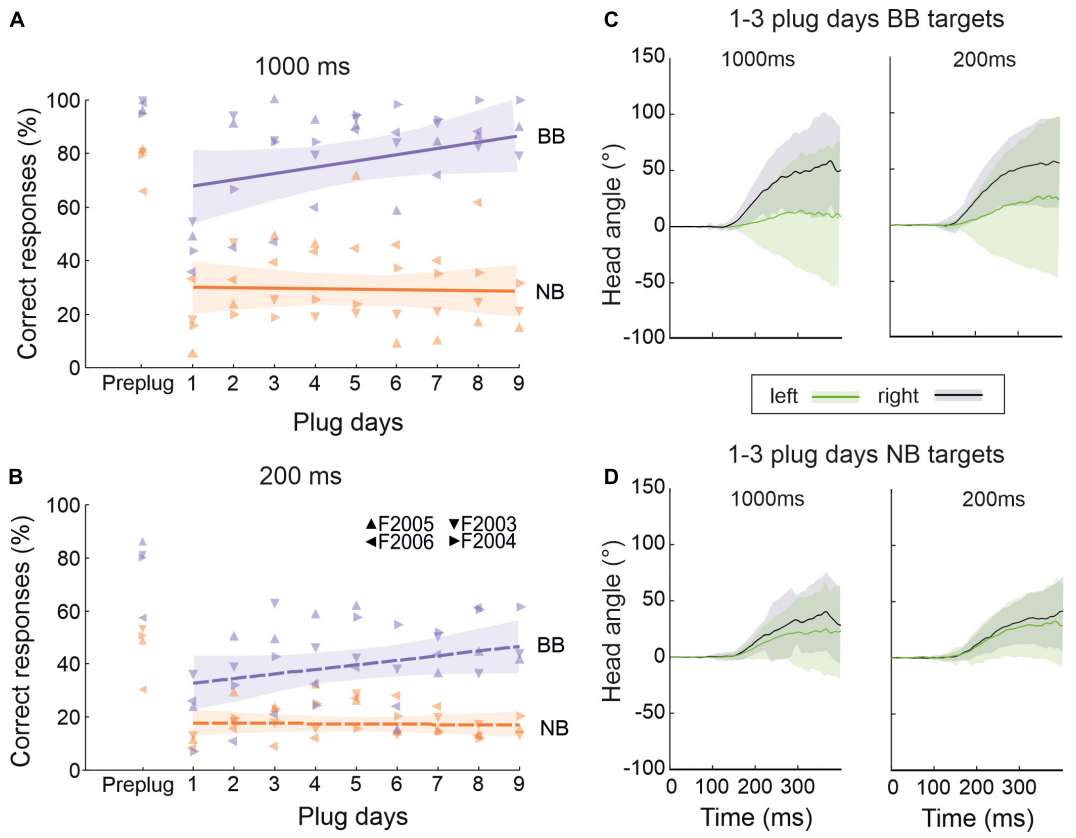



FIGURE 8
Role of dynamic and spectral cues in adaptation to unilateral conductive hearing loss. **(A)** Percentage of correct responses for 1,000-ms broadband noise (purple, BB) and 1/6 octave narrowband noise bursts with a center frequency of 15 kHz (orange, NB) over the course of daily training with one ear occluded. Colored lines are linear regressions, and the shaded areas correspond to the 95% confidence intervals. The symbols represent individual animals. **(B)** Same as panel **(A)** but for stimuli with a duration of 200 ms. **(C)** Change in the mean (±S.E.) horizontal angle of the head over time following BB target onset at time 0 ms. These measurements were taken at the start (days 1–3) of the ear-plugging run, combined across target locations in the left *(green)* and right *(black)* hemifields, and are plotted separately for 1000-ms (left) and 200-ms (right) noise bursts. **(D)** Same as panel **(C)** but for NB target sounds.

types and durations within the training block did not interfere with the expected adaptation to BB noise stimuli, with the percentage correct scores increasing throughout the period of monaural occlusion for both target durations (GLMM testing day, $p < 0.001$) (Figures 8A, B). By contrast, no improvement in performance was seen over time for NB targets that had a duration of either 1,000 ms (GLMM, testing day, $p = 0.071$) (Figure 8A) or 200 ms (GLMM, testing day, $p = 0.341$) (Figure 8B).

The sound-evoked head movements measured for these animals (combined across target locations in each hemifield for the first 3 days of monaural occlusion and plotted separately for BB and NB stimuli in Figures 8C, D, respectively) were very similar at each of the two stimulus durations used. These averaged measurements demonstrate that the head began to move near the end of the 200-ms noise bursts, and this movement continued during the longer stimulus until the animal left the central start platform [which typically occurs at ~600 ms after stimulus onset (Kacelnik et al., 2006; Nodal et al., 2008)]. Although the metrics of the head turns were little affected by stimulus duration, they differed with the bandwidth of the stimulus in a similar way to the approach-to-target response accuracy shown in Figures 8A, B.

These results therefore indicate that that 1/6-octave bandwidth noise bursts centered at 15 kHz do not contain sufficient spatial information to support the adaptive improvements in localization performance that take place during the training runs with BB noise. Furthermore, localization accuracy was reduced for both BB (GLMM, target duration, $p < 0.001$) and NB (GLMM, target duration, $p < 0.001$) stimuli when the duration used was too short for the animals to be able to use dynamic cues to any degree, suggesting that the ferrets may benefit from re-sampling the longer stimuli during and after the initial head turn. Nevertheless, some adaptation still took place with 200-ms noise bursts so long as they were sufficiently broadband to provide access to other auditory spatial cues (Figures 8A, B). Adaptive learning does not therefore depend on head movements, though it is possible that they enhance this process.

## Discussion

Perceptual learning enables auditory perceptual skills to be improved with practice or training (Irvine, 2018), and allows abnormal inputs to be accommodated in ways that are likely to be important in the treatment of hearing disorders (Kumpik and King, 2019; Glennon et al., 2020). Plasticity in the processing of auditory spatial information is essential for calibrating neural circuits through experience with the available cues, which vary with individual

differences in the geometry of the head and ears and change over time, particularly as these structures grow but potentially also following hearing loss. While this plasticity is most pronounced during development, many studies have shown that the adult brain can learn to utilize abnormal auditory space cues (Carlile, 2014; Mendonça, 2014). We used a well-established paradigm for investigating the adaptive capabilities of the auditory system by temporarily plugging one ear in order to induce a reversible binaural imbalance, so that we could explore the retention and retrieval of learning and its generalization beyond the standard stimuli conditions used for training.

The earplugs we used disrupt binaural localization cues by attenuating and delaying the sound at the occluded ear (Moore et al., 1989; Keating et al., 2013, 2016). The attenuation is frequency dependent and will therefore also distort the spectral cues available at the occluded ear, leaving the spectral filtering provided by the open ear as the only unchanged acoustic cue. Following an immediate decline in localization accuracy in the horizontal plane, we found that performance steadily improved with daily training toward the pre-plug scores, as shown in previous studies in ferrets (Kacelnik et al., 2006; Bajo et al., 2010, 2019; Nodal et al., 2010, 2012) and humans (Kumpik et al., 2010; Keating et al., 2016). Although an improvement in localization accuracy in monaurally plugged human subjects has been observed over the course of a single training session when the sound level was fixed and visually guided feedback provided (Zonooz and Van Opstal, 2019), other work in humans has shown that training has to be spread out over several days rather than compressed into a single block on 1 day in order for adaptation to occur (Kumpik et al., 2010). While this suggests that learning to localize sound after disrupting the cues in this fashion requires a period of consolidation between training sessions, daily training in monaurally occluded ferrets has been shown to produce faster and more complete adaptation than in animals that were trained every sixth day and which therefore experienced altered spatial cues over a longer period (Kacelnik et al., 2006). Related to this is the finding that human participants who have one ear plugged only during sound-localization training sessions that took place approximately every 3 days still show a gradual improvement in performance from one session to the next despite receiving normal binaural inputs in the intervening periods (Keating et al., 2016). These findings highlight the key role of training in the rapid adaptation to altered spatial cues, confirming that it represents a form of perceptual learning.

Compensatory changes in binaural cue sensitivity have been reported in adult humans who adapt to the presence of an earplug in one ear (Keating et al., 2016; Zonooz and Van Opstal, 2019) and have also been observed at a neurophysiological level in animals following several weeks of monaural occlusion during development (Keating et al., 2015) or in adulthood (Thornton et al., 2021). Nevertheless, the absence of an after-effect in the localization responses following earplug removal suggests that the auditory system primarily adapts by giving greater weight to the unchanged monaural spectral cues provided by the open ear. This conclusion is supported by the immediate reduction in percentage correct scores produced by inserting a mold in order to reshape the external ear contralateral to the earplug (Kacelnik et al., 2006) and by the greatly reduced adaptation observed when spectral cues were disrupted during training by randomizing the spectrum of the sounds (Keating et al., 2016). Monaural spectral cues also appear to contribute to the ability to localize sounds in the horizontal plane of at least some humans who are deaf in one ear (Slattery and Middlebrooks, 1994; Van Wanrooij and Van Opstal, 2004; Agterberg et al., 2014). Our results provide additional evidence for the importance of spectral cues in adaptation to unilateral hearing loss. While the ferrets exhibited an improvement in accuracy when trained to localize broadband noise stimuli, which provide access to the full range of acoustic spatial cues, no improvements were seen with narrowband sounds (1/6-octave bandwidth noise bursts centered at 15 kHz) presented during the training to restrict access to high-frequency interaural level differences.

Following earplug removal, the auditory system switches back to relying more on binaural cues for azimuthal localization. This is shown by our finding that subsequently replugging the same ear again degrades localization accuracy (see also Kacelnik et al., 2006; Bajo et al., 2019). A capacity to alternate between different sets of cues has also been demonstrated in adult humans who experience altered spectral cues by wearing ear molds for several weeks (Hofman et al., 1998; Van Wanrooij and Van Opstal, 2005; Carlile and Blackman, 2014; Trapeau et al., 2016; Watson et al., 2017). These manipulations impaired elevation responses, which then recovered as the participants learned to localize with their new ears, but when the molds were removed, restoring normal spectral cues, they were immediately able to localize as accurately as they did prior to insertion of the molds. Furthermore, upon reinsertion of the molds approximately 1 week later, localization performance was found to be no different from that measured at the end of the previous accommodation period (Carlile and Blackman, 2014). This suggests that after learning to localize with the molds in place, humans are able to interchangeably use their original own-ear spectral cues and the abnormal spectral cues they have adapted to. Brief periods of training with inappropriate spectral cues provided by using virtual auditory displays with non-individualized head-related transfer functions also result in improvements in localization performance that have been shown to persist for at least several weeks (Zahorik et al., 2006; Mendonça et al., 2013).

The effects of occluding one ear are different, however, with our results showing that additional training is required to achieve the same level of adapted performance whenever the same ear was re-plugged, with the magnitude of the initial drop in performance positively correlated with the duration of prior normal experience and greatest at the start of the first ear-plugging run. Thus, although it takes longer for the brain to adjust to the altered spatial cues resulting from monaural occlusion than to revert to normal inputs, previous ear-plugging experience leaves a "memory trace" so that less adaptation is required when the cues are changed again in the same manner. This was particularly the case for locations ipsilateral to the open ear, suggesting that upweighting of monaural spectral cues is facilitated by prior experience.

Switching the unilateral conductive hearing loss to the contralateral ear also resulted in an impairment in localization accuracy that partially recovered with training. However, the localization performance of the ferrets improved to a significantly lesser extent than when the hearing loss was previously experienced on the other side. This is consistent with evidence that ferrets raised with one ear occluded adapt by changing the relative weighting of the spectral cues in opposite directions for each ear (Keating et al., 2013). If this also happens when the animals adapt to a unilateral conductive hearing loss in adulthood, these opposing changes in cue weighting would then need to be reversed when the earplug is switched to the other ear. Alternatively, any remapping of binaural cues that accompanies adaptation to monaural conductive hearing loss, and

which may account for the more gradual improvements observed at frontal regions ipsilateral to the occluded ear, is also likely to hinder the capacity of the brain to recover localization accuracy when the side of the hearing loss is changed.

In these experiments, we measured the accuracy of the ferrets' head orienting responses following sound presentation as well as the target location they approached to receive a reward. Our previous work has shown that restricted lesions (Nodal et al., 2010) or reversible deactivation (Nodal et al., 2012) of the auditory cortex impair approach-to-target behavior, whilst preserving the accuracy of the initial sound-evoked head orienting responses, which are likely to be more dependent on midbrain circuits (Lomber et al., 2001; Isa et al., 2021). Nevertheless, the initial head turn and subsequent selection of target location are both components of the animals' behavioral response to sounds presented from different directions and are dependent on the same spatial cues. Indeed, we have previously reported that when ferrets with normal hearing approach an incorrect target location, the preceding head turn is more closely correlated with that response than with the target location (Nodal et al., 2008).

We observed a very similar pattern of training-dependent plasticity for each measure of sound localization, including the partial retention of learning when abnormal spatial cues are again experienced by replugging the same ear. Since silencing the primary auditory cortex impairs adaptation but not the retrieval of previous learning (Bajo et al., 2019), it is likely that the latter involves downstream circuits. Although we cannot rule out the involvement of other brain areas, our previous demonstration that the descending auditory corticocollicular pathway is essential for adaptive plasticity of both head-orienting and approach-to-target behavior (Bajo et al., 2010) suggests that the midbrain may be involved in the retention of learning. Future recording studies will be required to identify exactly where and how spatial information is stored to allow the partial retrieval of localization accuracy from one period of monaural occlusion to the next. It is interesting to note, however, that these findings are consistent with the reverse hierarchy theory of perceptual learning (Ahissar and Hochstein, 2004), whereby learning proceeds in a top-down fashion from higher to lower levels of processing, where the relevant stimulus features are represented more precisely.

Adaptation of the head-orienting responses also allowed us to address the potential contribution of sensorimotor behavior to auditory localization plasticity (Zahorik et al., 2006; Aytekin et al., 2008; Parseihian and Katz, 2012; Carlile et al., 2014). Head turning accuracy is likely to depend on the spatial cues available at the onset of the target sound and is not affected by its duration (Nodal et al., 2012). Our finding that adaptive changes in head-orienting responses during initial and subsequent periods of monaural occlusion mirrored those seen for the approach-to-target behavior suggests that head movements are not necessary for training-dependent plasticity to occur. This is supported by our finding that approach-to-target accuracy showed some recovery with one ear occluded when the duration of the broadband noise bursts was reduced to 200 ms, which is close to the head-orienting latency. Partial adaptation to unilateral conductive hearing loss has also been observed in ferrets for 40-ms broadband noise bursts (Kacelnik et al., 2006), and human participants wearing a plug in one ear can re-learn to correctly identify the target loudspeaker location without first turning toward it (Kumpik et al., 2010).

Although these findings indicate that head movements do not play an essential role in adjusting auditory localization to compensate for a conductive hearing loss in one ear, the recovery in accuracy is more complete when longer duration noise bursts are used for training (see also Kacelnik et al., 2006), suggesting that movement of the head may benefit adaptation by allowing multiple sampling of the stimulus. No improvement was observed, however, when 1/6-octave narrowband noise bursts of either 200 or 1,000 ms duration were presented during the training, suggesting that even when head movements during sound presentation were possible, these stimuli provide insufficient spatial information to provide a confident assessment of target location and therefore support learning. While this highlights the importance for adaptation to unilateral conductive hearing loss of using a broadband stimulus so that the full range of localization cues is available, it seems likely that further training with narrowband sounds would eventually lead to a remapping of high-frequency ILDs (Keating et al., 2015, 2016).

Investigating the generalization of perceptual learning to untrained stimulus properties can provide valuable insights into the neural substrates of learning and retrieval (Wright and Zhang, 2009; Dosher et al., 2013). In terms of adaptation to unilateral conductive hearing loss, generalization is clearly important if this form of training-dependent plasticity is to be relevant clinically to users of cochlear implants and hearing aids. In this respect, previous studies of adaptation to abnormal spatial cues in humans are promising in demonstrating that improvements in performance do transfer to untrained sound types (Mendonça et al., 2013; Watson et al., 2017) and locations (Zahorik et al., 2006; Mendonça et al., 2013; Watson et al., 2017; Zonooz and Van Opstal, 2019), and to reverberant conditions (Watson et al., 2017). Zonooz and Van Opstal (2019) reported that while adaptation to monaural ear-plugging leads to a generalized increase in azimuth accuracy, this was accompanied by degraded elevation responses, though individuals with single-sided deafness can benefit from monaural spectral cues for localization in both dimensions (Van Wanrooij and Van Opstal, 2004).

Our results show that the adaptation achieved after training ferrets to localize broadband noise with one ear occluded was positively correlated with their ability to generalize this performance improvement to untrained octave band noise. These narrowband stimuli would have provided high-frequency binaural cues and a subset of the directional spectral features produced by the filter properties of the head and ears (Keating et al., 2013). While these cues were sufficient to support the generalization of adaptation when the narrowband target sounds were presented in silence, the addition of background noise degraded the performance of the ferrets both before and after learning. However, when broadband target sounds were used, the spatial learning resulting from training ferrets with one ear plugged in quiet conditions was preserved when the animals were subsequently tested in a noisy environment. Indeed, following adaptation to monaural occlusion, they achieved significantly higher scores when background noise was added than in a silent background, raising the possibility that localization training may enhance the perception of targets in noise when hearing is perturbed. This indicates that the benefits of daily training for hearing-impaired individuals may translate to listening conditions that more closely resemble those of real-world environments.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The animal study was reviewed and approved by the Committee on Animal Care and Ethical Review, University of Oxford.

## Author contributions

AK and FN conceived and designed the study, with contributions from AS and VB. AS, FN, and VB acquired the data. AS led the data analysis, with contributions from KW and FN. AS, FN, VB, and AK wrote the manuscript. All authors contributed to results interpretation and final manuscript preparation.

## Funding

This research was supported by Wellcome through a Principal Research Fellowship (WT108369/Z/2015/Z) to AK and by a Ph.D. Studentship Grant from the Royal National Institute for Deaf People to AS (S55_King).

## Acknowledgments

We are grateful to Susan Spires for her expert assistance with the behavioral testing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Agterberg, M. J., Hol, M. K., Van Wanrooij, M. M., Van Opstal, A. J., and Snik, A. F. (2014). Single-sided deafness and directional hearing: contribution of spectral cues and high-frequency hearing loss in the hearing ear. *Front. Neurosci.* 8:188. doi: 10.3389/fnins.2014.00188

Ahissar, M., and Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends Cogn. Sci.* 8, 457–464. doi: 10.1016/j.tics.2004.08.011

Aytekin, M., Moss, C. F., and Simon, J. Z. (2008). A sensorimotor approach to sound localization. *Neural Comput.* 20, 603–635. doi: 10.1162/neco.2007.12-05-094

Bajo, V. M., Nodal, F. R., Korn, C., Constantinescu, A. O., Mann, E. O., Boyden, E. S. III, et al. (2019). Silencing cortical activity during sound-localization training impairs auditory perceptual learning. *Nat. Commun.* 10:3075. doi: 10.1038/s41467-019-10770-4

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Carlile, S. (1990). The auditory periphery of the ferret. I: directional response properties and the pattern of interaural level differences. *J. Acoust. Soc. Am.* 88, 2180–2195. doi: 10.1121/1.400115

Carlile, S. (2014). The plastic ear and perceptual relearning in auditory spatial perception. *Front. Neurosci.* 8:237. doi: 10.3389/fnins.2014.00237

Carlile, S., and Blackman, T. (2014). Relearning auditory spectral cues for locations inside and outside the visual field. *J. Assoc. Res. Otolaryngol.* 15, 249–263. doi: 10.1007/s10162-013-0429-5

Carlile, S., Balachandar, K., and Kelly, H. (2014). Accommodating to new ears: the effects of sensory and sensory-motor feedback. *J. Acoust. Soc. Am.* 135, 2002–2011. doi: 10.1121/1.4868369

Dosher, B. A., Jeter, P., Liu, J., and Lu, Z. L. (2013). An integrated reweighting theory of perceptual learning. *Proc. Natl. Acad. Sci. U.S.A.* 110, 13678–13683. doi: 10.1073/pnas.1312552110

Glennon, E., Svirsky, M. A., and Froemke, R. C. (2020). Auditory cortical plasticity in cochlear implant users. *Curr. Opin. Neurobiol.* 60, 108–114. doi: 10.1016/j.conb.2019.11.003

Grothe, B., Pecka, M., and McAlpine, D. (2010). Mechanisms of sound localization in mammals. *Physiol. Rev.* 90, 983–1012. doi: 10.1152/physrev.00026.2009

Hofman, P. M., Van Riswick, J. G., and Van Opstal, A. J. (1998). Relearning sound localization with new ears. *Nat. Neurosci.* 1, 417–421. doi: 10.1038/1633

Hogan, S. C., Stratford, K. J., and Moore, D. R. (1997). Duration and recurrence of otitis media with effusion in children from birth to 3 years: prospective study using monthly otoscopy and tympanometry. *BMJ* 314, 350–353. doi: 10.1136/bmj.314.7077.350

Irvine, D. R. F. (2018). Auditory perceptual learning and changes in the conceptualization of auditory cortex. *Hear. Res.* 366, 3–16. doi: 10.1016/j.heares.2018.03.011

Isa, T., Marquez-Legorreta, E., Grillner, S., and Scott, E. K. (2021). The tectum/superior colliculus as the vertebrate solution for spatial sensory integration and action. *Curr. Biol.* 31:R741–R762. doi: 10.1016/j.cub.2021.04.001

Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J. (2006). Training-induced plasticity of auditory localization in adult mammals. *PLoS Biol.* 4:e71. doi: 10.1371/journal.pbio.0040071

Keating, P., and King, A. J. (2013). Developmental plasticity of spatial hearing following asymmetric hearing loss: context-dependent cue integration and its clinical implications. *Front. Syst. Neurosci.* 7:123. doi: 10.3389/fnsys.2013.00123

Keating, P., and King, A. J. (2015). Sound localization in a changing world. *Curr. Opin. Neurobiol.* 35, 35–43. doi: 10.1016/j.conb.2015.06.005

Keating, P., Dahmen, J. C., and King, A. J. (2013). Context-specific reweighting of auditory spatial cues following altered experience during development. *Curr. Biol.* 23, 1291–1299. doi: 10.1016/j.cub.2013.05.045

Keating, P., Dahmen, J. C., and King, A. J. (2015). Complementary adaptive processes contribute to the developmental plasticity of spatial hearing. *Nat. Neurosci.* 18, 185–187. doi: 10.1038/nn.3914

Keating, P., Rosenior-Patten, O., Dahmen, J. C., Bell, O., and King, A. J. (2016). Behavioral training promotes multiple adaptive processes following acute hearing loss. *eLife* 5:e12264. doi: 10.7554/eLife.12264

Kumpik, D. P., and King, A. J. (2019). A review of the effects of unilateral hearing loss on spatial hearing. *Hear. Res.* 372, 17–28. doi: 10.1016/j.heares.2018.08.003

Kumpik, D. P., Kacelnik, O., and King, A. J. (2010). Adaptive reweighting of auditory localization cues in response to chronic unilateral earplugging in humans. *J. Neurosci.* 30, 4883–4894. doi: 10.1523/JNEUROSCI.5488-09.2010

Lomber, S. G., Payne, B. R., and Cornwell, P. (2001). Role of the superior colliculus in analyses of space: superficial and intermediate layer contributions to visual orienting, auditory orienting, and visuospatial discriminations during unilateral and bilateral deactivations. *J. Comp. Neurol.* 441, 44–57. doi: 10.1002/cne.1396

Mendonça, C. (2014). A review on auditory space adaptations to altered head-related cues. *Front. Neurosci.* 8:219. doi: 10.3389/fnins.2014.00219

Mendonça, C., Campos, G., Dias, P., and Santos, J. A. (2013). Learning auditory space: generalization and long-term effects. *PloS One* 8:e77900. doi: 10.1371/journal.pone.0077900

Moore, D. R., Hutchings, M. E., King, A. J., and Kowalchuk, N. E. (1989). Auditory brain stem of the ferret: some effects of rearing with a unilateral ear plug on the cochlea, cochlear nucleus, and projections to the inferior colliculus. *J. Neurosci.* 9, 1213–1222. doi: 10.1523/JNEUROSCI.09-04-01213.1989

Nodal, F. R., Bajo, V. M., and King, A. J. (2012). Plasticity of spatial hearing: behavioural effects of cortical inactivation. *J. Physiol.* 590, 3965–3986. doi: 10.1113/jphysiol.2011.222828

Nodal, F. R., Bajo, V. M., Parsons, C. H., Schnupp, J. W., and King, A. J. (2008). Sound localization behavior in ferrets: comparison of acoustic orientation and approach-to-target responses. *Neuroscience* 154, 397–408. doi: 10.1016/j.neuroscience.2007.12.022

Nodal, F. R., Kacelnik, O., Bajo, V. M., Bizley, J. K., Moore, D. R., and King, A. J. (2010). Lesions of the auditory cortex impair azimuthal sound localization and its recalibration in ferrets. *J. Neurophysiol.* 103, 1209–1225. doi: 10.1152/jn.00991.2009

Parseihian, G., and Katz, B. F. (2012). Rapid head-related transfer function adaptation using a virtual auditory environment. *J. Acoust. Soc. Am.* 131, 2948–2957. doi: 10.1121/1.3687448

Slattery, W. H. III, and Middlebrooks, J. C. (1994). Monaural sound localization: acute versus chronic unilateral impairment. *Hear. Res.* 75, 38–46. doi: 10.1016/0378-5955(94)90053-1

Thompson, G. C., and Masterton, R. B. (1978). Brain stem auditory pathways involved in reflexive head orientation to sound. *J. Neurophysiol.* 41, 1183–1202. doi: 10.1152/jn.1978.41.5.1183

Thornton, J. L., Anbuhl, K. L., and Tollin, D. J. (2021). Temporary unilateral hearing loss impairs spatial auditory information processing in neurons in the central auditory system. *Front. Neurosci.* 15:721922. doi: 10.3389/fnins.2021.721922

Trapeau, R., Aubrais, V., and Schönwiesner, M. (2016). Fast and persistent adaptation to new spectral cues for sound localization suggests a many-to-one mapping mechanism. *J. Acoust. Soc. Am.* 140:879. doi: 10.1121/1.4960568

Van Wanrooij, M. M., and Van Opstal, A. J. (2004). Contribution of head shadow and pinna cues to chronic monaural sound localization. *J. Neurosci.* 24, 4163–4171. doi: 10.1523/JNEUROSCI.0048-04.2004

Van Wanrooij, M. M., and Van Opstal, A. J. (2005). Relearning sound localization with a new ear. *J. Neurosci.* 25, 5413–5424. doi: 10.1523/JNEUROSCI.0850-05.2005

Van Wanrooij, M. M., and Van Opstal, A. J. (2007). Sound localization under perturbed binaural hearing. *J. Neurophysiol.* 97, 715–726. doi: 10.1152/jn.00260.2006

Watson, C., Carlile, S., Kelly, H., and Balachandar, K. (2017). The generalization of auditory accommodation to altered spectral cues. *Sci. Rep.* 7:11588. doi: 10.1038/s41598-017-11981-9

Wright, B. A., and Zhang, Y. (2009). A review of the generalization of auditory learning. *Phil. Trans. R. Soc. Lond. B Biol. Sci.* 364, 301–311. doi: 10.1098/rstb.2008.0262

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C. (2006). Perceptual recalibration in human sound localization: learning to remediate front-back reversals. *J. Acoust. Soc. Am.* 120, 343–359. doi: 10.1121/1.2208429

Zonooz, B., and Van Opstal, A. J. (2019). Differential adaptation in azimuth and elevation to acute monaural spatial hearing after training with visual feedback. *eNeuro* 6:ENEURO.219–ENEURO.219. doi: 10.1523/ENEURO.0219-19.2019

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 07 February 2023
DOI 10.3389/fnins.2023.1077455



OPEN ACCESS

EDITED BY
Huiming Zhang,
University of Windsor, Canada

REVIEWED BY
Hyunjoon Shim,
Eulji University, Republic of Korea
Shouqin Zhao,
Capital Medical University, China

*CORRESPONDENCE
Shusheng Gong
✉ gongss@ccmu.edu.cn
Ke Liu
✉ liuke@ccmu.edu.cn

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 23 October 2022
ACCEPTED 23 January 2023
PUBLISHED 07 February 2023

CITATION
Long Y, Wang W, Liu J, Liu K and Gong S (2023) The interference of tinnitus on sound localization was related to the type of stimulus.
*Front. Neurosci.* 17:1077455.
doi: 10.3389/fnins.2023.1077455



# The interference of tinnitus on sound localization was related to the type of stimulus

Yue Long[1,2], Wei Wang[1], Jiao Liu[1], Ke Liu[1]* and Shusheng Gong[1,2]*

[1]Department of Otolaryngology-Head and Neck Surgery, Beijing Friendship Hospital, Capital Medical University, Beijing, China, [2]Clinical Center for Hearing Loss, Capital Medical University, Beijing, China

Spatial processing is a major cognitive function of hearing. Sound source localization is an intuitive evaluation of spatial hearing. Current evidence of the effect of tinnitus on sound source localization remains limited. The present study aimed to investigate whether tinnitus affects the ability to localize sound in participants with normal hearing and whether the effect is related to the type of stimulus. Overall, 40 participants with tinnitus and another 40 control participants without tinnitus were evaluated. The sound source discrimination tasks were performed on the horizontal plane. Pure tone (PT, with single frequency) and monosyllable (MS, with spectrum information) were used as stimuli. The root-mean-square error (RMSE) score was calculated as the mean target response difference. When the stimuli were PTs, the RMSE scores of the control and tinnitus group were 11.77 ± 2.57° and 13.97 ± 4.18°, respectively. The control group performed significantly better than did the tinnitus group ($t$ = 2.841, $p$ = 0.006). When the stimuli were MS, the RMSE scores of the control and tinnitus groups were 7.12 ± 2.29° and 7.90 ± 2.33°, respectively. There was no significant difference between the two groups ($t$ = 1.501, $p$ = 0.137). Neither the effect of unilateral or bilateral tinnitus (PT: $t$ = 0.763, $p$ = 0.450; MS: $t$ = 1.760, $p$ = 0.086) nor the effect of tinnitus side (left/right, PT: $t$ = 0.389, $p$ = 0.703; MS: $t$ = 1.407, $p$ = 0.179) on sound localization ability were determined. The sound source localization ability gradually deteriorated with an increase in age (PT: $r^2$ = 0.153, $p$ < 0.001; MS: $r^2$ = 0.516, $p$ = 0.043). In conclusion, tinnitus interfered with the ability to localize PTs, but the ability to localize MS was not affected. Therefore, the interference of tinnitus in localizing sound sources is related to the type of stimulus.

KEYWORDS

**binaural hearing, sound localization, tinnitus, spectrum information, interaural time differences, interaural level differences**

## 1. Introduction

Tinnitus is an involuntary phantom percept of internally generated non-verbal noises and tones without any external acoustic input. Tinnitus is a common disorder, affecting 10–15% of the adult population, and 2–3% of these cases are severe (Hesser et al., 2015; Mohan et al., 2022). Given the increased exposure to damaging recreational noise, the prevalence of tinnitus is expected to continue to increase (Langguth et al., 2013). However, the mechanism underlying tinnitus remains unclear. Although sensorineural hearing loss and excessive noise exposure are considered common causes of tinnitus, there is no obvious or immediately identifiable cause in 65–98% of cases (Attarha et al., 2018). The primary cause of tinnitus is thought to be associated with cochlear dysfunction; however, it is now generally accepted that alterations in central

auditory system function also play a role in the pathogenesis of tinnitus (Kwee et al., 2017). Clinical studies have shown that subjects with tinnitus without obvious hearing loss have some form of dysfunction in the auditory pathway (Song et al., 2018; Xiong et al., 2019; Liu et al., 2020; Han et al., 2021). In the last two decades, concerted efforts in basic and clinical research have significantly advanced our understanding of tinnitus. However, the exact mechanisms underlying this disorder remain unclear (Henton and Tzounopoulos, 2021).

The ability to localize sound sources is important for human listeners to be aware of their surroundings. Sound localization is based on three types of cues: Two binaural cues [interaural time differences (ITD) and interaural level difference (ILD)] and one monaural spectral cue (Risoud et al., 2018). Listeners require access to both binaural differences and spectral cues to localize accurately (Wightman and Kistler, 1997; Martin et al., 2004; Carlile et al., 2005; Marks et al., 2018). Localization of sound sources is a complex process in the human brain. The spatial cues come from both ears and are analyzed in specific brainstem pathways. The medial superior olive (MSO) units are dominated by ITD cues, while the lateral superior olive (LSO) units are dominated by ILD information. Units in the dorsal cochlear nucleus are involved in the processing of spectral cues (Arle and Kim, 1991; Middlebrooks, 2015; Ryan and Bauer, 2016). Projections from these nuclei form various degrees of cue integration in the central nucleus of the inferior colliculus (ICC) (Chase and Young, 2005; Bender and Trussell, 2011). The ITD cues mainly ascend to type V neurons. ILD information ascends to Type I and O units, and spectral cues primarily ascend to Type O units (Davis et al., 2003). Accurate localization requires precise specification of the number and intensity of projected inputs. Tinnitus is thought to arise from increased spontaneous firing rates (SFR), dysregulated synchrony across neurons ensembles, and increased bursting along the auditory pathway (Weisz et al., 2007; Wu et al., 2016). These processes begin in the dorsal cochlear nucleus and convey to higher brainstem and cortical regions (Shore et al., 2016). In the inferior colliculus, increased synchrony across multi-unit clusters and bursting have also been observed in animal models of tinnitus (Bauer et al., 2008). An intact auditory pathway is indispensable for normal sound localization. A previous study focused on the auditory localization of subjects with unilateral tinnitus, suggesting that tinnitus-related activity in localization-sensitive areas may interfere with localization cues and result in degraded localization performance (Hyvärinen et al., 2016).

Tinnitus in patients with normal hearing may be subclinical and thus not captured by the traditional audiometric test battery (Diges et al., 2017). Sound source localization requires binaural auditory cues starting with the ventral cochlear nucleus. Therefore, it may be used to reflect the effects of tinnitus on the central auditory function. Rhee et al. (2020) reported that adolescents with tinnitus whose hearing loss was not detected complained of difficulty in sound localization. Kwee et al. (2017) performed a detailed structural analysis of a patient with a unilateral lesion of the inferior colliculus using magnetic resonance microscopy with a 7T system. They reported that ICC dysfunction might be the cause of tinnitus and lack of sound localization, but not hearing loss.

Currently, literature on sound localization in patients with tinnitus is scarce. However, the understanding of the effect of tinnitus on sound source localization remains limited. Hyvärinen et al. (2016) used pink noise burst as a stimulus and found that the accuracy of sound source discrimination was significantly worse in participants with unilateral tinnitus than in those with normal hearing. However, participants with tinnitus experienced hearing loss at a high frequency. Explorative analysis suggested that the results might not be directly related to tinnitus because of inter-individual differences in hearing abilities. An et al. (2012) found that participants with tinnitus performed worse in localizing pure tones (PTs) than did those without tinnitus. All the participants had normal hearing. However, this scoring method is uncommon, and they scored their participants one point for each 30-degree difference between the target speaker and the response speaker and used this error score to compare the accuracy of sound localization. The error score increased with the number of recognition errors regardless of the total number of stimulations. Playing each speaker five times may have amplified the errors in the tinnitus group. Although the effects of tinnitus on sound localization have not been clearly demonstrated in people with symmetrical hearing, studies on people with single-sided deafness have begun. Liu et al. (2018) reported that in patients with single-sided deafness, the degree of tinnitus was negatively correlated with sound localization. However, the binaural cues used for sound localization were lost in these patients, and the effects of tinnitus on sound source localization were not explicitly explained. Unilateral peripheral inputs are associated with central auditory changes over time (Bernstein et al., 2022). Asymmetric hearing complicates the effects of tinnitus on sound-source localization. The relationship between tinnitus and sound source localization in abnormal hearing patients can only be better studied based on the understanding of the relationship between tinnitus and sound source localization in normal hearing population.

Thus, this study aimed to investigate whether tinnitus affects the ability to localize sound in participants with normal hearing and whether the effect is related to the type of stimulus. Toward this goal, we performed sound source discrimination tasks in participants with normal hearing with and without tinnitus. The root mean square error (RMSE) score was calculated as the mean target-response difference to investigate whether tinnitus affected the ability to localize sound. PTs and monosyllables (MS) were used as stimuli to investigate whether the effect was related to the type of stimulus.

## 2. Materials and methods

### 2.1. Study design and participants

This study was approved by the Ethics Committee of Beijing Friendship Hospital, Capital Medical University (2021-P2-004-01) and adhered to the tenets of the Declaration of Helsinki. Written informed consent before recruitment was obtained from the participants after adequate explanation of the purpose and procedures of the study.

A total of 40 participants with tinnitus [22 males and 18 females aged 22–66 years (mean ± SD age: 35.53 ± 10.31 years)] and 40 participants without tinnitus [13 males and 27 females aged

Abbreviations: ICC, inferior colliculus; ILD, interaural level difference; ITD, interaural time difference; LSO, lateral superior olive; MS, monosyllable; PT, pure tone; PTAL, pure-tone average in the left ear; PTAR, pure-tone average in the right ear; RMSE, root-mean-square error; SFR, spontaneous firing rates; THI, Tinnitus Handicap Inventory.

17–43 years (mean ± SD: 28.15 ± 5.98 years)] were enrolled in the study. The participants with tinnitus were recruited from the ENT outpatient department of Beijing Friendship Hospital, Capital Medical University from June 2021 to September 2022. The inclusion criteria were as follows: (1) Subjective tinnitus; (2) persistent tinnitus; (3) duration ≥3 months; and (4) hearing threshold ≤25 dB HL at all frequencies (0.25–8 kHz) in both ears. Participants in the control group had normal hearing and no tinnitus. They were recruited through advertising. The inclusion criteria were (1) no tinnitus and (2) hearing threshold ≤25 dB HL at all frequencies (0.25–8 kHz) in both ears. The exclusion criteria for both groups were as follows: (1) Objective or pulsatile tinnitus; (2) difference in bilateral hearing threshold >10 dB HL at any frequency (0.25–8 kHz); (3) air-bone gap >10 dB HL; (4) history of hearing loss or vertigo; (5) diagnosis of depressive disorder or anxiety disorder; and (6) diagnosis of hypertension. The participants in the tinnitus group were slightly older than those in the control group ($t$ = 3.912, $p$ < 0.001). The hearing thresholds of the two groups were similar at all frequencies from 0.25 to 8 kHz, as shown in Figure 1. In the tinnitus group, the mean duration of tinnitus was 2.07 ± 3.81 years.

## 2.2. Tinnitus evaluation

Tinnitus matching included pitch and loudness using audiometers (Madsen Astera, Otometrics, USA) with headphones (ME70, Otometrics, USA). A two-alternative forced choice method was used in this process (Vernon and Fenwick, 1985). First, the test ear was given a pair of pure-tone signals starting with multiples of 1 kHz, and the patient was asked to identify which was closer to the tinnitus frequency. Once a pair of frequencies was identified, the frequency resolution was increased, becoming closer to the tinnitus frequency (the finest frequency was 1/48 octave). The intensities of the matched pitch were then increased in 5-dB increments, starting with an intensity below the hearing threshold and then gradually increasing or decreasing in 1 dB step until the loudness of tinnitus was matched. Unilateral and bilateral tinnitus was detected in 17 participants (6 in the right ear and 11 in the left ear) and 23 participants, respectively.

The Tinnitus Handicap Inventory (THI) was used to assess tinnitus severity. Briefly, THI is a self-reported questionnaire comprising 25 items that reflect the impact of tinnitus on daily life. Each question is answered with "yes, sometimes, or no," with each response counting for 4, 2, or 0 points, respectively. The total score was graded on five scales: Slight (0–16), mild (18–36), moderate (38–56), severe (58–76), and catastrophic (78–100) (Newman et al., 1996).

## 2.3. Sound source discrimination task

The sound-source discrimination task was conducted in an anechoic chamber (LSsx2021-21270). Thirty-seven loudspeakers were set in a 180° arc, 5° apart (Figure 2). The speakers were 1.2 m away from the participant and at the height of the subject's external auditory canal. The participants were instructed to face directly ahead until the stimuli stopped and to indicate the speaker number (1–37) on a touchscreen. The head movement away from the 0 azimuth was monitored by the experimenter while the stimuli were playing. No feedback was provided after each response. The test was divided into two conditions according to the different types of stimuli: PT and MS conditions. In the PT condition, stimuli were PTs of 0.25, 0.5, 1, 2, 4, and 8 kHz at 50 dB SPL for 0.5 s. Six times of 0.25, 0.5, 2, 4, and 8 kHz stimuli and seven times of 1 kHz stimuli were presented randomly from the 37 loudspeakers. Each speaker played the stimulus only once. In the MS condition, the stimulus was a monosyllabic word "song." A spectrogram of the words is shown in Figure 3. Stimuli were presented randomly from 37 loudspeakers, with each speaker playing the stimulus only once. The sequences of the two task conditions were generated randomly. The formal test began after participants attempted to respond 10 times and became familiar with the process. The RMSE score is calculated as the mean target-response difference as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\alpha_{RESP}^{i}-\alpha_{STIM}^{i}\right)^{2}}{n}}$$

where $n$ is the total number of stimuli, $i$ is the number of stimuli, $\alpha_{RESP}$ is the response azimuth, and $\alpha_{STIM}$ is the target azimuth angle. Lower scores indicated greater accuracy.

## 2.4. Statistical analysis

Continuous variables were expressed as the means and standard deviations. An independent sample $t$-test was performed to evaluate the differences between the two groups or subgroups. A paired-sample $t$-test was performed to evaluate the differences between the two conditions, and between the left and right sides. Pearson correlation was used to analyze the relationship between the RMSE scores of the two conditions and the relationship among sound localization, tinnitus-matched pitch and loudness, and THI score. A Mann–Whitney test was conducted to compare the ability between participants with tone-like tinnitus and with broadband noise-like tinnitus to localized the monosyllable stimuli. A general linear model was used to explore the effects of tinnitus, age, and hearing threshold on sound localization. All statistical analyses were performed using SPSS version 25 (IBM Corp., Armonk, NY, USA). Diagrams were drawn using GraphPad Prism 8.0 (San Diego, CA, USA). The spectrogram was drawn using the Praat software. $P$ < 0.05 was considered statistically significant.

# 3. Results

## 3.1. Sound localization behavior

When the stimuli were PTs, the RMSE scores of the control and tinnitus group were 11.77 ± 2.57° and 13.97 ± 4.18°, respectively. The control group performed significantly better than did the tinnitus group [$t_{(78)}$ = 2.841, $p$ = 0.006]. When the stimuli were monosyllable, the RMSE scores of the control and tinnitus group were 7.12 ± 2.29° and 7.90 ± 2.33°, respectively, with no significant between-group difference [$t_{(78)}$ = 1.501, $p$ = 0.137]. For the accuracy of sound source discrimination of the participants under different stimulus conditions, in both groups, sound source was more accurately localized under MS condition than under PT condition [control group: 7.12 ± 2.290 vs. 11.77 ± 2.57, $t_{(39)}$ = 11.979, $p$ < 0.001; tinnitus

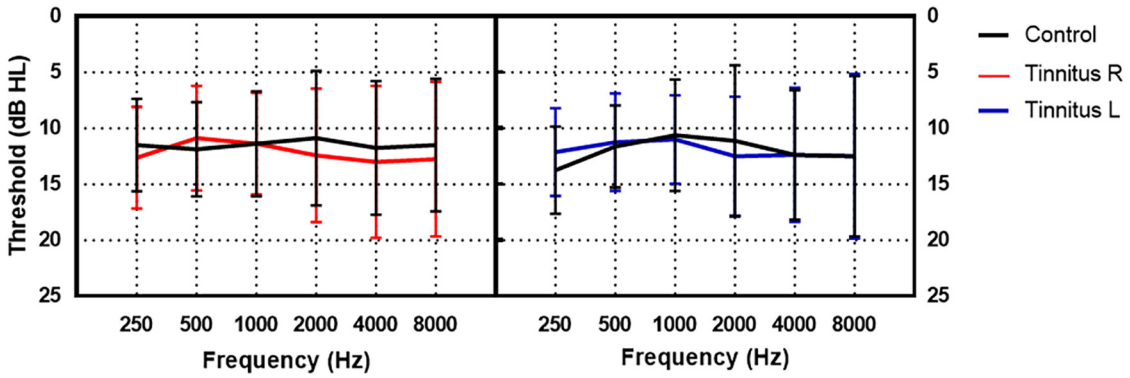



FIGURE 1
The hearing thresholds of the tinnitus and control group. The black lines indicate the mean hearing threshold of the control group. The red and blue lines indicate the mean hearing threshold of the right and left ear of the tinnitus group, respectively.

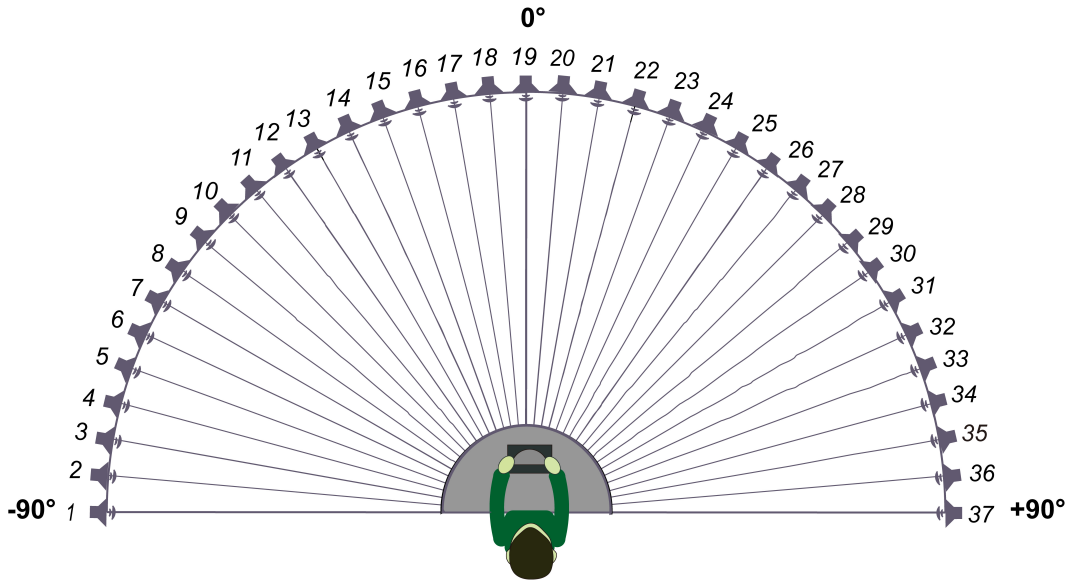



FIGURE 2
Schematic diagram of loudspeaker placement.

group: 7.90 ± 2.33 vs. 13.97 ± 4.18, $t_{(39)}$ = 9.250, $p < 0.001$]. The results are shown in Figure 4A. There was a positive correlation between the RMSE scores of the PT and MS conditions in the control group [$r^2_{(38)}$ 0.246, $p = 0.001$] but not in the tinnitus group [$r^2_{(38)}$ 0.084, $p = 0.068$]. The results are shown in Figures 4B, C.

## 3.2. Effect of tinnitus side on sound localization behavior

The participants with tinnitus were divided into two subgroups: Unilateral and bilateral. Comparison of the ability of sound source localization between the two subgroups showed no significant difference regardless of the stimulus type [PT: 14.56 ± 3.80 vs. 13.54 ± 4.48, $t_{(38)}$ = 0.763, $p$ = 0.450; MS: 8.63 ± 2.50 vs. 7.36 ± 2.09, $t_{(38)}$ = 1.760, $p$ = 0.086, Figure 5A]. For participants with unilateral tinnitus, we investigated whether sounds originating from the same side as tinnitus were more difficult to localize. Stimuli emitted by loudspeakers numbers 1–18 originated from the left, while those emitted by loudspeakers numbers 20–37 originated from the right. The RMSE scores for the left and right sounds were calculated separately for comparison purposes. There was no significant difference in the ability to localize sound originating from the same side of tinnitus and those from the opposite side of tinnitus regardless of stimulus type [PT: 13.97 ± 6.18 vs. 14.57 ± 2.99, $t_{(16)}$ = 0.389, $p$ = 0.703; MS: 7.67 ± 1.94 vs. 9.26 ± 4.23, $t_{(16)}$ = 1.407, $p$ = 0. 179, Figure 5B].

## 3.3. Effect of characteristic and severity of tinnitus on sound localization behavior

In 36 participants tinnitus sounds like a pure tone, and in 4 participants it sounds like broadband noise. For participants with tone-like tinnitus, there was no correlation between the tinnitus-matched pitch and the RMSE scores regardless the type of stimulus

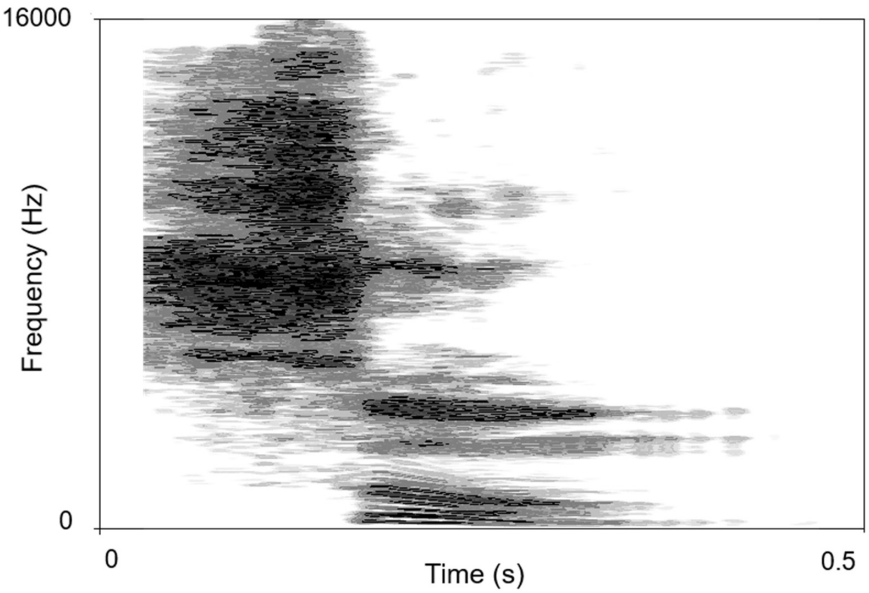



FIGURE 3
The spectrogram of "song". The abscissa is time, the ordinate is frequency, and the gray shade represents intensity.

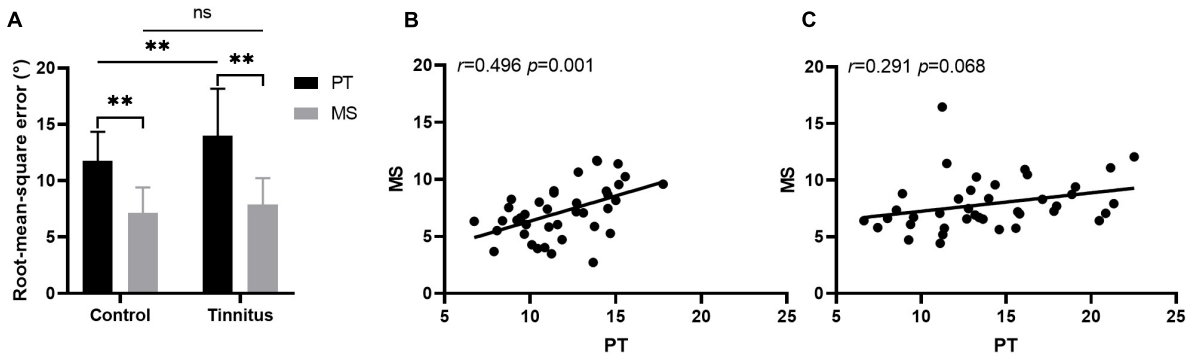



FIGURE 4
Sound localization behavior. **(A)** The accuracy of sound source discrimination of the participants under different stimulus conditions. PT, pure tone; MS, monosyllable. $^{**}P < 0.01$, ns, not significant. The correlation between the RMSE scores of the PT and MS conditions in the control **(B)** and tinnitus **(C)** group.

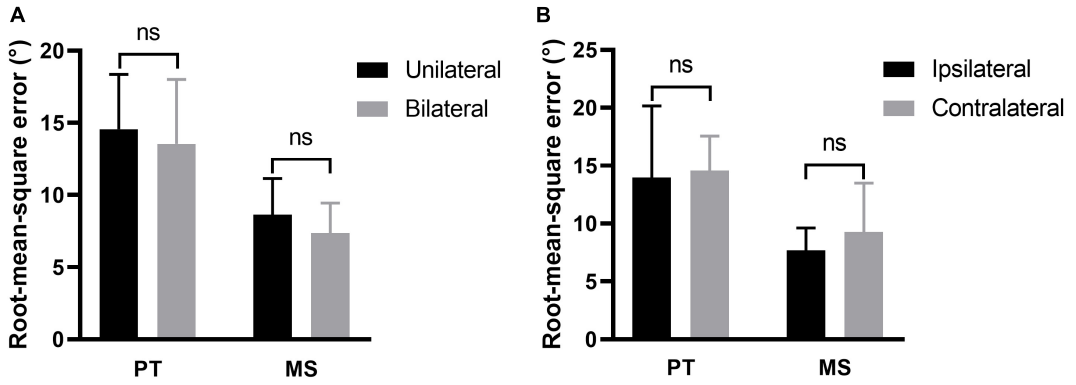



FIGURE 5
Effect of tinnitus side on sound localization. **(A)** The sound localization behavior of participants with unilateral and bilateral tinnitus. **(B)** The ability to localize sound originating from the same side of tinnitus and those from the opposite side of tinnitus. PT, pure tone; MS, monosyllable. Ns, not significant.

[PT: $r^2_{(34)}$ 0.001, $p$ = 0.910, Figure 6A; MS: $r^2_{(34)}$ 0.070, $p$ = 0.096, Figure 6B]. There was no difference of the ability to localized the monosyllable stimuli between participants with tone-like tinnitus and with broadband noise-like tinnitus either (7.74 ± 2.28 vs. 9.37 ± 2.64, $z$ = 1.285, $p$ = 0.199, Figure 6C). Tinnitus severity was evaluated by loudness and THI scores. To minimize the influence of hearing threshold, we used the sensation level to represent tinnitus loudness of the participants. The average of tinnitus loudness was 15.31 ± 13.42 dB HL. No correlation was found between loudness and RMSE scores [PT: $r^2_{(33)}$ 0.036, $p$ = 0.277, Figure 7A; MS: $r^2_{(33)}$ 0.027, $p$ = 0.348, Figure 7B]. The number of participants in the slight, mild, moderate, severe, and catastrophic grade of tinnitus were 11, 10, 10, 7, and 2, respectively. There was no correlation between the THI and RMSE scores [PT: $r^2_{(37)}$ 0.015, $p$ = 0.459, Figure 7C; MS: $r^2_{(37)}$ 0.001, p = 0.923, Figure 7D].

### 3.4. Explorative analysis of confounding factors

Although the hearing thresholds of the two groups were similar, there was a significant difference in age between them. An additional explorative analysis was performed to investigate the effects of inter-individual variability of these factors on the observed results. The RMSE score, group, age, and pure-tone average among 0.25–8 kHz of the left (PTAL) and right (PTAR) ears were included in a general linear model. The RMSE score was a fixed effect, while group category, age, PTAL, and PTAR were independent variables. For the PT condition, group category and age significantly affected the RMSE scores [group, $r^2_{(79)}$ 0.082, $p$ = 0. 006; age, $r^2$ = 0.153, $p$ < 0.001]. For the MS condition, the group category did not affect the RMSE scores [$r^2_{(79)}$ 0.016, $p$ = 0.137], whereas age significantly affected the ability to localize [$r^2_{(79)}$ 0.516, $p$ = 0.043]. Hearing thresholds did not affect the ability to localize [PT, PTAL: $r^2_{(79)}$ 0.098, $p$ = 0. 633, PTAR: $r^2_{(79)}$ 0.307, $p$ = 0.750; MS, PTAL: $r^2_{(79)}$ 0.277, $p$ = 0.800, PTAR: $r^2_{(79)}$ 0.129, $p$ = 0.383].

## 4. Discussion

Despite its high prevalence, the exact mechanisms underlying tinnitus remain unclear. The current study found that tinnitus interfered with the ability to localize sounds without spectrum information but not with the ability to localize sounds with spectral information. As the processing of ITD, ILD, and spectral cues were in different parts of the auditory pathway, this result suggests that tinnitus interfered with certain sections of localization-sensitive areas. These findings provide a new perspective on the relationship of tinnitus with sound localization ability.

Most sound localization studies have used artificial stimuli that listeners do not often encounter in their daily lives, such as PTs and noise bursts (Van Wanrooij and Van Opstal, 2007; Voss et al., 2015; Zhao et al., 2021). However, little is known regarding the localization of meaningful sounds (van der Heijden et al., 2019). The duplex theory of PT or narrowband noise is universally confirmed in both human and animal subjects. ITD is mainly used to localize low-frequency sounds, whereas ILD is mainly used for high frequencies (Tollin et al., 2013). Although ITD and ILD are the principal cues for localization in the horizontal plane (Middlebrooks, 2015), one study indicates an intriguing correlation between perceived lateral location and the weighting of spectral cues (Macpherson and Sabin, 2007). Tan et al. (2013) found that tinnitus patients had better frequency selectivity than those without tinnitus. Zeng et al. (2020) revealed that there was no significant difference in frequency discrimination between control and tinnitus participants. Moreover, Moon et al. (2015) reported that there were no significant differences in spectral-ripple discrimination between the tinnitus lateral and contralateral ears of unilateral tinnitus participants with symmetric hearing thresholds. Therefore, tinnitus does not affect spectral resolution of patients' hearing. In addition, several studies on cochlear implant receivers have used speech signals as stimuli for sound discrimination tasks (Dieudonné and Francart, 2018; Killan et al., 2019) because of their relevance to realistic listening conditions (Dieudonné and Francart, 2018). Imitating this, a MS word was used as another stimulus in our study. Previous studies confirmed that localization acuity is higher for broadband sounds than for narrowband sounds (Butler, 1986; Carlile et al., 1999; Tollin et al., 2013). In line with these studies (Butler, 1986; Carlile et al., 1999; Tollin et al., 2013), our results showed that in both groups, MS stimuli were localized more accurately than were PT stimuli. This was mainly because of the presence of spectral cues in the former.

The finding that the accuracy of PT localization was worse in tinnitus participants than in non-tinnitus participants was in line with the findings of An et al. (2012). However, the lack of difference in localizing sound with spectrum information between the two groups was inconsistent with the findings of Hyvärinen et al. (2016). This might be related to the worse hearing sensitivity of tinnitus participants in the study by Hyvärinen et al. (2016) which interfered with the accuracy of the sound source discrimination. Participants relied only on ITD and ILD cues to localize the sound source under PT conditions, whereas spectral information could also be used under MS conditions. As tinnitus only interfered with the ability to localize sound without spectrum information, tinnitus was more likely to affect the process of ITD and ILD. ITD is mainly processed in MSO, while ILD is mainly processed in LSO (Chase and Young, 2005). The stimulus-dependent dominance of binaural cues in the ICC could potentially result from the convergence of MSO and LSO inputs onto the same neuron. Both low and high characteristic frequency neurons in the ICC can exhibit dominance of ITD or ILD cues according to the spectrum of the stimulus (Dorkoski et al., 2020). Therefore, it is possible that changes in MSO and LSO contribute to the decreased ability to localize PTs in patients with tinnitus. Meanwhile, we speculated that because fewer nerves are involved in PT localization, tinnitus-induced changes in the auditory pathway might be easier to detect. Abnormal processing of ITD and ILD cues may be compensated for by the involvement of more neurons when localizing MS stimuli. A previous review supports this conjecture. It has been reported that the dorsal cochlear nucleus type IV unit exhibits excitation only around the characteristic frequency neurons when stimulated with PTs, whereas a wide range of neurons are involved when stimulated with broadband sound (Carlile et al., 2005). Moreover, the RMSE scores for PT and MS were positively correlated in the control group, whereas there was no such correlation between the two conditions in the tinnitus group. This also indicated that the effect of tinnitus on the localization of the two types of sounds was not completely consistent.

We did not find differences in the sound localization behavior of participants with unilateral and bilateral tinnitus, nor did we find an effect of tinnitus side on this behavior. However,

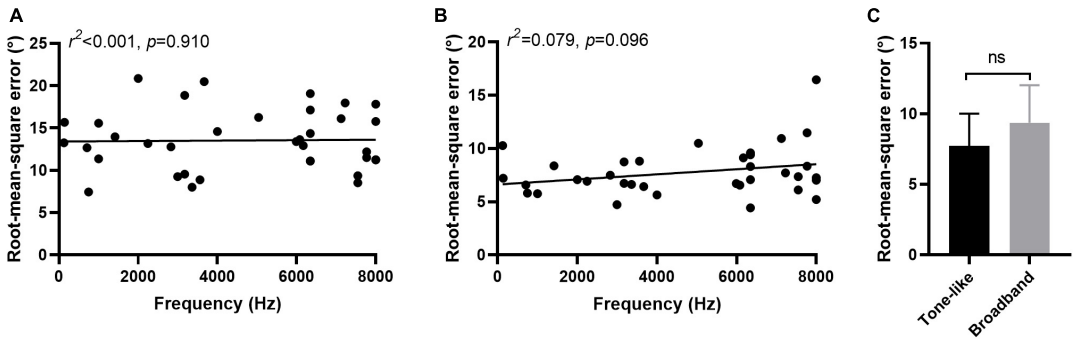



FIGURE 6
Effect of characteristic of tinnitus on sound localization behavior. No correlation was found between tinnitus-matched pitch and RMSE scores in PT **(A)** and MS **(B)** condition. **(C)** No difference was found of the ability to localized monosyllable stimuli between participants with tone-like tinnitus and with broadband noise-like tinnitus. Ns, not significant.

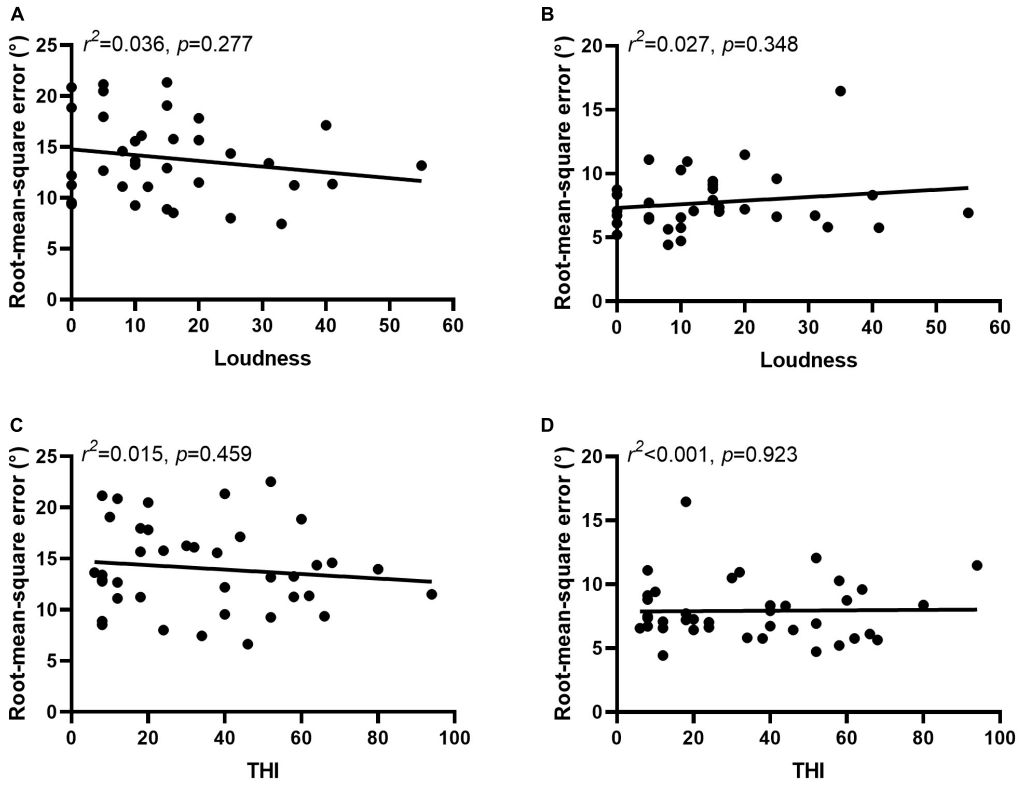



FIGURE 7
Severity of tinnitus was not related to sound localization behavior. No correlation was found between loudness and RMSE scores in PT **(A)** and MS **(B)** condition. No correlation was found between the THI and RMSE scores in PT **(C)** and MS **(D)** condition.

An et al. (2012) reported that when localizing sound sources from one side, patients with tinnitus on the same side performed worse than did those with tinnitus on the opposite side and those with bilateral tinnitus. They suggested that tinnitus interfered with ILD cues and degraded the localization performance. However, there was no conclusive evidence to confirm this finding as the data analysis did not consider the frequency and loudness of individual tinnitus matched (Hyvärinen et al., 2016). Tinnitus is a subjective feeling that lacks objective measurement (Henry, 2016). Investigations of the qualitative characteristics of tinnitus, such as pitch matching, loudness matching, and tinnitus suppression with acoustic stimulation, were not diagnostic and were not used in making management decisions (Bauer, 2018). The current study also did not find a correlation between tinnitus loudness and sound localization ability. Moreover, both our study and the study of An et al. (2012) showed that there was no correlation between the tinnitus-matched pitch and the sound localization behavior. Therefore, it may not be comprehensive to simply conclude that tinnitus alters the perceived ILD and thus affects the ability of sound localization. Accurate sound-source localization requires a complete auditory pathway. Tinnitus is associated with neuronal enhanced SFR or decreased SFR, changes in neuronal transfer functions (gain), and changes in neural synchrony (Henton and Tzounopoulos, 2021), and thus, it could affect the processing of binaural auditory cues regardless

of the affected side. As tinnitus may affect the processing of binaural auditory cues, it is reasonable to hypothesize that a greater degree of tinnitus may be associated with worse sound localization (Liu et al., 2018). THI is a widely used assessment tool that is sensitive to tinnitus severity (McCombe et al., 2001). However, the current study found no correlation between the THI and RMSE scores, indicating that worsening sound source localization ability was not directly related to the degree of tinnitus annoyance experienced by the participants.

Explorative analysis suggested that the sound source localization ability gradually deteriorated with increasing age. Previous studies have consistently shown that despite the generally normal hearing thresholds, sound localization ability generally decreases with age, both in humans (Dobreva et al., 2011; Goupell, 2022; Weissgerber et al., 2022) and in animals (Cheng et al., 2020). Increasing age adversely affects the processing of ITD cues, including both the temporal fine structure and the slowly varying envelope (Weissgerber et al., 2022). Anatomical studies have found that inhibitory inputs into the neuronal circuit responsible for sound localization are significantly reduced in aged animals (Ashida et al., 2021). In Sprague–Dawley rats, the number of inhibitory neurons in the medial nucleus of the trapezoid body begins to decrease at age 2–3 months (roughly equivalent to 10 years in humans) (Casey and Feldman, 1982). The medial nucleus of the trapezoid body is a sign-inverting relay nucleus that forms an inhibitory pathway to the LSO *via* glutamatergic neurons in the sound localization circuit of the brainstem. Age-related degradation of sound localization ability may also be caused by altered functions of higher stages in the auditory pathways above the brainstem level (Ashida et al., 2021). In the current study, the ability to localize PTs in participants with tinnitus was still worse than that in the control group even after controlling for age.

To our best knowledge, this is the first study to investigate whether the interference of tinnitus in localizing sound sources is related to the type of stimulus. The results showed that spectrum information could help tinnitus patients improve their ability to localize sound sources and reach the level of the non-tinnitus group. MS stimuli are more complex and meaningful than are PTs and are more similar to sounds encountered regularly in daily life. Thus, the results regarding the ability of localization in patients with tinnitus should be carefully interpreted and considered as a starting point for further studies. However, it should also be noted that there were a few limitations to this study. First, tinnitus sounds come in many varieties such as pure tones, hissing, buzzing, humming, and growling. However, we only included participants whose tinnitus sounds like tone and broadband noise, thus the results were only applicable to this subset of patients. We found that our participants' ability to locate pure tone stimuli was affected but their ability to locate monosyllable was not. Because 90% of participants had tone-like tinnitus and only four participants heard tinnitus like broadband noise, a question was raised that whether the change in the ability of sound localization is dependent on both the quality of the sound that is being localized and the type of tinnitus perception. Because of the small number of participants with broadband noise-like tinnitus and the different characteristic between tinnitus and stimulus sound, we could not answer this question well. Previous studies (An et al., 2012; Hyvärinen et al., 2016) of sound localization ability in tinnitus patients included only subjects with tone-like tinnitus as well. In the future, we need to expand the number of patients with other types of tinnitus sounds and use more kinds of stimuli to answer this question. Second, exploratory analysis showed that age affected sound localization ability. Differences in age may have led to differences in cognitive abilities and affected the results of the experiment. Therefore, cognitive assessments should be included in future studies. Third, the study only reported behavioral results. To better understand this phenomenon, cellular mechanisms need to be explored in animal experiments.

## 5. Conclusion

Tinnitus interferes with the ability to localize PTs but not the ability to localize MS stimuli. Therefore, the interference of tinnitus in localizing sound sources is related to the type of stimulus. Moreover, the relationship between tinnitus and sound localization behavior should be carefully interpreted.

## Data availability statement

The original contributions presented in this study are included in this article/supplementary material, further inquiries can be directed to the corresponding authors.

## Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of Beijing Friendship Hospital, Capital Medical University. The patients/participants provided their written informed consent to participate in this study.

## Author contributions

YL collected and analyzed the data and wrote the manuscript. WW and JL collected and analyzed the data. KL provided the technical support and reviewed the manuscript. SG conceived the study and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This work was supported by the National Key Scientific Research Project of China (2020YFC2005201).

## Acknowledgments

We thank all participants.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

An, Y. H., Lee, L. H., Yoon, S. W., Jin, S. Y., and Shim, H. J. (2012). Does tinnitus affect the sound localization ability? *Otol. Neurotol.* 33, 692–698. doi: 10.1097/MAO.0b013e31825952e9

Arle, J. E., and Kim, D. O. (1991). Neural modeling of intrinsic and spike-discharge properties of cochlear nucleus neurons. *Biol. Cybern.* 64, 273–283. doi: 10.1007/bf00199590

Ashida, G., Tollin, D. J., and Kretzberg, J. (2021). Robustness of neuronal tuning to binaural sound localization cues against age-related loss of inhibitory synaptic inputs. *PLoS Comput. Biol.* 17:e1009130. doi: 10.1371/journal.pcbi.1009130

Attarha, M., Bigelow, J., and Merzenich, M. M. (2018). Unintended consequences of white noise therapy for tinnitus-otolaryngology's cobra effect: A review. *JAMA Otolaryngol. Head Neck Surg.* 144, 938–943. doi: 10.1001/jamaoto.2018.1856

Bauer, C. A. (2018). Tinnitus. *N. Engl. J. Med.* 378, 1224–1231. doi: 10.1056/NEJMcp1506631

Bauer, C. A., Turner, J. G., Caspary, D. M., Myers, K. S., and Brozoski, T. J. (2008). Tinnitus and inferior colliculus activity in chinchillas related to three distinct patterns of cochlear trauma. *J. Neurosci. Res.* 86, 2564–2578. doi: 10.1002/jnr.21699

Bender, K. J., and Trussell, L. O. (2011). Synaptic plasticity in inhibitory neurons of the auditory brainstem. *Neuropharmacology* 60, 774–779. doi: 10.1016/j.neuropharm.2010.12.021

Bernstein, J. G. W., Phatak, S. A., Schuchman, G. I., Stakhovskaya, O. A., Rivera, A. L., and Brungart, D. S. (2022). Single-sided deafness cochlear implant sound-localization behavior with multiple concurrent sources. *Ear Hear* 43, 206–219. doi: 10.1097/aud.0000000000001089

Butler, R. A. (1986). The bandwidth effect on monaural and binaural localization. *Hear Res.* 21, 67–73. doi: 10.1016/0378-5955(86)90047-x

Carlile, S., Delaney, S., and Corderoy, A. (1999). The localisation of spectrally restricted sounds by human listeners. *Hear Res.* 128, 175–189. doi: 10.1016/s0378-5955(98)00205-6

Carlile, S., Martin, R., and McAnally, K. (2005). Spectral information in sound localization. *Int. Rev. Neurobiol.* 70, 399–434. doi: 10.1016/s0074-7742(05)70012-x

Casey, M. A., and Feldman, M. L. (1982). Aging in the rat medial nucleus of the trapezoid body I light microscopy. *Neurobiol. Aging.* 3, 187–195. doi: 10.1016/0197-4580(82)90039-2

Chase, S. M., and Young, E. D. (2005). Limited segregation of different types of sound localization information among classes of units in the inferior colliculus. *J. Neurosci.* 25, 7575–7585. doi: 10.1523/jneurosci.0915-05.2005

Cheng, Y., Zhang, Y., Wang, F., Jia, G., Zhou, J., Shan, Y., et al. (2020). Reversal of age-related changes in cortical sound-azimuth selectivity with training. *Cereb. Cortex.* 30, 1768–1778. doi: 10.1093/cercor/bhz201

Davis, K. A., Ramachandran, R., and May, B. J. (2003). Auditory processing of spectral cues for sound localization in the inferior colliculus. *J. Assoc. Res. Otolaryngol.* 4, 148–163. doi: 10.1007/s10162-002-2002-5

Dieudonné, B., and Francart, T. (2018). Head shadow enhancement with low-frequency beamforming improves sound localization and speech perception for simulated bimodal listeners. *Hear Res.* 363, 78–84. doi: 10.1016/j.heares.2018.03.007

Diges, I., Simón, F., and Cobo, P. (2017). Assessing auditory processing deficits in tinnitus and hearing impaired patients with the auditory behavior questionnaire. *Front. Neurosci.* 11:187. doi: 10.3389/fnins.2017.00187

Dobreva, M. S., O'Neill, W. E., and Paige, G. D. (2011). Influence of aging on human sound localization. *J. Neurophysiol.* 105, 2471–2486. doi: 10.1152/jn.00951.2010

Dorkoski, R., Hancock, K. E., Whaley, G. A., Wohl, T. R., Stroud, N. C., and Day, M. L. (2020). Stimulus-frequency-dependent dominance of sound localization cues across the cochleotopic map of the inferior colliculus. *J. Neurophysiol.* 123, 1791–1807. doi: 10.1152/jn.00713.2019

Goupell, M. J. (2022). Age-related changes in interaural-level-difference-based across-frequency binaural interference. *Front. Aging Neurosci.* 14:887401. doi: 10.3389/fnagi.2022.887401

Han, M. S., Jeong, Y. J., Im, G. J., Song, J. J., Chae, S. W., Chan Rah, Y., et al. (2021). Auditory brainstem response test results in normal hearing adolescents with subjective tinnitus. *Int. J. Pediatr. Otorhinolaryngol.* 146, 110775. doi: 10.1016/j.ijporl.2021.110775

Henry, J. A. (2016). "Measurement" of tinnitus. *Otol. Neurotol.* 37, e276–85. doi: 10.1097/mao.0000000000001070

Henton, A., and Tzounopoulos, T. (2021). What's the buzz? The neuroscience and the treatment of tinnitus. *Physiol. Rev.* 101, 1609–1632. doi: 10.1152/physrev.00029.2020

Hesser, H., Bånkestad, E., and Andersson, G. (2015). Acceptance of tinnitus as an independent correlate of tinnitus severity. *Ear Hear* 36, e176–182. doi: 10.1097/aud.0000000000000148

Hyvärinen, P., Mendonça, C., Santala, O., Pulkki, V., and Aarnisalo, A. A. (2016). Auditory localization by subjects with unilateral tinnitus. *J. Acoust. Soc. Am.* 139:2280. doi: 10.1121/1.4946897

Killan, C., Scally, A., Killan, E., Totten, C., and Raine, C. (2019). Factors affecting sound-source localization in children with simultaneous or sequential bilateral cochlear implants. *Ear Hear* 40, 870–877. doi: 10.1097/aud.0000000000000666

Kwee, I. L., Matsuzawa, H., Nakada, K., Fujii, Y., and Nakada, T. (2017). Inferior colliculus syndrome: Clinical magnetic resonance microscopy anatomic analysis on a 7 t system. *SAGE Open Med. Case Rep.* 5, 2050313x17745209. doi: 10.1177/2050313x17745209

Langguth, B., Kreuzer, P. M., Kleinjung, T., and De Ridder, D. (2013). Tinnitus: Causes and clinical management. *Lancet Neurol.* 12, 920–930. doi: 10.1016/s1474-4422(13)70160-1

Liu, Y. W., Cheng, X., Chen, B., Peng, K., Ishiyama, A., and Fu, Q. J. (2018). Effect of tinnitus and duration of deafness on sound localization and speech recognition in noise in patients with single-sided deafness. *Trends Hear* 22, 2331216518813802. doi: 10.1177/2331216518813802

Liu, Y. W., Wang, B., Chen, B., Galvin, J. J. III, and Fu, Q. J. (2020). Tinnitus impairs segregation of competing speech in normal-hearing listeners. *Sci. Rep.* 10:19851. doi: 10.1038/s41598-020-76942-1

Macpherson, E. A., and Sabin, A. T. (2007). Binaural weighting of monaural spectral cues for sound localization. *J. Acoust. Soc. Am.* 121, 3677–3688. doi: 10.1121/1.2722048

Marks, K. L., Martel, D. T., Wu, C., Basura, G. J., Roberts, L. E., Schvartz-Leyzac, K. C., et al. (2018). Auditory-somatosensory bimodal stimulation desynchronizes brain circuitry to reduce tinnitus in guinea pigs and humans. *Sci. Transl. Med.* 10:eaal3175. doi: 10.1126/scitranslmed.aal3175

Martin, R. L., Paterson, M., and McAnally, K. I. (2004). Utility of monaural spectral cues is enhanced in the presence of cues to sound-source lateral angle. *J. Assoc. Res. Otolaryngol.* 5, 80–89. doi: 10.1007/s10162-003-3003-8

McCombe, A., Baguley, D., Coles, R., McKenna, L., McKinney, C., and Windle-Taylor, P. (2001). Guidelines for the grading of tinnitus severity: The results of a working group commissioned by the british association of otolaryngologists, head and neck surgeons, 1999. *Clin. Otolaryngol. Allied Sci.* 26, 388–393. doi: 10.1046/j.1365-2273.2001.00490.x

Middlebrooks, J. C. (2015). Sound localization. *Handb. Clin. Neurol.* 129, 99–116. doi: 10.1016/b978-0-444-62630-1.00006-8

Mohan, A., Leong, S. L., De Ridder, D., and Vanneste, S. (2022). Symptom dimensions to address heterogeneity in tinnitus. *Neurosci. Biobehav. Rev.* 134:104542. doi: 10.1016/j.neubiorev.2022.104542

Moon, I. J., Won, J. H., Kang, H. W., Kim, D. H., An, Y. H., and Shim, H. J. (2015). Influence of tinnitus on auditory spectral and temporal resolution and speech perception in tinnitus patients. *J. Neurosci.* 35, 14260–14269. doi: 10.1523/jneurosci.5091-14.2015

Newman, C. W., Jacobson, G. P., and Spitzer, J. B. (1996). Development of the tinnitus handicap inventory. *Arch. Otolaryngol. Head Neck Surg.* 122, 143–148. doi: 10.1001/archotol.1996.01890140029007

Rhee, J., Lee, D., Suh, M. W., Lee, J. H., Hong, Y. C., Oh, S. H., et al. (2020). Prevalence, associated factors, and comorbidities of tinnitus in adolescents. *PLoS One* 15:e0236723. doi: 10.1371/journal.pone.0236723

Risoud, M., Hanson, J. N., Gauvrit, F., Renard, C., Lemesre, P. E., Bonne, N. X., et al. (2018). Sound source localization. *Eur Ann Otorhinolaryngol Head Neck Dis.* 135, 259–264. doi: 10.1016/j.anorl.2018.04.009

Ryan, D., and Bauer, C. A. (2016). Neuroscience of tinnitus. *Neuroimaging Clin. N. Am.* 26, 187–196. doi: 10.1016/j.nic.2015.12.001

Shore, S. E., Roberts, L. E., and Langguth, B. (2016). Maladaptive plasticity in tinnitus–triggers, mechanisms and treatment. *Nat. Rev. Neurol.* 12, 150–160. doi: 10.1038/nrneurol.2016.12

Song, K., Shin, S. A., Chang, D. S., and Lee, H. Y. (2018). Audiometric profiles in patients with normal hearing and bilateral or unilateral tinnitus. *Otol. Neurotol.* 39, e416–421. doi: 10.1097/mao.0000000000001849

Tan, C. M., Lecluyse, W., McFerran, D., and Meddis, R. (2013). Tinnitus and patterns of hearing loss. *J. Assoc. Res. Otolaryngol.* 14, 275–282. doi: 10.1007/s10162-013-0371-6

Tollin, D. J., Ruhland, J. L., and Yin, T. C. (2013). The role of spectral composition of sounds on the localization of sound sources by cats. *J. Neurophysiol.* 109, 1658–1668. doi: 10.1152/jn.00358.2012

van der Heijden, K., Rauschecker, J. P., de Gelder, B., and Formisano, E. (2019). Cortical mechanisms of spatial hearing. *Nat. Rev. Neurosci.* 20, 609–623. doi: 10.1038/s41583-019-0206-5

Van Wanrooij, M. M., and Van Opstal, A. J. (2007). Sound localization under perturbed binaural hearing. *J. Neurophysiol.* 97, 715–726. doi: 10.1152/jn.00260.2006

Vernon, J. A., and Fenwick, J. A. (1985). Attempts to suppress tinnitus with transcutaneous electrical stimulation. *Otolaryngol Head Neck Surg.* 93, 385–389. doi: 10.1177/019459988509300318

Voss, P., Tabry, V., and Zatorre, R. J. (2015). Trade-off in the sound localization abilities of early blind individuals between the horizontal and vertical planes. *J. Neurosci.* 35, 6051–6056. doi: 10.1523/jneurosci.4544-14.2015

Weissgerber, T., Müller, C., Stöver, T., and Baumann, U. (2022). Age differences in speech perception in noise and sound localization in individuals with subjective normal hearing. *Front. Psychol.* 13:845285. doi: 10.3389/fpsyg.2022.845285

Weisz, N., Müller, S., Schlee, W., Dohrmann, K., Hartmann, T., and Elbert, T. (2007). The neural code of auditory phantom perception. *J. Neurosci.* 27, 1479–1484. doi: 10.1523/jneurosci.3711-06.2007

Wightman, F. L., and Kistler, D. J. (1997). Monaural sound localization revisited. *J. Acoust. Soc. Am.* 101, 1050–1063. doi: 10.1121/1.418029

Wu, C., Martel, D. T., and Shore, S. E. (2016). Increased synchrony and bursting of dorsal cochlear nucleus fusiform cells correlate with tinnitus. *J. Neurosci.* 36, 2068–2073. doi: 10.1523/jneurosci.3960-15.2016

Xiong, B., Liu, Z., Liu, Q., Peng, Y., Wu, H., Lin, Y., et al. (2019). Missed hearing loss in tinnitus patients with normal audiograms. *Hear Res.* 384, 107826. doi: 10.1016/j.heares.2019.107826

Zeng, F. G., Richardson, M., and Turner, K. (2020). Tinnitus does not interfere with auditory and speech perception. *J. Neurosci.* 40, 6007–6017. doi: 10.1523/jneurosci.0396-20.2020

Zhao, C., Liu, Y., Yang, J., Chen, P., Gao, M., and Zhao, S. (2021). Sound-localisation performance in patients with congenital unilateral microtia and atresia fitted with an active middle ear implant. *Eur. Arch. Otorhinolaryngol.* 278, 31–39. doi: 10.1007/s00405-020-06049-w

frontiers | Frontiers in Neuroscience

TYPE Original Research
PUBLISHED 20 March 2023
DOI 10.3389/fnins.2023.1130892



OPEN ACCESS

EDITED BY
Jun Yan,
University of Calgary, Canada

REVIEWED BY
Ramesh Rajan,
Monash University, Australia
Michael Pecka,
Ludwig Maximilian University of Munich, Germany

*CORRESPONDENCE
Huiming Zhang
✉ hzhang@uwindsor.ca

SPECIALTY SECTION
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

RECEIVED 23 December 2022
ACCEPTED 15 February 2023
PUBLISHED 20 March 2023

CITATION
Asim SA, Tran S, Reynolds N, Sauve O and Zhang H (2023) Spatial-dependent suppressive aftereffect produced by a sound in the rat's inferior colliculus is partially dependent on local inhibition. *Front. Neurosci.* 17:1130892. doi: 10.3389/fnins.2023.1130892



# Spatial-dependent suppressive aftereffect produced by a sound in the rat's inferior colliculus is partially dependent on local inhibition

Syed Anam Asim, Sarah Tran, Nicholas Reynolds, Olivia Sauve and Huiming Zhang*

Department of Biomedical Sciences, University of Windsor, Windsor, ON, Canada

In a natural acoustic environment, a preceding sound can suppress the perception of a succeeding sound which can lead to auditory phenomena such as forward masking and the precedence effect. The degree of suppression is dependent on the relationship between the sounds in sound quality, timing, and location. Correlates of such phenomena exist in sound-elicited activities of neurons in hearing-related brain structures. The present study recorded responses to pairs of leading-trailing sounds from ensembles of neurons in the rat's inferior colliculus. Results indicated that a leading sound produced a suppressive aftereffect on the response to a trailing sound when the two sounds were colocalized at the ear contralateral to the site of recording (i.e., the ear that drives excitatory inputs to the inferior colliculus). The degree of suppression was reduced when the time gap between the two sounds was increased or when the leading sound was relocated to an azimuth at or close to the ipsilateral ear. Local blockage of the type-A γ-aminobutyric acid receptor partially reduced the suppressive aftereffect when a leading sound was at the contralateral ear but not at the ipsilateral ear. Local blockage of the glycine receptor partially reduced the suppressive aftereffect regardless of the location of the leading sound. Results suggest that a sound-elicited suppressive aftereffect in the inferior colliculus is partly dependent on local interaction between excitatory and inhibitory inputs which likely involves those from brainstem structures such as the superior paraolivary nucleus. These results are important for understanding neural mechanisms underlying hearing in a multiple-sound environment.

KEYWORDS

**binaural hearing, inferior colliculus, GABAergic inhibition, glycinergic inhibition, forward masking, free-field stimulation, suppressive aftereffect, sound location**

## Introduction

A natural acoustic environment typically contains multiple qualitatively different sounds that are generated at different time and locations. The perception of one sound can be affected by another sound, with the effect being dependent on the temporal and spatial relationships as well as the qualitative difference between the sounds (Bregman, 1990;

Feng and Ratnam, 2000). For instance, a speech sound can be masked by an interfering sound in a temporospatial-dependent manner (Cherry, 1953; Bronkhorst, 2000; Shinn-Cunningham et al., 2001; Arbogast et al., 2002; Jones and Litovsky, 2011; Warzybok et al., 2013; Yost, 2017).

Responses to multiple sounds have been recorded in brain structures to study neural bases of auditory perception in a natural acoustic environment. In the inferior colliculus (IC, a midbrain nucleus), neurons display two-tone suppression when two tone bursts are simultaneously presented (Zhang and Feng, 1998; Egorova et al., 2001; Alkhatib et al., 2006; Hurley et al., 2008). Action-potential firing elicited by a tone burst at a neuron's characteristic frequency (CF, the frequency at which the threshold of response is the lowest) is suppressed by another tone burst at a different frequency. When two sounds are temporally separated, the response to a trailing sound can be suppressed by a leading sound with the degree of suppression being dependent on the time gap between the sounds (Finlayson and Adam, 1997; Faure et al., 2003; Mei et al., 2006; Nelson et al., 2009; Zhang and Kelly, 2009; Singheiser et al., 2012; Gai, 2016). When two simultaneously presented sounds are spatially separated, the response to a sound with a fixed location can be suppressed by a relocated sound, with the effect being dependent on the angle of separation (Ratnam and Feng, 1998; Lane and Delgutte, 2005; Day et al., 2012). A suppressive effect can be observed even when two sounds are separated both temporally and spatially (Yin, 1994; Litovsky and Yin, 1998a,b; Tolnai et al., 2017; Chot et al., 2019, 2020). Knowledge about the neurobiological bases of such suppressive effects can greatly help us understand hearing in a natural acoustic environment.

The strength of the sound-driven response of an auditory neuron is dependent on excitatory/inhibitory inputs received by the neuron. For most IC neurons, major excitatory inputs are from the contralateral cochlear nucleus and lateral superior olivary nucleus (LSO), and the ipsilateral medial superior olivary nucleus (Loftus et al., 2004; Cant, 2005). Major inhibitory inputs are from the contralateral dorsal nucleus of the lateral lemniscus (DNLL) and the ipsilateral LSO, superior paraolivary nucleus (SPN), and dorsal as well as ventral nuclei of the lateral lemniscus (VNLL) (Helfert et al., 1989; Saint Marie et al., 1989; González-Hernández et al., 1996; Zhang et al., 1998; Kulesza and Berrebi, 2000; Loftus et al., 2004; Saldaña et al., 2009).

The dependence of sound-driven responses of IC neurons on excitatory/inhibitory inputs is supported by neurophysiological results (Xie et al., 2008; Liu et al., 2019). Characteristics of responses such as two-tone suppression can be affected by local blockage of inhibition (Hurley et al., 2008). When a sound that is presented in a free acoustic field is relocated from the contralateral to the ipsilateral ear, many neurons reduce firing over the duration of the sound due to decreased excitation and/or increased inhibition (Park and Pollak, 1994; Zhang et al., 1999; Poirier et al., 2003). In case another sound with a fixed location is presented simultaneously, such changes in excitation and/or inhibition can conceivably affect the responses of neurons to the second sound. A sound can activate inhibitory inputs to IC neurons even when the sound is presented at the contralateral ear. Such inhibitory inputs can be provided by the ipsilateral SPN (Saldaña et al., 2009). As neurons in this structure receive inputs from the contralateral cochlear nucleus and fire action potentials at the offset of a sound, outputs from these neurons to the IC can allow a contralaterally presented sound to suppress responses of IC neurons to a subsequent sound (Schofield, 1995; Kulesza et al., 2003; Kadner and Berrebi, 2008; Felix et al., 2015, 2017; Gai, 2016; Salimi et al., 2017). An inhibitory aftereffect can also be generated on IC neurons through the activation of neurons in the dorsal nucleus of the lateral lemniscus (Burger and Pollak, 2001; Pecka et al., 2007). Further studies have yet to be conducted to characterize the temporal and spatial dependences of the suppressive aftereffect produced by a sound on IC neurons.

In the present study, we used a pair of leading and trailing sounds to elicit local-field potentials (LFPs) in the IC. We studied how the suppressive aftereffect produced by a leading sound was dependent on the temporal and spatial separations between the two sounds. We used pharmacological agents to block $GABA_A$ and glycine receptors in the IC to study how local inhibition contributed to the suppressive aftereffect.

# Materials and methods

## Animal preparation

Eighteen adult male Wistar albino rats (*Rattus norvegicus*) were obtained from Charles River Canada Inc. (St. Constant, QC, Canada). They were 210–600 g when experiments were conducted. Surgical anesthesia was induced by ketamine hydrochloride (60 mg/kg, i.m.) and xylazine hydrochloride (10 mg/kg, i.m.) and maintained by ketamine hydrochloride (20 mg/kg, i.m.) and xylazine hydrochloride (3.3 mg/kg, i.m.).

A craniotomy was made in the skull over the right temporal lobe for placing a recording electrode into the IC. The skull was cemented onto a head bar attached to a custom-made holding instrument. A recording electrode was held by a custom-made clamp attached to the slave cylinder of a Model 650 micropositioner, which was fitted onto a Model 900 stereotaxic instrument (Kopf Instruments, Tujunga, CA, USA). The rat, along with the head-holding and electrode positioning instruments, was placed in a Model CL-15A LP acoustic chamber (Eckel Industries, Morrisburg, ON, Canada) when sound-driven responses were recorded. Instruments were positioned in such a way that acoustic shadows and reflections were minimized. All procedures were approved by the University of Windsor Animal Care Committee in accordance with the guidelines of the Canadian Council on Animal Care.

## Acoustic stimulation

Sound waveforms were generated using a System 3 real-time signal processing system controlled by a personal computer

Abbreviations: CF, Characteristic frequency; DNLL, Dorsal nucleus of the lateral lemniscus; GABA, γ-aminobutyric acid; $GABA_A$ receptor, Type-A γ-aminobutyric acid receptor; IC, Inferior colliculus; ISI, Inter-stimulus interval; LFP, Local-field potential; LSO, Lateral superior olivary nucleus; NAR, Normalized amplitude of response; SPN, Superior paraolivary nucleus; VNLL, Ventral nucleus of the lateral lemniscus.

running the OpenEx software (Tucker-Davis Technologies, Alachua, FL, USA). Sounds were presented using two FF1 free-field speakers (Tucker-Davis Technologies, Alachua, FL, USA) that could be positioned at any azimuthal location that was 50 cm away from the midpoint of the interaural line of a rat. Each speaker was calibrated over the range between 100 and 40,000 Hz using a model 2608 measuring amplifier and a model 4135 microphone (Brüel & Kjaer, Dorval, QC, Canada). The microphone was positioned at the location where the midpoint of the interaural line would be. Calibration was conducted with the speaker at each of the five locations used in the present study, which included the midline of the frontal field and 90° and 45° on the contra- and ipsilateral side of the recording site (**Figure 1A**, denoted by 0°, c90°, c45°, i45°, and i90°).

## Recording electrode and procedures

A "piggy-back" multi-barrel electrode assembly was used to record an LFP from an ensemble of neurons in the IC and to release neuropharmacological agents at the site of recording (Patel and Zhang, 2014). The electrode assembly consisted of a single-barrel recording glass pipette (filled with 0.5 M sodium acetate with 3% Chicago Sky Blue, tip diameter ∼15 μm, impedance 100–300 kΩ) and a 5-barrel "H-configuration" drug-release glass pipette (tip diameter 15–20 μm). The tip of the single-barrel pipette protruded beyond the tip of the multi-barrel pipette by 20–25 μm. Each of the side barrels of the 5-barrel pipette was filled with the $GABA_A$ receptor antagonist Gabazine or the glycine receptor antagonist strychnine (both 5 mM in physiological saline, pH 3.5). The central barrel was filled with physiological saline for the balance of electrical current.

Signals registered by the recording pipette were amplified by 1,000 times and bandpass filtered (0.3–300 Hz) by a model 2,400A preamplifier (Dagan, Minneapolis, MN, USA). The signals were sampled at 3.1 kHz using the System 3 real-time signal processing system. The five barrels of the drug-release electrode were connected to a BH-2 Neurophore microiontophoretic system (Harvard Apparatus, Holliston, MA, USA). To prevent drug leakage, a current around −20 nA was applied to each of the side barrels when a pharmacological agent was not being released.

An electrode assembly was inserted into the right IC while Gaussian noise bursts at 60 dB SPL were presented from a loudspeaker at c90° to search for a site of recording. Recordings were conducted only on one side of the IC to minimize the utilization of electrode holding devices and avoid repositioning of holding devices, which helped maintain the consistency of an acoustic field. Upon identification of a recording site, the CF and the threshold at CF were determined for the site using tone bursts presented at c90°. These tone bursts had an 8-ms duration (4-ms rising/falling phases, no plateau).

A train of leading-trailing tone-burst pairs (see **Figure 1B**) was created for the recording site using two 8-ms tone bursts. A leading sound was randomly chosen from two tone bursts named $T_L$ and $T_H$. The frequencies of $T_L$ and $T_H$ were $f_L$ (lower than CF) and $f_H$ (higher than CF) and were calculated based on the CF of the recording site. The center frequency of $f_L$ and $f_H$ [i.e., $(f_L \times f_H)^{1/2}$] was at CF while the frequency difference between $f_L$ and $f_H$ was 10% of the center frequency [i.e., $(f_H - f_L)/(f_H \times f_L)^{1/2} = 0.1$]. A trailing sound was a tone burst with a frequency at the CF of the recording site. The leading and trailing sounds had the same intensity, which was typically 30 dB above the threshold at CF. The interval between onsets of the two sounds (i.e., inter-stimulus interval or ISI) could be varied systematically. Each train had 60 pairs of leading-trailing tone bursts. The interval between the offset of a trailing sound of one pair and the onset of a leading sound of a subsequent pair was 1,500 ms.

Responses to a train of leading-trailing sound pairs were first recorded when the two sounds were colocalized at c90° (**Figure 1A** left panel). The ISI between leading and trailing sounds was systematically changed to examine how the suppressive effect of the leading sound was dependent on the temporal separation between the two sounds (i.e., the time course of the aftereffect). Recordings were then conducted when the leading sound was at other azimuths (i.e., i90°, c45°, i45°, and 0°) while the trailing sound remained at c90° to examine how the suppressive effect of the leading sound was dependent on the spatial separation between the sounds (see **Figure 1A** right panel for leading sound at c45°). At each angle of separation, ISI between the leading and trailing sounds was systemically varied. Sounds that were used to create a train of leading-trailing pairs were presented individually to elicit responses. These responses were used as references to evaluate whether/how the leading and training sounds in a train of sound pairs affected each other in eliciting responses. For this purpose, the response to a leading sound was recorded when the sound was at each of the five azimuths (i.e., c90°, i90°, c45°, i45°, and 0°) and the response to a trailing sound was recorded when the sound was at c90°.

Gabazine or strychnine was released microiontophoretically (current of injection at +3∼+140 nA) at the recording site to study whether/how the suppressive effect generated by a leading sound was dependent on local inhibition in the IC. The effect of a pharmacological agent was monitored by repeatedly recording responses to a leading and/or a trailing sound presented individually at c90°. For this purpose, a tone burst was presented at multiple (typically 7) intensities ranging from slightly below the threshold at the CF to well above the threshold (including the intensity used for creating a sound pair). At each intensity a tone burst was presented 20 times; and all presentations (i.e., 140 presentations for 7 intensities) had a random order. The rate of sound presentation was 1/sec. Responses elicited at different intensities were used to create an amplitude-intensity function (AIF, see data analysis for more details) and a latency-intensity function. When the effect of a drug on responses to individual sounds reached a stabilized level, responses to a train of sound pairs were recorded at various ISIs and angles of separation. AIFs for responses to the leading and the trailing sounds were also recorded. Following the evaluation of the effect of a drug, the injection current was terminated and a retention current was reapplied. Recovery was monitored by repeatedly recording responses to a leading or a trailing sound presented individually at c90°.

## Data analysis

Neural signals collected over presentations of leading and trailing sounds in a train of paired stimuli were averaged separately

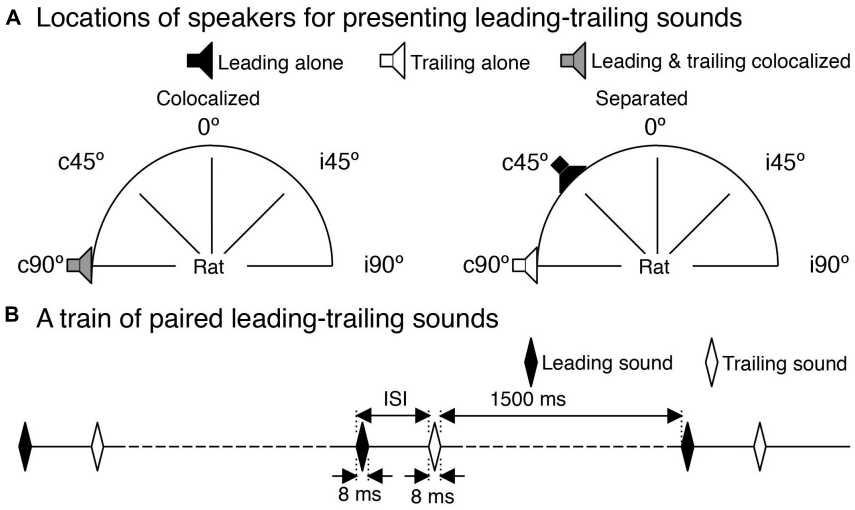



FIGURE 1
Speaker locations and a train of paired leading-trailing sounds. **(A)** Five azimuthal locations (c90°, c45°, 0°, i45°, and i90°) used in sound presentations. A pair of leading-trailing tone bursts were either colocalized at c90° (left panel) or separated with a leading sound being at a non-c90° azimuth while a trailing sound being at c90° (right panel for a leading sound at c45°). Two speakers used in the study were calibrated at these azimuths. **(B)** A train of leading-trailing tone-burst pairs. In panels **(A,B)**, a leading sound or a speaker that was used to present a leading sound is indicated by a black color. A trailing sound or a speaker that was used to present a trailing sound is indicated by a white color. A speaker that was used to present both a leading and a trailing sound (colocalized at c90°) is indicated by a gray color.

to obtain two mean waveforms for LFPs elicited by the two sounds, respectively. Signals collected over presentations of a single tone burst at multiple intensities (see the last paragraph of "*Recording electrode and procedures*") were averaged separately to obtain mean waveforms for LFPs elicited at these intensities, respectively. In agreement with our previous result (Patel et al., 2012), an LFP had a dominant negative peak that was preceded by a small positive peak and followed by an intermediate positive peak (**Figure 2A**). The amplitude and latency of the negative peak of an LFP were measured.

We evaluated whether the response to a sound (either leading or trailing) in a sound pair was affected by the other sound in the pair and how the response was dependent on the spatial location of a leading sound as well as the temporal separation between leading and trailing sounds. An index of the normalized amplitude of response (NAR) was calculated for these purposes:

$$NAR = A_{paired}(\theta, \Delta t)/A_{alone}(c90^\circ)$$

Where $A_{paired}$ ($\theta$, $\Delta$t) is the amplitude of the response to the sound under evaluation obtained when the sound was presented in a sound pair, with the leading and trailing sounds being at azimuth $\theta$ and c90°, respectively, and the interval between the two sounds being at $\Delta$t. $A_{alone}$ (c90°) is the amplitude of the response to the sound under evaluation obtained when the sound was presented alone at c90°. A NAR value of 0 indicates that the response to a sound was completely suppressed in comparison to the response to the same sound presented alone at c90°, while a NAR value of 1 indicates that the response to a sound was not affected.

To determine whether a pharmacological agent affected local inhibition at the site of recording, an AIF was created using amplitudes of responses elicited by a single sound presented at a single azimuth but various intensities (see the first paragraph of "*Data analysis*"). AIFs obtained before and during the application of the drug were compared. When a change in AIF produced by a drug stabilized, NAR values were calculated for responses elicited by leading and trailing sounds in a sound pair at each combination of temporal and angular separations. NAR values for responses to a trailing sound obtained before and during drug application were compared to find how the drug influenced a suppressive aftereffect.

Statistical analysis was conducted using the SPSS (version 23) software (IBM Corporation, Armonk, NY, USA).

## Results

LFPs elicited by single tone bursts and pairs of leading-trailing tone bursts were recorded in the right IC in each of the 18 rats used in the present study. The rats formed two groups, with $T_L$ presented as a leading sound in one group ($n$ = 9) while $T_H$ presented as a leading sound in another group ($n$ = 9). Data obtained from the two groups were not different from each other (not shown). Thus, the data are combined in this article.

### Response to a leading sound was not affected by a trailing sound over a train of sound pairs

Shown in **Figure 2** are LFPs obtained from the IC of an example rat when a leading and a trailing sound were colocalized at c90°. Regardless of the ISI between the two sounds, the waveform (including amplitude and latency) of the LFP elicited by a leading sound was very similar to that of the LFP elicited by the same sound presented individually at c90° (**Figure 2A1**). This similarity was observed at all combinations of ISI and angle of separation between

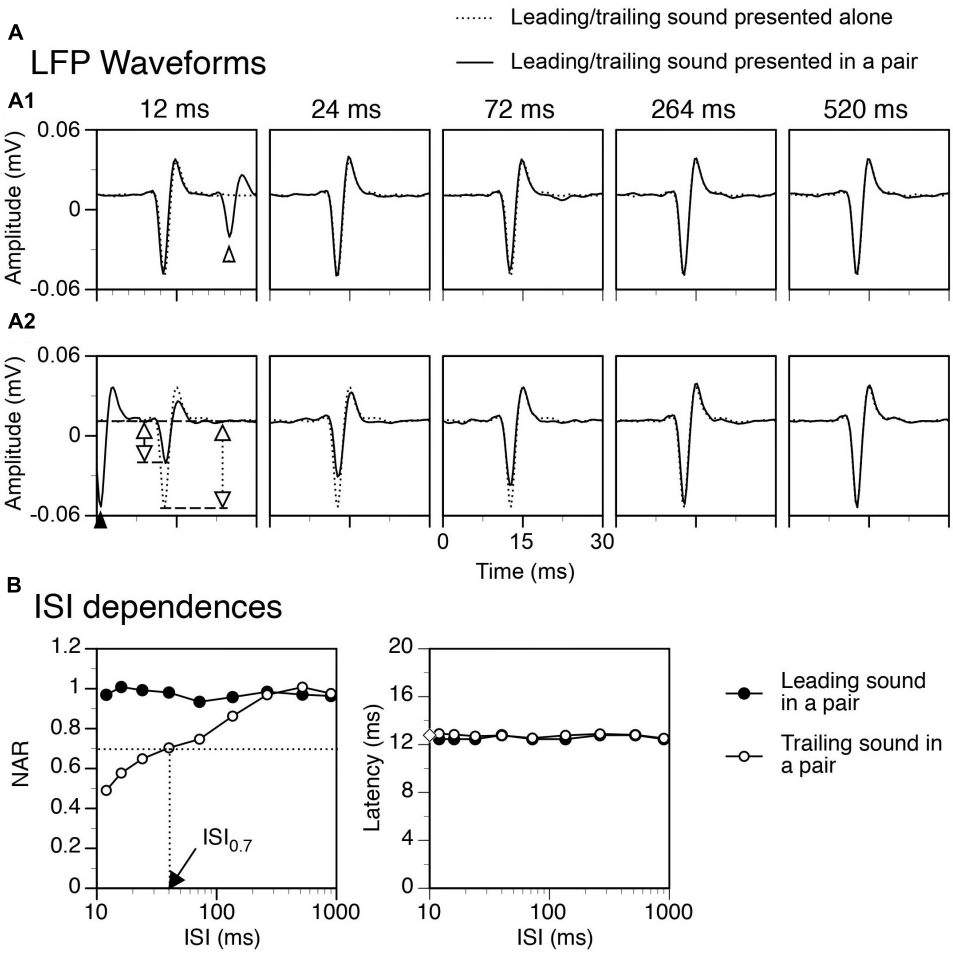



FIGURE 2
An example showing that a temporal separation between a leading and a trailing tone burst reduced the suppressive aftereffect produced by the leading sound. Results were obtained when the two sounds were colocalized at c90°. **(A)** Waveforms of local-field potentials (LFPs) elicited by a leading sound [solid lines in **(A1)**] and a trailing sound [solid lines in **(A2)**] in a leading-trailing sound pair. The waveforms shown in two corresponding panels in **(A1)** and **(A2)** were obtained at the same inter-stimulus interval (ISI) [indicated above a panel in **(A1)**]. The start of the x axis (0 ms) corresponds to the onset of a leading sound in each panel of **(A1)** and the onset of a trailing sound in each panel of **(A2)**. A part of the LFP elicited by a trailing sound is shown in the left panel of **(A1)** (pointed by an open upward triangle), while a part of the LFP elicited by a leading sound is shown in the left panel of **(A2)** (pointed by a filled upward triangle). LFPs elicited by the leading and the trailing sound when they were presented individually are shown in **(A1)** and **(A2)** (dotted lines) for comparison. In the left panel of **(A2)**, two vertical double arrows indicate amplitudes of the negative peak of LFP elicited by a trailing sound presented individually (dotted) and in a sound pair (solid). **(B)** Dependences of normalized amplitude of response (NAR, left panel) and latency (right panel) on the ISI for LFPs elicited by a leading (filled circle) and a trailing (open circle) sound. In the left panel, a horizontal dotted line indicates the level of NAR at 0.7. The vertical dotted line indicates the ISI associated with the 0.7 NAR. In the right panel, an open diamond on the Y-axis indicates the latency of the response to a trailing sound presented alone. The characteristic frequency (CF) of the recording site (hence the frequency of the trailing sound) was 13.0 kHz. The frequency of the leading sound was 13.458 kHz.

the two sounds and supported by group results from 18 animals. Shown in Figure 3 are group results obtained when two sounds were colocalized at c90° (black bars for responses to a leading sound, One-way ANOVA, $F = 0.410$, $p = 0.894$ for amplitude, and $F = 0.111$, $p = 0.998$ for latency, respectively). Thus, over a train of leading-trailing sound pairs the response to a leading sound was not affected by preceding presentations of a trailing sound.

## Suppression of the response to a trailing sound by a colocalized leading sound

The response to a trailing tone burst at c90° was greatly suppressed by a colocalized leading tone burst when the two sounds were separated by a small ISI (see Figure 2A2 left panel for an example). The degree of suppression was reduced when the ISI was increased (Figure 2A2 panels 2–5 and Figure 2B left panel). The NAR value was 0.7 (i.e., the amplitude of the response was suppressed by 30%) when ISI was close to 40 ms (Figure 2B left panel). Such an ISI value, namely $ISI_{0.7}$, was used in the article to reflect the duration of suppression. When the ISI was 264 ms, the response to a trailing sound was almost identical to that elicited by the sound presented alone. The latency of the response to a trailing sound was not affected by a leading sound at any ISIs (Figures 2A2, B right panel). Results from the entire group of animals confirmed that when leading and trailing sounds were colocalized at c90° the amplitude of the response to a trailing sound was dependent on the ISI while the latency of the response was not

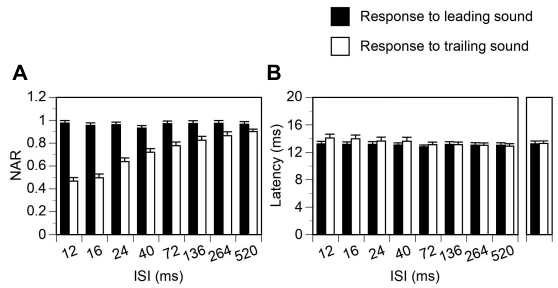



FIGURE 3
Group results showing that temporal separation between a leading and a trailing tone burst reduced the suppression of the response to a trailing sound but did not affect the response to a leading sound. Results were obtained when leading and trailing sounds were colocalized at c90°. **(A,B)** Show effects of separation on normalized amplitude of responses (NARs) and latencies of local-field potentials (LFPs), respectively. Latencies of LFPs elicited by a leading and a trailing sound when they were presented alone at c90° are shown on the right of panel **(B)**. An error bar indicates a standard error of the mean.

(Figure 3 white bars, One-way ANOVA, $F = 29.168$, $p < 0.001$ for amplitude and $F = 1.084$, $p = 0.377$ for latency).

## Relocation of a leading sound changed both responses to leading and trailing sounds

Relocating a leading sound from c90° to another azimuth reduced the amplitude of the response to the sound, as shown by an example in Figure 4 (Figure 4A and the left panel of Figure 4B, results obtained at an ISI at 24 ms). Such a reduction was not caused by presentations of a trailing sound in a train of sound pairs (see first subsection of "Results"). Rather, it reflected a direction-dependence of the response to the leading sound. A reduction of the response to a leading sound was accompanied by a reduction of suppression of the response to a trailing sound, which remained at c90°. The NAR value for the response was slightly below 0.50 when the leading sound was at c90° but above 0.55 when the leading sound was at i90°. For both responses to the leading and trailing sounds, the latency was minimally affected by the relocation of a leading sound (Figure 4B right panel).

Group results obtained when the ISI was at 24 ms confirmed that both NAR values for responses to a leading and a trailing sound were dependent on the location of the leading sound (Figures 5A, B left panels, One-way ANOVA, $F = 13.917$, $p < 0.001$ for response to leading sound, $F = 3.384$, $p = 0.014$ for response to trailing sound). The Tukey's test revealed that the NAR value for the response to a leading sound was significantly reduced when the sound was relocated from c90° to i45° ($p = 0.001$) or i90° ($p < 0.001$). The same test revealed that the NAR value for the response to a trailing sound was significantly increased when the leading sound was relocated from c90° to i90° ($p = 0.021$). Latencies of responses to leading and trailing sounds were not changed by the relocation of a leading sound (One-way ANOVA, $F = 1.464$, $p = 0.221$ for Figure 5A right panel and $F = 1.432$, $p = 0.233$ for Figure 5B right panel).

Time courses of suppression produced by a leading sound at c90° and i90° were compared in each animal. Shown in Figure 6A left panel are NAR-ISI curves for responses to a trailing sound obtained from an example animal. The degree of suppression at short ISIs was higher and suppression lasted longer when the leading sound was at c90° than at i90°. In agreement with these differences, the $ISI_{0.7}$ value was larger when the leading sound was at c90° than at i90°. Group results from 18 animals supported the difference between the suppressive effects produced by a leading sound at c90° and at i90° (Figure 6B1 top panel, two-way ANOVA, $F = 53.686$, $p < 0.001$). A *post-hoc* *t*-test indicated that the difference was significant at all ISIs between 12 and 72 ms (see Figure 6 caption for statistical results). The $ISI_{0.7}$ value was larger when the sound was at c90° than i90° (Figure 6B2, Paired *t*-test, $t = 3.230$, $p = 0.004$). The latency of the response to a trailing sound was not affected by the ISI and the location of a leading sound (see Figure 6A right panel for an example and Figure 6B1 bottom panel for group results, two-way ANOVA, $F = 0.365$, $p = 0.546$).

## Effects of pharmacological manipulations

Gabazine released at the site of recording enhanced both LFPs elicited by a leading and a trailing sound when they were presented in a pair (Figure 7A). It also enhanced responses to the sounds when they were presented individually (Figure 7B for response to a trailing sound). For the example shown in Figure 7, the suppressive aftereffect produced by a leading sound was reduced (i.e., the NAR value for the response to a trailing sound was increased) by gabazine at short ISIs when the leading sound was at c90° (Figure 7C left panel). The aftereffect was minimally changed by the drug when the leading sound was at i90° (Figure 7C right panel).

Group results ($n = 8$) support that gabazine changed the degree of suppression when a leading sound was at c90° (Figure 8A left panel, Two-way ANOVA, $F = 6.596$, $p = 0.012$). A *post-hoc* *t*-test indicated that the reduction was significant at the level of $p < 0.1$ when the ISI was at 40 ms ($p = 0.063$). An increase of NAR value at short ISIs was accompanied by a decrease of $ISI_{0.7}$ value at a statistical significance level at $p < 0.1$ (Figure 8B, Paired *t*-Test, $t = 1.958$, $p = 0.091$). When a leading sound was at i90°, gabazine did not change the degree of suppression (Figure 8A right panel, Two-way ANOVA, $F = 1.078$, $p = 0.305$). It did not change the $ISI_{70}$, either (Paired *t*-test, $t = 0.554$, $p = 0.604$).

Strychnine increased amplitudes of both responses to a leading and a trailing sound no matter whether the sounds were presented in a pair (Figure 9A) or individually (Figure 9B for the response to a trailing sound). NAR-ISI functions revealed that the suppressive effect of a leading sound was moderately reduced over a wide range of ISIs both when a leading sound was at c90° and at i90° (Figure 9C left and right panels).

Group results ($n = 7$) indicate that strychnine reduced the degree of suppression generated by a leading sound when the sound was at c90° (Two-way ANOVA, $F = 48.492$, $p < 0.001$) and

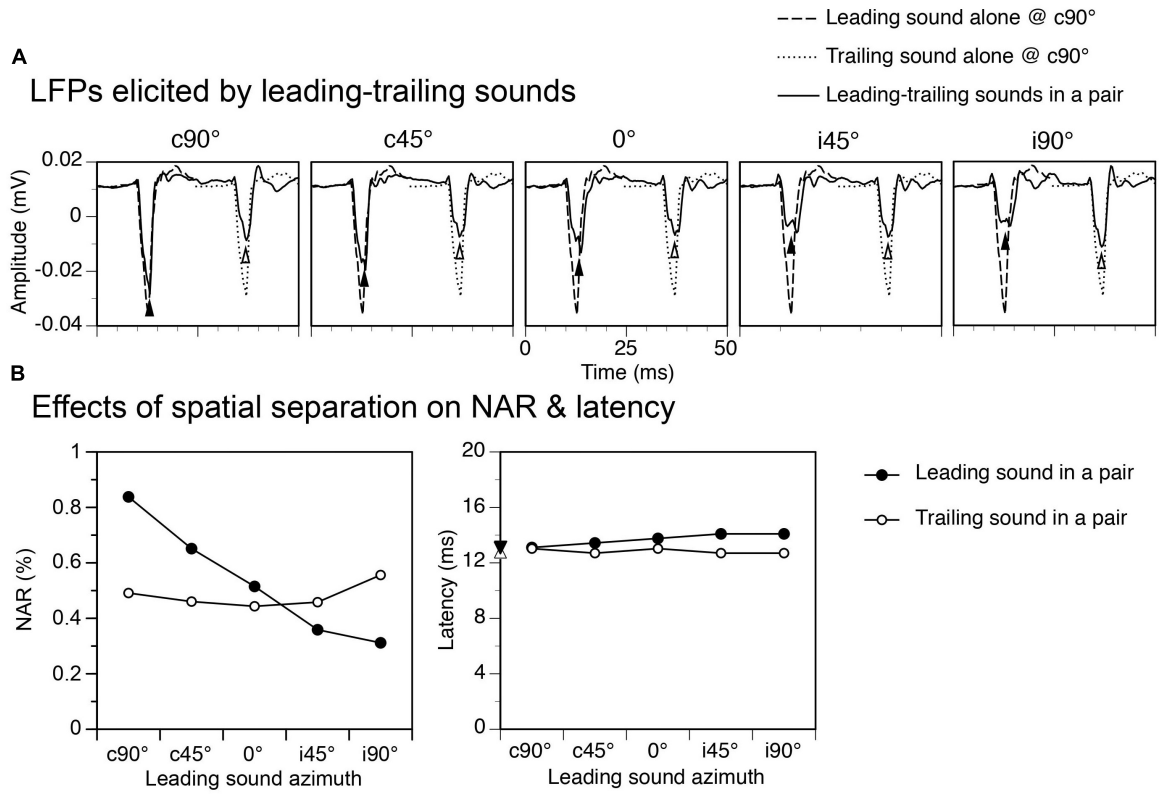



FIGURE 4
An example showing effects of spatial separation between a leading and a trailing tone burst on the local-field potentials (LFPs) elicited by the sounds. The time interval between leading and trailing tone bursts was 24 ms. **(A)** Waveforms of LFPs elicited by leading and trailing tone bursts. Results were obtained when a leading tone burst was at c90°, c45°, 0°, i45°, and i90° (indicated above each panel) while a trailing tone burst was at a fixed location at c90°. The start of the x axis (0 ms) corresponds to the onset of a leading sound in each panel of **(A)**. Filled and open upward triangles indicate LFPs elicited by the leading and the trailing sound, respectively. Waveforms of LFPs elicited by the leading and the trailing tone burst presented alone at c90° are shown in each panel for comparison. **(B)** Line charts showing effects of spatial separation between leading and trailing sounds on normalized amplitude of responses (NARs) (left panel) and latencies (right panel) of LFPs elicited by the sounds. Filled downward and open upward triangles on the y-axis of the right panel indicate the latencies of the responses elicited by a leading and a trailing sound alone at c90°, respectively. The characteristic frequency (CF) of the recording site (hence the frequency of the trailing sound) was 5.5 kHz. The frequency of the leading sound was 5.694 kHz.

when it was at i90° (Two-way ANOVA, $F = 16.650$, $p < 0.001$). A significant drug-induced change was observed at all ISIs between 16 and 264 ms when a leading sound was at c90° and at ISIs at 24 and 40 ms when a leading sound was at i90° (*Post-hoc t*-test, see **Figure 10** caption for statistical results). A reduced degree of suppression was accompanied by a decreased $ISI_{0.7}$ value when a leading sound was at c90° (Paired *t*-test, $t = 3.475$, $p = 0.013$) and i90° ($t = 3.024$, $p = 0.023$).

# Discussion

The present study revealed that a leading sound could produce an aftereffect to suppress an LFP elicited by a trailing sound in the IC. Such an aftereffect lasted for up to a few hundred milliseconds when the two sounds were co-localized at the ear contralateral to the site of recording. The degree and duration of suppression were reduced when a leading sound was relocated to an ipsilateral azimuth. The suppressive aftereffect was partially reduced by local application of $GABA_A$ and glycine receptor antagonists.

## Time course and directional dependence of a sound-elicited suppressive aftereffect

A sound-elicited suppressive aftereffect in the IC similar to that revealed by the present study was reported by previous papers (Park and Pollak, 1993; Covey et al., 1996; Finlayson and Adam, 1997; Litovsky and Yin, 1998a,b; Faure et al., 2003; Mei et al., 2006; Nelson et al., 2009; Singheiser et al., 2012). Despite similarities, quantitative differences existed among different studies in the time course of suppression. The time course revealed by the present study tended to be longer than those found by some other studies (e.g., Park and Pollak, 1993; Covey et al., 1996).

Factors causing differences in the time course of suppression likely include the type of acoustic field in which a sound is presented and parameters of the sound. When a sound is presented in a free field (e.g., the present study), both ears are stimulated which leads to the activation of neurons in all structures that provide inputs to the IC. These source structures include those driven by the contralateral ear only and those driven by both ears

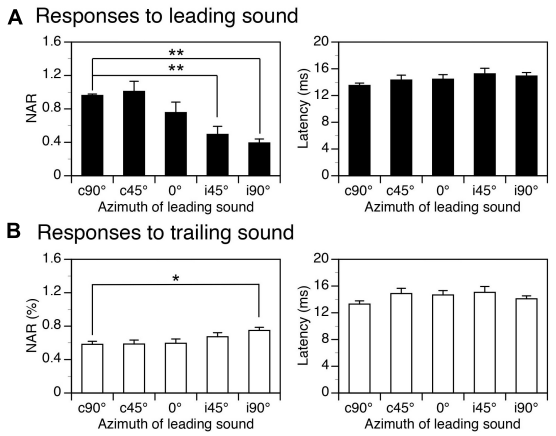



FIGURE 5
Group results showing effects of spatial separation between a leading and a trailing tone burst on the amplitudes and latencies of local-field potentials (LFPs) elicited by the sounds. **(A,B)** Bar charts based on responses to a leading and a trailing sound, respectively. In both panels **(A,B)**, the left and right panels show effects of spatial separation on the normalized amplitude of response (NAR) and the latency, respectively. An error bar indicates a standard error of the mean. "*" and "**" indicate statistical significance (Tukey's test) at the level of 0.05 and 0.005, respectively.

(Cant, 2005; Schofield, 2005). When a sound is presented only to the ear contralateral to the IC in a closed field (most previous studies), it can also activate neurons in both types of structures. However, activities of neurons in binaurally driven structures can be different from those elicited by the sound presented in a free field. Thus, inputs to the IC and responses of neurons in the structure can be different under the two conditions. Previous results have demonstrated that an ipsilaterally presented leading sound can reduce the firing elicited by a contralaterally presented trailing sound in some IC neurons (Zhang and Kelly, 2009). These results support that a sound presented in a free acoustic field (e.g., the present study) generates a stronger suppressive aftereffect than when the sound is presented in a closed field to the contralateral ear alone. Other factors causing differences between our and other studies in the time course of suppression likely include acoustic parameters such as the level (Faure et al., 2003; Nelson et al., 2009) and the duration of a sound (Faure et al., 2003).

The time course of a suppressive aftereffect may also be dependent on the type of neurophysiological signal that is used to reflect the aftereffect. In the present study, LFPs generated by a population of neurons rather than action potential discharges generated by single neurons were registered. It is generally believed that the generation of an LFP depends on current dipoles produced nearby a recording electrode, which primarily reflect postsynaptic potentials created by neurons (Mitzdorf, 1985; Logothetis, 2003; Burkard et al., 2006; Goense and Logothetis, 2008). It is conceivable that a disparity exists between suppressive aftereffects reflected by the amplitude of LFP and by the strength of action potential firing, as differences exist between neurophysiological processes underlying these signals. Differences in morphological/ neurophysiological/ biophysical properties exist among IC neurons. These differences may lead to dissimilarities among the neurons when they generate current dipoles. The contribution of dipoles generated by different neurons to an LFP is dependent on the spatial relationship between the neurons and a recording electrode. These factors along with variations in the site of recording across different animals can lead to differences among the animals in the waveform of LFP and consequently in the suppressive aftereffect reflected by the electric potential.

Among the major findings of the present study was the directional dependence of the suppressive aftereffect generated by a sound. Such an effect was the strongest when a sound was presented at the ear contralateral to the IC (**Figures 4–6**). This result agrees with findings from previous studies showing that in most individual IC neurons a sound produced the strongest suppressive aftereffect when it was presented at locations where it could elicit the strongest excitatory responses over its duration (Litovsky, 1998; Litovsky and Yin, 1998a,b). For most IC neurons, these locations were at or near c90° (Litovsky and Yin, 1998b; Kuwada et al., 2011; Chot et al., 2019, 2020). Such space tuning explains results from the present study showing that a sound elicited the largest LFP as well as the strongest suppressive aftereffect when it was presented at c90°.

The peak latency of an LFP elicited by a trailing sound was not substantially affected by the temporal and spatial relationship between a leading and a trailing sound. This might be because the latency was primarily dependent on the excitatory inputs that drove the LFP response (e.g., those from the contralateral cochlear nucleus) (Cant, 2005). Latencies of such inputs may not be substantially affected by a leading sound.

## Mechanisms responsible for the sound-elicited suppressive aftereffect in the IC

Local application of Gabazine or strychnine reduced the suppressive aftereffect produced by a leading sound (**Figures 7–10**). During application of Gabazine, a reduction was observed only when the sound was presented at the contralateral ear. During application of strychnine, the reduction appeared to be stronger when the sound was at the contralateral than at the ipsilateral ear. Thus, both local GABA- and glycinergic inhibition contributed to the suppressive aftereffect produced by a sound in the IC in a direction-dependent manner.

GABAergic innervations received by the IC include those from the ipsilateral SPN, which is driven by excitatory inputs from the contralateral cochlear nucleus (Schofield, 1995; Kulesza and Berrebi, 2000; Saldaña et al., 2009; Felix et al., 2017). This pathway along with offset patterns of firing displayed by SPN neurons allows the stimulation of the contralateral ear to produce a suppressive aftereffect on IC neurons (Kulesza et al., 2003; Kadner and Berrebi, 2008; Gai, 2016; Salimi et al., 2017). Our results obtained during application of Gabazine (**Figures 7, 8**) agree with these facts. Structures providing GABAergic inputs to the IC also include the ipsilateral DNLL and VNLL as well as the contralateral DNLL (Vater et al., 1992; González-Hernández et al., 1996; Zhang et al., 1998). Each of these source structures is driven by the ear on its opposite side (Markovitz and Pollak, 1994; Wu, 1999; Batra and Fitzpatrick, 2002; Zhang and Kelly, 2006). Additionally, the DNLL is inhibited by stimulation of the ear on its same side (Markovitz and Pollak, 1994). The lack of effect of Gabazine on

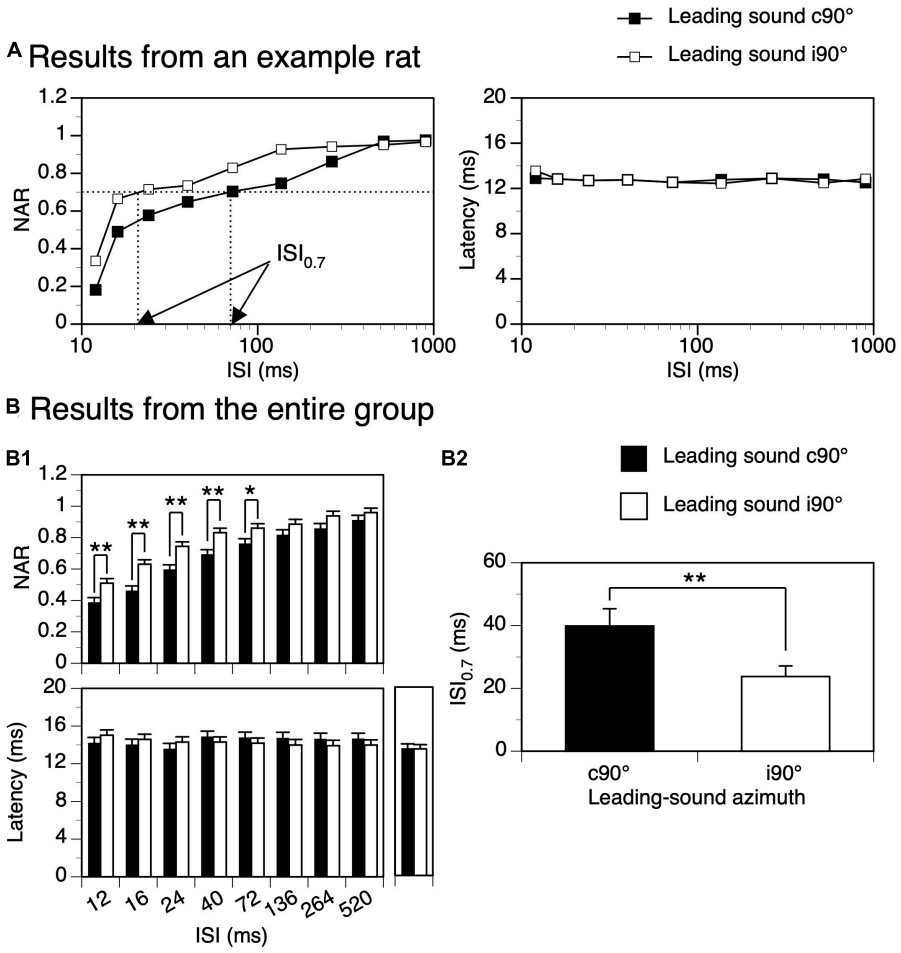



FIGURE 6
Difference between suppressive aftereffects produced by a leading tone burst presented at c90° and i90°. **(A)** Results from an example rat showing normalized amplitude of response (NAR)-inter-stimulus interval (ISI) (left panel) and latency-ISI (right panel) relationships for the response to a trailing sound obtained when the leading sound was at c90° and i90°, respectively. In the left panel, two arrows point toward the $ISI_{0.7}$ values obtained when the leading sound was at the two azimuths. The characteristic frequency (CF) of the recording site (hence the frequency of the trailing sound) was 13.0 kHz. The frequency of the leading sound was 13.458 kHz. **(B)** Results from the entire group of 18 rats comparing effects generated by a leading sound at c90° and i90°. (B1) Top and bottom panels show effects of a leading sound on the NAR and latency of the response to a trailing sound, respectively. A *post-hoc t*-test (see text for the result from a two-way ANOVA test) indicates that the difference in the suppressive effect was significant at all ISIs between 12 and 72 ms ($p = 0.004$ at 12 ms; $p < 0.001$ at 16 ms; $p < 0.001$ at 24 ms; $p = 0.001$ at 40 ms; $p = 0.020$ at 72 ms). Before responses to a leading-trailing sound pair were recorded at each angle of separation (c90° and i90°), the response to a trailing sound presented alone was recorded. Latencies of responses elicited by a trailing sound under these conditions are shown on the right side of the bottom panel. (B2) Bar chart comparing $ISI_{0.7}$ values obtained when the leading sound was at c90° and i90°. An error bar indicates a standard error of the mean. "*" and "**" indicate statistical significance at the level of 0.05 and 0.005, respectively.

the aftereffect produced by a sound presented at the ear ipsilateral to the IC (**Figures 7**, **8**) suggests that GABAergic inputs from the contralateral DNLL were not heavily involved in generating the aftereffect. It is likely that GABAergic inputs from the ipsilateral DNLL and VNLL were also not greatly involved, as synaptic responses elicited by these inputs presumably have similar time courses as those elicited by inputs from the contralateral DNLL. In contrast to the SPN, the DNLL and VNLL lack neurons with offset firing which might have prevented them from contributing to a long-lasting suppressive aftereffect in the IC. Our results regarding the function of the DNLL in generating a suppressive aftereffect seem to be in contrast with findings obtained from the Mexican free-tailed bat (Burger and Pollak, 2001) and the Mongolian gerbil (Pecka et al., 2007). These findings suggest that inputs from the DNLL can suppress responses in the IC for tens of milliseconds. Further studies have yet to be conducted to find factors that cause such a contrast.

Major glycinergic inputs to the IC are from the ipsilateral LSO and VNLL (Saint Marie et al., 1989; Yavuzoglu et al., 2011). Neurons in the LSO are excited by stimulation of the ear on the same side of the neurons while neurons in the VNLL are excited by stimulation of the ear on the opposite side (Park et al., 2004; Zhang and Kelly, 2006). Neurons in the LSO are inhibited by stimulation of the ear on the opposite side. Thus, glycinergic inputs to the IC can be activated regardless of the direction of a sound. Existing results from brain slice studies indicate that inhibitory postsynaptic potentials mediated by glycine receptors in the IC do not last longer than potentials mediated by $GABA_A$ receptors (Moore and Trussell, 2017). If these synaptic events contributed to a suppressive effect by directly counteracting excitatory postsynaptic

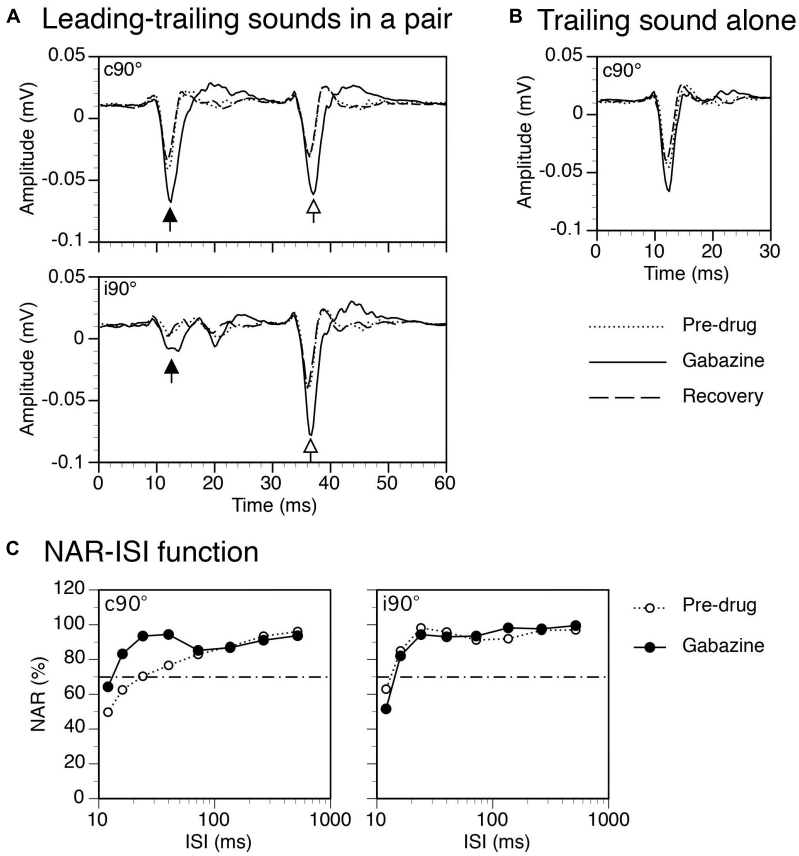



FIGURE 7
Results from an example rat showing that gabazine enhanced local-field potentials (LFPs) elicited by a leading and a trailing sound and reduced the suppressive aftereffect produced by a leading sound. **(A)** The effect of gabazine on waveforms of LFPs elicited by a leading (indicated by a filled upward arrowhead) and a trailing sound (indicated by an open upward arrowhead) in a pair (ISI = 24 ms). Results were obtained when the leading sound was presented at c90° (top panel) and i90° (bottom panel), respectively. **(B)** The effect of gabazine on the waveform of an LFP elicited by the trailing sound presented alone at c90°. **(C)** The effect of gabazine on the normalized amplitude of response (NAR)-ISI relationship obtained when the leading sound was at c90° (left panel) and i90° (right panel), respectively. The horizontal dash-and-dotted line indicates the value of NAR at 0.7. The characteristic frequency (CF) of the recording site (hence the frequency of the trailing sound) was 12.0 kHz. The frequency of the leading sound was 11.501 kHz.

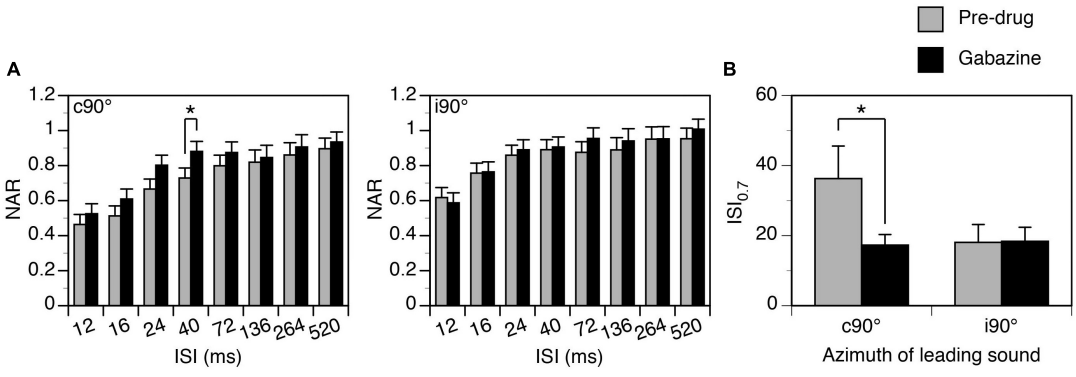



FIGURE 8
Group results showing that gabazine partially reduced the suppressive aftereffect produced by a leading sound. **(A)** Bar charts showing effects of gabazine on the normalized amplitude of response (NAR)-inter-stimulus interval (ISI) relationship. Results were obtained when a leading sound was presented at c90° (left panel) and i90° (right panel). The location of the leading sound is shown in the top left corner of each plot. **(B)** A bar chart showing the effect of gabazine on the $ISI_{0.7}$ obtained at c90° and i90°. An error bar indicates a standard error of the mean. "*" indicates statistical significance at the level of 0.1.

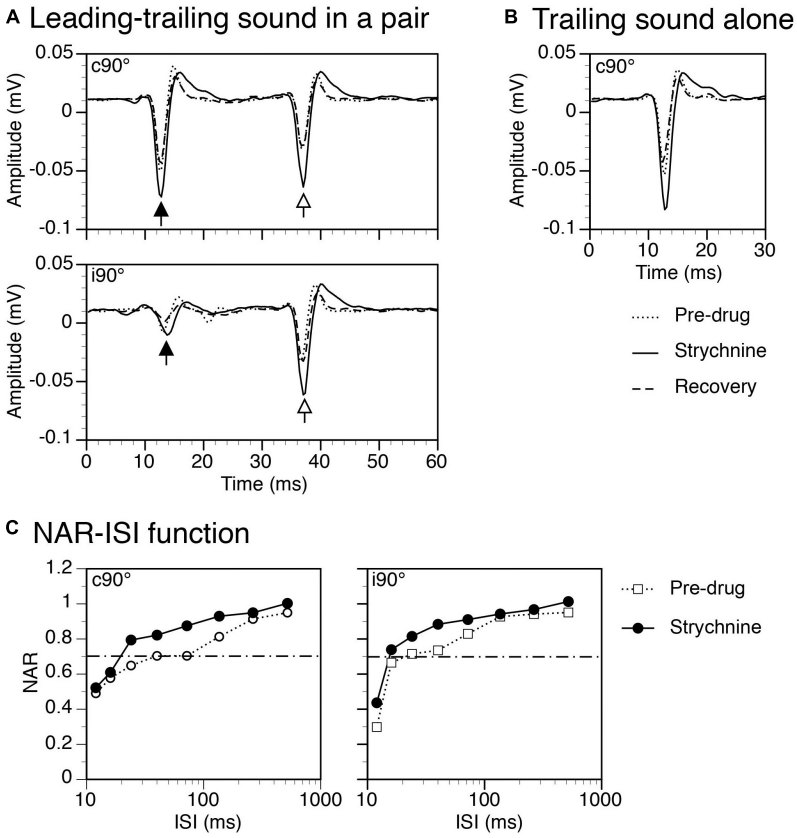



FIGURE 9
Results from an example rat showing that strychnine enhanced local-field potentials (LFPs) elicited by leading and trailing sounds and reduced the suppressive aftereffect produced by a leading sound. **(A)** The effect of strychnine on waveforms of LFPs elicited by a leading (indicated by a filled upward arrowhead) and a trailing sound (indicated by an open upward arrowhead) in a pair (ISI = 24 ms). Results were obtained when the leading sound was presented at c90° (top panel) and i90° (bottom panel), respectively. **(B)** The effect of strychnine on the waveform of an LFP elicited by the trailing sound presented alone at c90°. **(C)** The effect of strychnine on the normalized amplitude of response (NAR)-inter-stimulus interval (ISI) relationship obtained when the leading sound was at c90° (left panel) and i90° (right panel), respectively. The horizontal dash-and-dotted line indicates the value of NAR at 0.7. The characteristic frequency (CF) of the recording site (hence the frequency of the trailing sound) was 13.0 kHz. The frequency of the leading sound was 13.458 kHz.

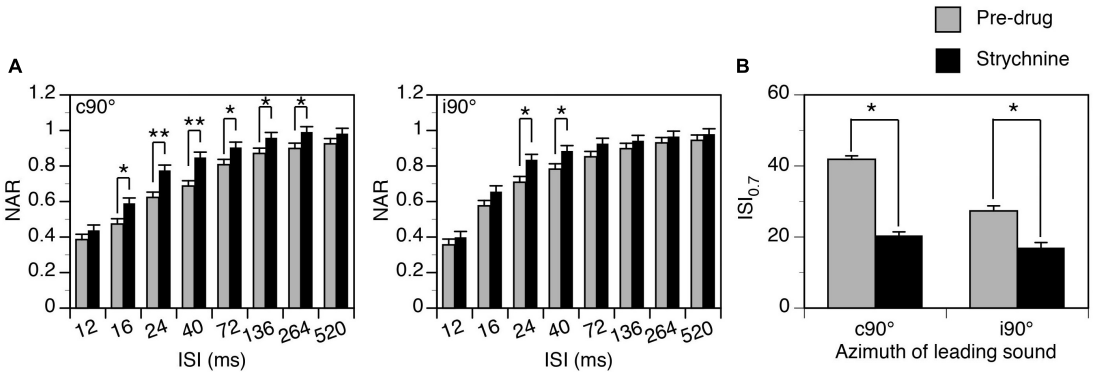



FIGURE 10
Group results showing that strychnine reduced the suppressive aftereffect produced by a leading sound at some inter-stimulus intervals (ISIs). **(A)** Bar charts showing the effect of glycine on the normalized amplitude of response (NAR)-ISI relationship. Results were obtained when a leading sound was presented at c90° (left panel) and i90° (right panel). **(B)** A bar chart showing the effect of glycine on the $ISI_{0.7}$ obtained at c90° and i90°. A *post-hoc t*-test (see text for the result from a two-way ANOVA test) indicates that strychnine significantly changed the NAR values at ISIs at 16 ms ($p = 0.007$), 24 ms ($p < 0.001$), 40 ms ($p < 0.001$), 72 ms ($p = 0.023$), 136 ms ($p = 0.039$), and 264 ms ($p = 0.030$) when a leading sound was at c90°. The drug significantly changed the NAR values at ISIs at 24 ms ($p = 0.005$) and 40 ms ($p = 0.018$) when a leading sound was at i90°. An error bar indicates a standard error of the mean. "*" indicates statistical significance at the level of 0.05, while "**" indicates statistical significance at the level of 0.005.

events, the effect would not last long and the time course of suppression observed in the present study would not have been affected by block of glycine receptors. Thus, the effect of strychnine observed in the present study (**Figures 9**, **10**) likely suggests that complex (e.g., reverberating) local circuits that were influenced by glycinergic inputs might have been involved in the generation of the suppressive aftereffect.

The effect of a pharmacological agent on an LFP is dependent on both the area over which the agent spreads and the area within which neurons contribute to the LFP. Currently, there are no effective methods that can be used to evaluate the sizes of these areas. Some existing results from the IC suggest that a drug released microiontophoretically can reach neurons that are a few hundred micrometers away (Burger and Pollak, 1998). Thus, in the present study a pharmacological agent likely affected at least a large percentage (if not all) of the neurons that contributed to an LFP. This likelihood is supported by our results showing that amplitudes of LFPs elicited by leading and trailing sounds were both substantially increased by a drug no matter whether these sounds were presented alone or in a pair (**Figures 7**, **9**). Despite large effects of Gabazine and strychnine on amplitudes of LFPs elicited by leading and trailing sounds, the drugs only mildly changed the suppressive effect caused by a leading sound (**Figures 7**–**10**). This result tended to suggest that mechanisms other than local inhibition had contributed to the aftereffect.

Previous findings indicate that a sound can produce suppressive aftereffects in lower auditory brain structures including the 8th nerve fibers and brainstem nuclei (Harris and Dallos, 1979; Kaltenbach et al., 1993; Finlayson and Adam, 1997). Such effects can be inherited by neurons in the IC through ascending pathways. Local mechanisms such as adaptation of excitatory response can also lead to a sound-elicited suppressive aftereffect in the IC (Singheiser et al., 2012). The contribution of adaptation is supported by the fact that for most IC neurons a sound produced the strongest suppressive aftereffect when the sound was presented at a location where it could elicit the strongest excitatory responses over its duration (Litovsky, 1998; Litovsky and Yin, 1998a,b). For most neurons in the structure, relocating a sound from c90° to i90° reduces the firing of these neurons (Chot et al., 2019, 2020). This could have lowered the degree of adaptation caused by the sound and reduced the degree of suppression of the response to a subsequent sound in the present study.

## Sound-elicited suppressive aftereffect in the IC and auditory functions

Previous studies based on activities of individual neurons have suggested that a sound-elicited suppressive aftereffect in the IC is involved in generating hearing phenomena such as forward masking (Finlayson, 1999; Faure et al., 2003; Furukawa et al., 2005; Nelson et al., 2009) and the precedence effect (Litovsky and Yin, 1998a,b; Litovsky and Delgutte, 2002; Tollin et al., 2004). Results from the present study based on activities of neural ensembles further support such involvement. One strong piece of evidence is the resemblance between the time course of suppression revealed by our study and those revealed by previous studies. The reduction of suppression caused by the relocation of a leading sound from the contralateral to the ipsilateral ear indicates that the IC is also involved in generating phenomena such as spatial release from masking. LFPs recorded in the present study did not provide information about action potential discharges in individual neurons and the threshold for eliciting these discharges. However, they provided important information related to excitatory/inhibitory interactions that underlay a suppressive aftereffect. The effect of local blockage of $GABA_A$ and glycine receptors on the suppressive effect indicates that local inhibitory neurotransmission contributes to the suppressive aftereffect displayed by neural responses in the IC. It implies that such neurotransmission contributes to behavioral/psychoacoustic phenomena of forward masking and the precedence effect.

A sound-elicited suppressive aftereffect may also contribute to other aspects of spatial hearing. To understand how the detection of a recurring sound is affected by another independently recurring sound, we previously used two tone bursts to create a train of stimuli with a random order (Chot et al., 2019, 2020). Stimuli in the train were presented at a constant rate (250 ms per stimulus). Such a train could be used to mimic a "novel sound" (10% probability) interleaved with a "standard sound" (90% probability) (Chot et al., 2020). It was found that the detection of a "novel sound" by IC neurons was suppressed by a colocalized "standard sound" and the suppression was released when the two sounds were spatially separated. Results from the present study suggest that the suppression caused by a "standard sound" could be attributed to the aftereffect produced by the sound. Each presentation of a "novel sound" in a train had a high probability to be preceded by multiple presentations of a "standard sound." A relatively short inter-stimulus interval (250 ms) enabled aftereffects produced by multiple presentations of the "standard sound" to be temporally summated to generate a large effect to suppress the response to a presentation of "novel sound" that followed. The suppressive aftereffects could be reduced when the "standard sound" was relocated from c90° to i90°. Such a change could reduce the suppression of the responses to a "novel sound" and lead to a separation-dependent enhancement in the neural detection of the sound.

In general, a sound-elicited suppressive aftereffect on neural responses in the IC is important for hearing in a natural acoustic environment in which qualitatively different sounds occurs at different time and locations. Local neural mechanisms including excitatory/inhibitory interactions contribute to the generation of such an aftereffect.

# Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

# Ethics statement

This animal study was reviewed and approved by University of Windsor Animal Care Committee.

## Author contributions

HZ conceived and designed the experiments. SA, ST, NR, and OS performed the experiments. SA and HZ analyzed the data and contributed to the writing of the manuscript. All authors contributed to the article and approved the submitted version.

## Funding

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the University of Windsor to HZ.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

Alkhatib, A., Biebel, U. W., and Smolders, J. W. (2006). Inhibitory and excitatory response areas of neurons in the central nucleus of the inferior colliculus in unanesthetized chinchillas. *Exp. Brain Res.* 174, 124–143. doi: 10.1007/s00221-006-0424-8

Arbogast, T. L., Mason, C. R., and Kidd, G. Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. *J. Acoust. Soc. Am.* 112, 2086–2098. doi: 10.1121/1.1510141

Batra, R., and Fitzpatrick, D. C. (2002). Monaural and binaural processing in the ventral nucleus of the lateral lemniscus: A major source of inhibition to the inferior colliculus. *Hear. Res.* 168, 90–97. doi: 10.1016/S0378-5955(02)00368-4

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/1486.001.0001

Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of speech intelligibility in multiple-talker condition. *Acta Acust. United Acust.* 86, 117–128.

Burger, R. M., and Pollak, G. D. (1998). Analysis of the role of inhibition in shaping responses to sinusoidally amplitude-modulated signals in the inferior colliculus. *J. Neurophysiol.* 80, 1686–1701. doi: 10.1152/jn.1998.80.4.1686

Burger, R. M., and Pollak, G. D. (2001). Reversible inactivation of the dorsal nucleus of the lateral lemniscus reveals its role in the processing of multiple sound sources in the inferior colliculus of bats. *J. Neurosci.* 21, 4830–4843. doi: 10.1523/JNEUROSCI.21-13-04830.2001

Burkard, R. F., Don, M., and Eggermont, J. J. (2006). *Auditory evoked potentials: Basic principles and clinical application*. Baltimore, MD: Lippencott Williams and Wilkins Press.

Cant, N. B. (2005). "Projections from the cochlear nuclear complex to the inferior colliculus," in *The inferior colliculus*, eds J. A. Winer and C. E. Schreiner (New York, NY: Springer Press), 115–131. doi: 10.1007/0-387-27083-3_3

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Chot, M. G., Tran, S., and Zhang, H. (2019). Responses of neurons in the rat's inferior colliculus to a sound are affected by another sound in a space-dependent manner. *Sci. Rep.* 9:13938. doi: 10.1038/s41598-019-50297-8

Chot, M. G., Tran, S., and Zhang, H. (2020). Spatial separation between two sounds of an oddball paradigm affects responses of neurons in the rat's inferior colliculus to the sounds. *Neuroscience* 444, 118–135. doi: 10.1016/j.neuroscience.2020.07.027

Covey, E., Kauer, J. A., and Casseday, J. H. (1996). Whole-cell patch-clamp recording reveals subthreshold sound-evoked postsynaptic currents in the inferior colliculus of awake bats. *J. Neurosci.* 16, 3009–3018. doi: 10.1523/JNEUROSCI.16-09-03009.1996

Day, M. L., Koka, K., and Delgutte, B. (2012). Neural encoding of sound source location in the presence of a concurrent, spatially separated source. *J. Neurophysiol.* 108, 2612–2628. doi: 10.1152/jn.00303.2012

Egorova, M., Ehret, G., Vartanian, I., and Esser, K. H. (2001). Frequency response areas of neurons in the mouse inferior colliculus. I. Threshold and tuning characteristics. *Exp. Brain Res.* 140, 145–161. doi: 10.1007/s002210100786

Faure, P. A., Fremouw, T., Casseday, J. H., and Covey, E. (2003). Temporal masking reveals properties of sound-evoked inhibition in duration-tuned neurons of the inferior colliculus. *J. Neurosci.* 23, 3052–3065. doi: 10.1523/JNEUROSCI.23-07-03052.2003

Felix, R. A. II, Gourévitch, B., Gómez-Álvarez, M., Leijon, S. C. M., Saldaña, E., and Magnusson, A. K. (2017). Octopus cells in the posteroventral cochlear nucleus provide the main excitatory input to the superior paraolivary nucleus. *Front. Neural Circuits* 11:37. doi: 10.3389/fncir.2017.00037

Felix, R. A. II, Magnusson, A. K., and Berrebi, A. S. (2015). The superior paraolivary nucleus shapes temporal response properties of neurons in the inferior colliculus. *Brain Struct. Funct.* 220, 2639–2652. doi: 10.1007/s00429-014-0815-8

Feng, A. S., and Ratnam, R. (2000). Neural basis of hearing in real-world situations. *Annu. Rev. Psychol.* 51, 699–725. doi: 10.1146/annurev.psych.51.1.699

Finlayson, P. G. (1999). Post-stimulatory suppression, facilitation and tuning for delays shape responses of inferior colliculus neurons to sequential pure tones. *Hear. Res.* 131, 177–194. doi: 10.1016/S0378-5955(99)00032-5

Finlayson, P. G., and Adam, T. J. (1997). Short-term adaptation of excitation and inhibition shapes binaural processing. *Acta Otolaryngol.* 117, 187–191. doi: 10.3109/00016489709117766

Furukawa, S., Maki, K., Kashino, M., and Riquimaroux, H. (2005). Dependency of the interaural phase difference sensitivities of inferior colliculus neurons on a preceding tone and its implications in neural population coding. *J. Neurophysiol.* 93, 3313–3326. doi: 10.1152/jn.01219.2004

Gai, Y. (2016). ON and OFF inhibition as mechanisms for forward masking in the inferior colliculus: A modeling study. *J. Neurophysiol.* 115, 2485–2500. doi: 10.1152/jn.00892.2015

Goense, J. B., and Logothetis, N. K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Curr. Biol.* 18, 631–640. doi: 10.1016/j.cub.2008.03.054

González-Hernández, T., Mantolán-Sarmiento, B., González-González, B., and Pérez-González, H. (1996). Sources of GABAergic input to the inferior colliculus of the rat. *J. Comp. Neurol.* 372, 309–326. doi: 10.1002/(SICI)1096-9861(19960819)372:2<309::AID-CNE11>3.0.CO;2-E

Harris, D. M., and Dallos, P. (1979). Forward masking of auditory nerve fiber responses. *J. Neurophysiol.* 42, 1083–1107. doi: 10.1152/jn.1979.42.4.1083

Helfert, R. H., Bonneau, J. M., Wenthold, R. J., and Altschuler, R. A. (1989). GABA and glycine immunoreactivity in the guinea pig superior olivary complex. *Brain Res.* 501, 269–286. doi: 10.1016/0006-8993(89)90644-6

Hurley, L. M., Tracy, J. A., and Bohorquez, A. (2008). Serotonin 1B receptor modulates frequency response curves and spectral integration in the inferior colliculus by reducing GABAergic inhibition. *J. Neurophysiol.* 100, 1656–1667. doi: 10.1152/jn.90536.2008

Jones, G. L., and Litovsky, R. Y. (2011). A cocktail party model of spatial release from masking by both noise and speech interferers. *J. Acoust. Soc. Am.* 130, 1463–1474. doi: 10.1121/1.3613928

Kadner, A., and Berrebi, A. S. (2008). Encoding of temporal features of auditory stimuli in the medial nucleus of the trapezoid body and superior paraolivary nucleus of the rat. *Neuroscience* 151, 868–887. doi: 10.1016/j.neuroscience.2007.11.008

Kaltenbach, J. A., Meleca, R. J., Falzarano, P. R., Myers, S. F., and Simpson, T. H. (1993). Forward masking properties of neurons in the dorsal cochlear nucleus: Possible role in the process of echo suppression. *Hear. Res.* 67, 35–44. doi: 10.1016/0378-5955(93)90229-T

Kulesza, R. J. Jr., and Berrebi, A. S. (2000). Superior paraolivary nucleus of the rat is a GABAergic nucleus. *J. Assoc. Res. Otolaryngol.* 1, 255–269. doi: 10.1007/s101620010054

Kulesza, R. J. Jr., Spirou, G. A., and Berrebi, A. S. (2003). Physiological response properties of neurons in the superior paraolivary nucleus of the rat. *J. Neurophysiol.* 89, 2299–2312. doi: 10.1152/jn.00547.2002

Kuwada, S., Bishop, B., Alex, C., Condit, D. W., and Kim, D. O. (2011). Spatial tuning to sound-source azimuth in the inferior colliculus of unanesthetized rabbit. *J. Neurophysiol.* 106, 2698–2708. doi: 10.1152/jn.00532.2011

Lane, C. C., and Delgutte, B. (2005). Neural correlates and mechanisms of spatial release from masking: Single-unit and population responses in the inferior colliculus. *J. Neurophysiol.* 94, 1180–1198. doi: 10.1152/jn.01112.2004

Litovsky, R. Y. (1998). Physiological studies of the precedence effect in the inferior colliculus of the kitten. *J. Acoust. Soc. Am.* 103, 3139–3152. doi: 10.1121/1.423072

Litovsky, R. Y., and Delgutte, B. (2002). Neural correlates of the precedence effect in the inferior colliculus: Effect of localization cues. *J. Neurophysiol.* 87, 976–994. doi: 10.1152/jn.00568.2001

Litovsky, R. Y., and Yin, T. C. (1998a). Physiological studies of the precedence effect in the inferior colliculus of the cat. I. Correlates of psychophysics. *J. Neurophysiol.* 80, 1285–1301. doi: 10.1152/jn.1998.80.3.1285

Litovsky, R. Y., and Yin, T. C. (1998b). Physiological studies of the precedence effect in the inferior colliculus of the cat. II. Neural mechanisms. *J. Neurophysiol.* 80, 1302–1316. doi: 10.1152/jn.1998.80.3.1302

Liu, Y., Zhang, G., Yu, H., Li, H., Wei, J., and Xiao, Z. (2019). Robust and intensity-dependent synaptic inhibition underlies the generation of non-monotonic neurons in the mouse inferior colliculus. *Front. Cell. Neurosci.* 13:131. doi: 10.3389/fncel.2019.00131

Loftus, W. C., Bishop, D. C., Saint Marie, R. L., and Oliver, D. L. (2004). Organization of binaural excitatory and inhibitory inputs to the inferior colliculus from the superior olive. *J. Comp. Neurol.* 472, 330–344. doi: 10.1002/cne.20070

Logothetis, N. K. (2003). The underpinnings of the BOLD functional magnetic resonance imaging signal. *J. Neurosci.* 23, 3963–3971. doi: 10.1523/JNEUROSCI.23-10-03963.2003

Markovitz, N. S., and Pollak, G. D. (1994). Binaural processing in the dorsal nucleus of the lateral lemniscus. *Hear. Res.* 73, 121–140. doi: 10.1016/0378-5955(94)90290-9

Mei, H., Guo, Y., Wu, F., and Chen, Q. (2006). Masking effect of different durations of forward masker sound on acoustical responses of mouse inferior collicular neurons to probe sound. *Front. Biol. China* 1:285–289. doi: 10.1007/s11515-006-0004-0

Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: Investigation of evoked potentials and EEG phenomena. *Physiol. Rev.* 65, 37–100. doi: 10.1152/physrev.1985.65.1.37

Moore, L. A., and Trussell, L. O. (2017). Corelease of inhibitory neurotransmitters in the mouse auditory midbrain. *J. Neurosci.* 37, 9453–9464. doi: 10.1523/JNEUROSCI.1125-17.2017

Nelson, P. C., Smith, Z. M., and Young, E. D. (2009). Wide-dynamic-range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking. *J. Neurosci.* 29, 2553–2562. doi: 10.1523/JNEUROSCI.5359-08.2009

Park, T. J., and Pollak, G. D. (1993). GABA shapes a topographic organization of response latency in the mustache bat's inferior colliculus. *J. Neurosci.* 13, 5172–5187. doi: 10.1523/JNEUROSCI.13-12-05172.1993

Park, T. J., and Pollak, G. D. (1994). Azimuthal receptive fields are shaped by GABAergic inhibition in the inferior colliculus of the mustache bat. *J. Neurphysiol.* 72, 1080–1102. doi: 10.1152/jn.1994.72.3.1080

Park, T. J., Klug, A., Holinstat, M., and Grothe, B. (2004). Interaural level difference processing in the lateral superior olive and the inferior colliculus. *J. Neurophysiol.* 92, 289–301. doi: 10.1152/jn.00961.2003

Patel, C. R., and Zhang, H. (2014). Local application of sodium salicylate enhances auditory responses in the rat's dorsal cortex of the inferior colliculus. *Front. Neurol.* 5:235. doi: 10.3389/fneur.2014.00235

Patel, C. R., Redhead, C., Cervi, A. L., and Zhang, H. (2012). Neural sensitivity to novel sounds in the rat's dorsal cortex of the inferior colliculus as revealed by evoked local field potentials. *Hear. Res.* 286, 41–54. doi: 10.1016/j.heares.2012.02.007

Pecka, M., Zahn, T. P., Saunier-Rebori, B., Siveke, I., Felmy, F., Wiegrebe, L., et al. (2007). Inhibiting the inhibition: A neuronal network for sound localization in reverberant environments. *J. Neurosci.* 27, 1782–1790. doi: 10.1523/JNEUROSCI.5335-06.2007

Poirier, P., Samson, F. K., and Imig, T. J. (2003). Spectral shape sensitivity contributes to the azimuth tuning of neurons in the cat's inferior colliculus. *J. Neurophysiol.* 89, 2760–2777. doi: 10.1152/jn.00640.2002

Ratnam, R., and Feng, A. S. (1998). Detection of auditory signals by frog inferior collicular neurons in the presence of spatially separated noise. *J. Neurophysiol.* 80, 2848–2859. doi: 10.1152/jn.1998.80.6.2848

Saint Marie, R. L., Ostapoff, E. M., Morest, D. K., and Wenthold, R. J. (1989). Glycine-immunoreactive projection of the cat lateral superior olive: Possible role in midbrain ear dominance. *J. Comp. Neurol.* 279, 382–396. doi: 10.1002/cne.902790305

Saldaña, E., Aparicio, M. A., Fuentes-Santamaría, V., and Berrebi, A. S. (2009). Connections of the superior paraolivary nucleus of the rat: Projections to the inferior colliculus. *Neuroscience* 163, 372–387. doi: 10.1016/j.neuroscience.2009.06.030

Salimi, N., Zilany, M. S. A., and Carney, L. H. (2017). Modeling responses in the superior paraolivary nucleus: Implications for forward masking in the inferior colliculus. *JARO* 18, 441–456. doi: 10.1007/s10162-016-0612-6

Schofield, B. R. (1995). Projections from the cochlear nucleus to the superior paraolivary nucleus in guinea pigs. *J. Comp. Neurol.* 360, 135–149. doi: 10.1002/cne.903600110

Schofield, B. R. (2005). "Superior olivary complex and lateral lemniscal connections of the auditory midbrain," in *The inferior colliculus*, eds J. A. Winer and C. E. Schreiner (New York, NY: Springer Press), 132–154. doi: 10.1007/0-387-27083-3_4

Shinn-Cunningham, B. G., Schickler, J., Kopèo, N., and Litovsky, R. (2001). Spatial unmasking of nearby speech sources in a simulated anechoic environment. *J. Acoust. Soc. Am.* 110, 1118–1129. doi: 10.1121/1.1386633

Singheiser, M., Ferger, R., von Campenhausen, M., and Wagner, H. (2012). Adaptation in the auditory midbrain of the barn owl (*Tyto alba*) induced by tonal double stimulation. *Eur. J. Neurosci.* 35, 445–456.

Tollin, D. J., Populin, L. C., and Yin, T. C. (2004). Neural correlates of the precedence effect in the inferior colliculus of behaving cats. *J. Neurophysiol.* 92, 3286–3297.

Tolnai, S., Beutelmann, R., and Klump, G. M. (2017). Effect of preceding stimulation on sound localization and its representation in the auditory midbrain. *Eur. J. Neurosci.* 45, 460–471. doi: 10.1111/ejn.13491

Vater, M., Kössl, M., and Horn, A. K. (1992). GAD- and GABA-immunoreactivity in the ascending auditory pathway of horseshoe and mustached bats. *J. Comp. Neurol.* 325, 183–206. doi: 10.1002/cne.903250205

Warzybok, A., Rennies, J., Brand, T., Doclo, S., and Kollmeier, B. (2013). Effects of spatial and temporal integration of a single early reflection on speech intelligibility. *J. Acoust. Soc. Am.* 133, 269–282. doi: 10.1121/1.4768880

Wu, S. H. (1999). Synaptic excitation in the dorsal nucleus of the lateral lemniscus. *Prog. Neurobiol.* 57, 357–375. doi: 10.1016/S0301-0082(98)00058-6

Xie, R., Gittelman, J. X., and Pollak, G. D. (2008). Whole cell recordings of intrinsic properties and sound-evoked responses from the inferior colliculus. *Neuroscience* 154, 245–256. doi: 10.1016/j.neuroscience.2008.02.039

Yavuzoglu, A., Schofield, B. R., and Wenstrup, J. J. (2011). Circuitry underlying spectrotemporal integration in the auditory midbrain. *J. Neurosci.* 31, 14424–14435. doi: 10.1523/JNEUROSCI.3529-11.2011

Yin, T. C. (1994). Physiological correlates of the precedence effect and summing localization in the inferior colliculus of the cat. *J. Neurosci.* 14, 5170–5186. doi: 10.1523/JNEUROSCI.14-09-05170.1994

Yost, W. A. (2017). Spatial release from masking based on binaural processing for up to six maskers. *J. Acoust. Soc. Am.* 141, 2093–2106. doi: 10.1121/1.4978614

Zhang, D. X., Li, L., Kelly, J. B., and Wu, S. H. (1998). GABAergic projections from the lateral lemniscus to the inferior colliculus of the rat. *Hear. Res.* 117, 1–12. doi: 10.1016/S0378-5955(97)00202-5

Zhang, H., and Feng, A. S. (1998). Sound direction modifies the inhibitory as well as the excitatory frequency tuning characteristics of single neurons in the frog torus semicircularis (inferior colliculus). *J. Comp. Physiol. A* 182, 725–735. doi: 10.1007/s003590050217

Zhang, H., and Kelly, J. B. (2006). Responses of neurons in the rat's ventral nucleus of the lateral lemniscus to monaural and binaural tone bursts. *J. Neurophysiol.* 95, 2501–2512. doi: 10.1152/jn.01215.2005

Zhang, H., and Kelly, J. B. (2009). Time-dependent effects of ipsilateral stimulation on contralaterally elicited responses in the rat's central nucleus of the inferior colliculus. *Brain. Res.* 1303, 48–60. doi: 10.1016/j.brainres.2009.09.059

Zhang, H., Xu, J., and Feng, A. S. (1999). Effects of GABA-mediated inhibition on direction-dependent frequency tuning in the frog inferior colliculus. *J. Comp. Physiol. A* 184, 85–98. doi: 10.1007/s003590050308

# Frontiers in Neuroscience

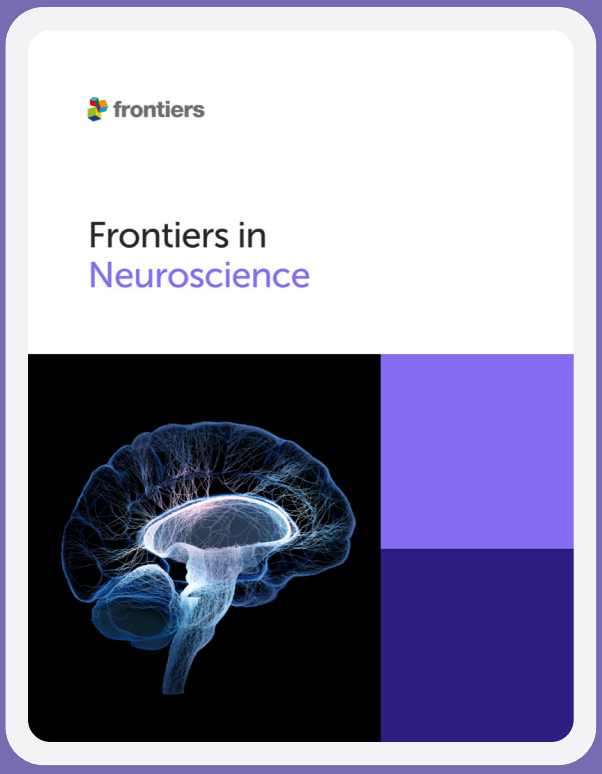
**Provides a holistic understanding of brain function from genes to behavior**

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

---

## Discover the latest Research Topics

See more →




**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact