LOUDNESS: FROM NEUROSCIENCE TO PERCEPTION

EDITED BY: Sabine Meunier, Maaike Van Eeckhoutte and
Brian Cecil Joseph Moore

PUBLISHED IN: Frontiers in Psychology and Frontiers in Neuroscience

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88971-994-5
DOI 10.3389/978-2-88971-994-5

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

LOUDNESS: FROM NEUROSCIENCE TO PERCEPTION

Topic Editors:
Sabine Meunier, Laboratoire de Mécanique et d’Acoustique - CNRS - Aix-Marseille University - Ecole Centrale Marseille, France
Maaike Van Eeckhoutte, Technical University of Denmark, Denmark
Brian Cecil Joseph Moore, University of Cambridge, United Kingdom

Citation: Meunier, S., Van Eeckhoutte, M., Moore, B. C. J., eds. (2021). Loudness: From Neuroscience to Perception. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88971-994-5

Table of Contents

frontiers in Psychology

EDITORIAL
published: 15 November 2021
doi: 10.3389/fpsyg.2021.785093

Editorial: Loudness: From Neuroscience to Perception

Sabine Meunier[1]*****, *Maaike Van Eeckhoutte*[2,3] *and Brian C. J. Moore*[4]

[1] Aix Marseille Univ, CNRS, Centrale Marseille, LMA, Marseille, France, [2] Hearing Systems, Department of Health Technology, Technical University of Denmark, Lyngby, Denmark, [3] Copenhagen Hearing and Balance Center, Ear, Nose, and Throat (ENT) and Audiology Clinic, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark, [4] Cambridge Hearing Group, Department of Psychology, University of Cambridge, Cambridge, United Kingdom

Keywords: loudness, psychophysics, temporal and frequency integration, binaural loudness, hearing aids and cochlear implant

Editorial on the Research Topic

Loudness: From Neuroscience to Perception

Loudness is the sensation that allows judgment of whether a sound is strong or soft. Sounds can be characterized by several perceptual features and among them loudness plays an important role. Loudness is very important for sound quality. Noise annoyance is mainly influenced by loudness, because, in most situations, the louder the sound, the more annoying it is. It is very important to control loudness for users of hearing aids and cochlear implants, for whom the loudness of sounds must be appropriate and the temporal fluctuations in loudness (particularly for speech) must be well-reproduced. Understanding how the percept of loudness is formed in the auditory system and how it is coded is therefore of great importance.

This special issue includes nine articles on loudness, mainly using psychoacoustical approaches, and ranging from theoretical issues to clinical applications. The issue explores psychophysics, loudness measurement, multisensory integration, the influence of the temporal and frequency characteristics of sounds on loudness, the way that loudness is combined across the two ears, and clinical applications to hearing aids and cochlear implants.

The article by Zeng presents a unified theory of psychophysical laws in auditory intensity perception. There has been a long history of psychophysical laws that attempt to relate the physical sound intensity of a stimulus to its perceived magnitude or loudness. The first approach was published by Fechner in 1860, who used just noticeable differences to infer that loudness is a logarithmic function of sound intensity. Over the years, Fechner's original assumption has been criticized and modified and a widely accepted view is that loudness is a compressive power function of sound intensity; this relationship is sometimes called Steven's power law. In this paper, Zeng reviews previous theories based on just noticeable differences and integrates them in a new unified theory, thereby also showing the validity of Fechner's original idea for a range of hearing situations.

The measurement of loudness is discussed in two articles. The article by Fultz et al. deals with categorical loudness scaling, a procedure that is often used for measuring the growth of loudness with increasing stimulus intensity. Some authors have proposed that categorical loudness scaling should be used in the fitting of hearing aids, but this requires time-efficient tests. Aiming to make categorical loudness scaling more efficient, this article describes a comparison of a "traditional" method using fixed stimulus levels with a method using Bayesian inference to select stimulus parameters that yield the maximum expected information gain during data collection. The article discusses methods for decreasing the test time, while maintaining test-retest reliability and accuracy, and it further discusses optimizations. In their study on the moment-by-moment

Edited and reviewed by:
Robert J. Zatorre,
McGill University, Canada

****Correspondence:***
Sabine Meunier
meunier@lma.cnrs-mrs.fr

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

Received: *28 September 2021*
Accepted: *14 October 2021*
Published: *15 November 2021*

Citation:
Meunier S, Van Eeckhoutte M and Moore BCJ (2021) Editorial: Loudness: From Neuroscience to Perception. Front. Psychol. 12:785093. doi: 10.3389/fpsyg.2021.785093

loudness assessment of time-varying sounds, Schlittenlacher and Ellermeier used continuous cross-modality matching between line length and loudness (and vice versa) for musical excerpts of either rock or classical music. They found that line length is highly correlated with long-term loudness calculated using the time-varying loudness (TVL) model of loudness (Moore et al., 2018), showing the reliability of the method. Their results provide some support for the time constant (temporal portions of the sound that affect momentary judgment) of 750 ms used in the TVL model. As expected, because of the regression effect, the line-length adjustment task yielded an exponent of the loudness function smaller than predicted by Steven's power law.

The article by Fischenich et al. explores spectro-temporal processes that influence loudness. Using the correlations between loudness judgments and the levels of brief temporal segments within longer sounds, they show that temporal weights in loudness judgments are frequency specific. This result suggests that temporal integration precedes spectral integration. This is consistent with the most recent version of the TVL model of loudness (Moore et al., 2016, 2018) and with recent neurophysiological data (Thwaites et al., 2017).

Two papers deal with the way that loudness combines across ears, often referred to as binaural loudness. In one paper, Denk et al. explore the "missing 6 dB" (the often-reported but sometimes disputed claim that a sound presented *via* headphones needs to have a 6 dB higher level at the eardrum than the same sound presented *via* a loudspeaker in order to elicit the same loudness). In a task where the listener adjusted the level of the sound presented *via* headphones to match the loudness of the same sound presented *via* a loudspeaker, they found that this mismatch was large at low frequencies but largely disappeared at high frequencies. The mismatch decreased when the interaural coherence (a measure of the correlation of the sound across the two ears) decreased, i.e., when the sound appeared to be more diffuse. Surprisingly, the mismatch was different in two different anechoic rooms whereas there was no difference between two non-anechoic rooms. Thus, the different results found in the literature concerning the "missing 6 dB" may be related to differences in the experimental conditions (reproduction mode, room, stimuli). The paper of Pieper et al. is concerned with Individualized Loudness Models (ILMs), which might help in the fitting of hearing aids in order to improve audibility, comfort and naturalness. Loudness models applied to impaired hearing take into account individual frequency-dependent reductions of cochlear gain and compression produced by hearing loss. Pieper et al. argue that, in addition, ILMs should take into account individual differences in binaural loudness summation. They propose an extension of a monaural loudness model "toward an individual binaural loudness model for hearing aid fitting and development."

The paper by McKay describes the application of three loudness models to the perception of loudness by people with cochlear implants. One model is applied in the simple case of electrical stimuli applied to a single electrode. In this model, cochlear neural excitation is integrated over time using a central temporal integration window similar to that used in models of loudness for normal hearing, such as the TVL model. The other more complex model (the "Detailed" model) is applied when multiple electrodes are stimulated within a short time interval. This model includes the effects of interaction between different electrodes. McKay also presents a "Practical" model, which is a simplified version of the "Detailed" model, and which can be used to predict the loudness of pulsatile electrical stimuli applied to multiple electrodes. The models have been applied to the development of novel signal processing strategies that aim to provide users of cochlear implants with a more natural perception of loudness.

In the paper by Sun et al., the authors use both behavioral experiments and electro-encephalography (EEG) to measure subtle multi-modal effects in loudness perception. Specifically, in four behavioral and EEG experiments, the authors show that visual-motor information from manual gestures modulates the loudness perception of consecutive sounds whose intensity changes, as well as the early auditory neural responses that correspond to the changes in loudness perception.

The paper by Berthomieu et al. describes mounting evidence that the loudness of sounds is influenced not only by their physical characteristics at the eardrum (intensity, spectrum temporal pattern, and binaural differences) but also by the manner of presentation, for example whether or not the sound source is visible, whether the sounds are presented *via* headphones or loudspeakers, or from "live" sources, such as a person talking, and whether or not the sounds are meaningful. Berthomieu et al. argue that loudness appears to depend on how listeners interpret the sound sources, notably whether they focus on the sound that reaches their ears (the proximal stimulus) or the sound as produced by the source (the distal stimulus). This distinction was made many years ago by Helmholtz who stated "...we are exceedingly well-trained in finding out by our sensations the objective nature of the objects around us, but we are completely unskilled in observing these sensations *per se*" [quoted in Warren (1981)]. Berthomieu et al. argue that whether the listener focusses on the proximal or distal stimulus depends on the instruction to the listener and on how the sound is interpreted. Many experiments on loudness perception have been set up so as to promote listening to the proximal stimulus, whereas in everyday life loudness may be more related to the distal stimulus.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

BM was supported by the Engineering and Physical Sciences Research Council (UK, Grant Number RG78536).

REFERENCES

Moore, B. C. J., Glasberg, B. R., Varathanathan, A., and Schlittenlacher, J. (2016). A loudness model for time-varying sounds incorporating binaural inhibition. *Trends Hear*. 20, 1–16. doi: 10.1177/2331216516682698

Moore, B. C. J., Jervis, M., Harries, L., and Schlittenlacher, J. (2018). Testing and refining a loudness model for time-varying sounds incorporating binaural inhibition. *J. Acoust. Soc. Am*. 143, 1504–1513. doi: 10.1121/1.5027246

Thwaites, A., Schlittenlacher, J., Nimmo-Smith, I., Marslen-Wilson, W. D., and Moore, B. C. J. (2017). Tonotopic representation of loudness in the human cortex. *Hear. Res*. 344, 244–254. doi: 10.1016/j.heares.2016.11.015

Warren, R. M. (1981). Measurement of sensory intensity. *Behav. Brian Sci.* 4, 175–189. doi: 10.1017/S0140525X00008256

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.

frontiers in Psychology

HYPOTHESIS AND THEORY
published: 26 June 2020
doi: 10.3389/fpsyg.2020.01459

A Unified Theory of Psychophysical Laws in Auditory Intensity Perception

*Fan-Gang Zeng**

Center for Hearing Research, Department of Anatomy and Neurobiology–Department of Biomedical Engineering–Department of Cognitive Sciences–Department of Otolaryngology – Head and Neck Surgery, University of California, Irvine, Irvine, CA, United States

Psychophysical laws quantitatively relate perceptual magnitude to stimulus intensity. While most people have accepted Stevens's power function as the psychophysical law, few believe in Fechner's original idea using just-noticeable-differences (jnd) as a constant perceptual unit to educe psychophysical laws. Here I present a unified theory in hearing, starting with a general form of Zwislocki's loudness function (1965) to derive a general form of Brentano's law. I will arrive at a general form of the loudness-jnd relationship that unifies previous loudness-jnd theories. Specifically, the "slope," "proportional-jnd," and "equal-loudness, equal-jnd" theories, are three additive terms in the new unified theory. I will also show that the unified theory is consistent with empirical data in both acoustic and electric hearing. Without any free parameters, the unified theory uses loudness balance functions to successfully predict the jnd function in a wide range of hearing situations. The situations include loudness recruitment and its jnd functions in sensorineural hearing loss and simultaneous masking, loudness enhancement and the midlevel hump in forward and backward masking, abnormal loudness and jnd functions in cochlear implant subjects. Predictions of these loudness-jnd functions were thought to be questionable at best in simultaneous masking or not possible at all in forward masking. The unified theory and its successful applications suggest that although the specific form of Fechner's law needs to be revised, his original idea is valid in the wide range of hearing situations discussed here.

Keywords: loudness, intensity discrimination, just-noticeable-differences (jnd), Weber's law, Fechner's law, Stevens's law, Zwislocki, auditory

OPEN ACCESS

Edited by:
Maaike Van Eeckhoutte,
Technical University of Denmark,
Denmark

Reviewed by:
Daniel Oberfeld,
Johannes Gutenberg University
Mainz, Germany
Wolfgang Ellermeier,
Darmstadt University of Technology,
Germany

****Correspondence:***
Fan-Gang Zeng
fzeng@uci.edu

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

Received: *17 March 2020*
Accepted: *02 June 2020*
Published: *26 June 2020*

Citation:
Zeng F-G (2020) A Unified Theory
of Psychophysical Laws in Auditory
Intensity Perception.
Front. Psychol. 11:1459.
doi: 10.3389/fpsyg.2020.01459

INTRODUCTION

Psychophysical laws attempt to relate the amplitude of a physical stimulus to its perceived magnitude, such as loudness as a function of sound pressure or brightness as a function of luminance. The classic approach to uncovering psychophysical laws was advanced by Fechner (1966) in the mid 18th century (original work published in 1860). Fechner assumed that the just-noticeable-difference (jnd), expressed as the Weber fraction ($\Delta I/I$), where I is a standard sound intensity and ΔI is the intensity change required for the jnd, produced an equal increment in loudness sensation (ΔL). Integrating this equation, namely $\Delta L = \Delta I/I$, he produced what is known as Fechner's law: loudness is a logarithmic function of sound intensity ($L = log\, I$).

Not only was Fechner's logarithmic law replaced by Stevens's power law or $L = I^{\theta}$, where θ is a constant (Stevens, 1961), his general approach was also questioned due to failure to integrate the jnd functions of two different sounds to predict their respective loudness functions (Newman, 1933;

Miller, 1947). Thus, it was not too surprising that the Fechnerian approach in relating the stimulus jnd to the subjective magnitude was abandoned by some researchers. What was surprising is the grounds on which the Fechnerian approach was abandoned. For example, Stevens (1961) argued that the direct magnitude estimation technique obsolesced intensity discrimination as a measure of the stimulus-sensation relationship. He viewed the discrimination measure as "an engineer talking…the scatter of some dial settings." In a completely opposing view, Viemeister and Bacon (1988) stated that loudness estimation data were a measure with "probably strong involvement of non-sensory factors, (and) we did not attempt to relate these data to those for intensity discrimination."

There have been other researchers who continued to advance the Fechnerian approach in searching for a unified theory relating intensity discrimination to the loudness function. Fechner's original assumption was sometimes referred to as the "slope" theory, because it predicted that the steeper the loudness function, the smaller the jnd or Weber fraction for a constant increment in loudness. This simple slope prediction turned out to be not true at least in cases of loudness recruitment, where cochlear hearing loss or partial masking elevated the hearing threshold but produced abnormally steep loudness growth so that normal loudness was perceived at high sound levels (Fowler, 1937). To account for the failure of Fechner's slope theory, several researchers proposed a "proportional-jnd" theory, in which the jnd size needed to be normalized by the total jnd number within a stimulus's dynamic range (Riesz, 1933; Teghtsoonian, 1971; Lim et al., 1977). On the other hand, the "equal-loudness, equal-jnd" theory argued that the jnd had no relation to the slope of the loudness function, but rather was determined by the total loudness (Zwislocki and Jordan, 1986). Despite significant effort in testing these loudness-jnd relationships, no consensus has been reached yet (Houtsma et al., 1980; Hellman et al., 1987; Schlauch and Wier, 1987; Rankovic et al., 1988; Johnson et al., 1993; Stillman et al., 1993; Schlauch et al., 1995; Allen and Neely, 1997; Hellman and Hellman, 2001).

Here I present a unified theory, starting with a general form of Zwislocki's (1965) loudness function to derive a general form of Brentano's law, and I will arrive at a general form of the loudness-jnd relationship that unifies previous loudness-jnd theories. Specifically, I find that the previous "slope," "proportional-jnd," and "equal-loudness, equal-jnd" theories, are three additive terms in the new unified theory. I also show that the new theory is capable of predicting loudness and jnd data across a wide range of hearing situations, including sensorineural hearing loss, simultaneous masking, forward masking, and electric hearing.

DERIVATION OF A UNIFIED THEORY

Derivation of a General Form of Brentano's or Ekman's Law

I start with the general form of a loudness function proposed by Zwislocki (1965; Eq. 212):

$$L = k[(I + cI_0)^{\theta} - (cI_0)^{\theta}] \tag{1}$$

where I_0 is the detection threshold for a particular type of sound, c represents an internal noise scaling factor, and k is a constant.

Generality and symmetry are the two reasons for choosing Zwislocki's loudness function. First, at high intensities ($I >> I_o$), Zwislocki's function can be simplified as Stevens's power law, namely, $L = kI^{\theta}$. At low intensities, Zwislocki made an implicit but important assumption to account for loudness recruitment near threshold: The slope (θ) of the loudness function does not increase as initially thought (Fowler, 1937), instead the loudness at threshold is increased. Setting $I = I_o$ in Eq. (1), the loudness at threshold, or $L_o = k[(I_o + cI_o)^{\theta} - (cI_o)^{\theta}] = k\ [(1/c + 1)^{\theta} - 1)]\ (cI_o)^{\theta} \sim k\ [\theta\ (1/c)^{1-\theta}]\ (I_o)^{\theta}$, is directly proportional to the threshold and "must be greater than zero (Zwislocki, 1965; p. 87)." Mathematically, the loudness at threshold is infinite when the internal noise is zero ($c = 0$), and vice versa. This is a fundamental argument for why the brain has or needs internal noise because infinite loudness is clearly biologically unacceptable. Zwislocki's internal noise concept was also expanded to form the basis for treating loudness recruitment as "softness imperception" (Buus and Florentine, 2002) and tinnitus as "additive central noise" (Zeng, 2013). In the interest of simplicity, I define loudness at threshold as: $L_o = k(cI_o)^{\theta}$ (or $c = 0.125$ for $\theta = 0.27$).

Second, the mathematical symmetry can be shown by differentiating Eq. (1):

$$\frac{\Delta L}{\Delta I} = \theta k(I + cI_0)^{\theta - 1} = \theta k\frac{(I + cI_0)^{\theta}}{I + cI_0} \tag{2}$$

Adding and subtracting the same component in the above equation, I obtain:

$$\frac{\Delta L}{\Delta I} = \theta k\frac{(I + cI_0)^{\theta} - (cI_0)^{\theta} + (cI_0)^{\theta}}{I + cI_0} = \theta\frac{L + L_0}{I + cI_0} \tag{3}$$

Rewriting the above equation, I obtain the general form of Brentano's law or Ekman's law, namely, $\frac{\Delta L}{L} = \frac{\Delta I}{I}$, (see Stevens, 1961, for discussion of these laws):

$$\frac{\Delta L}{L + L_0} = \theta\frac{\Delta I}{I + cI_0} \tag{4}$$

Equation (4) is mathematically symmetrical and balanced, having a general form of Weber's law including a threshold-correction term in both the sensation domain (L_o) and the stimulus domain (cI_o).

To the first-order approximation, Weber's law in the stimulus domain has been "replicated in hundreds of studies across all sensory modalities and many animal species over the last two centuries (Pardo-Vazquez et al., 2019)." In auditory intensity discrimination, the Weber fraction is constant for broadband noise but decreases slightly with increasing intensity, resulting in a "near miss" to Weber's law (McGill and Goldberg, 1968). Therefore, Eq. (4) can be written as:

$$\frac{\Delta L}{L + L_0} = wI^{\alpha} \tag{5}$$

where w and α are both constants, with $\alpha = 0$ indicating perfect conformity to Weber's law.

According to the "proportional-jnd" theory (Lim et al., 1977), the constant w is inversely proportional to the number of jnds (N) within the stimulus dynamic range. In other words, $w = 1/N$, which can be considered as a scaling factor to account for the fact that different subjects or different types of stimuli may have a different number of discriminable steps within their respective dynamic range (e.g., a normal-hearing listener has 100 steps but a cochlear-implant user has only 10), but they all have similar loudness growth from soft at the threshold to uncomfortably loud at the upper limit of the range. The "proportional-jnd" theory states that 10 jnd steps in the normal-hearing listener would produce the same amount of loudness change as one jnd step in the cochlear-implant user. Although the "proportional-jnd" theory did not assume or require any specific jnd-loudness function, Lim et al. (1977) hinted that Brentano's law "is nearly the correct one" (see footnote 7 on p. 1264 in Lim et al., 1977). In this case, a relative change in loudness is inversely proportional to the number of jnds with an intensity correction term, whose origin will be considered in section "Discussion":

$$\frac{\Delta L}{L + L_0} = \frac{1}{N} I^{\alpha} \tag{6}$$

Prediction of the jnd Function From the Loudness Balance Function

Suppose that the loudness function for a tone in quiet is: $L = f(I)$, and that the loudness balance function between the tone in quiet and the tone in masking has been obtained: $I = g(I_m)$. By definition, at $I = g(I_m)$, loudness is balanced so that the loudness function can be derived for a partially masked tone:

$$L_m = L = f(I) = f[g(I_m)] \tag{7}$$

Differentiating the above equation to obtain:

$$\frac{\Delta L_m}{\Delta I_m} = f'(I) g'(I_m) = \frac{\Delta L}{\Delta I} g'(I_m) \tag{8}$$

Rewrite the above equation:

$$\Delta I_m = \Delta I \frac{1}{g'(I_m)} \frac{\Delta L_m}{\Delta L} \tag{9}$$

Replace ΔL_m and ΔL with Eq. (6) to obtain:

$$\Delta I_m = \Delta I \frac{1}{g'(I_m)} \frac{N}{N_m} \frac{I_m^{\alpha}}{I^{\alpha}} \frac{L_m + L_{mo}}{L + L_o} \tag{10}$$

To predict the jnd in the form of the Weber fraction at the same intensity, that is, $I_m = I$ so that one can cancel out the intensity correction term (I_m^{α}/I^{α}) and divide the above equation by (I):

$$\frac{\Delta I_m}{I} = \frac{\Delta I}{I} \frac{1}{g'(I_m)} \frac{N}{N_m} \frac{L_m + L_{mo}}{L + L_o} \tag{11}$$

Taking a logarithmic transformation, one can calculate the jnd in terms of the Weber fraction in dB (WFdB):

$$WF_m dB(I) = WFdB(I) - 10 \log g'(I_m) + 10 \log \frac{N}{N_m} + 10 \log \frac{L_m + L_{mo}}{L + L_o} \tag{12}$$

where $WF_m dB(I) = 10\log(\Delta I_m/I)$, which is the log Weber fraction for a masked tone and $WFdB(I) = 10\log(\Delta I/I)$, which is the log Weber fraction for a tone in quiet.

Equation (12) indicates that, if $WFdB(I)$ is known at a given intensity (I), then one can predict $WF_m dB(I)$ at the same intensity from three additional measures: (1) the local slope of the loudness balance function [$g'(I_m)$], (2) a scaling factor (N/N_m), and (3) the local loudness ratio between the masked tone and the tone in quiet [$(L_m + L_{mo})/(L + L_o)$]. Interestingly, in theory, there is no need to know explicitly the detection threshold, nor the exact form of loudness growth or intensity discrimination function for the tone in quiet.

I consider Eq. (12) as a unified theory of psychophysical laws in auditory intensity perception because the last three terms in the equation contain the three previous theories that attempted to relate the jnd function to the loudness function. The $10\log g'(I_m)$ term represents Fechner's original "slope" theory; the $10\log(N/N_m)$ term represents Riesz's "proportional-jnd" theory; and the final term represents Zwislocki's "equal-loudness, equal-jnd" theory.

VALIDATION OF THE UNIFIED THEORY

Prediction of the jnd Functions in Simultaneous Masking

Simultaneous masking not only elevates a pure tone's threshold but also affects its loudness perception, similar to loudness recruitment in sensorineural hearing loss. Both loudness balance and intensity discrimination functions have been measured in the same group of listeners for pure tones in quiet and in simultaneous noise maskers (Houtsma et al., 1980; Rankovic et al., 1988; Schlauch et al., 1995).

Here, I use the Schlauch et al. (1995) data to predict the masked jnd from the quiet jnd because Schlauch et al. (1995) had the most complete set of data. **Figure 1** illustrates the relative contributions of the three special terms in Eq. (12) to predictions of the jnd data in simultaneous masking. **Figure 1A** shows three loudness balance functions: the solid line represents a hypothetical condition where the same tone is perfectly balanced in loudness (i.e., 1:1 ratio) between two ears in quiet, the dashed line represents the measured balance function for a masked tone in a 15-SPL/Hz broadband noise and the dotted line for a masked tone in a 40-dB SPL/Hz broadband noise (from Figure 3 in Schlauch et al., 1995). An interpolation of the loudness balance function is then differentiated to derive the slopes as a function of intensity (X's represent the 15 dB SPL/Hz masking and O's represent the 40 dB SPL/Hz masking condition). **Figure 1B** shows the loudness growth function for a 1000-Hz tone in quiet

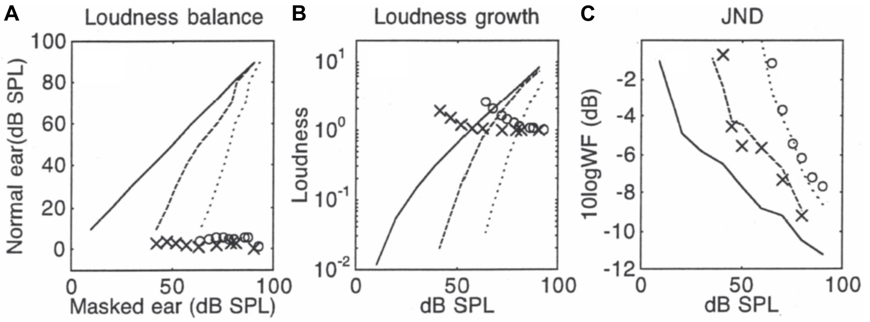

FIGURE 1 | Predictions in simultaneous masking, with data (lines) being from Schlauch et al. (1995). Panel **(A)** shows loudness balance functions between a tone in quiet (*y*-axis) and a tone in noise (*x*-axis): The solid line represents the control condition where the same tone was balanced between the two ears in quiet, the dashed line represents the balance function for a tone being masked by a 15-dB SPL/Hz broadband noise, and the dotted line represents the loudness balance function for a tone by a 40-dB SPL/Hz noise. The symbols represent slope values for the balance function. The slope values use the same scale as the balance function from 0 to 100, except the slopes are unitless. Panel **(B)** shows derived loudness growth functions. The symbols represent loudness ratio values between quiet and masked tones and tones in quiet. Panel **(C)** shows the measured jnd functions (lines) and predicted jnd values (symbols).

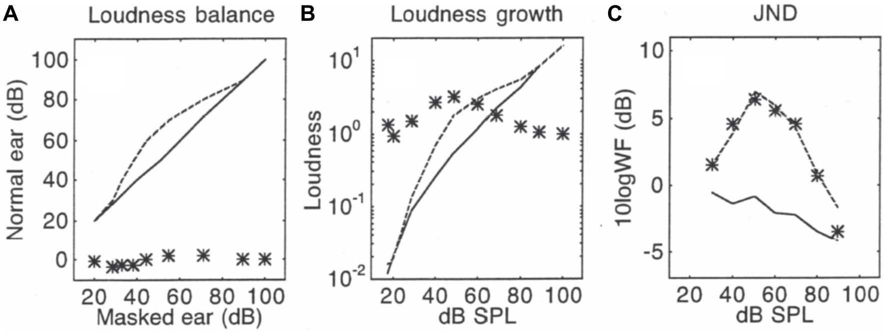

FIGURE 2 | Predictions in forward masking, with data (lines) from Zeng (1994). Panel **(A)** shows loudness balance functions between a tone in quiet (*y*-axis) and a tone in forward masking (*x*-axis): The solid line represents the control condition where the same tone was balanced between the two ears in quiet, while the dashed line represents the balance function for a tone in forward masking. The * symbols represent slope values for the balance function, which uses the same scale as the balance function from 0 to 100, except the slopes are unitless. Panel **(B)** shows derived loudness growth functions. The symbols represent loudness ratio values between the masked tone and the tone in quiet. Panel **(C)** shows the measured jnd functions (lines) and predicted jnd values (symbols).

(solid line) based on Zwislocki's model [Eq. (1), using $k = 3.1$; $\theta = 0.27$; $c = 2.5$; $I_o = 10^{-12}$ W/m^2 or 0 dB SPL], as well as the two masked loudness growth functions obtained by applying the loudness balance functions in **Figure 1A** to the loudness growth function in quiet. The X's and O's represent the loudness ratio between the corresponding quiet and masking conditions. **Figure 1C** shows measured jnd functions in quiet (solid line), 15-dB masking (dashed line), and 40-dB masking (dotted line). The X's and O's represent the predicted jnd values in the above two masking conditions based on Eq. (12). In addition to using the slope values in **Figure 1A** and loudness ratio values in **Figure 1B**, Eq. (12) uses a normalization factor of 4 dB and 8 dB for the 15-dB and 40-dB masking conditions, respectively. The 4-dB and 8-dB normalization factor was estimated from the both the dynamic range and the jnd values (Nelson et al., 1996; see their Figure 9), with the quiet condition having 2.5 times and 6.3 times more jnd steps than the 15-dB and 40-dB masking condition, respectively. There was no free parameter in this prediction. In terms of relative contributions to the successful prediction, the "equal-loudness, equal-jnd" theory was essential to the prediction of the overall trend (the same downward pattern in **Figures 1B,C**), while the slope theory (the relatively flat pattern of the X and O symbols in **Figure 1A**) behaved similarly to the proportional jnd theory as a constant to shift the predicted function up or down.

Prediction of the jnd Function in Forward Masking

Loudness and its jnd functions of a stimulus can also be affected by forward and backward masking. Loudness is enhanced and intensity discrimination is degraded in forward and backward masking, particularly at middle intensities (Zeng et al., 1991; Plack and Viemeister, 1992; Zeng and Turner, 1992). Although an early attempt to relate the "midlevel hump" (the jnd function) to loudness enhancement was not successful (Zeng, 1994), Oberfeld (2008) found a significant correlation between the elevated jnd and enhanced loudness when a wide range of masker-to-signal level differences was tested.

Using the same processing steps as in **Figures 1**, **2** shows the loudness balance function between a 25-ms tone in quiet and in

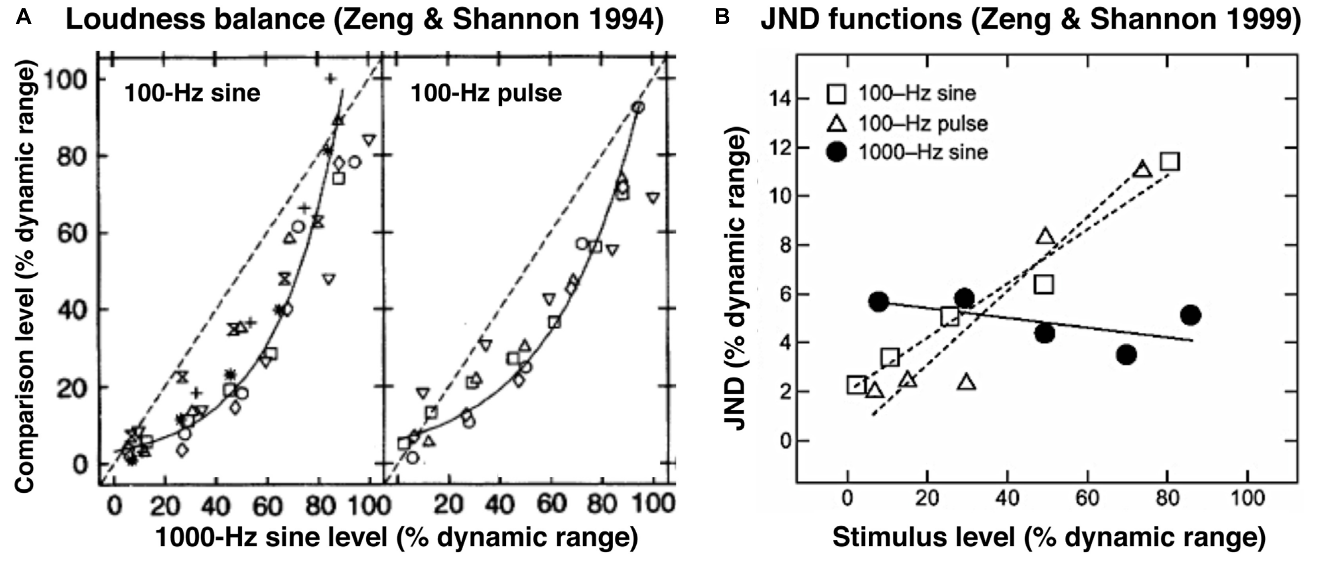

FIGURE 3 | Loudness balance **(A)** and JND functions **(B)** in cochlear implant users. **(A)** Loudness balance functions were obtained between 100-Hz sine or 100-Hz pulse and 1000-Hz sine electric stimuli, adapted from Figures 2D,E in Zeng and Shannon (1994). Reprinted with permission from AAAS. Symbols represent individual data and the solid line represents a logarithmic balance function. The dashed line represents a linear balance function, which clearly was not the true. **(B)** JND data (symbols) and predicted functions (lines) using the same stimuli from the same subjects in **(A)**, adapted from Figure 4 in Zeng and Shannon (1999). Reprinted with permission from Wolters Kluwer Health.

the presence of a 90-dB SPL, 100-ms forward masker (**Figure 2A**), the derived loudness growth function (**Figure 2B**), and the measured as well as predicted jnd functions in quiet and masking (**Figure 2C**). The slope theory (**Figure 2A**) predicted that forward masking would produce smaller than normal jnds for standard levels below 50 dB SPL but larger jnds for levels above 50 dB SPL. The "equal-loudness, equal-jnd" theory (**Figure 2B**) predicted the midlevel hump jnd function due to enhanced loudness in forward masking. A 7-dB normalization factor, or five times less jnd steps in forward masking, was used in the final successful prediction (**Figure 2C**) that combined all three special theories in Eq. (12). The similar pattern between **Figures 2B,C** is generally consistent with the observed correlation between enhanced loudness and elevated jnd (Oberfeld, 2008), but the quantitative prediction needs further investigation. It would be also interesting to know if the present unified theory could predict a similar jnd function observed for brief high-frequency tones under notched noise conditions (Carlyon and Moore, 1984). Oxenham and Moore (1995) hinted such a possibility by proposing "a new theory [that] explain[s] the severe departure from Weber's law in terms of both the variance. . . and the loudness of partially masked signals."

Predictions of the jnd Functions in Electric Hearing

In electric hearing where hair cells are missing and the auditory nerve fibers are directly stimulated by electric currents, loudness generally has a narrow dynamic range of 10–20 dB (Zeng and Galvin, 1999). Zeng and Shannon (1994) found that, in cochlear implant users, loudness grows as a traditional power function of electric current for stimulus frequencies lower than 300 Hz, but as an exponential function for stimulus frequencies higher than 300 Hz. These two different loudness growth functions would produce a logarithmic loudness balance function between low- and high-frequency electric stimuli. **Figure 3A** shows, indeed, such a logarithmic balance function (solid lines) between a 100-Hz stimulus (sinusoid or pulse amplitude on *y*-axis) and a 1000-Hz sinusoid (*x*-axis).

$$E_{1000\ Hz} = \theta \log E_{100\ Hz} \tag{13}$$

where θ is the slope of the logarithmic loudness balance function. Differentiating the above equation to derive the following JND function between the high- and low-frequency electric stimuli:

$$\Delta E_{1000\ Hz} = \theta \frac{\Delta E_{100\ Hz}}{E_{100\ Hz}} \tag{14}$$

Zeng and Shannon (1999) measured jnds of these stimuli in the same implant subjects (symbols in **Figure 3B**) and found that not only did this jnd function hold but more importantly the jnd function was nearly constant (the solid line in **Figure 3B**). Given the same power loudness growth function for the 100-Hz electric stimuli, it is not surprising that their Weber fraction was also constant. But why was the absolute difference ($\Delta E_{1000\ Hz}$) constant for the 1000-Hz stimulus? Zeng and Shannon (1999) showed that this constant absolute difference was a result of the exponential loudness growth function.

$$L_{1000\ Hz} = \exp(E_{1000\ Hz}) \tag{15}$$

Differentiating the above equation to obtain:

$$\frac{\Delta L_{1000\ Hz}}{\Delta E_{1000\ Hz}} = \exp(E_{1000\ Hz}) = L_{1000\ Hz} \tag{16}$$

Rewriting the above equation to obtain:

$$\frac{\Delta L_{1000\ Hz}}{L_{1000\ Hz}} = \Delta E_{1000\ Hz} \tag{17}$$

Equation (17) means that Brentano's ratio is also constant in electric stimulation. The only difference between Eqs. (17) and (4) is that (17) does not contain a threshold term, probably due to a lack of spontaneous neural activity in the deafened ear (Kiang and Moxon, 1972).

DISCUSSION

None of the individual components in the present unified theory is new. Previous studies have proposed these individual theories and evaluated them separately (e.g., Zwislocki and Jordan, 1986; Hellman and Hellman, 1990, 2001; Schlauch, 1994; Schlauch et al., 1995; Allen and Neely, 1997). The present study is novel in two respects. First, the present study integrates the previously disconnected individual components through a unified theoretical framework, namely, the general form of Brentano's law in Eq. (4). Second, the present study offers a new formula, namely, Eq. (12), which specifically combines these individual terms to successfully predict the loudness and jnd relationships in simultaneous and forward masking, as well as in cochlear implant users. The present unified theory and its successful applications suggest that although Weber's law needs to be replaced by the general form of Brentano's law, Fechner's original idea using jnds to derive psychophysical laws is valid at least in the wide range of hearing situations examined here.

The general form of Brentano's law can be used to examine how close the actual jnd data follow Weber's law and its potential mechanisms by combining Eqs. (4) and (5):

$$\frac{\Delta L}{L + L_0} = \theta \frac{\Delta I}{I + cI_0} = wI^{\alpha} \ or \ \frac{\Delta I}{I + cI_0} = w'I^{\alpha} \tag{18}$$

where both $w'(= w/\theta)$ and α are free parameters to be estimated, with $\alpha = 0$ indicating perfect conformity to Weber's law. **Figure 4** shows the jnd data and the model estimation for a 1-kHz tone (Schlauch et al., 1995), 8-kHz broadband noise (6–14 kHz) and the same noise in a notched noise background (Viemeister, 1983). All three sets of data can be modeled by a two-stage function, with a steep first stage (~10–20 dB SPL) reflecting the threshold influence and a shallower second stage (~20–100 dB SPL) with its slope being α in Eq. (16). All three sets of data follow the near-miss to Weber's law (McGill and Goldberg, 1968), with α being -0.09 for the tone, -0.03 for the noise, and 0.04 for the noise in a notched noise background. The near-miss ranges from -9% to 4% and has an average of 3% for the three stimuli considered here.

To provide a solution to the near-miss to Weber's law, McGill and Goldberg adopted a Poisson-like process, in which the loudness mean (L) and its variance (σ^2) are equal, where σ is the standard deviation. To achieve 75% correct detection in a jnd task, the signal detection theory requires: $d' = \frac{\Delta L}{\sigma} = \frac{\Delta L}{L^{0.5}} = 1$ (Green and Swets, 1966). Replacing $\Delta L = L^{0.5}$ in Eq. (19) to produce:

$$\frac{\Delta L}{L + L_0} = \frac{L^{0.5}}{L + L_0} \propto L^{-0.5} \propto (I^{0.27})^{-0.5} \propto I^{-0.14} \tag{19}$$

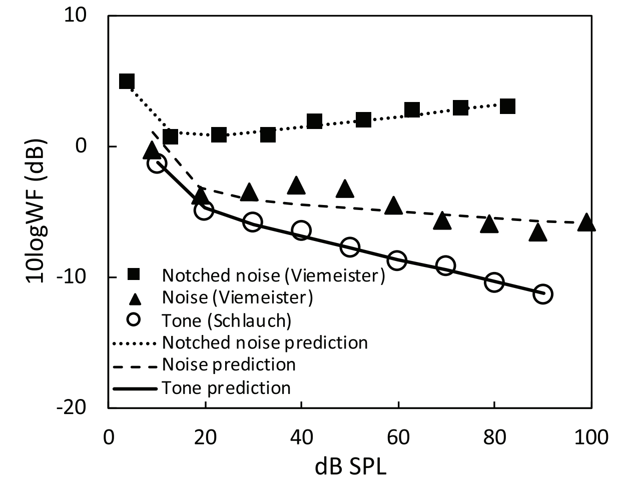

FIGURE 4 | Prediction of JND for noise and tone stimuli. The JND data for a broadband noise (solid triangles) and the same noise in a notched-noise background (solid squares) were from Viemeister (1983; the same symbols in his **Figure 1**) and the 1000-Hz tone JND data (open circles) were from Schlauch et al. (1995; circles in their **Figure 2** bottom-right panel). The dashed line represents prediction of the noise JND function, the dotted line represents the noise in a notched-noise background, and the solid line represents the tone JND function.

Compared with the -0.14 slope predicted by the Poisson-like process, the estimated slope was is 5% off for the tone, 11% off for the noise and 18% off for the noise in a notched-noise background. As an overcorrection, McGill and Goldberg's solution has created a much greater difference (average = 11%) than the original problem, i.e., the near-miss (average = 3%) to Weber's law. Alternatively, the use of spread of excitation cue is the more likely mechanism underlying the near-miss to Weber's law (Florentine and Buus, 1981; Viemeister, 1983), but a quantitative treatment of its predictive accuracy is still lacking. At least as a first-order approximation, Weber's law holds for sound intensity discrimination.

While it is challenging, the search for a unified psychophysical law has continued to attract attention, especially on its biological basis (e.g., Shepard, 1987; Nieder and Miller, 2003; Dehaene et al., 2008; Dzhafarov and Colonius, 2011; Teghtsoonian, 2012; Pardo-Vazquez et al., 2019). In an influential paper, which drew 30 open peer commentaries, Krueger (1989) attempted to reconcile Fechner and Stevens by proposing a unified psychophysical law, in which (1) "each jnd has the same subjective magnitude for a given modality," (2) "subjective magnitude increases as approximately a power function of physical magnitude," and (3) "subjective magnitude depends primarily on peripheral sensory processes, that is, no non-linear central transformations occur." With regard to (1), Krueger preferred ΔS or in the present term $\Delta L = c$ (constant) for the law of parsimony, but was willing to accept $\Delta L/L = c$ (Brentano's Law) or even $\Delta L/L = L^{-0.5}$ (McGill and Goldberg's Poisson process). The present study favors Brentano's Law with a threshold correction factor. The second point was the primary concern of Kruger's unified law, in which not only did he attempt to reconcile the different ways to measure sensation magnitude (e.g., magnitude estimate versus categorical rating), but also derive the subjective magnitude function from the jnd data. He explicitly examined the "proportional-jnd theory" (p. 260), implicitly discussed the "slope" theory (his Table 1 on p. 261), but probably didn't know about the "equal-loudness, equal-jnd" theory, letting alone consider them as three independent factors that collectively contribute to the jnd-loudness function (the present study). Kruger's third point treating the brain as a linear device is wrong, because not only does the present study (B3) show that electric stimulation of the auditory nerve, which bypasses the auditory hair cells, produces an exponential loudness function in cochlear implant users, but more importantly many studies on neuroplasticity have found abnormally increased gain in the brain in response to reduced input in the periphery (e.g., Qiu et al., 2000; Norena, 2011; Chambers et al., 2016).

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethical approval was not required for this study as the human data were from previously published studies. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

F-GZ is solely responsible for the work presented here.

ACKNOWLEDGMENTS

Joe Zwislocki stimulated my interest in this subject when I was a graduate student taking his classes at Syracuse University. I learned a great deal, and also benefited from intellectual discussion with Chris Turner, Neal Viemeister, Bert Schlauch, Adrian Houtsma, Bob Shannon, Walt Jesteadt, Jont Allen, Brian Moore, Bob Carlyon, Chris Plack, Duncan Luce, Bill McGill, Bert Scharf, Rhona, and Bill Hellman. Maaike Van Eeckhoutte motivated the discussion on Weber's law (**Figure 4**), Daniel Oberfeld and Wolfgang Ellermeier made critical comments, which have helped tie up several loose ends and broaden the scope of the discussion in this article. However, I am solely responsible for the work presented here.

REFERENCES

Allen, J. B., and Neely, S. T. (1997). Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise. *J. Acoust. Soc. Am.* 102, 3628–3646. doi: 10.1121/1.420150

Buus, S., and Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: recruitment reconsidered. *J. Assoc. Res. Otolaryngol.* 3, 120–139. doi: 10.1007/s101620010084

Carlyon, R. P., and Moore, B. C. (1984). Intensity discrimination: a severe departure from Weber's law. *J. Acoust. Soc. Am.* 76, 1369–1376. doi: 10.1121/1.391453

Chambers, A. R., Resnik, J., Yuan, Y., Whitton, J. P., Edge, A. S., Liberman, M. C., et al. (2016). Central gain restores auditory processing following near-complete cochlear denervation. *Neuron* 89, 867–879. doi: 10.1016/j.neuron.2015.12.041

Dehaene, S., Izard, V., Spelke, E., and Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* 320, 1217–1220. doi: 10.1126/science.1156540

Dzhafarov, E. N., and Colonius, H. (2011). The fechnerian idea. *Am. J. Psychol.* 124, 127–140. doi: 10.5406/amerjpsyc.124.2.0127

Fechner, G. T. (1966). *Elemente der Psychophysik [Elements of Psychophysics]*. New York, NY: Holt, Rinehart and Winston, Inc.

Florentine, M., and Buus, S. (1981). An excitation-pattern model for intensity discrimination. *J. Acoust. Soc. Am.* 70, 1646–1654. doi: 10.1121/1.387219

Fowler, E. P. (1937). Measuring the sensation of loudness a new approach to the physiology of hearing and the functional and differential diagnostic tests. *Arch. Otolaryngol.* 26, 514–521. doi: 10.1001/archotol.1937.00650020568002

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Hellman, R., Scharf, B., Teghtsoonian, M., and Teghtsoonian, R. (1987). On the relation between the growth of loudness and the discrimination of intensity for pure tones. *J. Acoust. Soc. Am.* 82, 448–453. doi: 10.1121/1.395445

Hellman, W. S., and Hellman, R. P. (1990). Intensity discrimination as the driving force for loudness. Application to pure tones in quiet. *J. Acoust. Soc. Am.* 87, 1255–1265. doi: 10.1121/1.398801

Hellman, W. S., and Hellman, R. P. (2001). Revisiting relations between loudness and intensity discrimination. *J. Acoust. Soc. Am.* 109(5 Pt 1), 2098–2102. doi: 10.1121/1.1366373

Houtsma, A. J., Durlach, N. I., and Braida, L. D. (1980). Intensity perception XI. Experimental results on the relation of intensity resolution to loudness matching. *J. Acoust. Soc. Am.* 68, 807–813. doi: 10.1121/1.384819

Johnson, J. H., Turner, C. W., Zwislocki, J. J., and Margolis, R. H. (1993). Just noticeable differences for intensity and their relation to loudness. *J. Acoust. Soc. Am.* 93, 983–991. doi: 10.1121/1.405404

Kiang, N. Y., and Moxon, E. C. (1972). Physiological considerations in artificial stimulation of the inner ear. *Ann. Otol. Rhinol. Laryngol.* 81, 714–730. doi: 10.1177/000348947208100513

Krueger, L. E. (1989). Reconciling fechner and stevens - toward a unified psychophysical law. *Behav. Brain Sci.* 12, 251–267. doi: 10.1017/S0140525x0004855x

Lim, J. S., Rabinowtiz, W. M., Braida, L. D., and Durlach, N. I. (1977). Intensity perception. VIII. Loudness comparisons between different types of stimuli. *J. Acoust. Soc. Am.* 62, 1256–1267. doi: 10.1121/1.381641

McGill, W. J., and Goldberg, J. (1968). A study of the near-miss involving Weber's law and pure-tone intensity discrimination. *Percept. Psychophys.* 4, 105–109. doi: 10.3758/bf03209518

Miller, G. (1947). Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *J. Acoust. Soc. Am.* 19, 609–619. doi: 10.1121/1.1916528

Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (1996). Intensity discrimination as a function of stimulus level with electric stimulation. *J. Acoust. Soc. Am.* 100(4 Pt 1), 2393–2414. doi: 10.1121/1.417949

Newman, E. (1933). The validity of the just noticeable difference as a unit of psychological magnitude. *Trans. Kansas Acad. Sci.* 36, 172–175. doi: 10.2307/3625353

Nieder, A., and Miller, E. K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37, 149–157. doi: 10.1016/s0896-6273(02)01144-1143

Norena, A. J. (2011). An integrative model of tinnitus based on a central gain controlling neural sensitivity. *Neurosci. Biobehav. Rev.* 35, 1089–1109. doi: 10.1016/j.neubiorev.2010.11.003

Oberfeld, D. (2008). The mid-difference hump in forward-masked intensity discrimination. *J. Acoust. Soc. Am.* 123, 1571–1581. doi: 10.1121/1.2837284

Oxenham, A. J., and Moore, B. C. J. (1995). Overshoot and the severe departure from webers law. *J. Acoust. Soc. Am.* 97, 2442–2453. doi: 10.1121/1.411965

Pardo-Vazquez, J. L., Castineiras-de Saa, J. R., Valente, M., Damiao, I., Costa, T., Vicente, M. I., et al. (2019). The mechanistic foundation of Weber's law. *Nat. Neurosci.* 22, 1493–1502. doi: 10.1038/s41593-019-0439-7

Plack, C. J., and Viemeister, N. F. (1992). Intensity discrimination under backward masking. *J. Acoust. Soc. Am.* 92, 3097–3101. doi: 10.1121/1.404205

Qiu, C., Salvi, R., Ding, D., and Burkard, R. (2000). Inner hair cell loss leads to enhanced response amplitudes in auditory cortex of unanesthetized chinchillas: evidence for increased system gain. *Hear. Res.* 139, 153–171. doi: 10.1016/s0378-5955(99)00171-9

Rankovic, C. M., Viemeister, N. F., Fantini, D. A., Cheesman, M. F., and Uchiyama, C. L. (1988). The relation between loudness and intensity difference limens for tones in quiet and noise backgrounds. *J. Acoust. Soc. Am.* 84, 150–155. doi: 10.1121/1.396981

Riesz, R. (1933). The relationship between loudness and the minimum perceptible increment of intensity. *J. Acoust. Soc. Am.* 4, 211–216. doi: 10.1121/1.1915601

Schlauch, R. S. (1994). Intensity resolution and loudness in high-pass noise. *J. Acoust. Soc. Am.* 95, 2171–2179. doi: 10.1121/1.410017

Schlauch, R. S., Harvey, S., and Lanthier, N. (1995). Intensity resolution and loudness in broadband noise. *J. Acoust. Soc. Am.* 98, 1895–1902. doi: 10.1121/1.413375

Schlauch, R. S., and Wier, C. C. (1987). A method for relating loudness-matching and intensity-discrimination data. *J. Speech Hear. Res.* 30, 13–20. doi: 10.1044/jshr.3001.13

Shepard, R. N. (1987). Toward a Universal law of generalization for psychological science. *Science* 237, 1317–1323. doi: 10.1126/science.3629243

Stevens, S. S. (1961). To honor fechner and repeal his law: a power function, not a log function, describes the operating characteristic of a sensory system. *Science* 133, 80–86. doi: 10.1126/science.133.3446.80

Stillman, J. A., Zwislocki, J. J., Zhang, M., and Cefaratti, L. K. (1993). Intensity just-noticeable differences at equal-loudness levels in normal and pathological ears. *J. Acoust. Soc. Am.* 93, 425–434. doi: 10.1121/1.405622

Teghtsoonian, R. (1971). On the exponents in Stevens' law and the constant in Ekman's law. *Psychol. Rev.* 78, 71–80. doi: 10.1037/h0030300

Teghtsoonian, R. (2012). The standard model for perceived magnitude: a framework for (almost) everything known about it. *Am. J. Psychol.* 125, 165–174. doi: 10.5406/amerjpsyc.125.2.0165

Viemeister, N. F. (1983). Auditory intensity discrimination at high frequencies in the presence of noise. *Science* 221, 1206–1208. doi: 10.1126/science.6612337

Viemeister, N. F., and Bacon, S. P. (1988). Intensity discrimination, increment detection, and magnitude estimation for 1-kHz tones. *J. Acoust. Soc. Am.* 84, 172–178. doi: 10.1121/1.396961

Zeng, F. G. (1994). Loudness growth in forward masking: relation to intensity discrimination. *J. Acoust. Soc. Am.* 96, 2127–2132. doi: 10.1121/1.410154

Zeng, F. G. (2013). An active loudness model suggesting tinnitus as increased central noise and hyperacusis as increased nonlinear gain. *Hear. Res.* 295, 172–179. doi: 10.1016/j.heares.2012.05.009

Zeng, F. G., and Galvin, J. J. III (1999). Amplitude mapping and phoneme recognition in cochlear implant listeners. *Ear Hear.* 20, 60–74. doi: 10.1097/00003446-199902000-00006

Zeng, F. G., and Shannon, R. V. (1994). Loudness-coding mechanisms inferred from electric stimulation of the human auditory system. *Science* 264, 564–566. doi: 10.1126/science.8160013

Zeng, F. G., and Shannon, R. V. (1999). Psychophysical laws revealed by electric hearing. *Neuroreport* 10, 1931–1935. doi: 10.1097/00001756-199906230-00025

Zeng, F. G., and Turner, C. W. (1992). Intensity discrimination in forward masking. *J. Acoust. Soc. Am.* 92(2 Pt 1), 782–787. doi: 10.1121/1.403947

Zeng, F. G., Turner, C. W., and Relkin, E. M. (1991). Recovery from prior stimulation. II: effects upon intensity discrimination. *Hear. Res.* 55, 223–230. doi: 10.1016/0378-5955(91)90107-k

Zwislocki, J. (1965). "Analysis of some auditory characteristics," in *Handbook of Mathematical Psychology*, eds R. Luce, R. R. Bush, and E. Galanter (New York, NY: John Wiley and Sons, Inc), 79–97.

Zwislocki, J. J., and Jordan, H. N. (1986). On the relations of intensity jnd's to loudness and neural noise. *J. Acoust. Soc. Am.* 79, 772–780. doi: 10.1121/1.393467

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

ORIGINAL RESEARCH
published: 17 November 2020
doi: 10.3389/fpsyg.2020.578352

Maximum Expected Information Approach for Improving Efficiency of Categorical Loudness Scaling

Sara E. Fultz*, Stephen T. Neely, Judy G. Kopun and Daniel M. Rasetshwane

Center for Hearing Research, Boys Town National Research Hospital, Omaha, NE, United States

Categorical loudness scaling (CLS) measures provide useful information about an individual's loudness perception across the dynamic range of hearing. A probability model of CLS categories has previously been described as a multi-category psychometric function (MCPF). In the study, a representative "catalog" of potential listener MCPFs was used in conjunction with maximum-likelihood estimation to derive CLS functions for participants with normal hearing and with hearing loss. The approach of estimating MCPFs for each listener has the potential to improve the accuracy of the CLS measurements, particularly when a relatively low number of data points are available. The present study extends the MCPF approach by using Bayesian inference to select stimulus parameters that are predicted to yield maximum expected information (MEI) during data collection. The accuracy and reliability of the MCPF-MEI approach were compared to the standardized CLS measurement procedure (ISO 16832:2006, 2006). A non-adaptive, fixed-level, paradigm served as a "gold-standard" for this comparison. The test time required to obtain measurements in the standard procedure is a major barrier to its clinical uptake. Test time was reduced from approximately 15 min to approximately 3 min with the MEI-adaptive procedure. Results indicated that the test–retest reliability and accuracy of the MCPF-MEI adaptive procedures were similar to the standardized CLS procedure. Computer simulations suggest that the reliability and accuracy of the MEI procedure were limited by intrinsic uncertainty of the listeners represented in the MCPF catalog. In other words, the MCPF provided insufficient predictive power to significantly improve adaptive-tracking efficiency under practical conditions. Concurrent optimization of both the MCPF catalog and the MEI-adaptive procedure have the potential to produce better results. Regardless of the adaptive-tracking method used in the CLS procedure, the MCPF catalog remains clinically useful for enabling maximum-likelihood determination of loudness categories.

Keywords: loudness, loudness perception, psychoacoustics, maximum likelihood, categorical loudness scaling

OPEN ACCESS

Edited by:
*Maaike Van Eeckhoutte,
Technical University of Denmark,
Denmark*

Reviewed by:
*Philippe Fournier,
Aix-Marseille Université, France
Jesko Lars Verhey,
University Hospital Magdeburg,
Germany
Dirk Oetting,
HörTech gGmbH, Germany*

***Correspondence:**
*Sara E. Fultz
Sara.Fultz@boystown.org*

Specialty section:
*This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Psychology*

Received: *30 June 2020*
Accepted: *29 October 2020*
Published: *17 November 2020*

Citation:
*Fultz SE, Neely ST, Kopun JG and
Rasetshwane DM (2020) Maximum
Expected Information Approach
for Improving Efficiency of Categorical
Loudness Scaling.
Front. Psychol. 11:578352.
doi: 10.3389/fpsyg.2020.578352*

INTRODUCTION

Loudness is the perceptual correlate of the physical intensity of a sound (Fletcher and Munson, 1933). A variety of psychometric procedures may be used to quantify loudness in humans, including but not limited to: loudness matching, magnitude estimation, cross-modality matching, and loudness scaling (Cox, 1989; Kollmeier and Hohmann, 1995). Categorical loudness scaling

(CLS) is a procedure in which listeners assign meaningful labels to stimuli of varying intensities as a means of estimating loudness growth with increasing stimulus level (Brand and Hohmann, 2002; ISO 16832:2006, 2006). Measurements of loudness perception offer insight into auditory health because they become altered when the cochlea is damaged (e.g., Allen, 2008). CLS has often been used for studying loudness perception in listeners with sensorineural hearing loss due to both its ease of testing and validity (Al-Salim et al., 2010; Rasetshwane et al., 2015, 2018; Oetting et al., 2016). CLS measurements have been used to assess loudness perception in patients with tinnitus (Hébert et al., 2013) and hyperacusis, which is a reduced tolerance to loud sounds (Noreña and Chery-Croze, 2007). CLS procedures have also been used to evaluate abnormalities in loudness perception in patients with autism (Khalfa et al., 2004) and in concussed athletes (Assi et al., 2018).

New hearing aid users often have complaints about the loudness and annoyance of certain sounds. Although abnormal loudness perception is a driving factor in dissatisfaction with hearing aids (Blamey and Martin, 2009), loudness is not typically measured during the clinical hearing aid fitting process. This is in part due to concerns related to the reliability, accuracy, and test time required to obtain loudness measures, and because the nature of suprathreshold variability across listeners is not yet fully understood (Elberling, 1999; Al-Salim et al., 2010).

Several procedures have been used in previous studies to calculate a CLS function from trial-by-trial data. These include (1) fitting a loudness model (two segment straight lines) to the trial-by-trial data (e.g., Brand and Hohmann, 2002; Heeren et al., 2013; Oetting et al., 2014), and (2) fitting a model to the median of the trial-by-trial data (Al-Salim et al., 2010; Rasetshwane et al., 2015). It has been noted that these procedures can lead to over-smoothing of the data (Trevino et al., 2016a; Wròblewski et al., 2017) and that using the median of trial-by-trial data may produce more reliable results. In the current study, we follow the method described in Trevino et al. (2016a).

We previously developed a probability model of CLS that characterizes loudness-category selection as a multi-category psychometric function (MCPF) (Trevino et al., 2016a), which is a generalization of the commonly used two-category psychometric function. The MCPF provides a more comprehensive characterization of the variability associated with listener responses because it combines all categories into a single framework. The MCPF provides a statistical basis for smoothing listener responses across categories that supports a maximum-likelihood determination of loudness-category boundaries for a given set of responses. The MCPF adds a new dimension to CLS data and facilitates parameterization of suprathreshold variability across listeners. In the present study, we extend the MCPF approach by using Bayesian inference to select stimulus parameters that are predicted to yield maximum expected information (MEI) during data collection.

We then assess the test–retest reliability and accuracy of an adaptive procedure that utilizes a limited number of trials for MCPF-MEI. Test–retest reliability was assessed across two visits. For assessment of accuracy, the International Standards Organization (ISO) fixed-level procedure, which utilizes numerous trials, served as the reference procedure for estimating a listener's CLS function (Brand and Hohmann, 2002; Kinkel, 2007). Improving the reliability and accuracy of CLS procedures may enhance the clinical acceptability of loudness measurements and potentially improve hearing aid fitting methods.

Entropy is an information-theoretic concept that quantifies the randomness (or uncertainty) of a system that has many possible states. The entropy of any system has its maximum value when all possible states are equally likely. Entropy is reduced when information becomes available that makes some states more likely than other states. Thus, entropy and information have a complementary relationship. Information increase is always associated with an equal amount of entropy reduction. In the context of CLS measurements, each trial, which consists of a listener's response to a particular stimulus, provides a small amount of new information about the listener's loudness perception. When listener responses are reliable (e.g., when listener responses are monotonic functions of stimulus level), the accumulated information increases, and the entropy is reduced, as the number of trials increases. This study investigated the idea that the efficiency of a CLS test could be improved by selecting the stimulus for each trial that is expected to provide the maximum amount of information from the response portion of that trial.

In this study, we compared two different adaptive-tracking methods: (1) the standard CLS method described by ISO 16832:2006 (2006) and (2) the MEI method. The "gold-standard" for this comparison was a non-adaptive, fixed-level method, which was not considered to be clinically viable because it required too much time. A further comparison was included in the method used to construct the MEI loudness functions from the trial-by-trial data: (1) median sound pressure level (SPL) within each loudness category and (2) maximum likelihood (ML) MCPF.

MATERIALS AND METHODS

Participants

Forty-five adults participated in this study (23 female). The demographic makeup of our sample was 91.9% Not Hispanic, 4.4% Hispanic, and 4.4% Not Reported. The participants were 77.8% White, 11.1% Black, 0% American Indian and Alaska Native, 0% Asian, 0% Native Hawaiian and Other Pacific Islander, 4.4% Two or More Races, and 6.7% Not Reported. According to the United States Census 2018 American Community Survey, the demographic makeup of our local community, Omaha, NE is 85.3% Not Hispanic and 14.7% Hispanic. The city population is 77.0% White, 12.1% Black, 0.9% American Indian and Alaska Native, 3.7% Asian, 0.0% Native Hawaiian and Other Pacific Islander, and 3.6% Two or More Races. The demographic makeup of the United States is 81.5% Not Hispanic and 18.5% Hispanic. The population is 72.2% White, 12.7% Black, 0.9% American Indian and Alaska Native, 5.6% Asian, 0.2% Native Hawaiian and Other Pacific Islander, and 3.4% Two or More Races (American Community Survey 2018). All participants reported English as their primary language.

All participants were recruited from a database of potential research participants that is maintained by Boys Town National Research Hospital (BTNRH). Data collection was conducted under a protocol that was approved by the BTNRH Institutional Review Board. Informed consent was obtained prior to testing and participants were compensated for their participation.

Audiometric thresholds were measured at eight frequencies (0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz) with an audiometer (GSI AudioStar Pro, Grason-Stadler, Eden Prairie, MN, United States) using ER3A headphones (Etymotic Research, Elk Grove Village, IL, United States) following the Hughson-Westlake procedure (American Speech-Language-Hearing Association [ASHA], 1978). Participants were classified as having normal hearing when thresholds in the test ear were ≤15 dB HL at all audiometric frequencies. Participants were classified as having sensorineural hearing loss when thresholds in the test ear were ≥20 dB HL at both of the test frequencies used for the CLS procedures, 1 and 4 kHz. Fifteen participants had normal hearing (age range 21–74, mean 43 years) and thirty participants had hearing loss (age range 23–74, mean 55 years). Participants with sensorineural hearing loss had audiometric thresholds ≤75 dB HL at the test frequencies for the CLS procedures. The distribution of audiometric thresholds is displayed in **Figure 1**.

All participants had normal middle-ear status in the test ear based on normal otoscopic inspection, normal 226-Hz tympanogram, and air-bone gaps ≤10 dB from 0.5 to 4 kHz. The inclusion criteria for tympanometry (Madsen Otoflex 100, GN Otometrics, Denmark) required peak-compensated static acoustic admittance between 0.3 and 2.5 mmhos and peak tympanometric pressure between −100 and +50 daPa.

All CLS testing was conducted monaurally. If both ears met the inclusion criteria, the better ear was selected for testing. If the thresholds were symmetrical, the test ear was selected randomly, though there was an attempt to balance the number of left and right ears. Overall, data were collected from 21 right and 24 left ears.

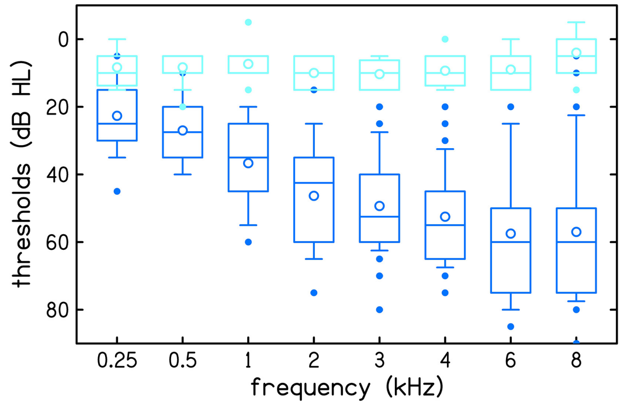

FIGURE 1 | Audiometric thresholds of 15 normal hearing participants (light blue) and 30 participants with hearing loss (dark blue). Boxes represent the interquartile range and whiskers represent the 10th and 90th percentiles. Outliers, defined as data points that are outside the 10th to 90th percentile range, are plotted using filled circles. Within each box, lines represent the median and open circles represent the mean.

Procedures

Participants were seated in a sound-treated room. Pure-tone stimuli (1 and 4 kHz) at levels ranging from 0 to 110 dB SPL were presented monaurally for each of three CLS procedures: (1) fixed-level procedure, (2) slope-adaptive procedure, and (3) MEI-adaptive procedure. Pure tones were 1000 ms in duration with a 20 ms rise/fall time. Stimuli were generated using custom-designed software (MATLAB) that controlled a 24-bit soundcard (Babyface Pro, RME) and were presented to the participants' ear with an insert earphone (ER3A; Etymotic Research, Elk Grove, MN).

The CLS procedure closely followed the ISO standard (ISO 16832:2006, 2006), though it was not our intention to replicate it exactly as described. The procedure determined the level of sounds that corresponded to 11 different loudness categories, with seven of these categories assigned meaningful labels ("Can't Hear," "Very Soft," "Soft," "Medium," "Loud," "Very Loud," and "Too Loud"). The categories were graphically displayed on a computer monitor as colored horizontal bars that increased in length from bottom ("Can't Hear") to top ("Too Loud"). The response window is displayed in **Figure 2**. After listening to each stimulus, participants selected a category that best represented their perception of the loudness of the sound. Participants were instructed to select "Too Loud" if the sound was loud enough that they wouldn't want to hear it again and "Very Soft" if the sound was just detectable. The labels used for boundary categories are different than the labels used in the ISO 16832:2006 (2006) but matched those used in our previous studies (Rasetshwane et al., 2015, 2018). However, it should be noted that the ISO standard is open to the use of different labels, including symbols. For the purpose of numerical representation, the 11 loudness categories were assigned categorical units (CUs) ranging from 0 ("Can't Hear") to 50 ("Too Loud") in steps of 5. This numerical representation was not shown to participants.

Participants completed one practice run at one frequency, 1.5 kHz, of either the slope-adaptive or MEI-adaptive procedure, selected randomly. Six conditions were then collected (1 and 4 kHz for each of the three procedures), with the procedure and test frequency randomized for each participant. Data collection was repeated over two visits separated by at least 1 day and up to 42 days. The average number of days between visits was 10.

The CLS test included two stages. The participant's dynamic range was determined in the first stage, in the first stage, and a loudness function was measured in the second stage. The procedure for determining the dynamic range was the same for all three CLS methods. In this procedure, two sequences of stimuli were interleaved, one sequence ascending in level and the other descending in level. The lower end of a participant's dynamic range was based on the last audible level ("Very Soft" category) of the descending sequence, while the upper end was based on the last level of the ascending sequence that was not judged as "Too Loud." The starting level was set equal to the midpoint of the participant's dynamic range.

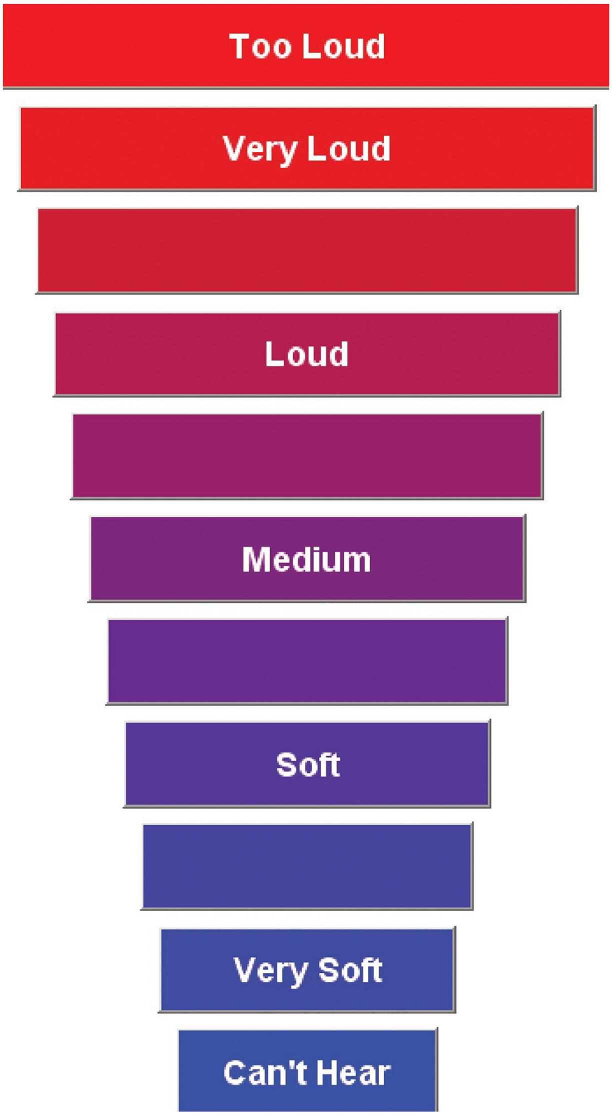

FIGURE 2 | Display of the categorical loudness scale with 11 response categories. This is displayed on a computer monitor used by participants to rate the loudness of the signal. The horizontal bars increase in width from the softest level to the loudest level. This figure appeared previously in Rasetshwane et al. (2015).

Procedures for measuring the loudness function differed by CLS method. For the non-adaptive, fixed-level procedure, up to 22 distinct levels spanning the dynamic range were presented in 5 dB steps. The exact number of levels depended on the listener's dynamic range. Each level was repeated 10 times, for a total of up to 220 trials. Levels were randomized with restrictions that the same level was never presented consecutively and differences between consecutive levels never exceeded 45 dB.

For the adaptive procedures (slope-adaptive and MEI-adaptive), nine levels within the dynamic range were presented. The run of nine trials was repeated five times, for a total of 45 trials. In the slope-adaptive procedure, the nine levels evenly spanned the dynamic range. In the MEI-adaptive approach, MEI was used to select the next stimulus level as the one that minimized entropy based on the MCPF catalog. In contrast to the fixed-level procedure, the dynamic range of the presentation levels was not fixed during the test for the adaptive procedures. The listeners were instructed to select "Too Loud" if they felt the sound was loud enough that they did not want to hear it again. Thus, whenever a listener responded with "Too Loud," the upper limit was reduced by 5 dB for the next run to avoid presenting uncomfortable loud sounds. If a listener did not respond with "Too Loud" to any of the nine levels within a run, then the upper limit of the dynamic range was increased by 5 dB for the next run, but never exceeded the 110 dB SPL limit.

A catalog of MCPFs that represent a wide range of potential listeners was created based on fixed-level trial-by-trial data obtained at two frequencies (1 and 4 kHz) from 16 listeners with normal hearing and 25 listeners with sensorineural hearing loss (Trevino et al., 2016a). MCPF generalizes the concept of a psychometric function (the probability of a particular response in a two-alternative paradigm as a function of an experimental variable) to more than two possible responses and represents the probability distribution across multiple response categories as a function of an experimental variable (e.g., Torgerson, 1958). Within the context of CLS data, a MCPF described how loudness category probabilities change with stimulus level. The Trevino et al. catalog has a total of 1460 MCPFs entries. The MEI-adaptive procedure used MEI to select the next stimulus level as the one that minimized entropy based on the MCPF catalog.

Entropy is an information-theoretic measure of how much information is needed to determine an unknown variable (i.e., the *uncertainty* of the variable). In this case, the listener's CLS function is the unknown variable. At the beginning of the experiment, no prior information is known, many CLS functions are equally probable, and thus the entropy is at a maximum. With each stimulus-response trial, some CLS functions can be determined to be more statistically probable than others, and the entropy is reduced. For each additional trial, the stimulus level that leads to the most entropy reduction is the one that provides the maximum information. MEI is an iterative algorithm that uses the catalog of parameterized CLS psychometric functions to calculate entropy. With each stimulus-response trial, the probability of each potential CLS psychometric function is updated. After updating, the probability-weighted expected entropy of all experimental stimulus levels is computed. The stimulus level with the greatest expected entropy reduction (i.e., provides the MEI) is selected as the level for the following stimulus presentation.

The calculation of entropy is based on posterior probability distribution. At the start of each track, prior to the first trial, each catalog entry is assumed to be equally likely. With each stimulus-response trial, the probability of each potential CLS psychometric function is updated which alters the distribution of probabilities associated with MCPF catalog entries. The procedure for updating the likelihood of each entry after each trial was described by Trevino et al. (2016a). A probability for each entry was calculated by dividing the likelihood for each entry by the sum of the likelihoods for all entries. Prior probabilities are transformed into posterior probabilities by applying relevant conditional probabilities contained in the catalog. The entropy of each posterior probability distribution was calculated by the usual definition as minus the expected value of the log (base 2) of the entry probability. Thus, this entropy decreases with each additional stimulus-response trial.

Analyses

Loudness-growth functions (loudness in CU as a function of SPL) were generated for each participant from their trial-by-trial responses. The term *trial* refers to a single stimulus/response pair. For all three procedures, CLS functions were obtained by calculating the median SPL for each CU. Unlike our previous procedures (Al-Salim et al., 2010; Rasetshwane et al., 2015), outliers were not removed. However, the drawback is the median value may be based on a single response for sparse data. In addition to calculating a loudness function based on the median SPL for each CU, ML estimation was used to select one MCPF from the catalog that was the best fit to each listener's responses. Each MCPF describes all boundaries between adjacent loudness categories as individual psychometric functions. The 50% on each of these boundary functions was used to construct conventional CLS loudness growth functions. Although the MEI-ML procedure was intended to be an update of the MEI-Med method, it was not known prior to the study how the two methods would compare, therefore, both methods were applied to the data. Thus, there were a total of four CLS functions: fixed-level, slope-adaptive, MEI-Median (Med) and MEI-ML. See Trevino et al. (2016a) for detailed descriptions of the MCPF procedure and its development.

For analysis purposes, CUs were converted to phons based on the conversion function of Rasetshwane et al. (2015). Besides being the international standard unit for loudness level, phon has the advantage (over CUs) of being a continuous function of stimulus level, which is desirable when computing slopes (ISO 226:2003, 2003). Data for 0 and 50 CUs were not included in analysis because the levels corresponding to these loudness categories are unbounded. For example, if a listener judged 100 dB SPL as "Too Loud" (50 CU), then we would expect that listener to also judge all levels >100 dB SPL as "Too Loud."

Estimates of hearing threshold were derived from the CLS functions as the stimulus level corresponding to 2.5 CUs through simple linear regression using data for CU $\leq$ 20. This portion of the loudness function varies linearly with level, as was previously demonstrated (Al-Salim et al., 2010; Oetting et al., 2014). Because 2.5 CU is midway between 0 CU ("Can't Hear") and 5 CU ("Very Soft"), the estimate of threshold is equivalent to a condition in which the stimulus was audible 50% of the time. This definition of threshold is consistent with that used by Trevino et al. (2016a), in which threshold was defined as the inflection point between 0 and 5 CU. There were instances for the MEI-Med procedure when the CLS function did not have any data for CU $\leq$ 20. When this occurred, the lowest level that the participant responded was used as the estimate of CLS threshold. This occurred for two participants at 1 kHz and three participants at 4 kHz.

Audiometric thresholds, obtained in dB HL, were converted to dB SPL for analysis based on reference level equivalents for insert earphones (American National Standards Institute, 2010).

Reliability was assessed by comparing CLS functions between the first visit and second visit for each of the four procedures. Accuracy was assessed by comparing CLS functions for the adaptive procedures to CLS functions for the fixed-level procedure including data from both visits. The fixed-level procedure was the reference for accuracy assessment because it had a larger number of trials compared to the adaptive procedures. Both reliability and accuracy were quantified using a comprehensive set of statistical methods including (1) Bland-Altman bias, (2) Cronbach's α, and (3) root mean square error (rmse).

Bland-Altman plots show the distribution of differences between two sets of measurements. The bias represents systematic error and should be close to 0 for repeatable measurements. The plots also show 95% limits of agreement (LOA) between measurements, calculated as mean ±1.96 standard deviation (SD) when the differences are uniformly distributed and as mean ±2 SD when the differences are not uniformly distributed (Bland and Altman, 1986, 1999). The Kolmogorov-Smirnov test (Massey, 1951) indicated that differences for all conditions were normally distributed. Thus, the 95% LOA were calculated as mean ±1.96 SD. A 95% confidence interval of bias that does not include the line of equality (zero line) indicates a significant systematic error. It is worth noting that, although the Bland-Altman method is a useful tool for assessing similarities between two data sets, it does not provide criterion for acceptable bias or LOA. Interpretation of the Bland-Altman plots often requires some *a priori* information or assumptions related to the clinical or research question.

Cronbach's α is a coefficient of reliability that measures how closely a set of measurements are related (Cronbach, 1951). Values of Cronbach's α can be interpreted as follows: α: ≥ 0.9 = excellent, ≥ 0.8 = good, ≥ 0.7 = acceptable, ≥ 0.6 = questionable, $\geq$ 0.5 = poor, and <0.5 = unacceptable (George and Mallery, 2003).

Although the participants were encouraged to use all 11 response categories in their loudness judgments, some participants did not use all categories. In those cases, there were missing data for the categories that were not utilized by the listener. Most of these instances occurred for CUs of 40 and 45. Some conditions in the dataset were missing due to tester error in data collection. These included the MEI procedure at 4 kHz for one participant with normal hearing and the slope-adaptive procedure at 4 kHz for two participants with hearing loss. Additionally, one participant with hearing loss did not return for the second visit. These conditions were excluded from analysis. Overall, 3.4 and 4.6% of data were missing for the fixed-level procedure at 1 and 4 kHz, respectively; 5.9 and 8.4% of data were missing for the slope-adaptive procedure at 1 and 4 kHz, respectively; 9.8 and 11.9% of data were missing for the MEI-Med procedure at 1 and 4 kHz, respectively; and 5.6 and 1.1% of data were missing for the MEI-ML procedure.

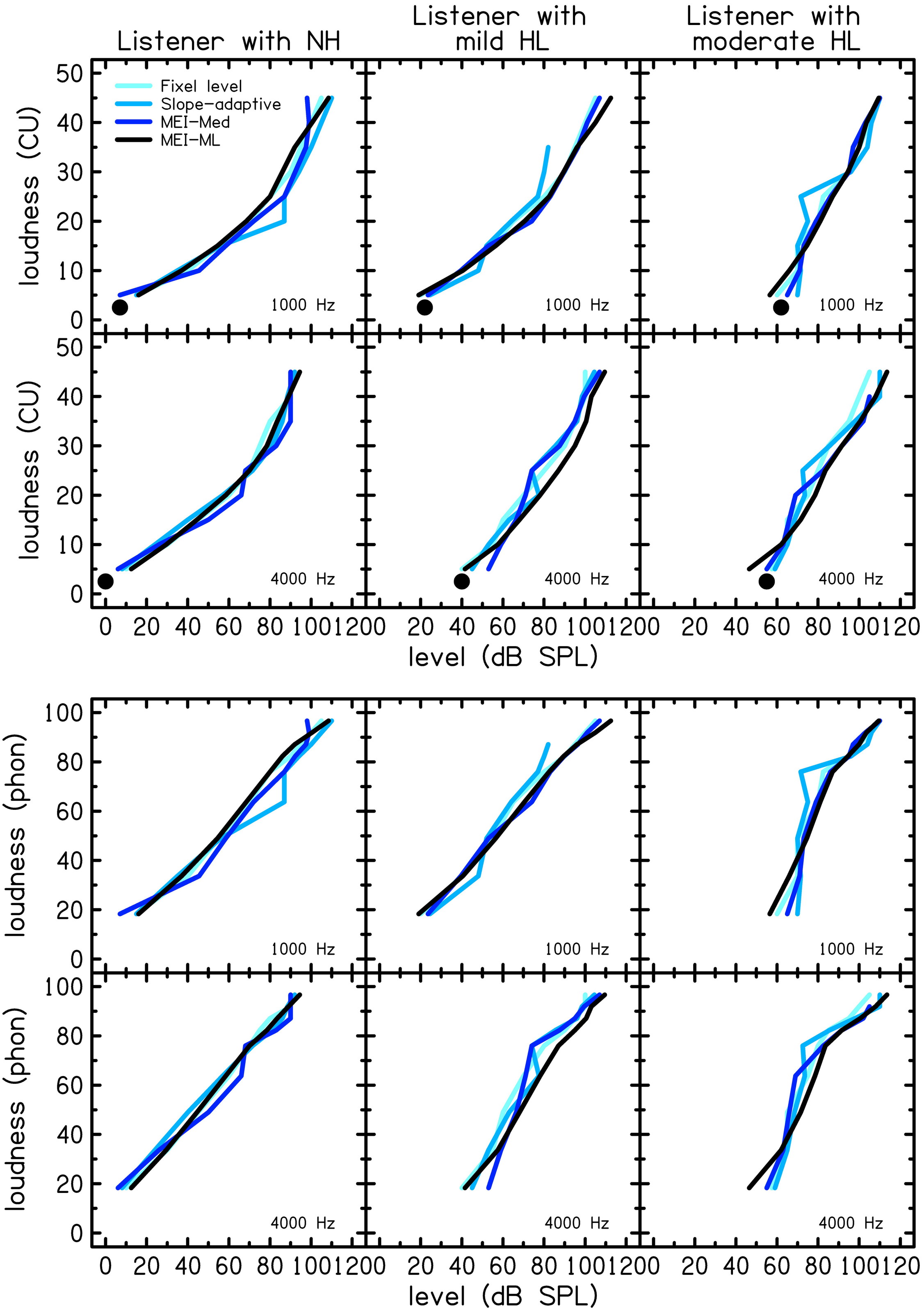

FIGURE 3 | CLS functions for each of the procedures at 1 (top row) and 4 kHz (bottom row) for three individual (representative) participants with normal hearing (NH; left column), mild hearing loss (HL, middle column), and moderate HL (right column). The top set of six panels show loudness in categorical units and the bottom set of six panels show loudness level in phons. The participants' audiometric thresholds are indicated by a black filled circle.

RESULTS

The test time for the adaptive procedures was significantly less than required for the fixed-level procedure. The mean test time for the fixed-level procedure was 15 min, 0 s (range 6 min, 3 s to 26 min, 34 s). The mean test time for the slope-adaptive procedure was reduced to 2 min 47 s (range 2 min, 0 s to 4 min, 7 s). The mean test time for the MEI procedure was reduced to 2 min, 35 s (range 1 min, 49 s to 4 min, 33 s).

Figure 3 shows CLS functions for each of the procedures at 1 kHz (top rows) and 4 kHz (bottom rows) for three individual participants: one with normal hearing (NH), one with mild sensorineural hearing loss (HL), and one with moderate sensorineural hearing loss. The functions are created from averages of measurements collected over two visits. The top set of six panels display loudness in CUs and the bottom set of six panels display loudness level in phons. The participants' audiometric thresholds are indicated by a solid circle. CLS functions are shifted to the right with increasing degrees of hearing loss. The MEI-ML method is thought likely to be more reliable than MEI-Med because its estimates are smoothed across categories.

Figure 4 shows mean CLS functions for each of the procedures at 1 (top row) and 4 kHz (bottom row) for the group of participants with normal hearing (NH; solid lines) and the group with hearing loss (HL; dashed lines). The left panels display loudness in CUs and the right panels display loudness level in phons. CLS functions were calculated from the mean of median SPL for each CU. There were similarities between the procedures. On average, participants with hearing loss have a reduced dynamic range compared to participants with normal hearing. CLS functions are shifted to the right for the group of listeners with hearing loss. The variability of the loudness function was assessed using SD, calculated separately for each CU. To avoid clutter, the SDs are presented in **Tables 1–4** instead of as error bars in **Figure 4**. Specifically, **Tables 1, 2** show SDs for participants with normal hearing at 1 and 4 kHz,

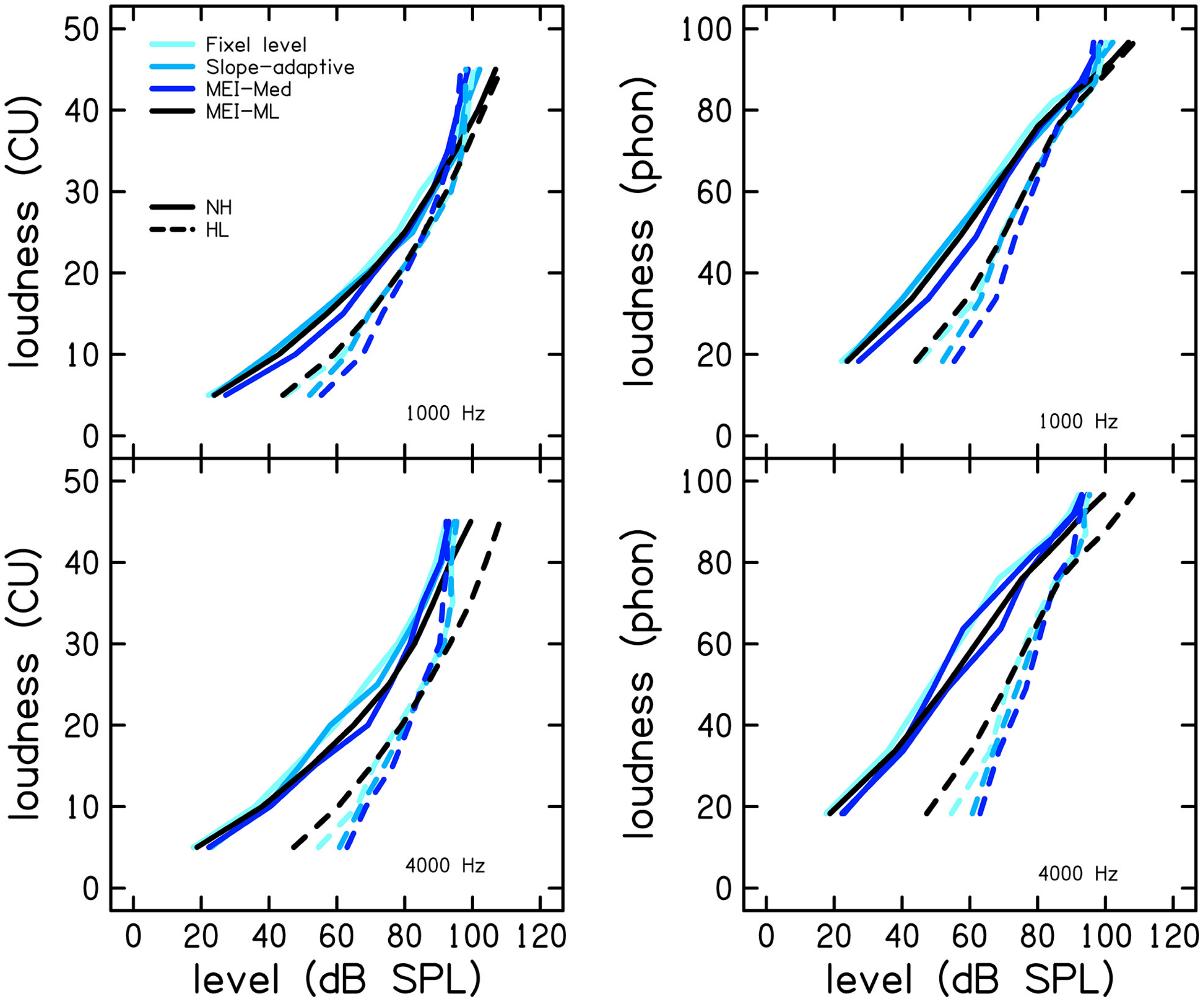

FIGURE 4 | Mean CLS functions for each of the procedures at 1 (top row) and 4 kHz (bottom row) for the group of participants with normal hearing (NH; solid lines) and the group with hearing loss (HL; dashed lines). The left column shows loudness in categorical units and the right column shows loudness level in phons. CLS functions were calculated based on the mean sound pressure level (SPL) per category (CU).

TABLE 1 | Standard deviations of sound pressure level (SPL) for each categorical unit (CU) for participants with normal hearing for each of the four CLS procedures at 1 kHz.

CU	5	10	15	20	25	30	35	40	45	Mean
Fixed-level	7.79	10.68	10.01	8.54	9.08	11.93	7.35	9.12	6.50	9.00
Slope-adaptive	11.76	11.70	11.90	14.07	10.99	12.17	9.43	8.48	7.55	10.90
MEI-Med	14.52	13.13	9.64	12.78	9.89	8.59	6.28	6.17	4.88	9.54
MEI-ML	12.27	13.42	12.44	10.74	8.87	7.51	6.91	7.12	7.43	9.63
Mean	11.59	12.24	11.00	11.53	9.71	10.05	7.49	7.72	6.59	9.77

Values are given in dB.

TABLE 2 | Standard deviations of sound pressure level (SPL) for each categorical unit (CU) for participants with normal hearing for each of the four CLS procedures at 4 kHz.

CU	5	10	15	20	25	30	35	40	45	Mean
Fixed-level	8.14	11.39	10.62	11.21	9.78	9.39	8.11	7.94	6.38	9.22
Slope-adaptive	14.08	12.06	13.29	12.27	12.37	11.36	8.73	7.58	6.07	10.87
MEI-Med	11.63	13.52	16.14	10.16	9.40	7.54	6.30	7.55	4.94	9.69
MEI-ML	8.02	12.46	13.10	11.85	10.07	8.38	7.35	6.57	6.13	9.32
Mean	10.47	12.36	13.29	11.37	10.40	9.17	7.62	7.41	5.88	9.77

Values are given in dB.

TABLE 3 | Standard deviations of sound pressure level (SPL) for each categorical unit (CU) for participants with hearing loss for each of the four CLS procedures at 1 kHz.

CU	5	10	15	20	25	30	35	40	45	Mean
Fixed-level	11.19	9.74	8.70	11.04	10.91	10.42	10.31	9.70	8.53	10.06
Slope-adaptive	13.88	10.76	10.62	11.16	10.59	10.82	10.28	9.25	8.63	10.67
MEI-Med	12.62	9.94	10.76	9.90	11.21	12.16	10.80	11.06	10.00	10.94
MEI-ML	15.15	13.35	12.41	11.24	10.15	9.43	9.70	9.70	9.07	11.13
Mean	13.21	10.95	10.62	10.84	10.72	10.71	10.27	9.93	9.06	10.70

Values are given in dB.

TABLE 4 | Standard deviations of sound pressure level (SPL) for each categorical unit (CU) for participants with hearing loss for each of the four CLS procedures at 4 kHz.

CU	5	10	15	20	25	30	35	40	45	Mean
Fixed-level	12.43	11.86	12.59	13.05	13.62	13.47	12.87	10.63	11.68	12.47
Slope-adaptive	14.46	13.80	13.16	14.56	14.03	14.00	12.90	12.40	11.96	13.47
MEI-Med	11.52	12.00	11.08	11.53	13.75	13.02	12.91	11.25	12.80	12.21
MEI-ML	16.83	15.41	15.40	15.53	15.31	14.73	14.23	13.74	12.14	14.81
Mean	13.81	13.27	13.06	13.67	14.18	13.81	13.23	12.00	12.14	13.24

Values are given in dB.

respectively, and **Tables 3**, **4** show SDs for participants with hearing loss at 1 and 4 kHz, respectively. Values are given in dB. Across procedures, SDs were higher for lower CUs compared to higher CUs. The variability was similar between participants with NH and HL at 1 kHz but was increased for participants with hearing loss at 4 kHz.

Reliability was assessed by comparing CLS functions from the first visit to those obtained on the second visit. Test–retest reliability for the fixed-level, slope-adaptive, MEI-Med and MEI-ML procedures are displayed in **Figure 5**. Panels are Bland-Altman plots for each CLS procedure. Values of Bland-Altman bias, Cronbach's α, and rmse are displayed as insets in each panel and in **Table 5**. Bland-Altman bias was $<|4|$ and values for Cronbach's α were ≥0.9 for all procedures, indicating excellent reliability. As expected, the fixed-level was the most reliable CLS procedure because it utilized a larger number of trials compared to the adaptive procedures.

Accuracy was assessed by comparing the slope-adaptive and MEI-adaptive procedures to the fixed-level procedure. Bland-Altman plots are shown in **Figure 6** for each CLS procedure. Values of Bland-Altman bias, Cronbach's α, and rmse are displayed as insets in each panel and in **Table 6**. As with the reliability analysis, values for Cronbach's α were ≥0.9 for all procedures, indicating excellent internal consistency. Bland-Altman bias was $<|3|$. The accuracy was best for the slope-adaptive procedure. The accuracy was better for MEI-ML

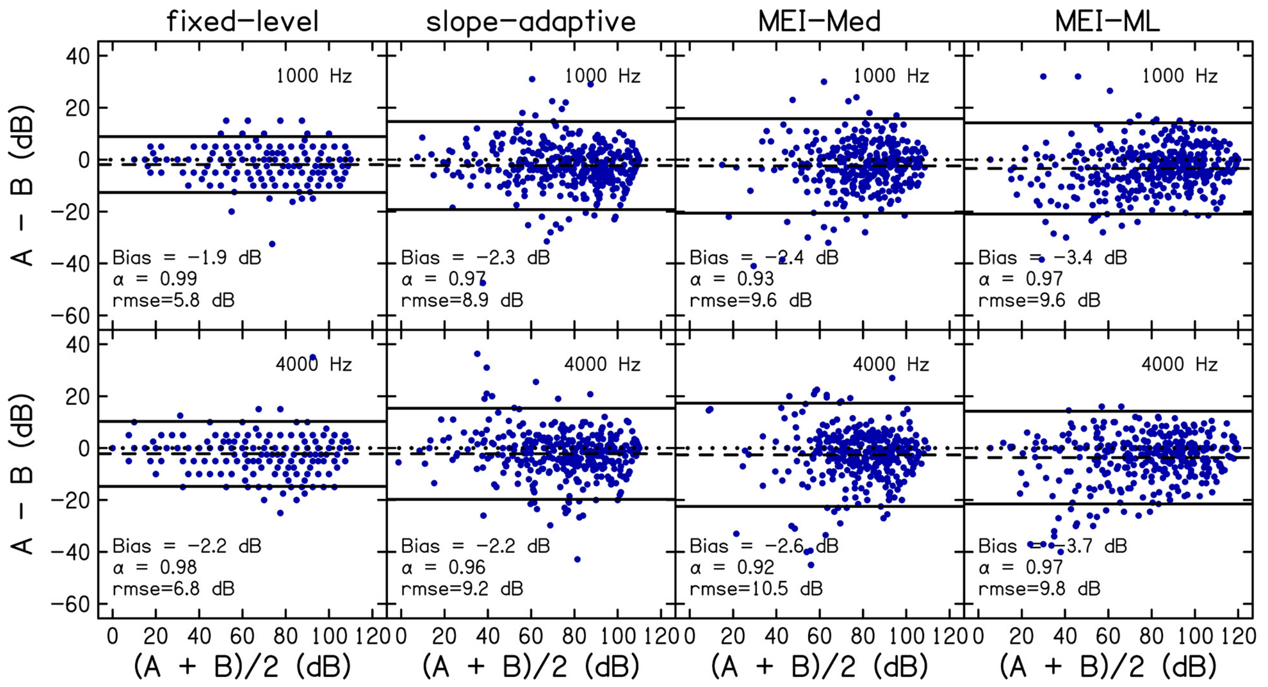

FIGURE 5 | Reliability was assessed by comparing data for visit 1 to visit 2. Panels are Bland-Altman plots for each CLS procedure at 1 (top row) and 4 kHz (bottom row). A = visit 1; B = visit2. Data points represent the difference between visits (A–B; *y*-axis) compared to the mean of both visits [(A + B)/2; *x*-axis]. The dashed line represents the bias. For references, a difference of zero is indicated using a dotted line. The solid lines represent the 95% limits of agreement. Values of Bland-Altman bias, Cronbach's α, and root-mean-square errors (rmse) are displayed as insets in each panel and in **Table 5**.

TABLE 5 | Reliability was assessed by comparing data for each of the four CLS procedures from visit one to visit two.

	1 kHz			4 kHz		
	B&A bias	**α**	**rmse**	**B&A bias**	**α**	**rmse**
Fixed-level	−1.90	0.99	5.81	−2.21	0.98	6.75
Slope-adaptive	−2.29	0.97	8.92	−2.19	0.96	9.20
MEI-Med	−2.43	0.93	9.58	−2.61	0.92	10.46
MEI-ML	−3.38	0.967	9.57	−3.66	0.97	9.81

Bland-Altman (B&A) bias, Cronbach's α, and root-mean-square errors (rmse) are reported.

than for MEI-Med, but neither is as good as the slope adaptive-procedure.

Figure 7 shows the difference between CLS estimates of threshold and audiometric thresholds for each CLS procedure. The difference in thresholds were calculated by subtracting audiometric threshold from the CLS threshold. CLS thresholds were higher than audiometric thresholds for all four procedures (difference >0 in **Figure 7**). However, error bars included zero for all four CLS procedures.

DISCUSSION

The evaluation and diagnosis of abnormalities in loudness perception in a variety of patient populations may benefit from improvements in the reliability and accuracy of CLS measurement procedures. Cochlear damage, including sensorineural hearing loss, leads to reduced dynamic range, and in some cases, hyperacusis and/or tinnitus. An attractive feature of adaptive procedures for CLS is that levels that are too uncomfortable that one would not like to listen to again are not presented, allowing for measurement of loudness in listeners who may have hyperacusis. Incorporating individual loudness measures in the hearing aid fitting may improve listener satisfaction and device acceptance. However, CLS measurements have not been accepted by clinicians, partly due to the time required to obtain them. On average, the standard fixed-level CLS procedure took approximately 15 min per frequency. The test time for the MEI-adaptive procedure was, on average, reduced to approximately 3 min per frequency, increasing the feasibility of including loudness measures in clinical practice.

Overall, reliability and accuracy were excellent at both 1 and 4 kHz (Cronbach's α > 0.9). Both accuracy and reliability were better at 1 kHz than 4 kHz (higher α and lower absolute bias). This perhaps reflects the fact that our listeners had greater hearing loss at 4 kHz than 1 kHz (see **Figure 1**).

The CLS functions plotted in **Figures 3**, **4** are averages of measurements collected over two visits. The participant with mild HL represented in **Figure 3** did not use CU 40 or 45 in the 1 kHz slope-adaptive procedure on the first visit (though they rated 110 dB SPL as "Too Loud" on visit 2), thus reducing the data for those CUs. This variability is common in human behavioral data.

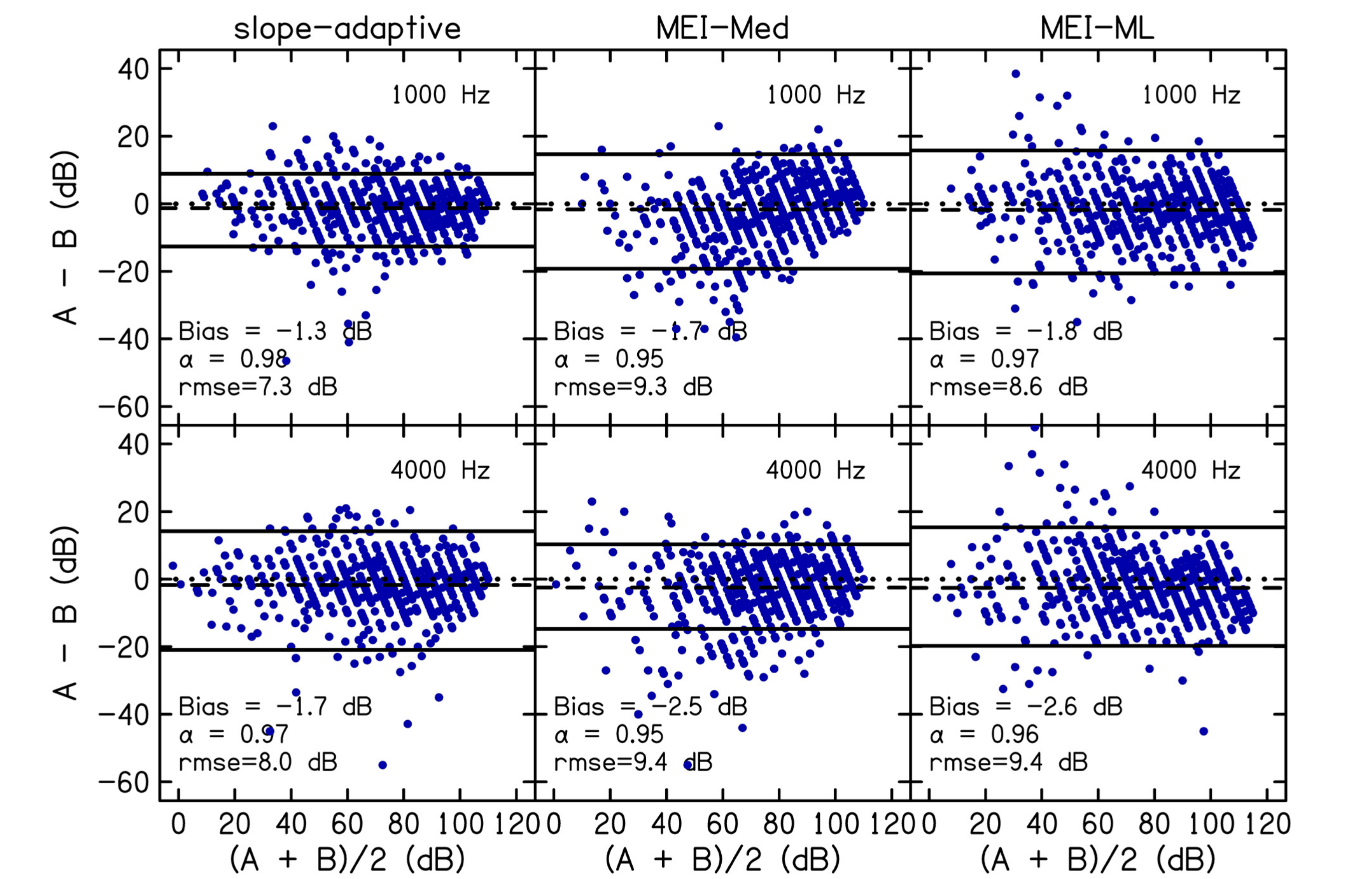

FIGURE 6 | Accuracy was assessed by comparing slope-adaptive and MEI-adaptive procedures to the fixed-level procedure. Panels are Bland-Altman plots for each CLS procedure at 1 (top row) and 4 kHz (bottom row). A = fixed-level procedure; B = adaptive procedure. Data points represent the difference between the fixed-level and adaptive procedure (*y*-axis) compared to the mean of the fixed-level and adaptive procedure (*x*-axis). The dashed line represents the bias. The solid lines represent the 95% limits of agreement. Bland-Altman bias, Cronbach's α, and root-mean-square errors (rmse) are displayed as insets in each panel and in **Table 6**.

TABLE 6 | Accuracy was assessed by comparing the three adaptive procedures to the Fixed-Level procedure across both visits.

	1 kHz			4 kHz		
	B&A bias	**α**	**rmse**	**B&A bias**	**α**	**rmse**
Slope-adaptive	−1.34	0.98	7.28	−1.73	0.97	8.02
MEI-Med	−1.70	0.95	9.31	−2.49	0.95	9.44
MEI-ML	−1.83	0.97	8.64	−2.57	0.96	9.44

Bland-Altman (B&A) bias, Cronbach's α, and root-mean-square errors (rmse) are reported.

Overall, the trends in the group data (**Figure 4**) were consistent with those of the individual data (**Figure 3**).

Across procedures, SDs were higher for lower CUs compared to higher CUs, similar to trends noted in Trevino et al. (2016a). The variability was similar between participants with NH and HL at 1 kHz but was increased for participants with hearing loss at 4 kHz. This contrasts with previous studies that observed higher variability for participants with NH than for participants with hearing loss (Brand and Hohmann, 2002; Rasetshwane et al., 2015). Larger variability of CLS functions for participants with NH compared to participants with sensorineural HL is expected because participants with NH have a wider dynamic range and thus a wider range of possible SPLs that they can assign to a particular loudness category. The observed discrepancy remains unexplained.

In the Bland-Altman analysis of reliability and accuracy (**Figures 5**, **6**), the distribution of differences between two sets of measurements (A−B) is plotted against the mean of the measurements [(A + B)/2]. Measurement bias, which represents systematic error, is calculated as the mean of the differences, and should be close to zero for repeatable measurements. Whether the bias is negative or positive is not important. A bias <0 simply means that measurement B is larger in magnitude/amplitude compared to measurement A.

Interpretation of the LOA for the Bland-Altman plot requires prior information regarding what is considered a significant change in the measurement being analyzed. As an example, hearing conservation programs consider a change in audiometric threshold of 10 dB as a significant threshold shift. Thus, a Bland-Altman analysis for accuracy or repeatability of

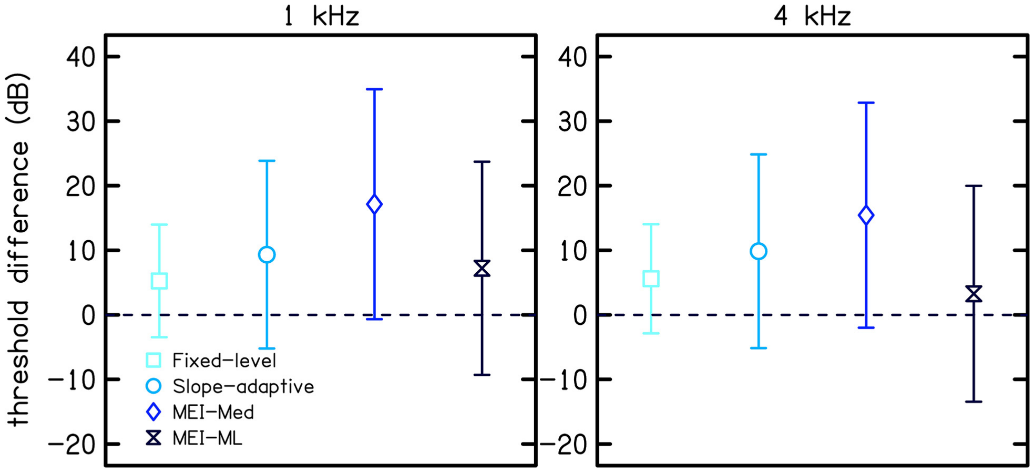

FIGURE 7 | Threshold differences between CLS thresholds and audiometric thresholds for each of the four CLS procedures. The difference in threshold was calculated by subtracting the audiometric threshold from the CLS threshold. Symbols represent the mean difference and error bars represent one standard deviation.

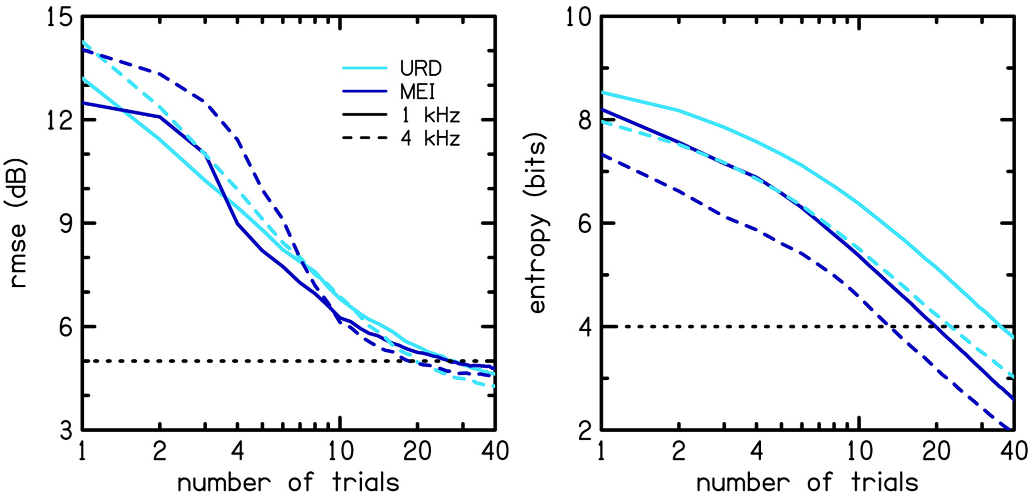

FIGURE 8 | Model simulation from a catalog created in previous work (Trevino et al., 2016a). Root-mean-square error (rmse; left panel) and entropy (right panel) for the MEI algorithm (dark blue lines) are compared to those for the Uniform Random Distribution (URD; light blue lines) procedure. The solid lines represent 1 kHz and the dashed lines represent 4 kHz.

audiometric threshold can utilize 10 dB to interpret the LOA. Unfortunately, prior work has not defined a significant change in CLS data that can be applied to interpret Bland-Altman analyses. Therefore, the analyses were complemented with Cronbach's alpha. An attractive feature of Cronbach's alpha is that there are published guidelines for interpreting the outcome, and the interpretation is not dependent on the type of measurement.

Figure 7 compares CLS estimates of hearing thresholds (i.e., the stimulus level corresponding to 2.5 CU obtained by extrapolation using linear regression) to audiometric thresholds for each CLS procedure. In particular, **Figure 7** shows the mean difference between CLS and audiometric thresholds across participants. Threshold differences were greater than zero for all four CLS procedures, indicating that CLS estimates of thresholds were higher than audiometric thresholds. Of the adaptive procedures, the MEI-ML method resulted in the best estimate of thresholds (difference = 7 and 3 dB at 1 and 4 kHz, respectively). This is likely due to the smoothing across loudness categories that was done for this procedure. Our observation of higher thresholds for CLS compared to audiometric testing is in contrast with that of Trevino et al. who reported that, on average, CLS thresholds were lower than audiometric thresholds, with differences up to 20 dB. The discrepancy between the two studies may be due to the differences in the study populations.

Consistent with Moore (2004), our procedure for estimating CLS threshold is not thought to be related to the concept of softness imperception (abnormally large loudness at absolute threshold; Buus and Florentine, 2002). Unlike other procedures for measuring loudness, CLS relates more to a listener's experience and informal descriptions of their loudness percepts

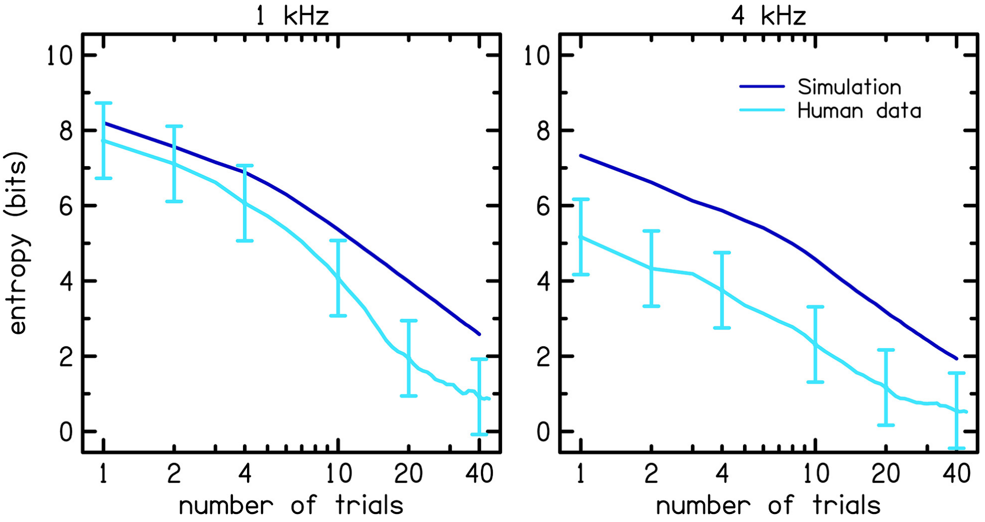

FIGURE 9 | Comparison of the entropy between the simulation (dark blue) and human data (light blue). Error bars for the human data represent one standard deviation from the mean.

and includes loudness descriptors such as "Very Soft" and "Soft." Thus, because listeners did perceive "Soft" sounds in CLS, the concept of softness imperception is not applicable to CLS.

In order to better understand the poorer performance of the MEI procedure, model simulations were conducted by using the same empirical distributions that were used to construct the MCPF catalog. Hundreds of simulated listeners were selected from these empirical distributions and simulated responses in each simulated trial were generated according to the expected performance of the simulated listener. Estimation of simulated CLS functions has multiple advantages. The most common use for probabilistic listener models is to support the development of experiments or listening devices. They also allow for the application of concepts from detection, information, and estimation theory to the analysis of results and methodology of the experiment (Trevino et al., 2016b). The simulations were implemented using Monte Carlo methods, and therefore accounted for the randomness of individual listeners.

Figure 8 shows rmse (left panel) and entropy (right panel) for the MEI algorithm (dark blue lines) compared to those for the Uniform Random Distribution (URD; light blue lines) procedure. URD was used to simulate the adaptive ISO procedure by randomly selecting the next stimulus level from a uniform distribution, or range of possible levels that each have equal probability. The solid lines represent 1 kHz and the dashed lines represent 4 kHz. The fact that MEI consistently outperforms URD in terms of entropy reduction (right panel) tells us that the tracking implementation is performing as well as expected. However, the fact that MEI is not consistently better than URD in terms of rmse reduction tells us that the MCPF catalog lacks sufficient information to improve the accuracy of the adaptive tracking procedure. Comparison of the simulated entropy reduction with the human data in **Figure 9** further validates the MEI implementation by showing greater entropy reduction in the human listeners (light blue) compared to the simulation (dark blue) for 1 (left panel) and 4 kHz (right panel). Error bars for the human data represent one SD from the mean. Entropy is lower for the human data compared to the simulation. This result rules out the possibility that the observed poorer performance was due to flawed implementation of the MEI tracking methods, which implicates intrinsic uncertainty in the MCPF catalog as the factor that currently limits MEI performance.

In summary, the MEI tracking apparently produced less accurate CLS functions compared to the other tracking methods because of inherent uncertainty in the MCPF which reflects the uncertainty of the listeners on whom the MCPF catalog was based, and not because the MEI procedure was improperly implemented or lacked the ability to reduce catalog entropy. The results of this study indicate that our measure of entropy was not sufficiently correlated to rmse to produce more reliable CLS functions.

Further investigation is warranted to understand how the MCPF catalog could be modified to achieve closer correspondence with catalog entropy. The MCPF catalog would be improved by reconstructing it from new fixed-level CLS data at a larger number of stimulus frequencies and a more uniform representation of hearing-loss categories. However, such an improved MCPF catalog would not necessarily improve MEI efficiency. Relaxing the restriction on large level transitions (45 dB for the current study), which can bias listener responses, or by including catch trials where listener biases are expected, may improve performance of the MEI-adaptive method. During a CLS test, large transitions in SPL as well as presentations of multiple consecutive trials at similar SPLs are avoided as these can bias listener responses. For example, if a presentation of 10 dB SPL is followed by a presentation of 80 dB SPL, listeners will perceive the 80 dB SPL signal as louder than if it followed a 50 dB SPL signal. Thus, changes were made to our MEI-adaptive approach to accommodate listener effects that can bias CLS data. Unfortunately, these changes, although necessary for practical purposes, resulted in a suboptimal MEI-adaptive procedure. Thus, there is potential to improve the performance of the MEI-adaptive procedure. Such modifications could improve both MEI tracking efficiency and ML estimation of CLS functions. Further improvements in the reliability and accuracy of CLS could enhance the clinical acceptability of loudness measurements and potentially improve hearing aid fitting methods.

DATA AVAILABILITY STATEMENT

All datasets and analysis code generated for this study are located online at OSF (Open Science Framework) https://osf.io/xwb6p/.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board, Boys Town National Research Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SF collected the data and wrote the manuscript. JK contributed to experiment design. SN and DR designed the study, reviewed the data, and provided interpretive analysis and critical revisions. All authors discussed the results and implications and commented on the manuscript at all stages.

FUNDING

This research was supported by grants from the National Institutes of Health: NIGMS Award Number 1U54GM115458, NIGMS Award Number P20GM109023, and NIDCD Award Number 5R01DC016348. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

ACKNOWLEDGMENTS

The authors thank Tessa Culbertson for assistance with data collection.

REFERENCES

Allen, J. B. (2008). "Nonlinear Cochlear Signal Processing and Masking in Speech Perception," in *Springer Handbook of Speech Processing. Springer Handbooks*, eds J. Benesty, M. M. Sondhi, and Y. A. Huang (Berlin: Springer), 27–60. doi: 10.1007/978-3-540-49127-9_3

Al-Salim, S. C., Kopun, J. G., Neely, S. T., Jesteadt, W., Stiegemann, B., and Gorga, M. P. (2010). Reliability of categorical loudness scaling and its relation to threshold. *Ear Hear.* 31, 567–578. doi: 10.1097/AUD.0b013e3181da4d15

American National Standards Institute (2010). *Specifications for audiometers (ANSI S3. 6-2010)*. New York, NY: American National Standards Institute.

American Speech-Language-Hearing Association [ASHA] (1978). Guidelines for manual pure-tone threshold audiometry. *ASHA* 20, 297–301.

Assi, H., Moore, R. D., Ellemberg, D., and Hébert, S. (2018). Sensitivity to sounds in sport-related concussed athletes: a new clinical presentation of hyperacusis. *Sci. Rep.* 8, 1–8. doi: 10.1038/s41598-018-28312-1

Blamey, P. J., and Martin, L. F. (2009). Loudness and satisfaction ratings for hearing aid users. *J. Am. Acad. Audiol.* 20, 272–282. doi: 10.3766/jaaa.20.4.7

Bland, J. M., and Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1, 307–310. doi: 10.1016/s0140-6736(86)90837-8

Bland, J. M., and Altman, D. G. (1999). Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* 8, 135–160. doi: 10.1191/096228099673819272

Brand, T., and Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *J. Acoust. Soc. Am.* 112, 1597–1604. doi: 10.1121/1.1502902

Buus, S., and Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: recruitment reconsidered. *J. Assoc. Res. Otolaryngol.* 3, 120–139. doi: 10.1007/s101620010084

Cox, R. M. (1989). Comfortable loudness level: stimulus effects, long-term reliability, and predictability. *J. Speech Hear. Res.* 32, 816–828. doi: 10.1044/jshr.3204.816

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334. doi: 10.1007/bf02310555

Elberling, C. (1999). Loudness scaling revisited. *J. Am. Acad. Audiol.* 10, 248–260.

Fletcher, H., and Munson, W. (1933). Loudness: its definition, measurement, and calculations. *J. Acoust. Soc. Am.* 5, 82–108. doi: 10.1121/1.1915637

George, D., and Mallery, P. (2003). *SPSS for Windows Step by Step: A Simple Guide and Reference. 11.0 Update*, 4th Edn. Boston, MA: Allyn & Bacon.

Hébert, S., Fournier, P., and Noreña, A. (2013). The auditory sensitivity is increased in tinnitus ears. *J. Neurosci.* 33, 2356–2364. doi: 10.1523/jneurosci.3461-12.2013

Heeren, W., Hohmann, V., Appell, J. E., and Verhey, J. L. (2013). Relation between loudness in categorical units and loudness in phons and sones. *J. Acoust. Soc. Am.* 133, EL314–EL319. doi: 10.1121/1.4795217

ISO 226:2003 (2003). *Acoustics-Normal Equal-Loudness-Level Contours*. Geneva: International Organization for Standardization.

ISO 16832:2006 (2006). *Acoustics-Loudness Scaling by Means of Categories*. Geneva: International Organization for Standardization.

Khalfa, S., Bruneau, N., Rogé, B., Georgieff, N., Veuillet, E., Adrien, J.-L., et al. (2004). Increased perception of loudness in autism. *Hear. Res.* 198, 87–92. doi: 10.1016/j.heares.2004.07.006

Kinkel, M. (2007). "The new ISO 16832 "Acoustics–loudness scaling by means of categories," in *Proceedings of the 8th EFAS Congress/10th Congress of the German Society of Audiology*, (Heidelberg).

Kollmeier, B., and Hohmann, V. (1995). "Loudness estimation and compensation employing a categorical scale," in *Advances in Hearing Research*, eds G. Manley, G. Klump, C. Köppl, H. Fasti, and H. Oeckinghaus (Singapore: World Scientific), 441–451.

Massey, F. J. Jr. (1951). The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Statist. Assoc.* 46, 68–78. doi: 10.1080/01621459.1951.10500769

Moore, B. C. (2004). Testing the concept of softness imperception: loudness near threshold for hearing-impaired ears. *J. Acoust. Soc. Am.* 115, 3103–3111. doi: 10.1121/1.1738839

Noreña, A. J., and Chery-Croze, S. (2007). Enriched acoustic environment rescales auditory sensitivity. *Neuroreport* 18, 1251–1255. doi: 10.1097/wnr.0b013e3282202c35

Oetting, D., Brand, T., and Ewert, S. D. (2014). Optimized loudness-function estimation for categorical loudness scaling data. *Hear. Res.* 316, 16–27. doi: 10.1016/j.heares.2014.07.003

Oetting, D., Hohmann, V., Appell, J. E., Kollmeier, B., and Ewert, S. D. (2016). Spectral and binaural loudness summation for hearing-impaired listeners. *Hear. Res.* 335, 179–192. doi: 10.1016/j.heares.2016.03.010

Rasetshwane, D. M., High, R. R., Kopun, J. G., Neely, S. T., Gorga, M. P., and Jesteadt, W. (2018). Influence of suppression on restoration of spectral loudness summation in listeners with hearing loss. *J. Acoust. Soc. Am.* 143:2994. doi: 10.1121/1.5038274

Rasetshwane, D. M., Trevino, A. C., Gombert, J. N., Liebig-Trehearn, L., Kopun, J. G., Jesteadt, W., et al. (2015). Categorical loudness scaling and equal-loudness contours in listeners with normal hearing and hearing loss. *J. Acoust. Soc. Am.* 137, 1899–1913. doi: 10.1121/1.4916605

Torgerson, W. (1958). *Theory and Methods of Scaling*. New York,NY: Wiley, 460.

Trevino, A. C., Jesteadt, W., and Neely, S. T. (2016a). Development of a multi-category psychometric function to model categorical loudness measurements. *J. Acoust. Soc. Am.* 140:2571. doi: 10.1121/1.4964106

Trevino, A. C., Jesteadt, W., and Neely, S. T. (2016b). Modeling the individual variability of loudness perception with a multi-category psychometric function. *Adv. Exp. Med. Biol.* 894, 155–164. doi: 10.1007/978-3-319-25474-6_17

Wròblewski, M., Rasetshwane, D. M., Neely, S. T., and Jesteadt, W. (2017). Deriving loudness growth functions from categorical loudness scaling data. *J. Acoust. Soc. Am.* 142, 3660–3669. doi: 10.1121/1.5017618

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.

frontiers in Psychology

REVIEW
published: 13 January 2021
doi: 10.3389/fpsyg.2020.611517

Applications of Phenomenological Loudness Models to Cochlear Implants

Colette M. McKay[1,2]*

[1] Bionics Institute, Melbourne, VIC, Australia, [2] Department of Medical Bionics, University of Melbourne, Melbourne, VIC, Australia

Cochlear implants electrically stimulate surviving auditory neurons in the cochlea to provide severely or profoundly deaf people with access to hearing. Signal processing strategies derive frequency-specific information from the acoustic signal and code amplitude changes in frequency bands onto amplitude changes of current pulses emitted by the tonotopically arranged intracochlear electrodes. This article first describes how parameters of the electrical stimulation influence the loudness evoked and then summarizes two different phenomenological models developed by McKay and colleagues that have been used to explain psychophysical effects of stimulus parameters on loudness, detection, and modulation detection. The Temporal Model is applied to single-electrode stimuli and integrates cochlear neural excitation using a central temporal integration window analogous to that used in models of normal hearing. Perceptual decisions are made using decision criteria applied to the output of the integrator. By fitting the model parameters to a variety of psychophysical data, inferences can be made about how electrical stimulus parameters influence neural excitation in the cochlea. The Detailed Model is applied to multi-electrode stimuli, and includes effects of electrode interaction at a cochlear level and a transform between integrated excitation and specific loudness. The Practical Method of loudness estimation is a simplification of the Detailed Model and can be used to estimate the relative loudness of any multi-electrode pulsatile stimuli without the need to model excitation at the cochlear level. Clinical applications of these models to novel sound processing strategies are described.

Keywords: Cochlear implants, loudness, intensity, temporal resolution, models

OPEN ACCESS

Edited by:
Brian Cecil Joseph Moore,
University of Cambridge,
United Kingdom

Reviewed by:
Fan-Gang Zeng,
University of California, Irvine,
United States
Wiebke Lamping,
University of Cambridge,
United Kingdom

****Correspondence:***
Colette M. McKay
cmckay@bionicsinstitute.org

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

Received: *29 September 2020*
Accepted: *11 December 2020*
Published: *13 January 2021*

Citation:
McKay CM (2021) Applications
of Phenomenological Loudness
Models to Cochlear Implants.
Front. Psychol. 11:611517.
doi: 10.3389/fpsyg.2020.611517

INTRODUCTION

Cochlear implants (CIs) have been one of the most successful medical devices developed over the last 40 years, now approaching a million users worldwide. CIs restore hearing sensation to severely or profoundly deaf people by electrically stimulating residual hearing nerves in the cochlea. Although there are many variations of signal processing strategies, which encode features of sounds into patterns of electrical stimulation, all are based upon a simple principle: amplitude variations in different acoustic frequency bands are encoded as current amplitude variations of electrical pulse trains (or rarely sinusoids) on tonotopically assigned intracochlear electrodes. Thus, in addition to the tonotopic assignment of frequency bands to intra-cochlear electrode position, intensity coding is the main means of transferring acoustic stimulus feature information to the electrical stimulus and hence to the perception of the CI user. This article summarizes features of intensity and loudness coding in CIs and places this knowledge in the context of two phenomenological loudness models developed and validated by McKay and collaborators. These models throw light on how the perception of loudness and temporal information are modulated by parameters of electrical stimulation and how the neural processing of sounds differs from that for acoustic stimulation. It should be noted that the psychophysical perception of loudness can vary with the context in which a sound is heard (Schneider and Parker, 1990; Wang and Oxenham, 2016) and with slow acting changes in central gain (Pieper et al., 2018; Auerbach et al., 2019). However, this review focuses on the influence of electrical stimulus parameters on perceived loudness and on the transmission of temporal features in sounds.

SINGLE-ELECTRODE STIMULI

Loudness of Simple Single-Electrode Stimuli

The electrical stimuli in the majority of commercial CI systems are composed of cathodic-first biphasic pulse trains. The biphasic pulses are defined by pulse duration (PD), current amplitude (i), interphase gap (IPG) (**Figure 1**), and the mode of stimulation. The mode defines the current return path from the activated intracochlear electrode: monopolar (MP) mode (the most common) uses a return electrode, or electrodes, situated outside the cochlea; bipolar (BP) mode uses a nearby intracochlear electrode; and multipolar modes use a combination of return-path and/or active electrodes. The mode of stimulation controls the spatial specificity of the current path. To complete the description of a pulse train on a single active electrode, the interpulse intervals (IPIs) are required. All of these five parameters (i, PD, IPG, mode, and IPI) influence the loudness evoked by the stimulus. Although commercial systems generally use cathodic-first biphasic pulses in MP or BP modes, researchers have evaluated the effect on neural excitation of alternative pulse shapes and multipolar modes (e.g., Bonnet et al., 2004; Macherey et al., 2010; Srinivasan et al., 2010; Undurraga et al., 2012; Fielden et al., 2013; Marozeau et al., 2015; Carlyon et al., 2017, 2018). Different pulse shapes and multipolar modes influence both the amount of excitation induced by a current pulse and the spatial specificity of the neural activation. In general, multipolar modes can improve the spatial specificity of activated neural populations, but at the expense of higher currents being required to achieve the same loudness (Srinivasan et al., 2010; Fielden et al., 2013; Marozeau et al., 2015). Anodic-first biphasic pulses, triphasic pulses, and pseudo-monophasic pulses have all been compared to biphasic pulses in studies that have shown that different pulse shapes can affect place specificity, the location of the peak excitation, and loudness (Macherey et al., 2010, 2011; Undurraga et al., 2012; Carlyon et al., 2017). However, these alternative pulse shapes and modes are not yet used in commercial systems, and this review will mostly not consider their effects in detail, except where specified.

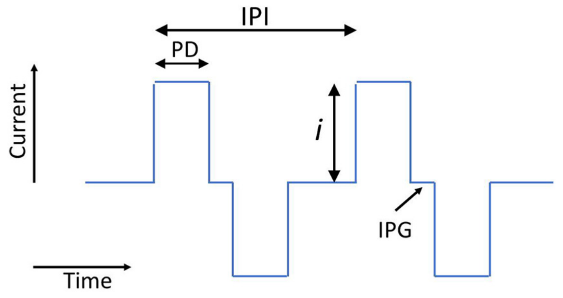

FIGURE 1 | Schematic showing two biphasic current pulses and the parameters current (i), pulse duration (PD), interphase gap (IPG), and interpulse interval (IPI).

In general sound processor usage, with few exceptions, the value of the current amplitude (i) is used to control the loudness evoked by the stimulus and to convey amplitude modulations of temporal envelopes within each frequency band, while other stimulus parameters are fixed (Wouters et al., 2015). Over the relatively small current range between hearing threshold and maximum loudness for a simple pulse train on a single electrode, the relation between current and loudness can be well described by either a power or exponential function (Kwon and van den Honert, 2006). However, as described and explained in more detail in section "Multi-Electrode Stimuli," the relation is more complex over the wider range of current amplitudes that can be used in complex multi-electrode or high-rate stimuli, with a power function describing the relation for low currents and an expansive function needed at high levels (McKay et al., 2003).

Since electrical charge is the means by which neurons are activated, it could be expected that changes in PD would have the same effect on loudness as changes in current (since both are linearly related to the total charge delivered). However, longer pulses are less effective at activating neurons than shorter pulses of equal total charge (Pfingst et al., 1991; Moon et al., 1993). This reduction in efficiency is well explained by the neural "leaky integrator" model (Miller et al., 2001). The ability with which neurons integrate charge on their membranes depends on the site of activation (dendrite, cell body or axon) and physical attributes of the neurons such as size and health, for example presence or absence of myelin (Parkins and Colombo, 1987; Horne et al., 2016). These neural properties lead to the amount of PD change versus current change for equivalent loudness change being different at different absolute current amplitudes and PDs (McKay and McDermott, 1999; Carlyon et al., 2005), and between different people and different electrode positions in the same person (Schvartz-Leyzac and Pfingst, 2016). The dependence of the effect of changing PD on neural health status has led to several proposals to use this effect in psychophysical or electrophysiological measures to evaluate neural health in individual CI users (Moon et al., 1993; McKay and McDermott, 1998; Prado-Guitierrez et al., 2006; Ramekers et al., 2014). In a similar way, an increase of the IPG between the two phases of

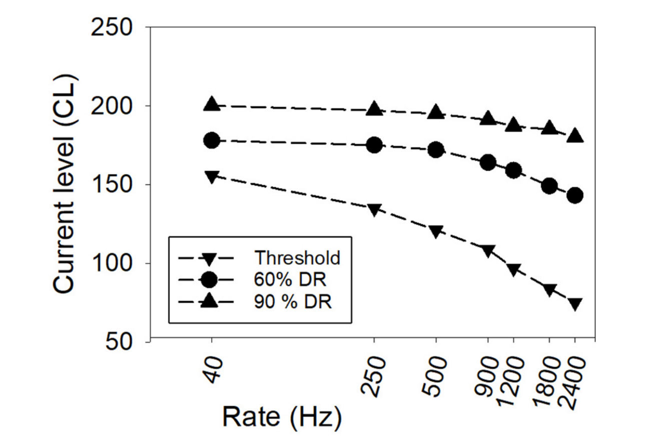

FIGURE 2 | An example from one CI user showing threshold and equal-loudness functions versus rate of stimulation. Currents are depicted in clinical current level (CL) units, where one CL is 0.176 dB (Data from McKay et al., 2013a).

the biphasic pulse leads to more effective activation of neurons (McKay and Henshall, 2003; Carlyon et al., 2005), possibly because the second phase can remove charge from the neuron before it fires. The influence of the IPG has also been shown in animal studies to be correlated with neural health (Prado-Guitierrez et al., 2006; Ramekers et al., 2014; Schvartz-Leyzac and Pfingst, 2016; Hughes et al., 2018), and the effect has been proposed as a measure of neural health in humans, in a similar way to the PD effect (Hughes et al., 2018; He et al., 2020; Schvartz-Leyzac et al., 2020).

The rate of stimulation (controlled by the IPI) also affects the loudness evoked by a stimulus, with loudness increasing with increasing rate (Shannon, 1985). **Figure 2** shows representative data for one CI user, illustrating how hearing threshold and equally loud currents typically change with rate of stimulation for biphasic pulse trains. Given that the phase duration and IPG are generally fixed for individuals in clinical use, the loudness of stimuli depends on the currents used, the time intervals between pulses, and the duration of the pulse train. The response state of auditory neurons (changing the probability of firing, and altering the total excitation elicited by an individual electrical pulse) depends on what has already occurred in the time leading up to a particular electrical pulse, with refractoriness reducing firing probability for neurons that have recently fired, facilitation increasing firing probability for very short IPIs, and adaptation lowering firing probability over sustained durations of stimulation (Tang et al., 2006; Boulet et al., 2016).

The Temporal Model

A phenomenological model was developed by McKay and McDermott (1998) to explain the effect on loudness of IPIs in 2-pulse-per-period stimuli, and was later generalized by McKay et al. (2013b) to model the effects of rate of stimulation and stimulus duration on loudness or hearing threshold, effects of modulation frequency on modulation detection, and effects of masker stimulus features on forward masked thresholds. This model, designated here as the Temporal Model, describes how temporal factors in single-electrode stimuli influence psychophysical data. The model was based on similar acoustic models (Oxenham and Moore, 1994, 1995; Moore et al., 1996; Oxenham, 2001; Plack et al., 2002) in which the cochlear excitation evoked by a stimulus is integrated by a sliding temporal integration window and perceptual decisions (e.g., equal loudness, discrimination, and detection) are made by applying criteria to the output of the integrator. These authors showed that, if the integration occurred after the non-linear cochlear processes (instead of on the acoustic waveform), the integration window is invariant with acoustic level and frequency. Plack et al. (2002) argued that the linear integration window should act upon the intensity of basilar membrane vibration, which in turn may be linearly related to auditory nerve firing rate (Muller et al., 1991). Therefore, in the development of the Temporal Model applied to electric stimulation, the same central temporal integration window was applied to peripheral neural activity evoked by electrical current pulses, on the assumption that processing in the central auditory system is largely unaffected by peripheral hearing loss. Similar central decision criteria to those used in acoustic hearing could then be applied to the integrator output.

The integration window used in the Temporal Model has the following form:

$$W(t) = (1 - w) \times \exp(t/T_{b1}) + w \times exp(t/T_{b2}), \quad t < 0$$
$$W(t) = \exp(-t/T_a), \quad t \geq 0 \tag{1}$$

where T_a and T_{b1} together define the short time constant associated with temporal resolution, T_{b2} defines a longer tail of the window associated with forward masking and the effect of stimulus duration, and w is the weighting of the long versus short time constants. For example, Oxenham (2001) derived the integration window shape to best fit forward masking data for normally hearing listeners: the best fitting values of the parameters were T_a = 3.5 ms, T_{b1} = 4.6 ms, T_{b2} = 16.6 ms, and w = 0.17.

To predict the effect of a stimulus parameter on detection, loudness, or discrimination using the Temporal Model the following four steps are used:

1. Using a reference stimulus, calculate the excitation evoked by each pulse relative to the first pulse. In practice this step involves modeling the peripheral effects of refractoriness, facilitation, adaptation, or amplitude modulation to describe how neural excitation changes with each pulse.
2. Integrate the excitation with the sliding temporal integration window in Eq. 1, the output of which is a function of integrated excitation versus time.
3. Apply the desired decision criterion to the integrator output. Such criteria will depend on the experiment being undertaken.
4. Repeat with different values of the stimulus parameter under investigation to achieve the aimed-for criterion at the integrator output. The adjustment of the input

stimulus current, when required to achieve the criterion value, requires the application of a scaling factor, *S*, to transform changes of input current in dB to changes of excitation in dB.

Given psychophysical data showing the effects of the stimulus parameter under investigation, the Temporal Model can be used to infer the physiological effects of the parameter on neural excitation in step 1, and the scaling factor in step 4, that are needed to fit the predictions to the actual data. Thus, the Temporal Model potentially provides insights into how individual peripheral neural factors can influence temporal effects on loudness. Some examples of this process are described below.

McKay and McDermott (1998) applied the Temporal Model in three experiments that investigated the effect of IPI on detection and loudness. In these experiments, IPI was varied and equal-loudness or threshold functions were measured by adjusting the stimulus current. In experiments 1 and 2, a second pulse was inserted into each period of a 50 or 250 Hz pulse train, respectively, with a varying IPI between the two pulses in each stimulus period, and in experiment 3, constant-rate stimuli were varied in rate. **Figure 3** shows representative results of experiment 1 for two CI users, illustrating both the non-monotonic effect of IPI on loudness and inter-listener differences. The non-monotonic effect of IPI on loudness is a result of the counteracting influences of refractoriness on the second of each pulse pair and the shape of the integration window. A smaller IPI reduces the excitation evoked by the second pulse, but also increases the weighting of the second pulse in the integration window. The Temporal Model was used to fit the predicted effect of IPIs for each individual in experiment 1 to the measured data by modeling the relative excitation evoked by the second pulse of each pulse pair compared to that evoked by the first (step 1 of the model). It was found that the differences in the shapes of the functions of current adjustment for equal loudness in experiment 1 (as seen in **Figure 3**) could be successfully modeled by fitting parameters relating to peripheral neural factors in step 1 (the average refractory recovery time, and the proportion of available neurons that fired on the first pulse), with the scaling factor in step 4 adjusting the vertical scale of the functions. The central decision criterion applied in step 3 for equal loudness or threshold was equal maximum output of the integrator. The fitted scaling factor, *S*, in step 4 ranged between 1 and 6 and was significantly larger at higher current levels. Individual scaling factors from experiment 1 were successfully re-used for application of the model to the data for experiments 2 and 3. On average across CI users, the values of the predicted individual neural factors were consistent with a large proportion of neurons being activated close to their individual thresholds for the current ranges used – with low spike probabilities (around 0.7) and long mean relative refractory times (average 5.5 ms). The variation of these factors between subjects can be hypothesized to be associated with neural survival density and the health of the surviving neurons.

In McKay et al. (2013b), the Temporal Model was further successfully applied to psychophysical data from CI users to understand the effects of modulation frequency on modulation detection (i.e., temporal resolution), the effect of stimulus duration on loudness, and the influence of masker-probe time interval on probe threshold in forward masking experiments. The decision criterion applied for the effect of modulation frequency on modulation detection was a fixed modulation depth of the integrator output for different modulation frequencies. For the effect of duration on loudness, the decision criterion was that the maximum integrator output for different durations was equal to that for the first pulse on its own. For the effect of masker-probe time interval on forward masked probe thresholds, the criterion was a fixed maximum difference between integrator outputs with and without the probe stimulus (which occurred near the probe offset). It is notable that all of the data

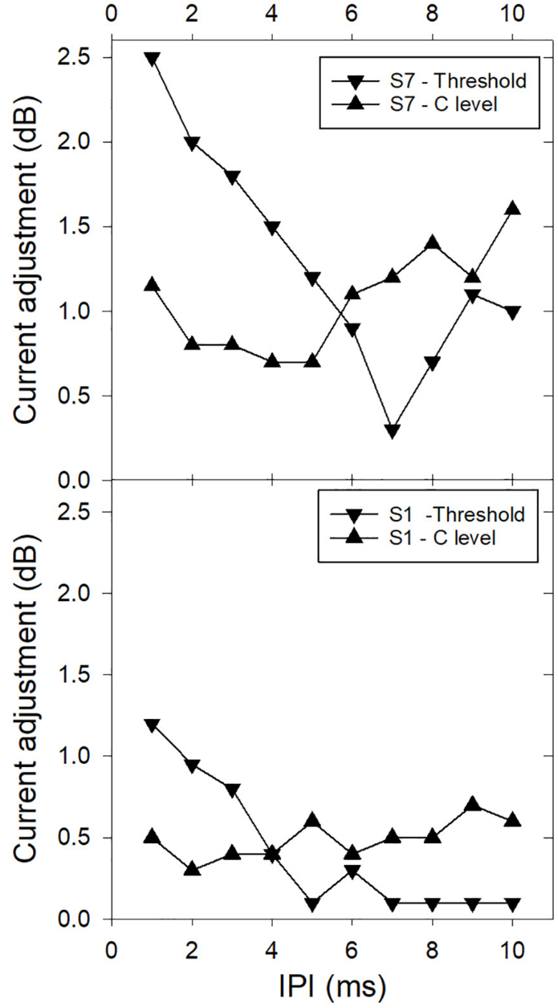

FIGURE 3 | Examples from two CI users showing the effect of interpulse interval (IPI) on loudness summation. The vertical axis shows the current reduction (in dB) needed to make the 2-pulse-per-period stimulus the same loudness (or threshold precept) as the single-pulse-per-period stimulus. The period was 20 ms. The two examples illustrate the non-monotonic effects that are variable between subjects and loudness levels (threshold or comfortable level – C) (Data redrawn from McKay and McDermott, 1998).

across the different psychophysical experiments in CI users were successfully predicted by the model using the central integration window identical to that used to predict data in similar acoustic experiments, and with consistent model fitting parameter values across experiments. As in McKay and McDermott (1998), it was clear that the scaling factor of current to excitation (in dB/dB) needed to fit the experimental data increased for stimuli with higher absolute current levels (i.e., excitation was not a fixed power function of current over an extended range of currents). The increase in *S* at higher levels is likely to be due to the higher currents accessing more tightly packed but distant axonal processes compared to the sparse peripheral processes in the deaf cochlea, as also proposed by Nelson et al. (1996) based on intensity discrimination experiments. The fact that the normal-hearing central temporal integration window could be used without adjustment to explain the measured data implies that temporal resolution is essentially normal in CI users, as measured by the low-pass cut-off frequency of temporal modulation transfer functions, which is determined by the integration window shape.

In contrast, by applying the same phenomenological model to data from the same psychophysical experiments for users of the auditory mid-brain implant, McKay et al. (2013b) demonstrated that electrically stimulated neurons in the inferior colliculus must behave quite differently to peripheral auditory neurons (a higher average spike probability, close to 1, and shorter average recovery time of 1–2 ms) and that the normal-hearing central integration window needed to be considerably widened to explain the psychophysical data. Additionally, a large degree of adaptation had to be included in the first model step to explain the effects of masker duration on forward masking (an inclusion that was not necessary for CI users).

Clinical Application of the Temporal Model: Objective Fitting of CIs

All modern implant designs enable the measurement of electrically evoked compound action potentials (ECAPs) – the whole-nerve response of the auditory nerve to individual current pulses – using implanted intracochlear electrodes as measurement electrodes. The use of ECAPs in automatic or objective programing of CIs has been limited by the very modest correlation between ECAP thresholds and the psychophysical data required for programing. The latter data are the current levels on individual electrodes required to attain hearing threshold and comfortably loud sensations for pulse trains at the sound processor stimulation rate (usually at least 500 Hz). Although hearing thresholds of single-pulse stimuli, or pulse trains with very low rate (e.g., 40 Hz), are highly correlated with ECAP thresholds (Brown et al., 1996), the correlation reduces as the rate of stimulation for the psychophysical measurement increases (Brown et al., 2000; Hughes et al., 2000; Cafarelli Dees et al., 2005). The decrease occurs because the slope of the threshold (or equal loudness) versus rate function (see **Figure 2**) varies across people in a way that cannot be predicted from the ECAP measurement for an isolated pulse. Therefore, ECAP thresholds for isolated pulses cannot be used on their own for totally objective programing. To achieve objective programing using ECAP thresholds, additional objective information about the shape of the behavioral threshold versus rate function is needed.

The relation between the total excitation evoked by an isolated current pulse and the loudness evoked by a high-rate pulse train (the latter needed for CI programing) can be predicted for an individual by the Temporal Model if we know how the evoked excitation varies for each pulse in a high-rate pulse train for that individual. If we could objectively measure the latter (instead of modeling it in step 1) then the slope of the individual behavioral threshold versus rate function could be predicted by the Temporal Model. The slope, in turn, would allow the high-rate threshold to be estimated given the low-rate threshold predicted from the low-rate ECAP threshold. McKay et al. (2013a) used a high-rate subtraction technique (Hay-McCutcheon et al., 2005) that allows ECAP amplitudes to be measured for individual pulses within an ongoing high-rate pulse train. They hypothesized that the relative excitation evoked by each pulse in the pulse train (see example in **Figure 4**) is linearly correlated with the relative ECAP amplitudes evoked by the same pulses, and that therefore these subject-specific relative ECAP amplitudes can be inserted into step 1 of the Temporal Model to predict individual differences in the slope of the behavioral threshold versus rate functions. The results showed that, for rates above 500 pps, where refractory effects and temporal integration have the most influence on loudness, the average ECAP amplitude changes (averaged across subjects) predicted the average behavioral slope well, but neither varied significantly between participants. Instead, for rates below 500 pps, where very little reduction in excitation occurs after the first pulse (**Figure 4**), there was large variability between participants in the slope of the behavioral threshold versus rate function. The differences between subjects could be fitted by the Temporal Model by adjusting the scaling factor, S, between current and excitation to increase more steeply with level in individuals with a flatter threshold function below 500 Hz. Based on the idea that a steep increase in S may be associated with activation of more distant axonal processes, it was hypothesized that individuals with a flatter behavioral function below 500 Hz were those with poorer survival of peripheral processes (thus needing higher currents to achieve the same loudness compared to those with better neural survival). Indeed, animal studies have shown that the effect of rate on threshold for low rates is correlated with cochlear health (Pfingst et al., 2011).

Based on the results of McKay et al. (2013a) it was hypothesized that an objective measure of neural health might be combined with standard ECAP thresholds to improve the prediction of high rate behavioral thresholds for objective programing. McKay and Smale (2017) tested this hypothesis, by measuring the current offset (in dB) between ECAP amplitude growth functions evoked by stimulus pulses differing in phase duration or IPG duration. These objective measurements have been correlated with spiral ganglion cell survival in animal studies (Prado-Guitierrez et al., 2006; Ramekers et al., 2014). Brochier et al. (2020) have presented a theoretical model to explain the effects of IPG on ECAPs, and applied it to previous

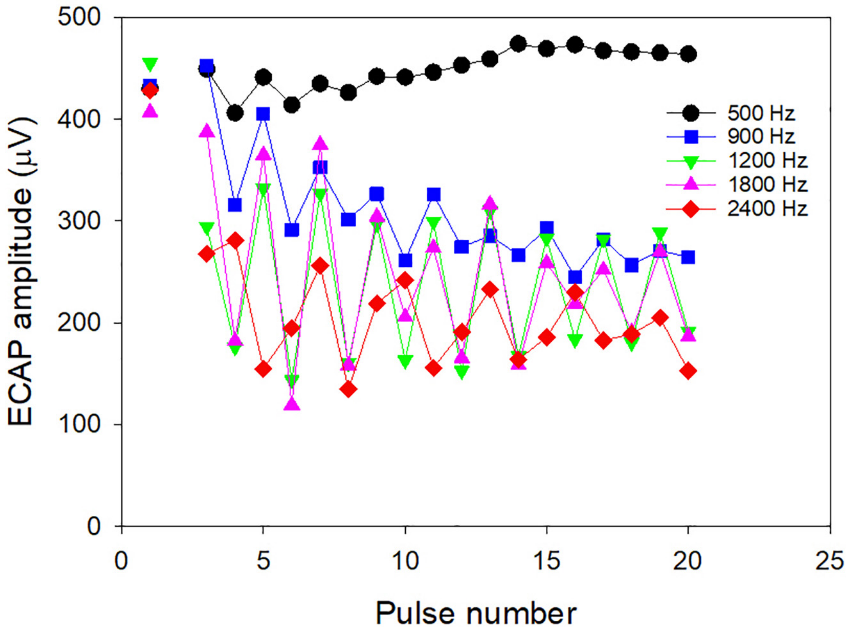

FIGURE 4 | ECAP amplitudes to individual pulses in continuous pulse trains with different rates of stimulation. The unconnected symbols on the left are amplitudes to individual pulses as per the usual clinical measurements of ECAPs. The data was collected by McKay et al. (2013a).

animal and human data. They argued that the ECAP function offset measurement (as used by McKay and Smale) is correlated with neural health (i.e., the health status of surviving neurons) as distinct from neural density (or number of surviving neurons), although these two aspects of cochlear health are likely to be correlated with each other, particularly in animal studies, due to the deafening techniques used. Consistent with their own hypothesis, McKay and Smale (2017) showed that the ECAP function offset (averaged across electrodes) was modestly correlated across subjects with the average slope of the behavioral thresholds versus rate function for rates between 40 and 1,000 Hz, but not the slopes for rates higher than 1,000 Hz. Thus, subjects with flatter low-rate function slopes on average across the electrode array were those with poorer health of surviving neurons, as measured by the ECAP offset.

With regard to the slopes of the ECAP amplitude growth functions, McKay and Smale (2017) found that, within individual subjects, electrodes with higher behavioral thresholds had greater ECAP slopes (expressed in μV/dB). This result is consistent with the observation of McKay et al. (2013b) that high current levels for high-rate stimuli are associated with a faster increase with level of the scaling factor S (excitation growth with current on a dB/dB scale). It is interesting to note that Brochier et al. (2020) argue that the ECAP amplitude growth function slope measured in dB/dB is not related to either neural survival density or health of the surviving neurons. The same would apply to the ECAP slopes in μV/dB measured in McKay and Smale (2017) since they were calculated over identical ranges of ECAP amplitudes for different stimulus conditions. Consistent with this observation, the ECAP offset measurement was not correlated with the ECAP amplitude growth function slope, and the ECAP slopes did not predict any across-subject variations in absolute behavioral thresholds or slopes of the threshold versus rate functions. Overall, use of both measures together improved the prediction of high-rate behavioral thresholds using ECAP measures alone. For example, for behavioral thresholds at rates of 1,000 Hz, the correlation between predicted and actual thresholds increased from $r = 0.47$ ($p = 0.12$) to $r = 0.70$ ($p < 0.001$) when the ECAP offset and ECAP slope were used as predictors in addition to the ECAP threshold.

MULTI-ELECTRODE STIMULI

Loudness of Multi-Electrode Stimuli and the Detailed Model

In normal CI use, multiple electrodes are activated in quick succession. It is therefore important to consider how loudness is summed across different places in the cochlea for interleaved electrical pulse trains. McKay et al. (2001) studied loudness summation for two interleaved pulse trains, measuring the influence on loudness summation of electrode separation, pulse repetition rate, and overall current level. In the experiment, two pulse trains on two different electrodes were first loudness balanced, and then interleaved. The current reduction (in dB) in the dual-electrode stimulus needed to equate its loudness to that of each component single-electrode stimulus was used as the (relative) measure of loudness summation. Surprisingly, the effect of electrode separation was very small, and, in addition, varied in direction, with some participants showing a reduction in loudness as the electrode separation was decreased and some showing an increase in loudness. Analogs to the effect of temporal separation described in section "Single-Electrode Stimuli," the results were consistent with two counteracting effects of spatial electrode separation. A phenomenological model (labeled here as the "Detailed Model") was proposed to explain the results of these experiments, in which the loudness of stationary (time invariant) electrical stimuli is determined by three steps as follows:

1. Using the Temporal Model, neural activity at each cochlear place is integrated using the sliding central temporal integration window. The output of this step is a spatial "excitation density" function that can vary over time, but will be relatively constant for a stationary stimulus.
2. The excitation density function from step 1 is transformed to an instantaneous "specific loudness" function (i.e., loudness arising from each place in the cochlea at that instant). The function that performs this transform relates neural activity to loudness.
3. The specific loudness is then integrated across cochlear place, similarly to the integration of specific loudness in acoustic models of loudness (Moore and Glasberg, 1997), the result of which is the overall loudness of the stimulus.

When electrodes are in close proximity, the overlap of the neural populations stimulated by each electrode is increased, leading to reduced overall neural activation in step 1 due to neural refractoriness. If loudness were linearly related to the total amount of evoked neural activity (i.e., the transform in step 2 was linear), then loudness would always decrease as electrode separation is decreased. The finding that loudness does not systematically decrease, however, leads to the conclusion that the transform in step 2 is non-linear and expansive (e.g., a power or exponential function). In that case, excitation density functions that are more localized (same total excitation but over a smaller area) would produce a greater loudness than ones that are more spatially spread. Thus, if neural refractoriness was not present in step 1, loudness would always systematically increase as electrode separation decreased. The two effects together lead to no, or little, effect of separation on loudness, as seen in the psychophysical data.

The application of the Detailed Model requires knowledge of individual characteristics of the spread of activation and the response properties of the activated neurons, both of which are likely to vary considerably between different people and places in the cochlea. However, these properties can be inferred from physiological data or psychophysical data, as described in section "Single-Electrode Stimuli," to apply the model in different conditions to explain how loudness varies for different stimuli. A second, practical, way of applying the model without the need to find the details needed in step 1 can be derived from the fact that there was very little effect of electrode separation on loudness in McKay et al. (2001). This method of applying the Detailed Model, which we will designate the "Practical Method" (McKay et al., 2003), is described below.

The Practical Method for Predicting the Relative Loudness of Electrical Stimuli

The development of the Practical Method used the approximation that there is no effect of electrode separation on loudness, together with the assumption that individual current pulses of a complex stimulus that do not produce spatially overlapping effects in the cochlea contribute independently to the overall loudness. The latter assumption is based on acoustic models of loudness (Zwicker and Scharf, 1965; Moore and Glasberg, 1997) in which loudness contributions from non-overlapping cochlear filters contribute additively to the total loudness. Since the loudness-addition step of acoustic models refers to loudness processing at stages more central than the cochlea, it is reasonable to presume that the same central process applies in electrical hearing. If pulses evoking non-overlapping neural excitation patterns contribute independently to loudness, and the overall loudness does not change with the degree of overlap, then electrical pulse trains must always behave *as if* the loudness contributions from different current pulses are independent, regardless of whether they are widely or closely spaced in the cochlea. In other words, if the effect of overlapping neural activation patterns on loudness is not significant, and can be approximated as zero, then the loudness evoked by the different pulses must always add similarly to the case when the activation patterns do not overlap, and the pulses must contribute additively and independently to the overall loudness, no matter where they occur on the electrode array.

The Practical Method proposes that a running estimate of loudness (defined here as "instantaneous loudness") relative to the loudness of a reference stimulus can be obtained by summing the loudness contributions of each pulse in small reference time windows (e.g., a 2 ms rectangular window). The loudness contribution of each pulse (L) is calculated from a loudness growth function of log(L) versus clinical current level (c). The loudness growth function for each electrode can be determined experimentally using the assumption that a stimulus that has two equal-current pulses in one period has twice the loudness of a stimulus with one pulse per period. The slope of the loudness growth function at that particular current level is then determined by the current adjustment needed to loudness balance the two stimuli. By measuring the slope at multiple absolute current levels and using different rates of stimulation, a complete growth function can be derived. An example of such a loudness growth function is shown in **Figure 5**, and is characterized by Eq. 2:

$$Log\,(L) = a \times c + [0.03 \times b \times e^{\frac{(c-c_0)}{b}}] + k, \tag{2}$$

where a, b, and c_0 are fitting parameters. The parameter a is the slope of the linear portion of the function and applies when c is less than c_0, the latter defining the knee-point above which the function becomes expansive. The arbitrary constant, k, can be used to set the loudness of a reference stimulus to an arbitrary loudness value. In the experiment to derive Eq. 2 (McKay et al., 2003), clinical current levels were used, which are equal to logarithmic steps of 0.176 dB for the CI24M implant used. The relation between current level (c) and current (i) in μA is given by the formula (provided by the manufacturer):

$$i = 10 \times 175^{c/255} \tag{3}$$

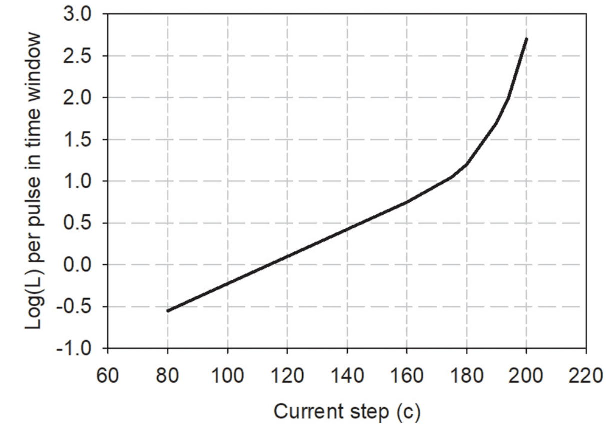

FIGURE 5 | Loudness growth function for one CI user derived from loudness summation experiments as described in the text. To use the Practical Method to estimate the loudness of any electrical pulse train, the loudness contribution of each pulse in 2 ms windows is obtained from the graph and summed to estimate the loudness of the stimulus relative to a reference stimulus of loudness 100 (Data redrawn from McKay et al., 2003).

It can be seen that the relation between loudness, L, and current, i, can be described as a power function (with exponent a) for low currents ($c << c_0$), when the second term in the equation becomes essentially zero. Low currents will usually apply when the rate of stimulation is high, for example, at the output of most clinically used sound processors. McKay et al. (2003) found that the slope a did not vary very much between participants. When the current is expressed in dB instead of clinical current units, the linear slope, a, had a mean of 0.1 *log(L)* per dB; in other words, loudness increased by a ratio of 1.26 for every dB increase in current in the linear part of the loudness growth function. This value of a can also be estimated from the slope of threshold versus rate functions for rates above 900 Hz (where absolute current values are low). For example, analysis of average high-rate slopes in the threshold data in Figure 2 of McKay et al. (2013a) produces the same value of $a = 0.1$, when current is expressed in dB.

An extended simplification is possible when predicting the relative loudness of high-rate stimuli, where the first term can be used on its own with $a = 0.10$ if expressing current in dB, without the need to generate participant-specific values of the other fitting parameters. The exponential term, which only becomes significant at higher current levels, is likely related to the increase at higher current levels of the scaling factor (S) described above that is needed to fit psychophysical data using the Temporal and Detailed Models. If we assume that loudness is a power function of neural excitation, as is common when relating psychophysical percepts to physiological data, then it can be inferred from Eq. 2 that the transform from current to excitation is also a power function for low currents (i.e., a constant exponent, S), but that for currents past the kneepoint, c_0, S will increase with increasing absolute current.

In McKay et al. (2003), two psychophysical experiments were carried out to validate the Practical Method using multi-electrode periodic stimuli with a period of 2 ms (which can be considered perceptually time invariant). In the first experiment, dual electrode 2-pulse-per-period stimuli were created in which the relative currents of the two pulses were varied and the stimuli loudness balanced against the reference stimulus, which comprised equally loud pulses on the two electrodes. The predicted loudness (derived from the Practical Method) of the balanced stimuli relative to the reference stimulus was constant, as expected, as the relative currents were varied. In the second experiment, 54 arbitrary stimuli of differing overall loudness were created, which had from 1 to 8 pulses in the 2 ms period, and where each pulse could be on an arbitrary electrode with arbitrary current value (within the dynamic range of the participant). A reference stimulus on a central electrode was balanced against each of the 54 stimuli and the balanced current of the reference was compared to that predicted by the Practical Method. The average difference between predicted and actual balanced current of the reference stimulus was very small, being only 0.2 clinical current levels (0.035 dB).

A third validation experiment was carried out by McKay and Henshall (2010), who investigated the effect of amplitude modulation on the loudness of single-electrode stimuli. In that experiment, modulated stimuli had different carrier rates (0.5, 1 or 8 kHz), different modulation rates (500 or 250 Hz), different modulation depths, and different overall levels (threshold, 60 and 90% of the dynamic range). The Practical Method was used to predict the effects of carrier rate, modulation frequency, and overall level on the current of the unmodulated stimulus of the same carrier rate that was equal in loudness to the modulated stimulus. The model correctly predicted that, for stimuli with low currents (the 8 kHz carrier rate stimuli at all levels in the dynamic range, and the threshold stimuli with lower carrier rates), the equally loud unmodulated stimulus had a current equal to the average current in the modulated stimulus. This finding is consistent with these stimuli having low enough currents to fall onto the linear part (in log/log coordinates) of the loudness growth function (Eq. 2 and **Figure 5**). The other stimuli (500- and 1,000-Hz carrier rates at 60 or 90% DR) comprised pulses with higher currents that fell into the non-linear expansive part of the loudness growth function, and both model and psychophysical data showed that the current of the equally loud non-modulated stimulus was greater than the average current of the modulated stimulus and moved closer to the peak modulated current as the absolute level of the stimulus increased (carrier rate decreasing or level in the dynamic range increasing). The insights provided by this study showed that it was important, when determining modulation detection ability in CI users, to take into account systematic differences in loudness between modulated and unmodulated stimuli, as loudness differences will provide confounding cues to the presence of modulation, leading to overestimation of modulation detection abilities.

This overestimation of modulation detection ability was demonstrated by Fraser and McKay (2012), who measured a series of temporal modulation transfer functions (modulation detection threshold versus modulation frequency) while limiting the use of loudness cues. In the experiment, the target (modulated) stimulus was loudness balanced with the standard (unmodulated) stimulus, and level jitter was used to additionally limit use of loudness cues. Previously studies investigating modulation detection in CI users had set the current in the reference unmodulated stimulus to the average current in the modulated stimulus. The loudness cues in the latter case would become more salient as the modulation frequency is increased, when larger modulation depths are needed. Thus, loudness cues led to overestimation of modulation detection ability, particularly for high-frequency modulations, thus underestimating the low-pass characteristics of the modulation transfer functions. The functions measured by Fraser and McKay (2012) had low-pass cut-off frequencies broadly consistent with those for normal hearing subjects. The facts that low-frequency cut-off frequencies are broadly in the normal range, and that the central temporal integration window used in the Temporal Model is the same as for normal hearing, suggest that temporal resolution is largely unaffected in CI users. These results suggest that the differences between CI users in absolute measures of modulation detection ability at low modulation frequencies, which have been related to differences in speech perception ability (Luo et al., 2008; Arora et al., 2011; Won et al., 2011; Brochier et al., 2017), are related more to variance across subjects in intensity difference limens (McKay et al., 2018) or modulation sensitivity than to variance in temporal resolution.

Extensions of the Practical Method

The Practical Method as derived by McKay et al. (2003) is able to output a running estimate of loudness in small increments of time by summing loudness contribution from each pulse. For a perceptually stationary stimulus, this estimate will suffice to deduce the overall loudness of the stimulus (relative to that of a reference stimulus). However, if the stimulus is dynamically changing over time, a further question would be how to derive the overall loudness perceived from the time-varying estimates output by the Practical Method. This question has been addressed in a study by Francart et al. (2014), who investigated how existing acoustic models for predicting the loudness of time-varying signals can be adapted to extend the Practical Method to predict the overall loudness of time-varying electrical signals in CIs. Two methods were described that well predicted the psychophysical data, both of which first calculated the "instantaneous loudness" by integrating the individual pulse loudness contributions (as defined by the Practical Method) over a sliding temporal integration window. In both cases, the shape of the integration window was defined as in Eq. 1, and the Equivalent Rectangular Duration (ERD) of the window was used as a fitting parameter. The first method investigated by Francart et al. (2014) that fitted the experimental data well used an integration window with ERD of 2 ms and then calculated long-term loudness from the varying instantaneous loudness following the method of Glasberg and Moore (2002), which entailed application of an automatic gain control like circuit to the instantaneous loudness values, with an attack time of 22 ms and a release time of 50 ms to obtain short-term loudness values, followed by application of a second automatic gain control like circuit to obtain long-term loudness values. The second successful method described by Francart et al. (2014) was simpler than the first, and used a temporal integration window with ERD of 4.3 ms to obtain the "instantaneous loudness" and then defined the 99th percentile of instantaneous loudness as the long-term loudness. Note that these integration windows have a smaller ERD than that used in the Detailed Method. These ERDs are not inconsistent with the Detailed Method, since the latter integrates peripheral neural activity, while the practical methods integrate loudness contributions. Since the transform between neural activity and specific loudness in the Detailed Method is non-linear and expansive, it would be expected that the ERD that best fits loudness integration data would be smaller than that which fits neural activity integration data.

The Practical Method also cannot be directly applied to pulsatile stimuli in which the pulses occur simultaneously rather than sequentially, for example, in certain signal processing strategies or in simultaneous analog stimuli. For these stimuli, an additional effect must be included when predicting loudness: the direct summation of simultaneous currents at the neural interface (Shannon, 1983; Tang et al., 2011). This effect is highly dependent on the distance between electrodes and the spatial spread of currents in individual cochleae. For example, Marozeau et al. (2015) compared simultaneous with sequential stimuli using monopolar and focused multipolar modes of stimulation. They found that stimuli in the multipolar mode, which is designed to produce a highly focused current field, produced only small differences in loudness between simultaneous and sequential conditions, whereas the monopolar stimuli needed current adjustments of up to 4 dB to make the simultaneous and sequential stimuli the same loudness.

In the case of stimuli with simultaneous biphasic pulses, the Practical Method could still be used if psychophysical loudness summation data due to current summation for the stimulus conditions used (e.g., mode of stimulation and electrode distance) were obtained and included in the model. An example of such an adaptation of the Practical Method was demonstrated by Langner et al. (2020), who measured loudness summation caused by current interaction of simultaneously activated pairs of virtual channels. Virtual channels simultaneously activate two adjacent intracochlear electrodes to steer the peak of the current field to positions between the physical electrodes. Paired virtual channels therefore activate four intracochlear electrodes simultaneously. Such paired virtual channels are used in the "Optima-Paired" sound coding strategy of Advanced Bionics. In the adaption of the Practical Method, Langner et al. (2020) balanced the loudness of paired-channel stimuli to those of the component single virtual channels to create a model of how channel distance, and relative currents in the component channels of each pair, influence the loudness. This additional model was then incorporated into the Practical Method to predict the loudness of paired-channel stimulation strategies compared to strategies that sequentially activated virtual channels. To do this prediction, the loudness contribution of each paired-channel pulse pair was replaced in the Practical Method calculation of the loudness by an equivalently loud single-channel pulse with current determined by the model derived from loudness balancing data. This method of Langner et al. (2020) provides a way for clinicians to automatically adjust the program of the sound processors when switching between paired-pulse and fully sequential signal processing strategies. To do this adjustment when changing to the paired strategy from a sequential strategy, clinicians can lower the current range assigned to each virtual electrode (which is determined for sequential stimulation using each virtual channel separately) by an amount predicted by the model calculation, so that the simultaneous stimulation does not produce sounds that are too loud.

Clinical Applications of the Loudness Models to Signal Processing Strategies

The Practical Method of loudness estimation has been applied to several novel signal processing strategies that aim to create more control of overall loudness and frequency-specific contributions to loudness (specific loudness). Current clinically used processing strategies assign a fixed electrical dynamic range to each electrode, based on single-electrode psychophysical measures of loudness. However, this technique does not take into account the loudness summation that occurs when multiple electrodes are activated concurrently in normal implant use, leading to sounds of different bandwidth or overall levels producing loudness percepts that vary in ways that are quite different to what an acoustic listener would hear.

The first signal processing strategy to use the Practical Method to control loudness was the SpeL strategy (McDermott et al., 2003), which utilized the acoustic loudness model of Moore and Glasberg (1996, 1997) to convert the incoming signal into specific loudness in each cochlear equivalent rectangular bandwidth (ERB), following which the specific loudness was converted using the Practical Method to the required current values on electrodes across the array (**Figure 6**). Cochlear ERBs divide the cochlea into non-overlapping sections with characteristic frequency ranges related to the width of cochlear filters at the same frequencies, and each electrode was assigned a constant 1.3 contiguous ERBs. Thus, in the SpeL strategy, the specific loudness pattern of the incoming acoustic signal (calculated for a person with normal hearing) was replicated as the specific loudness pattern produced by electrical pulses across the electrode array, effectively "normalizing" the relative overall loudness of incoming sounds, and the relative loudness contributions of different frequencies within the sound. McDermott et al. (2003) implemented SpeL in a wearable research processor and used a loudness estimation psychophysical task for participants wearing the research processor to compare the predicted and estimated loudness of acoustic noise bands of various bandwidths and levels. The results confirmed that SpeL restored the relative loudness of different bandwidths and different intensities to that experienced by normal-hearing listeners. McDermott et al. (2005) showed that, after 4 weeks trial use of SpeL, CI users had equivalent speech understanding in quiet and noise to their clinical strategy (ACE), while improving the audibility of soft sounds by an average of 5 dB. In the ACE strategy, soft speech will activate fewer electrodes than louder speech, as frequency bands with very low levels produce no stimulation. This drop in number of activated electrodes leads to an uncompensated reduction in loudness summation across electrodes, causing the soft speech to be too difficult to hear. In contrast, the SpeL strategy calculates the correct (or "normal") overall loudness of the speech and automatically adjusts the currents to produce the correct loudness.

The SpeL strategy required individual loudness growth functions (like **Figure 5**) to be measured on each electrode and also required the frequency-to-electrode allocation to be altered away from that which the participants were used to in the ACE strategy, so that each electrode received information from an a constant 1.3 contiguous ERBS. Although the total range of assigned frequencies across the electrode array were as closely matched as possible to the participant's usual range of assigned frequencies, there remained a significant shift in assignment

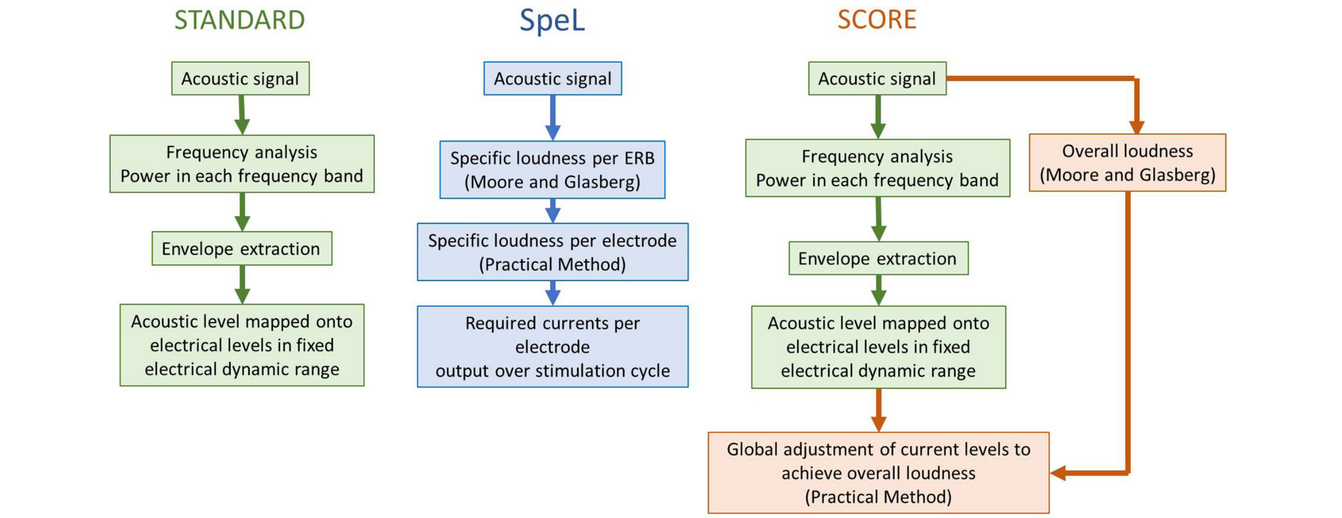
STANDARD
Acoustic signal
Frequency analysis
Power in each frequency band
Envelope extraction
Acoustic level mapped onto electrical levels in fixed electrical dynamic range
SpeL
Acoustic signal
Specific loudness per ERB (Moore and Glasberg)
Specific loudness per electrode (Practical Method)
Required currents per electrode output over stimulation cycle
SCORE
Acoustic signal
Frequency analysis
Power in each frequency band
Envelope extraction
Acoustic level mapped onto electrical levels in fixed electrical dynamic range
Overall loudness (Moore and Glasberg)
Global adjustment of current levels to achieve overall loudness (Practical Method)

FIGURE 6 | Schematic that illustrates the essential principles of the signal processing strategies SpeL and SCORE in relation to standard clinically used signal processing strategies. The flowchart illustrates the differences in processes that occur for each stimulation cycle or within each update window. The reader is referred to the original articles (McDermott et al., 2003; Varsavsky and McDermott, 2013) for a detailed technical description of the strategies.

toward the middle of the array. Thus, failure to adapt fully to the change of electrode assignment may have influenced the results of the speech test in McDermott et al. (2005). These considerations led to the development of a second strategy based on the Practical Method – SCORE (Varsavsky and McDermott, 2013). Instead of replicating the acoustic *specific* loudness pattern in the electrical stimulation across electrodes, SCORE aimed to only control the instantaneous *overall* loudness (**Figure 6**). It did this by estimating the incoming instantaneous loudness using the acoustic models of Moore and Glasberg (1996, 1997), and adjusting the output current levels (equally across electrodes) of the ACE strategy to match the acoustic instantaneous overall loudness, using the Practical Method. Since SCORE only acts upon the output of a signal processing strategy, it can be combined with any signal processing strategy (not solely ACE, as used by Varsavsky and McDermott, 2013) to control overall loudness. It can therefore take advantage of features of processing strategies (such as the noise reduction benefit of maxima selection in ACE) while normalizing overall loudness percepts. Varsavsky and McDermott (2013) implemented SCORE for experienced users of the ACE strategy and demonstrated that soft speech (50 dB SPL) was more intelligible with SCORE than with the ACE strategy (a mean increase of 8.8 percentage points). Since SCORE matches instantaneous acoustic loudness with instantaneous electric loudness, it has an ideal application in bimodal hearing, in which CI users use a hearing aid on the non-implanted ear. SCORE-Bimodal was developed and tested by Francart and McDermott (2012b). It has the same SCORE processing as described above for the CI side, so that the instantaneous loudness (measured in time frames of 6.9 ms) of the electrical signal matches the instantaneous loudness of the acoustic signal at the CI microphone as predicted for normal hearing by the model of Moore and Glasberg (1997). On the hearing aid side, the predicted difference in loudness for normal hearing and hearing impaired listeners is computed by the model of Moore and Glasberg (1997) and used to adjust the gain of the hearing aid to match the normal-hearing loudness. Clinical assessment of SCORE-Bimodal (Francart and McDermott, 2012a) showed that it improved localization ability while maintaining speech perception ability in quiet and noise.

The Temporal and Detailed Models use the output of the sliding temporal integration window (integrated excitation) to predict perceptual decisions about modulation detection. Based on the model, modulation of rate of stimulation would lead to similar modulation of the integrator output as modulation of current amplitude. Brochier et al. (2018a) compared rate modulation detection with amplitude modulation detection and investigated the effects of modulation frequency and presentation level. They found that the two types of modulation detection were affected similarly by level and modulation frequency and were correlated with each other across the subject group. Following this result, Brochier et al. (2018b) devised a novel sound coding strategy (ARTmod) that coded amplitude modulations of the acoustic signal onto simultaneous rate and amplitude modulation in the electrical signal. They hypothesized that the two types of modulation would independently contribute to perception of amplitude modulations in acoustic speech signals, and thus it would be possible to use the added rate modulation to improve speech understanding. They found that speech perception improved with increasing amounts of rate modulation, which is consistent with rate and amplitude modulation being processed similarly and additively to transmit the acoustic amplitude modulation in the speech signal.

Finally, an adaptation of the Temporal Model was used by Lamping et al. (2020) to devise a signal processing strategy (designated TIPS) that removed pulses that were more likely to be masked by preceding pulses. The authors used the sliding integration window of Eq. 1 and applied it directly to the currents of the pulses in a continuous interleaved sampling (CIS) strategy, followed by a decision criterion that compared the integrator output with and without the pulse at the center of each window to decide whether to omit that pulse. Criteria of less than 1, 1.3, and 1.8 dB difference in integrator output were used to remove 25, 50, and 75% of current pulses, respectively. It should be noted that, since excitation is a power function of current (the scaling factor, S, in the model), applying the integrator to the current should lead to less variation in the integrator output than applying it to the excitation: therefore the criteria differences in output would be larger than those used in the study if the Temporal Model was used, and closer to the 3 dB criterion for detection used in acoustic studies of forward masking (Plack et al., 2002). However, since the criteria were used as an experimental variable, this difference does not have relevance to the results, which showed that the TIPS strategy improved speech perception in noise by 2.4 dB signal-to-noise ratio when removing 50% of the masked pulses.

CONCLUSION

The application of phenomenological loudness models to psychophysical data of CI users has led to improved understanding of the influence of individual peripheral neural response behavior and neural health status on the transmission of features of the acoustic signal to the perception of the CI user. The knowledge gained has led to better understanding of differences in outcomes between CI users, and novel ways of determining cochlear health in CI users. The models have been applied to the development of novel signal processing strategies that aim to provide CI users with a more natural perception of loudness and better localization ability and to a novel way to improve the transmission of important amplitude modulations in speech to the CI listener.

AUTHOR CONTRIBUTIONS

CM is the sole contributor to this review.

ACKNOWLEDGMENTS

The Bionics Institute acknowledges the support it received from the Victorian Government through its Operational Infrastructure Support Program.

REFERENCES

Arora, K., Vandali, A., Dowell, R., and Dawson, P. (2011). Effects of stimulation rate on modulation detection and speech recognition by cochlear implant users. *Int. J. Audiol.* 50, 123–132. doi: 10.3109/14992027.2010.527860

Auerbach, B. D., Radziwon, K., and Salvi, R. (2019). Testing the central gain model: loudness growth correlates with central auditory gain enhancement in a rodent model of hyperacusis. *Neuroscience* 407, 93–107. doi: 10.1016/j.neuroscience.2018.09.036

Bonnet, R. M., Frijns, J. H., Peeters, S., and Briaire, J. J. (2004). Speech recognition with a cochlear implant using triphasic charge-balanced pulses. *Acta Otolaryngol.* 124, 371–375. doi: 10.1080/00016480410031084

Boulet, J., White, M., and Bruce, I. C. (2016). Temporal considerations for stimulating spiral ganglion neurons with cochlear implants. *J. Assoc. Res. Otolaryngol.* 17, 1–17. doi: 10.1007/s10162-015-0545-5

Brochier, T., McDermott, H. J., and McKay, C. M. (2017). The effect of presentation level and stimulation rate on speech perception and modulation detection for cochlear implant users. *J. Acoust. Soc. Am.* 141:4097. doi: 10.1121/1.4983658

Brochier, T., McDermott, H. J., and McKay, C. M. (2018a). Rate modulation detection thresholds for cochlear implant users. *J. Acoust. Soc. Am.* 143, 1214–1222. doi: 10.1121/1.5025048

Brochier, T., McKay, C., and McDermott, H. (2018b). Encoding speech in cochlear implants using simultaneous amplitude and rate modulation. *J. Acoust. Soc. Am.* 144:2042. doi: 10.1121/1.5055989

Brochier, T., McKay, C. M., and Carlyon, R. P. (2020). Interpreting the effect of stimulus parameters on the electrically evoked compound action potential and on neural health estimates. *J. Assoc. Res. Otolaryngol.* 20, 431–448.

Brown, C. J., Abbas, P. J., Borland, J., and Bertschy, M. R. (1996). Electrically evoked whole nerve action potentials in Ineraid cochlear implant users: responses to different stimulating electrode configurations and comparison to psychophysical responses. *J. Speech Hear. Res.* 39, 453–467. doi: 10.1044/jshr.3903.453

Brown, C. J., Hughes, M. L., Luk, B., Abbas, P. J., Wolaver, A., and Gervais, J. (2000). The relationship between EAP and EABR thresholds and levels used to program the nucleus 24 speech processor: data from adults. *Ear. Hear.* 21, 151–163. doi: 10.1097/00003446-200004000-00009

Cafarelli Dees, D., Dillier, N., Lai, W. K., von Wallenberg, E., van Dijk, B., Akdas, F., et al. (2005). Normative findings of electrically evoked compound action potential measurements using the neural response telemetry of the Nucleus CI24M cochlear implant system. *Audiol. Neurootol.* 10, 105–116. doi: 10.1159/000083366

Carlyon, R. P., Cosentino, S., Deeks, J. M., Parkinson, W., and Arenberg, J. A. (2018). Effect of stimulus polarity on detection thresholds in cochlear implant users: relationships with average threshold, gap detection, and rate discrimination. *J. Assoc. Res. Otolaryngol.* 19, 559–567. doi: 10.1007/s10162-018-0677-5

Carlyon, R. P., Deeks, J. M., Undurraga, J., Macherey, O., and van Wieringen, A. (2017). Spatial selectivity in cochlear implants: effects of asymmetric waveforms and development of a single-point measure. *J. Assoc. Res. Otolaryngol.* 18, 711–727. doi: 10.1007/s10162-017-0625-9

Carlyon, R. P., van Wieringen, A., Deeks, J. M., Long, C. J., Lyzenga, J., and Wouters, J. (2005). Effect of inter-phase gap on the sensitivity of cochlear implant users to electrical stimulation. *Hear. Res.* 205, 210–224. doi: 10.1016/j.heares.2005.03.021

Fielden, C. A., Kluk, K., and McKay, C. M. (2013). Place specificity of monopolar and tripolar stimuli in cochlear implants: the influence of residual masking. *J. Acoust. Soc. Am.* 133, 4109–4123. doi: 10.1121/1.4803909

Francart, T., Innes-Brown, H., McDermott, H. J., and McKay, C. M. (2014). Loudness of time-varying stimuli with electric stimulation. *J. Acoust. Soc. Am.* 135, 3513–3519. doi: 10.1121/1.4874597

Francart, T., and McDermott, H. (2012a). Speech perception and localisation with SCORE bimodal: a loudness normalisation strategy for combined cochlear implant and hearing aid stimulation. *PLoS One* 7:e45385. doi: 10.1371/journal.pone.0045385

Francart, T., and McDermott, H. J. (2012b). Development of a loudness normalisation strategy for combined cochlear implant and acoustic stimulation. *Hear. Res.* 294, 114–124. doi: 10.1016/j.heares.2012.09.002

Fraser, M., and McKay, C. M. (2012). Temporal modulation transfer functions in cochlear implantees using a method that limits overall loudness cues. *Hear. Res.* 283, 59–69. doi: 10.1016/j.heares.2011.11.009

Glasberg, B. R., and Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.* 50, 331–342.

Hay-McCutcheon, M. J., Brown, C. J., and Abbas, P. J. (2005). An analysis of the impact of auditory-nerve adaptation on behavioral measures of temporal integration in cochlear implant recipients. *J. Acoust. Soc. Am.* 118, 2444–2457. doi: 10.1121/1.2035593

He, S., Xu, L., Skidmore, J., Chao, X., Jeng, F. C., Wang, R., et al. (2020). The effect of interphase gap on neural response of the electrically stimulated cochlear nerve in children with cochlear nerve deficiency and children with normal-sized cochlear nerves. *Ear. Hear.* 41, 918–934. doi: 10.1097/aud.0000000000000815

Horne, C. D., Sumner, C. J., and Seeber, B. U. (2016). A phenomenological model of the electrically stimulated auditory nerve fiber: temporal and biphasic response properties. *Front. Comput. Neurosci.* 10:8. doi: 10.3389/fncom.2016.00008

Hughes, M. L., Brown, C. J., Abbas, P. J., Wolaver, A. A., and Gervais, J. P. (2000). Comparison of EAP thresholds with MAP levels in the nucleus 24 cochlear implant: data from children. *Ear. Hear.* 21, 164–174. doi: 10.1097/00003446-200004000-00010

Hughes, M. L., Choi, S., and Glickman, E. (2018). What can stimulus polarity and interphase gap tell us about auditory nerve function in cochlear-implant recipients? *Hear Res*. 359, 50–63. doi: 10.1016/j.heares.2017.12.015

Kwon, B. J., and van den Honert, C. (2006). Effect of electrode configuration on psychophysical forward masking in cochlear implant listeners. *J. Acoust. Soc. Am.* 119, 2994–3002. doi: 10.1121/1.2184128

Lamping, W., Goehring, T., Marozeau, J., and Carlyon, R. P. (2020). The effect of a coding strategy that removes temporally masked pulses on speech perception by cochlear implant users. *Hear. Res.* 391:107969. doi: 10.1016/j.heares.2020.107969

Langner, F., McKay, C. M., Buchner, A., and Nogueira, W. (2020). Perception and prediction of loudness in sound coding strategies using simultaneous electric stimulation. *Hear. Res.* 398:108091. doi: 10.1016/j.heares.2020.108091

Luo, X., Fu, Q. J., Wei, C. G., and Cao, K. L. (2008). Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. *Ear. Hear.* 29, 957–970. doi: 10.1097/aud.0b013e3181888f61

Macherey, O., Deeks, J. M., and Carlyon, R. P. (2011). Extending the limits of place and temporal pitch perception in cochlear implant users. *J. Assoc. Res. Otolaryngol.* 12, 233–251. doi: 10.1007/s10162-010-0248-x

Macherey, O., van Wieringen, A., Carlyon, R. P., Dhooge, I., and Wouters, J. (2010). Forward-masking patterns produced by symmetric and asymmetric pulse shapes in electric hearing. *J. Acoust. Soc. Am.* 127, 326–338. doi: 10.1121/1.3257231

Marozeau, J., McDermott, H. J., Swanson, B. A., and McKay, C. M. (2015). Perceptual interactions between electrodes using focused and monopolar cochlear stimulation. *J. Assoc. Res. Otolaryngol.* 16, 401–412. doi: 10.1007/s10162-015-0511-2

McDermott, H. J., McKay, C. M., Richardson, L. M., and Henshall, K. R. (2003). Application of loudness models to sound processing for cochlear implants. *J. Acoust. Soc. Am.* 114, 2190–2197. doi: 10.1121/1.1612488

McDermott, H. J., Sucher, C. M., and McKay, C. M. (2005). Speech perception with a cochlear implant sound processor incorporating loudness models. *Acoust. Res. Lett. Online* 6, 7–13. doi: 10.1121/1.1809152

McKay, C. M., and Henshall, K. R. (2003). The perceptual effects of interphase gap duration in cochlear implant stimulation. *Hear. Res.* 181, 94–99. doi: 10.1016/s0378-5955(03)00177-1

McKay, C. M., and Henshall, K. R. (2010). Amplitude modulation and loudness in cochlear implantees. *J. Assoc. Res. Otolaryngol.* 11, 101–111. doi: 10.1007/s10162-009-0188-5

McKay, C. M., Henshall, K. R., Farrell, R. J., and McDermott, H. J. (2003). A practical method of predicting the loudness of complex electrical stimuli. *J. Acoust. Soc. Am.* 113, 2054–2063. doi: 10.1121/1.1558378

McKay, C. M., Chandan, K., Akhoun, I., Siciliano, C., and Kluk, K. (2013a). Can ECAP measures be used for totally objective programming of cochlear implants? *J. Assoc. Res. Otolaryngol.* 14, 879–890. doi: 10.1007/s10162-013-0417-9

McKay, C. M., Lim, H. H., and Lenarz, T. (2013b). Temporal processing in the auditory system: insights from cochlear and auditory midbrain implantees. *J. Assoc. Res. Otolaryngol.* 14, 103–124. doi: 10.1007/s10162-012-0354-z

McKay, C. M., and McDermott, H. J. (1998). Loudness perception with pulsatile electrical stimulation: the effect of interpulse intervals. *J. Acoust. Soc. Am.* 104, 1061–1074. doi: 10.1121/1.423316

McKay, C. M., and McDermott, H. J. (1999). The perceptual effects of current pulse duration in electrical stimulation of the auditory nerve. *J. Acoust. Soc. Am.* 106, 998–1009. doi: 10.1121/1.428052

McKay, C. M., Remine, M. D., and McDermott, H. J. (2001). Loudness summation for pulsatile electrical stimulation of the cochlea: effects of rate, electrode separation, level, and mode of stimulation. *J. Acoust. Soc. Am.* 110, 1514–1524. doi: 10.1121/1.1394222

McKay, C. M., Rickard, N., and Henshall, K. (2018). Intensity discrimination and speech recognition of cochlear implant users. *J. Assoc. Res. Otolaryngol.* 19, 589–600. doi: 10.1007/s10162-018-0675-7

McKay, C. M., and Smale, N. (2017). The relation between ECAP measurements and the effect of rate on behavioral thresholds in cochlear implant users. *Hear. Res.* 346, 62–70. doi: 10.1016/j.heares.2017.02.009

Miller, C. A., Robinson, B. K., Rubinstein, J. T., Abbas, P. J., and Runge-Samuelson, C. L. (2001). Auditory nerve responses to monophasic and biphasic electric stimuli. *Hear. Res.* 151, 79–94. doi: 10.1016/s0300-2977(00)00082-6

Moon, A. K., Zwolan, T. A., and Pfingst, B. E. (1993). Effects of phase duration on detection of electrical-stimulation of the human cochlea. *Hear. Res.* 67, 166–178. doi: 10.1016/0378-5955(93)90244-u

Moore, B. C. J., and Glasberg, B. R. (1996). A revision of Zwicker's loudness model. *Acustica* 82, 335–345.

Moore, B. C. J., and Glasberg, B. R. (1997). A model of loudness perception applied to cochlear hearing loss. *Auditory Neurosci.* 3, 289–311.

Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1996). Detection of decrements and increments in sinusoids at high overall levels. *J. Acoust. Soc. Am.* 99, 3669–3677. doi: 10.1121/1.414964

Muller, M., Robertson, D., and Yates, G. K. (1991). Rate-versus-level functions of primary auditory nerve fibres: evidence for square law behaviour of all fibre categories in the guinea pig. *Hear. Res.* 55, 50–56. doi: 10.1016/0378-5955(91)90091-m

Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (1996). Intensity discrimination as a function of stimulus level with electric stimulation. *J. Acoust. Soc. Am.* 100, 2393–2414. doi: 10.1121/1.417949

Oxenham, A. J. (2001). Forward masking: adaptation or integration? *J. Acoust. Soc. Am.* 109, 732–741. doi: 10.1121/1.1336501

Oxenham, A. J., and Moore, B. C. J. (1994). Modeling the additivity of nonsimultaneous masking. *Hear. Res.* 80, 105–118. doi: 10.1016/0378-5955(94)90014-0

Oxenham, A. J., and Moore, B. C. J. (1995). Additivity of masking in normally hearing and hearing-impaired subjects. *J. Acoust. Soc. Am.* 98, 1921–1934. doi: 10.1121/1.413376

Parkins, C. W., and Colombo, J. (1987). Auditory-nerve single-neuron thresholds to electrical stimulation from scala tympani electrodes. *Hear. Res.* 31, 267–285. doi: 10.1016/0378-5955(87)90196-1

Pfingst, B. E., Colesa, D. J., Hembrador, S., Kang, S. Y., Middlebrooks, J. C., Raphael, Y., et al. (2011). Detection of pulse trains in the electrically stimulated cochlea: effects of cochlear health. *J. Acoust. Soc. Am.* 130, 3954–3968. doi: 10.1121/1.3651820

Pfingst, B. E., DeHaan, D. R., and Holloway, L. A. (1991). Stimulus features affecting psychophysical detection thresholds for electrical stimulation of the cochlea. I: phase duration and stimulus duration. *J. Acoust. Soc. Am.* 90, 1857–1866. doi: 10.1121/1.401665

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. doi: 10.1121/1.5050518

Plack, C. J., Oxenham, A. J., and Drga, V. (2002). Linear and nolinear processes in temporal masking. *Acta Acust. United Acust.* 88, 348–358.

Prado-Guitierrez, P., Fewster, L. M., Heasman, J. M., McKay, C. M., and Shepherd, R. K. (2006). Effect of interphase gap and pulse duration on electrically evoked potentials is correlated with auditory nerve survival. *Hear. Res.* 215, 47–55. doi: 10.1016/j.heares.2006.03.006

Ramekers, D., Versnel, H., Strahl, S. B., Smeets, E. M., Klis, S. F., and Grolman, W. (2014). Auditory-nerve responses to varied inter-phase gap and phase duration of the electric pulse stimulus as predictors for neuronal degeneration. *J. Assoc. Res. Otolaryngol.* 15, 187–202. doi: 10.1007/s10162-013-0440-x

Schneider, B., and Parker, S. (1990). Does stimulus context affect loudness or only loudness judgments? *Percept. Psychophys.* 48, 409–418. doi: 10.3758/bf03211584

Schvartz-Leyzac, K. C., Holden, T. A., Zwolan, T. A., Arts, H. A., Firszt, J. B., Buswinka, C. J., et al. (2020). Effects of electrode location on estimates of neural health in humans with cochlear implants. *J. Assoc. Res. Otolaryngol.* 21, 259–275. doi: 10.1007/s10162-020-00749-0

Schvartz-Leyzac, K. C., and Pfingst, B. E. (2016). Across-site patterns of electrically evoked compound action potential amplitude-growth functions in multichannel cochlear implant recipients and the effects of the interphase gap. *Hear. Res.* 341, 50–65. doi: 10.1016/j.heares.2016.08.002

Shannon, R. V. (1983). Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction. *Hear Res.* 12, 1–16. doi: 10.1016/0378-5955(83)90115-6

Shannon, R. V. (1985). Threshold and loudness functions for pulsatile stimulation of cochlear implants. *Hear. Res.* 18, 135–143. doi: 10.1016/0378-5955(85)90005-x

Srinivasan, A. G., Landsberger, D. M., and Shannon, R. V. (2010). Current focusing sharpens local peaks of excitation in cochlear implant stimulation. *Hear. Res.* 270, 89–100. doi: 10.1016/j.heares.2010.09.004

Tang, Q., Benitez, R., and Zeng, F. G. (2011). Spatial channel interactions in cochlear implants. *J Neural Eng.* 8:046029. doi: 10.1088/1741-2560/8/4/046029

Tang, Q., Liu, S., and Zeng, F. G. (2006). Loudness adaptation in acoustic and electric hearing. *J. Assoc. Res. Otolaryngol.* 7, 59–70. doi: 10.1007/s10162-005-0023-6

Undurraga, J. A., Carlyon, R. P., Macherey, O., Wouters, J., and van Wieringen, A. (2012). Spread of excitation varies for different electrical pulse shapes and stimulation modes in cochlear implants. *Hear. Res.* 290, 21–36. doi: 10.1016/j.heares.2012.05.003

Varsavsky, A., and McDermott, H. J. (2013). Application of real-time loudness models can improve speech recognition for cochlear implant users. *IEEE Trans. Neural. Syst. Rehabil. Eng.* 21, 81–87. doi: 10.1109/tnsre.2012.2213841

Wang, N., and Oxenham, A. J. (2016). Effects of auditory enhancement on the loudness of masker and target components. *Hear. Res.* 333, 150–156. doi: 10.1016/j.heares.2016.01.012

Won, J. H., Drennan, W. R., Nie, K., Jameyson, E. M., and Rubinstein, J. T. (2011). Acoustic temporal modulation detection and speech perception in cochlear implant listeners. *J. Acoust. Soc. Am.* 130, 376–388. doi: 10.1121/1.3592521

Wouters, J., McDermott, H. J., and Francart, T. (2015). Sound coding in cochlear implants. *IEEE Signal Process. Mag.* 32, 67–80.

Zwicker, E., and Scharf, B. (1965). A model of loudness summation. *Psychol. Rev.* 72, 3–26. doi: 10.1037/h0021703

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Psychology

MINI REVIEW
published: 28 January 2021
doi: 10.3389/fpsyg.2021.583690

Does Loudness Relate to the Strength of the Sound Produced by the Source or Received by the Ears? A Review of How Focus Affects Loudness

Gauthier Berthomieu, Vincent Koehl and Mathieu Paquier*

Univ Brest, Lab-STICC, CNRS, UMR 6285, Brest, France

Loudness is the magnitude of the auditory sensation that a listener experiences when exposed to a sound. Several sound attributes are reported to affect loudness, such as the sound pressure level at the listener's ears and the spectral content. In addition to these physical attributes of the stimulus, some subjective attributes also appear to affect loudness. When presented with a sound, a listener interacts with an auditory object and can focus on several aspects of the latter. Loudness appears to differ depending on how listeners apprehend this object, notably whether they focus on the sound that reaches their ears or that is produced by the source. The way listeners focus on the auditory object may depend on the stimulus itself. For instance, they might be more likely to focus on the sound emitted by the source if the latter is visible. The instructions given by the experimenters can also explicitly direct the listener's focus on the sound reaching the ears or emitted by the source. The present review aims at understanding how listeners focus on the auditory object depending on the stimuli and instructions they are provided with, and to describe how loudness depends on this focus.

Keywords: loudness, auditory perception, hearing, instructions, experimental setup

Edited by:
Brian Cecil Joseph Moore,
University of Cambridge,
United Kingdom

Reviewed by:
Clara Suied,
Armed Forces Biomedical Research Institute (IRBA), France
Michael Akeroyd,
University of Nottingham,
United Kingdom

****Correspondence:***
Gauthier Berthomieu
gauthier.berthomieu@univ-brest.fr

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

Received: *15 July 2020*
Accepted: *04 January 2020*
Published: *28 January 2021*

Citation:
Berthomieu G, Koehl V and Paquier M (2021) Does Loudness Relate to the Strength of the Sound Produced by the Source or Received by the Ears? A Review of How Focus Affects Loudness. Front. Psychol. 12:583690. doi: 10.3389/fpsyg.2021.583690

1. INTRODUCTION

According to Florentine (2011, pp. 4–5), loudness is the perceptual strength of a sound that ranges from very soft (or quiet) to very loud. The author noted that "most definitions of loudness are somewhat vague, but most people behave in a consistent manner when judging loudness". Loudness is known to depend on multiple factors such as the at-ear sound pressure level and the spectral content of the sound. For instance, the higher the sound pressure level is at the listener's ears, the greater its loudness generally is Stevens (1957). There is no absolute loudness value for a given sound. Rather, its assessment might vary from one listener to another, or even for the same listener during two different presentations of the sound (Algom and Marks, 1984) and depending on their mood (Siegel and Stefanucci, 2011). Loudness can also be assessed indirectly by measuring the reaction time to signal detection (Kohfeld et al., 1981), which appears to be a less subjective method but still exhibiting some variability (Schlittenlacher et al., 2014). Loudness can be estimated through models that analyze the physical properties of sounds in order to determine their typical loudness, i.e., the loudness value that would generally match the loudness values reported by a large group of human listeners (see Sivonen and Ellermeier, 2008; Moore, 2014, for examples of loudness models).

However, the link between the physical properties of a sound and the loudness experienced by the listener is not straightforward. Because loudness is a subjective experience, it depends on the way the listener interacts with the auditory object (a sound that can be assigned to a particular source following the definition of Bizley and Cohen, 2013). The environment and conditions in which the sound is presented to the listener are likely to affect the perception of this auditory object. As an example, when the source is identifiable, listeners may focus on the sound emitted by the latter (the distal stimulus) rather than on the signal reaching their ears (the proximal stimulus). This is likely to explain that loudness does not necessarily evolve in the same manner as what could be expected from variations of physical properties at the listener's ears (Zahorik and Wightman, 2001).

Listeners are able to focus on the proximal or distal stimulus when they are explicitly instructed to do so. Loudness can differ for the two cases. The assessments reported with these two distinct instructions have been described as "loudness at the ear" (Mershon et al., 1981) and "loudness at the source" (Sivonen and Ellermeier, 2011).

Loudness studies usually ask the participants to estimate the loudness of the sounds they hear without giving further specifications (see Sivonen and Ellermeier, 2006; Glasberg and Moore, 2010; Epstein and Florentine, 2012; Meunier et al., 2016). This can lead to the inter-individual variability inherent to loudness assessment (Algom and Marks, 1984; Siegel and Stefanucci, 2011; Schlittenlacher et al., 2014). By comparing the results found in the literature with different instructions and stimuli, this paper aims at understanding on what listeners focus when assessing loudness and how this focus affects their judgments. This might also help to understand differences observed in loudness assessments reported for studies that provide listeners with similar signals but with different presentation methods (Epstein and Florentine, 2009, 2012; Berthomieu et al., 2019a).

2. STIMULUS-DRIVEN FOCUS

The extent to which listeners focus on the proximal or distal stimulus while estimating loudness appears to depend on the stimulus itself. It will enable the listener to focus on its source if it contains enough information about the latter. If the stimulus does not include any information about its source, the listeners only focus on the proximal stimulus while estimating loudness. As an example, Stevens and Guirao (1962) asked their participants to estimate the loudness, softness, and apparent distance of noises and pure tones presented through headphones without any visual stimulus. Since no other information about the source was provided to the listeners, loudness, and distance estimates were solely dependent on the at-ear sound pressure level and varied inversely with each other.

2.1. Reverberation Cues

In reverberant environments, the direct-to-reverberant energy ratio (DRR) is an absolute distance cue (Mershon and King, 1975). This is mostly true in rooms that are sufficiently large that the reverberant energy is almost independent of sound source distance. As an example, Zahorik and Wightman (2001) measured a decrease of the diffuse reverberant energy of about 1 dB for each doubling of distance in a small auditorium with reverberation time RT_{60} of approximately 0.7 s. Since the direct energy decreases linearly with the square of distance, the difference between the direct energy and the reverberant energy is a direct cue to the source distance. Moreover, the reverberant energy is proportional to the energy delivered by the sound source and could be a direct cue to the latter. Thus, reverberant environments simultaneously provide the listener with distance and power information about the source. When listeners evaluate loudness, they might focus on the loudness of the distal stimulus by following two distinct approaches: directly focusing on the source power via the reverberant energy or combining the source distance perceived through the direct-to-reverberant energy ratio and the perceived level of the proximal stimulus. Zahorik and Wightman (2001) observed what they defined as loudness constancy (loudness remained constant despite physical changes in the stimulus) using noise bursts presented virtually at several distances from the listening point in the aforementioned environment. The stimuli were presented over headphones after being binaurally recorded in the environment and were thus not visible during the experiment. Listeners gave constant loudness estimates for sounds played at different distances by a source of constant power despite at-ear sound pressure level differences, in agreement with Altmann et al. (2013) who reported that reverberation cues are used to achieve constant loudness across distance. Zahorik and Wightman (2001) suggested the hypothesis that loudness constancy is not related to perceived distance on the basis of two arguments. Firstly, they asked the participants to verbally estimate the distance of the sound sources for which loudness constancy was observed and obtained discrepancies between the estimates and the actual distances. Nevertheless, such discrepancies could be accounted for by the distance assessment method (verbal report) which is reported to lead to systematic underestimation (Paquier et al., 2016) and to be less accurate than proprioceptive methods such as blind walking (Andre and Rogers, 2006). Secondly, loudness constancy was not observed at low source power levels, for which the reverberant field fell below the absolute threshold of hearing. However, the absence of a perceptible reverberant field might not only have removed the information about the power of the source, but also about its distance.

2.2. Timbral Cues

For stimuli such as speech or music, intrinsic information about the sound source can be conveyed through the sound timbre. Speech perception is not solely based on the extraction of simple physical parameters conveyed in the speech waveform (Moore, 2012). The perceived vocal effort of a speaker can give information about the source power (Rosenblum and Fowler, 1991), allowing the listeners to evaluate the strength of the emitted speech at the position of the speaker regardless of the level of the sound reaching their ears. Mohrmann (1939) asked listeners to adjust the output level of two sound sources positioned at different distances so that the two sources appeared to be equally loud. The sounds included speech, music, tones and noises, and the sources could be either visible or hidden. The results showed that the output levels set by the listeners were less dependent on the source distance for speech and music than for tones and noises. Thus, listeners focused more on the source for speech and music than for tones and noises, for which loudness estimates were more related to the strength of the sound reaching their ears. The output levels were also less dependent on the sources distances when the sources were visible. The distance cues provided by vision are reported to enhance the accuracy of distance judgements (Anderson and Zahorik, 2014) and are likely to help the participants to focus on the sound source by giving more accurate information about it, as discussed in the following subsection. Pollack (1952) and Warren (1973) asked their participants to compare the loudness of two sounds (that could be noises, pure tones, and speech) played at different levels. The results showed a weaker dependence of the loudness on the at-ear sound pressure level for speech than for noises and pure tones. Loudness comparisons for noises and pure tones were highly dependent on the level of the sounds reaching the listeners ears. The speech stimuli were always the same recording played at several levels. Thus, the at-ear sound pressure varied accordingly to the output level but the timbre was the same regardless of the output level. Since loudness estimates depended less on the at-ear sound pressure level, listeners might have taken into account the perceived invariant level of the original stimulus (whose constant strength was perceived via the vocal effort regardless of the at-ear sound pressure level) in their loudness estimates. Even though listener's focus was not explicitly driven on the distal or proximal stimulus, this focus was likely to have been more spontaneously put on the source for speech stimuli than for noises or tones.

Epstein and Florentine (2009) observed stronger loudness constancy in the binaural-to-monaural loudness ratio for speech than for pure tones, despite similar physical variations in the sound properties. Loudness estimates were gathered for pure tones and speech stimuli played to either one or both ears. Pure tones were perceived as significantly louder when presented binaurally than monaurally, in agreement with Fletcher and Munson (1933). The binaural-to-monaural loudness ratio was significantly smaller for speech stimuli. The intrinsic source information conveyed by speech could have led to the perception of an auditory object that naturally directs the focus toward the source, which strength might be acknowledged by the listeners to be independent on whether it is heard monaurally or binaurally (Culling and Dare, 2016).

2.3. Visual Cues

In a follow-up study using the same procedure as for their 2009 paper, Epstein and Florentine (2012) reported that binaural-to-monaural loudness ratio was significantly smaller for speech stimuli when the speaker was visible. Thus, visual cues might help the listeners to focus on the source. Rosenblum and Fowler (1991) gathered loudness estimates using graphic ratings. Videotapes of a speaker producing consonant-vowel utterances and of hand claps were presented to listeners, whose task was to adjust the position of a vertical slash mark on an horizontal line in a location that corresponded with their impression of loudness, with increasing loudness corresponding to increasing distance from the left end of the line. The auditory and visual stimuli were produced at four degrees of efforts, and could be presented with or without a discrepancy between the auditory and visual efforts. The loudness estimates were affected significantly by the effort apparent in the visual stimuli. Thus, listeners focused on the source thanks to non-auditory information while estimating loudness. Shigenaga (1965) asked listeners to adjust the output level of sources positioned at different distances so that they appeared to play sounds as loud as for a reference source positioned at a fixed distance. The sources were visible, in an environment with low reverberant energy (the experiment took place on a roof, with participants sitting on elevated chairs so that their heads were 3.3 m above the roof surface). The output powers of the sources adjusted this way were similar despite the at-ear sound pressure variations induced by the distance differences, showing loudness constancy with source distance.

Namba et al. (1997) gathered loudness ratings for car interior sounds presented with different videos filmed through a front car window. The videos showed different ways of driving (e.g., busy roads with a high amount of traffic or clear mountain areas), giving different information about how the car was running. The loudness ratings were highly dependent on the videos that were used. Videos of comfortable driving led to lower loudness. According to Menzel et al. (2008), the color of a car also has a small influence on its loudness for German listeners as the presentation of a red car produced higher loudness ratings compared to other colors. Suzuki et al. (2000) asked listeners to evaluate broadband noises that were difficult to identify with no visual information (such as the roaring of a waterfall). The noises were presented alone or with visual or verbal information about their source. The evaluations were made with pairs of verbal attributes. Based on the use of adjectives relative to loudness, such as powerful, loud, and noisy, the authors suggested that loudness was affected by the visual and verbal information provided about the sources. Berthomieu et al. (2019a) evaluated the directional loudness (i.e., the variation of loudness with the direction of the source) of narrow-band noise bursts in a sound-attenuated room. Loudness assessments were made using two experimental setups, one where the sounds were presented by visible loudspeakers, and one where the sounds were binaurally recorded and played through headphones, with no visual information about the sources. The loudness varied more with the source direction when the sounds were played through headphones (with no visual information about the sources) than when the sounds were played by the visible sources positioned around the listener. When no visual information about the source was available, estimates might have been made only with regard to the proximal stimulus. When information about the source was available through vision, listeners could have focused on the source and evaluated the distal stimulus.

3. INSTRUCTION-DRIVEN FOCUS

In some studies, the experimenters chose to explicitly drive the focus of the participants on the proximal or distal stimulus. These studies are rather sparse, but show a strong influence of the instructions on loudness.

The instructions given by Mohrmann (1939) that led to the aforementioned data were to adjust the output levels of the sources so that "the two sources—or else the two impacts—appeared to be equally loud," either based on "attitude toward loudness of sound emitted at the source" and "attitude toward loudness of impact at the ear" (as translated by Brunswik, 1956, p. 71). The adjustments made with the "attitude toward loudness of impact at the ear" were highly dependent on the sound source distance (and thus on the level of the proximal stimulus), which was not the case when the attitude was "toward loudness of sound emitted at the source."

The aforementioned data obtained by Zahorik and Wightman (2001, p. 83) were collected "using a free-modulus magnitude estimation procedure in which listeners were carefully instructed to make their judgments based on the sound source power." As described above, the loudness estimates gathered in this way showed that loudness did not vary with source distance. Listeners were able to take into account the power of the source, which was the same at every distance. When the reverberant energy fell under the absolute threshold of hearing, listeners could not focus on the source anymore and loudness estimates varied with the source distance.

Honda et al. (2019) asked participants to match the loudness of a target sound (2-s tones produced by an actual musical instrument performance at different distances from the listeners) by using two adjustment methods. They were either instructed to play a musical instrument (a melodica) as loudly as the target ("sound production") or to adjust the sound emitted by a loudspeaker so that it had the same loudness as the target ("sound level adjustment"). The loudness obtained through the sound production method depended less on the source distance than the sound level adjustment method, especially when visual cues about musical performance were available. This suggests that the sound production method combined with visual cues enabled the participants to focus on the source.

Rosenblum and Fowler (1991) gathered the aforementioned loudness estimates using graphic ratings (where listeners adjusted the position of a vertical slash mark on an horizontal line as described above). The sounds were consonant-vowel utterances and hand claps produced with different degrees of effort. They instructed their listeners to base their loudness judgments only on what they heard, despite the sound sources (the speaker or the person clapping their hands) being visible. Listeners were this way asked to focus on the sound only, but with no particular focus on the proximal or distal stimulus. Visual effort still affected loudness estimates, showing that listeners interacted with the audiovisual object despite being asked to focus on the sound only.

Listeners are nevertheless able to evaluate the loudness of the proximal stimulus, when instructed to do so, by ignoring the available information about the source. Berthomieu et al. (2019b) asked listeners to estimate the distance and loudness of sounds played at distances ranging from 1 to 16 m by both visible and hidden sound sources in both anechoic and reverberant environments. Listeners were explicitly instructed to report the apparent loudness of the sound reaching their ears using an absolute magnitude estimation. The perceived distance was estimated in meters. Loudness estimates depended on distance and thus on the at-ear sound pressure level. Moreover, no difference was observed between loudness estimates for visible and hidden sources (in either the anechoic or reverberant environment). Distance estimates were closer to the physical sound source distances for visible sources than for hidden sources. Thus, although visual cues provided the listeners with additional information about the sources that improved their distance estimates, loudness estimates were unchanged. Listeners might then have focused on the proximal stimulus whether or not the stimuli provided information about the source.

4. DISCUSSION

Since the definition of loudness itself is somewhat vague (Florentine, 2011) as the perceptual strength of the "sound," it may vary from one listener to another or from one experimental setup to another. Some experimenters have assumed to be more specific by focusing the listeners toward the sound emitted by the source (the distal stimulus) or reaching the ears (the proximal stimulus). This focus can be obtained from the stimulus through the information it conveys about its source (reverberation cues, visual cues, timbral cues) and from the related instructions. However, listeners might not be able to follow such instructions. As an example, even though the listeners are instructed to evaluate the loudness of the proximal stimulus, the judgments may still be influenced by available information about the source. Rosenblum and Fowler (1991) reported that listeners failed to focus on the proximal stimulus when provided with visual cues to the source in presentation conditions that could exhibit discrepancies between the visual and auditory stimuli.

Loudness experiments usually do not require the listeners to specifically focus on the proximal or distal stimulus. Rather, instructions are often free (e.g., assess the perceptual strength of the sound), for example in studies of directional loudness. Such studies show loudness variations according to the direction from which the sounds reach the listener (Sivonen and Ellermeier, 2006; Kopčo and Shinn-Cunningham, 2011; Koehl and Paquier, 2015; Meunier et al., 2016). Most of these variations are accounted for by physical binaural parameters such as the interaural time or level differences. However, significant individual differences were observed by Sivonen and Ellermeier (2006) and Meunier et al. (2016) and were hypothesized to be accounted for by different degrees of loudness constancy. Provided that the sources were visible in these experiments, some listeners would have assessed the (constant) loudness of the distal stimuli while others judged the proximal stimuli. This is supported by recent results (Berthomieu et al., 2020) that reported loudness constancy when explicitly asking the listeners to assess the loudness of the distal stimulus, but not when explicitly asking the listeners to assess the loudness of the proximal stimulus.

An example that highlights such a difference is a listener who evaluates the loudness of a siren played at different distances. If the listener is asked to focus on the sound reaching the ears, the estimates might strongly depend on the alarm distance since the latter induces at-ear sound pressure level variations of the stimulus on which they focus. If the listener is asked to focus on the source, the estimates might be constant with the alarm distance (and with the at-ear sound pressure level) since its timbre gives the listener an impression of the source power, which does not depend on the source distance. If the instructions do not ask to focus on the proximal or distal stimulus, listeners might focus on either while estimating loudness. If the siren is distant, the at-ear sound pressure level might be weak and the listener might assign a low loudness value to this stimulus. On the other hand, when the sound is recognized as a siren alarm—which is known by experience to be intense—the listener might assign a high loudness as this intensity is part of the source identity (Traer et al., 2020).

The way listeners focus on the sound when asked to evaluate loudness with no further specification is difficult to evaluate. Instructions might also be differently understood by various listener panels because of cultural differences. As an example, the loudness of passing-by train noises obtained with a magnitude estimation protocol appeared to be influenced by the train color for German and Japanese listeners (Patsouras et al., 2002; Rader et al., 2004), but not for French ones (Parizet and Koehl, 2011).

5. CONCLUSION

The results reviewed show that loudness assessments depend on what the listener focuses on when estimating loudness. According to the instructions they are given and to the quantity and quality of information provided about the sound source, loudness might relate to the strength of the sound emitted by the source (the distal stimulus) or received by the ears (the proximal stimulus). These two percepts do not depend on the physical attributes of the sound in the same way, and the listener's focus might vary from one listener to another in a same experiment. These observations could thus account for results in the literature according to which some parameters (sound pressure level, source position, monaural vs. binaural listening...) have a weaker effect on the loudness of sounds whose source is identifiable by the listener and where individual differences are observed.

AUTHOR CONTRIBUTIONS

This bibliographic review was part of GB's Ph.D. thesis focusing on the influence on source position on loudness. VK and MP supervised and directed GB's work during his Ph.D. thesis. All authors contributed to the article and approved the submitted version.

REFERENCES

Algom, D., and Marks, L. E. (1984). Individual differences in loudness processing and loudness scales. *J. Exp. Psychol. Gen.* 113, 571–593. doi: 10.1037/0096-3445.113.4.571

Altmann, C. F., Ono, K., Callan, A., Matsuhashi, M., Mima, T., and Fukuyama, H. (2013). Environmental reverberation affects processing of sound intensity in right temporal cortex. *Eur. J. Neurosci*. 38, 3210–3220. doi: 10.1111/ejn.12318

Anderson, P. W., and Zahorik, P. (2014). Auditory/visual distance estimation: accuracy and variability. *Front. Psychol.* 5:1097. doi: 10.3389/fpsyg.2014.01097

Andre, J., and Rogers, S. (2006). Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Percept. Psychophys*. 68, 353–361. doi: 10.3758/BF03193682

Berthomieu, G., Koehl, V., and Paquier, M. (2019a). Directional loudness of low-frequency noises actually presented over loudspeakers and virtually presented over headphones. *J. Audio Eng. Soc*. 67, 655–665. doi: 10.17743/jaes.2019.0018

Berthomieu, G., Koehl, V., and Paquier, M. (2019b). "Loudness and distance estimates for noise bursts coming from several distances with and without visual cues to their source," in *Proceedings of the 23rd International Congress on Acoustics, Integrating 4th EAA Euroregio 2019* (Aachen), 3897–3904.

Berthomieu, G., Koehl, V., and Paquier, M. (2020). "Loudness of speech pronounced by a visible or hidden speaker located at several distances," in *Proceedings of the 9th Forum Acusticum* (Lyon), 3045–3046.

Bizley, J. K., and Cohen, Y. E. (2013). The what, where and how of auditory-object perception. *Nat. Rev. Neurosci*. 14, 693–707. doi: 10.1038/nrn3565

Brunswik, E. (ed.). (1956). "Loudness constancy with distance variant," in *Perception and the Representative Design of Psychological Experiments, 2nd Edn* (Berkeley; Los Angeles, CA: University of California Press), 70–72.

Culling, J. F., and Dare, H. (2016). "Binaural loudness constancy," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing, Vol. 894*, eds P. van Dijk, D. Başkent, E. Gaudrain, E. de Kleine, A. Wagner, and C. Lanting (Cham: Springer International Publishing), 65–72. doi: 10.1007/978-3-319-25474-6_8

Epstein, M., and Florentine, M. (2009). Binaural loudness summation for speech and tones presented via earphones and loudspeakers. *Ear Hear*. 30, 234–237. doi: 10.1097/AUD.0b013e3181976993

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

Fletcher, H., and Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *J. Acoust. Soc. Am*. 5, 82–108. doi: 10.1121/1.1915637

Florentine, M. (2011). *Loudness*. New York, NY: Springer New York. doi: 10.1007/978-1-4419-6712-1_1

Glasberg, B. R., and Moore, B. C. J. (2010). The loudness of sounds whose spectra differ at the two ears. *J. Acoust. Soc. Am*. 127, 2433–2440. doi: 10.1121/1.3336775

Honda, A., Yasukouchi, A., and Sugita, Y. (2019). Sound production shows robust loudness constancy when visual cues of musical performance are available. *Psychol. Mus*. 47, 436–443. doi: 10.1177/0305735618755885

Koehl, V., and Paquier, M. (2015). Loudness of low-frequency pure tones lateralized by interaural time differences. *J. Acoust. Soc. Am*. 137, 1040–1043. doi: 10.1121/1.4906262

Kohfeld, D. L., Santee, J. L., and Wallace, N. D. (1981). Loudness and reaction time: I. *Percept. Psychophys*. 29, 535–549. doi: 10.3758/BF03207370

Kopčo, N., and Shinn-Cunningham, B. G. (2011). Effect of stimulus spectrum on distance perception for nearby sources. *J. Acoust. Soc. Am*. 130, 1530–1541. doi: 10.1121/1.3613705

Menzel, D., Fastl, H., Graf, R., and Hellbrück, J. (2008). Influence of vehicle color on loudness judgments. *J. Acoust. Soc. Am*. 123, 2477–2479. doi: 10.1121/1.2890747

Mershon, D. H., Desaulniers, D. H., Kiefer, S. A., Amerson, T. L., and Mills, J. T. (1981). Perceived loudness and visually-determined auditory distance. *Perception* 10, 531–543. doi: 10.1068/p100531

Mershon, D. H., and King, L. E. (1975). Intensity and reverberation as factors in the auditory perception of egocentric distance. *Percept. Psychophys*. 18, 409–415. doi: 10.3758/BF03204113

Meunier, S., Savel, S., Chatron, J., and Rabau, G. (2016). Interindividual differences in directional loudness. *J. Acoust. Soc. Am*. 140:3268. doi: 10.1121/1.4970368

Mohrmann, K. (1939). Lautheitskonstanz im entfernungswechsel. *Zeitsch. Psychol.* 145, 145–199.

Moore, B. C. J. (2012). "Speech perception," in *An Introduction to the Psychology of Hearing* (Leiden: BRILL), 299–332.

Moore, B. C. J. (ed.). (2014). Development and current status of the "Cambridge" loudness models. *Trends Hear*. 18:233121651455062. doi: 10.1177/2331216514550620

Namba, S., Kuwano, S., Kinoshita, A., and Hayakawa, Y. (1997). Psychological evaluation of noise in passenger cars–the effect of visual monitoring and the measurement of habituation. *J. Sound Vib*. 205, 427–433. doi: 10.1006/jsvi.1997.1008

Paquier, M., Côté, N., Devillers, F., and Koehl, V. (2016). Interaction between auditory and visual perceptions on distance estimations in a virtual environment. *Appl. Acoust*. 105, 186–199. doi: 10.1016/j.apacoust.2015.12.014

Parizet, E., and Koehl, V. (2011). Influence of train colour on loudness judgments. *Acta Acust*. 97, 347–349. doi: 10.3813/AAA.918414

Patsouras, C., Filippou, T., and Fastl, H. (2002). "Influences of color on the loudness judgement," in *Proceedings of the 3rd Forum Acusticum* (Sevilla).

Pollack, I. (1952). On the measurement of the loudness of speech. *J. Acoust. Soc. Am*. 24, 323–324. doi: 10.1121/1.1906900

Rader, T., Morinaga, M., Matsiu, T., Fastl, H., Kuwano, S., and Namba, S. (2004). "Crosscultural effects in audio-visual interactions," in *Proceedings of the Meeting of the Technical Committee on Noise and Vibration of the Acoustical Society of Japan* (Tokyo).

Rosenblum, L. D., and Fowler, C. A. (1991). Audiovisual investigation of the loudness-effort effect for speech and nonspeech events. *J. Exp. Psychol. Hum. Percept. Perform*. 17, 976–985. doi: 10.1037/0096-1523.17.4.976

Schlittenlacher, J., Ellermeier, W., and Arseneau, J. (2014). Binaural loudness gain measured by simple reaction time. *Atten. Percept. Psychophys*. 76, 1465–1472. doi: 10.3758/s13414-014-0651-1

Shigenaga, S. (1965). The constancy of loudness and of acoustic distance. *Bull. Faculty Lit. Kyushu Univ*. 9, 289–333.

Siegel, E. H., and Stefanucci, J. K. (2011). A little bit louder now: negative affect increases perceived loudness. *Emotion* 11:1006. doi: 10.1037/a0024590

Sivonen, V. P., and Ellermeier, W. (2006). Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. *J. Acoust. Soc. Am*. 119, 2965–2980. doi: 10.1121/1.2184268

Sivonen, V. P., and Ellermeier, W. (2008). Binaural loudness for artificial-head measurements in directional sound fields. *J. Audio Eng. Soc*. 56, 452–461.

Sivonen, V. P., and Ellermeier, W. (2011). "Binaural loudness," in *Loudness*, eds M. Florentine, A. N. Popper, and R. R. Fay (New York, NY: Springer New York), 169–197. doi: 10.1007/978-1-4419-6712-1_7

Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev*. 64, 153–181. doi: 10.1037/h0046162

Stevens, S. S., and Guirao, M. (1962). Loudness, reciprocality, and partition scales. *J. Acoust. Soc. Am*. 34, 1466–1471. doi: 10.1121/1.1918370

Suzuki, Y., Abe, K., Ozawa, K., and Sone, T. (2000). "Factors for perceiving sound environments and the effects of visual and verbal information on these factors," in *Contributions to Psychological Acoustics* (Oldenburg), 209–232.

Traer, J., Norman-Haignere, S. V., and McDermott, J. H. (2020). Causal inference in environmental sound recognition. *bioRxiv*. doi: 10.1101/2020.07.13.200949

Warren, R. M. (1973). Anomalous loudness function for speech. *J. Acoust. Soc. Am*. 54, 390–396. doi: 10.1121/1.1913590

Zahorik, P., and Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nat. Neurosci*. 4, 78–83. doi: 10.1038/82931

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.

ORIGINAL RESEARCH
published: 19 March 2021
doi: 10.3389/fpsyg.2021.588571

Temporal Loudness Weights Are Frequency Specific

Alexander Fischenich[1]*, *Jan Hots*[2], *Jesko Verhey*[2] *and Daniel Oberfeld*[1]*

[1] Department of Psychology, Johannes Gutenberg-Universität Mainz, Mainz, Germany, [2] Department of Experimental Audiology, Otto von Guericke University Magdeburg, Magdeburg, Germany

Previous work showed that the beginning of a sound is more important for the perception of loudness than later parts. When a short silent gap of sufficient duration is inserted into a sound, this primacy effect reoccurs in the second sound part after the gap. The present study investigates whether this temporal weighting occurs independently for different frequency bands. Sounds consisting of two bandpass noises were presented in four different conditions: (1) a simultaneous gap in both bands, (2) a gap in only the lower frequency band, (3) a gap in only the higher frequency band, or (4) no gap. In all conditions, the temporal loudness weights showed a primacy effect at sound onset. For the frequency bands without a gap, the temporal weights decreased gradually across time, regardless of whether the other frequency band did or did not contain a gap. When a frequency band contained a gap, the weight at the onset of this band after the gap was increased. This reoccurrence of the primacy effect following the gap was again largely independent of whether or not the other band contained a gap. Thus, the results indicate that the temporal loudness weights are frequency specific.

Keywords: loudness, frequency specific, intensity discrimination, temporal weights, auditory

OPEN ACCESS

Edited by:
Sabine Meunier,
Laboratoire de Mécanique et d'Acoustique – CNRS – Aix-Marseille University – Ecole Centrale Marseille, France

Reviewed by:
Josef Schlittenlacher,
The University of Manchester, United Kingdom
Léo Varnet,
Ecole Normale Supérieure, PSL Research University, France

Correspondence:
Alexander Fischenich
alexander.fischenich@uni-mainz.de
Daniel Oberfeld
oberfeld@uni-mainz.de

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

Received: *29 July 2020*
Accepted: *24 February 2021*
Published: *19 March 2021*

Citation:
Fischenich A, Hots J, Verhey J and Oberfeld D (2021) Temporal Loudness Weights Are Frequency Specific. Front. Psychol. 12:588571. doi: 10.3389/fpsyg.2021.588571

INTRODUCTION

In loudness judgments of time-varying sounds, higher perceptual weights are assigned to the first few hundred milliseconds of a sound compared to later temporal portions (e.g., Namba et al., 1976; Ellermeier and Schrödl, 2000; Plank, 2005; Pedersen and Ellermeier, 2008; Dittrich and Oberfeld, 2009; Rennies and Verhey, 2009; Ponsot et al., 2013). This primacy effect can be described by an exponential decay function with a time constant of about 300 ms (Hots et al., 2018; Oberfeld et al., 2018; Fischenich et al., 2019). The temporal weighting was reported to be largely independent of the spectral weighting (Oberfeld et al., 2012). Pedersen and Ellermeier (2008) showed that when the spectrum changes abruptly within a contiguous sound, a second primacy effect is observed on the second sound part. In a recent study, Fischenich et al. (2020) reported that such a reoccurrence of the primacy effect is also obtained when a silent gap is inserted into the sound. Their data showed that after a gap of at least 350 ms, a significant primacy effect reoccurred on the second sound part. The initial primacy effect on the first temporal segments of the sound was reduced, and at the onset of the sound after the silent gap, the weights on the first two to three segments (segment duration 100 ms) following the gap were increased relative to the weights assigned to the subsequent segments. This primacy effect on the second sound part became more pronounced when the gap duration was further increased.

In the present study, we investigated whether the effects of a silent gap inserted into a sound on the temporal loudness weights occur specifically for each presented frequency band, or if a gap in one of the frequency bands affects the temporal weights for the entire sound. Put differently, are temporal loudness weights frequency specific? The answer to this question is, among others, important for modeling temporal loudness weights. In a model, one could apply temporal weights independently for each auditory filter (i.e., before spectral integration of loudness), or to the sound as a whole (i.e., after spectral integration). How temporal weights are assigned is also important for the understanding of the everyday sound perception. In our acoustic environment, in many situations background sounds produced by other people, animals, technical devices, or weather phenomena are present most of the time. In this context, it is an interesting question whether different spectral components interact with each other in terms of the temporal loudness weights when the overall loudness of a time-varying sound is judged.

In most previous experiments on temporal loudness weights, a broadband noise (Pedersen and Ellermeier, 2008; Dittrich and Oberfeld, 2009; Oberfeld et al., 2018), a single narrowband noise band (Rennies and Verhey, 2009; Fischenich et al., 2019) or a pure tone (Ponsot et al., 2013) was presented. The corresponding data thus do not provide an answer to the question of whether the temporal weights are assigned on a frequency-specific basis.

Studies on temporal loudness weights (Oberfeld and Plank, 2011; Oberfeld, 2015; Fischenich et al., 2019) discussed the possibility that the temporal weighting pattern observed for loudness judgments of time varying sounds is caused by the response characteristics of auditory nerve (AN) fibers. AN fibers show an initial peak in their firing rate at the onset of a sound (Kiang et al., 1965). With a preceding masker a pronounced peak also occurs at the onset of a sound after a certain silent interval. The necessary duration of that silent interval for a pronounced peak to occur varies between different fibers (e.g., Rhode and Smith, 1985; Relkin and Doucet, 1991). As the inner hair cells that innervate the AN fibers are frequency specific, the recovery of the firing rate is also frequency specific (e.g., Harris and Dallos, 1979) and thus would support the assumption of frequency-specific temporal weights.

Another potential source of the temporal weighting patterns, which also predicts frequency-specific temporal weights, are forward-masking effects on the intensity resolution (e.g., Zeng et al., 1991). Such forward-masking effects might result in higher intensity resolution for the first few temporal segments of a longer sound compared to later segments. In a loudness judgment task, this could induce a strategy of attending primarily to the beginning of the sound where the intensity resolution is higher (for an in-depth discussion see Fischenich et al., 2020). Because Zeng and Turner (1992) found that maskers with frequency components two to three octaves away from the signal frequency did not affect the intensity resolution for the signal, this explanation for the primacy effect would also predict frequency-specific weights.

In contrast, a potential argument for an *interaction* of the temporal weights across frequency is that the bands may suppress each other. Two-tone suppression is observed over large spectral distances (e.g., Houtgast, 1974; Ernst et al., 2010). During a silent gap in one of the bands, the other band is no longer suppressed, and thus the auditory fibers encoding this frequency range might be more strongly activated during the gap in the other band compared to those positions in time where both bands are presented simultaneously. In perceptual terms, the loudness of a given frequency band could be reduced by suppression caused by a simultaneously presented band. In such a case, the loudness of the ongoing band should be higher during the gap in the other band. A phenomenon that could cause a change in the temporal weighting patterns in such a situation is *loudness dominance* (Berg, 1990), which has been shown in several studies on temporal loudness weights (e.g., Lutfi and Jesteadt, 2006; Oberfeld, 2008a, 2015; Ponsot et al., 2013). Loudness dominance describes the effect that temporal portions of a sound that are, on average, higher or lower in level or loudness compared to the rest of the sound receive higher or lower weights, respectively. If the effect of loudness dominance precedes spectral loudness summation, then the release from suppression during the gap in the other band might render the segments of the band that contains no gap presented during the gap in the other band *louder* than the segments presented simultaneously with the other band. In this case, higher weights on the temporal segments of the band that does not contain a gap can be expected during the gap in the other band.

In contrast, if loudness dominance takes place *after* spectral loudness summation, one may expect the opposite pattern. During the gap in one of the bands, the *overall* loudness of the sound (across frequencies) is *reduced*. Thus, the loudness dominance effect would cause the weights for the temporal segments of the band that did not contain a gap to be *reduced* during the duration of the gap in the other band. After the gap, when both bands are presented again, the weights on the band that contained no gap should increase again because the overall loudness of the sound increases.

To answer the question of whether temporal loudness weights are frequency specific or not, the present study used stimuli consisting of two frequency bands that were separated by more than two critical bands in order to minimize simultaneous masking. The two frequency bands were presented in four conditions. Either none of the bands, only the lower band, only the higher band, or both bands contained a silent gap in the temporal center of the sound. Using a behavioral reverse-correlation approach, temporal perceptual weights were measured for each of the two frequency bands. To this end, independent level variations were imposed on temporal segments in the two bands (Oberfeld et al., 2012). The rationale was that if the temporal weights are frequency specific, then the temporal weights on a given frequency band should not be affected by the presence or absence of a temporal gap in the other frequency band.

The study was organized into two experiments. Experiment 1 presented a gap duration of 500 ms. Experiment 2 was conducted to replicate the findings of Experiment 1 in an independent group of participants. Also, we presented slightly longer gap duration of 700 ms compared to the 500 ms in Experiment 1. Two gap durations were used to assess potential differences in the pattern of the weights due to the gap duration. Such differences were observed in previous work on temporal loudness weights of sounds including a temporal gap (Fischenich et al., 2020).

EXPERIMENT 1

Method

Listeners

Eight normal hearing listeners (four female, four male, age 18–29 years) participated in this experiment. Their hearing thresholds were measured by Békésy audiometry with pulsed 270-ms pure tones and were lower than or equal to 15 dB HL on both ears in the frequency range between 125 Hz and 8 kHz. All participants were students from the Johannes Gutenberg-Universität Mainz and received partial course credit for their participation. The experiment was conducted according to the principles expressed in the Declaration of Helsinki. All listeners participated voluntarily and provided informed written consent, after the topic of the study and potential risks had been explained to them. They were uninformed about the experimental hypotheses. The Ethics Committee of the Institute of Psychology of the Johannes Gutenberg-Universität Mainz approved the study (reference number 2016-JGU-psychEK-002).

Stimuli and Apparatus

The stimuli consisted of two level-fluctuating noise bands, each comprising 10 or 15 bandpass-filtered temporal noise segments. The total number of segments depended on the condition, as outlined below. To reduce the intrinsic envelope fluctuations of the noise within a segment, low-noise noise was used (Hartmann and Pumplin, 1988). The present study generated low-noise noise using the first method of Kohlrausch et al. (1997) with two iterations. To generate low-noise noise, first, a Gaussian white noise was generated and filtered with a fast Fourier transform (FFT) based bandpass filter. The amplitudes of all frequency components outside the desired frequency range were set to zero. The cutoff frequencies were 200 Hz (2 Bark) and 510 Hz (5 Bark) for the lower noise band (referred to as LB), and 3150 Hz (16 Bark) and 5300 Hz (19 Bark) for the higher noise band (referred to as HB). Second, the following steps were iterated two times: (i) The Hilbert envelope was calculated, (ii) the stimulus was divided by its Hilbert envelope, and (iii) it was filtered using the FFT-based bandpass filtering, as described above. For each temporal segment, a new random Gaussian noise was generated, and the signal processing steps described above were applied to it. Each noise segment had a duration of 120 ms including 20-ms $\cos^2$ ramps at segment on- and offset. Contiguous segments were presented, with a temporal overlap of 20 ms. Random level fluctuations were created by assigning a sound pressure level drawn independently and at random from a normal distribution to each temporal segment on each trial (see section "Procedure" for details).

All sounds were generated digitally with a sampling frequency of 44.1 kHz and a resolution of 24 bit, D/A-converted by an RME ADI/S, attenuated by a TDT PA5 programmable attenuator, buffered by a TDT HB7 headphone buffer, and presented diotically via Sennheiser HDA 200 circumaural headphones. The reproducing system was calibrated according to IEC 318 (1970), and free-field equalized as specified in ISO 389-8 (2017). Participants were tested in a double-walled sound-insulated chamber. Instructions were presented on a computer screen.

Experimental Conditions

The two noise bands were presented simultaneously in four different conditions, which are displayed in **Figure 1**. In the first condition, each of the two noise bands consisted of 15 contiguous segments. This condition is referred to as LB_0HB_0, where 0 indicates a gap duration of 0 ms (no gap). In the second condition, referred to as condition $LB_{500}HB_0$, a gap of 500 ms was inserted between segments 5 and 6 of LB, while no gap was presented in HB, so that the latter noise band contained 15 contiguous temporal segments. In the third condition (LB_0HB_{500}), a gap was inserted in the middle of HB whereas LB did not contain a gap. Finally, in the fourth condition ($LB_{500}HB_{500}$), both LB and HB were presented with a gap of 500 ms duration.

Procedure

To estimate temporal loudness weights, we used an established experimental paradigm from previous experiments (e.g., Pedersen and Ellermeier, 2008; Oberfeld and Plank, 2011). On each trial, the two noise bands were presented. Depending on the experimental condition (see **Figure 1**), each noise band consisted of either 10 or 15 100-ms segments. For each trial, the segment levels of both bands were drawn independently and at random from a truncated normal distribution. With equal probability and uniformly for both bands, either a level distribution with a lower mean or a distribution with a higher mean was selected on each trial. The main aim of the introduction of two different mean levels was to adjust the difficulty of the task and to motivate the listeners by giving feedback about the "correctness" of their response. The level difference between the two distribution means was selected so that the listeners were able to respond with roughly 70% correct.

For LB, the level distribution with higher mean had a mean level of μ_{H_low} = 52.75 dB SPL and the distribution with lower mean had a mean level of μ_{L_low} = 51.25 dB SPL. In an initial session, HB was loudness-matched to LB for each listener in an adaptive two-interval forced-choice procedure (see **Supplementary Material** "Loudness matching" for details of the matching procedure). This was done to eliminate the effect of "loudness dominance," i.e., the effect that stimulus components with on average higher loudness receive higher weights (e.g., Berg, 1990; Oberfeld, 2008c; Oberfeld and Plank, 2011; Oberfeld et al., 2013). Averaged across the eight listeners, the level difference between HB and LB at equal loudness was −0.27 dB (*SD* = 4.11 dB) and the resulting mean sound pressure levels of HB were μ_{H_high} = 52.48 dB SPL and μ_{L_high} = 50.98 dB SPL. The individual sound pressure level differences between HB and LB are displayed in **Supplementary Table 1**. In the final session, loudness matches were obtained again for each listener, to assess if the matches remained stable across time. The test–retest reliability was high, indicating adequate stability across time (see **Supplementary Material** "Loudness matching" for information on stability).

The standard deviation of all level distributions was $\sigma = 2.5$ dB. Overly loud or soft segments were avoided by limiting the range of possible sound pressure levels to $\mu \pm 3 \cdot \sigma$. On

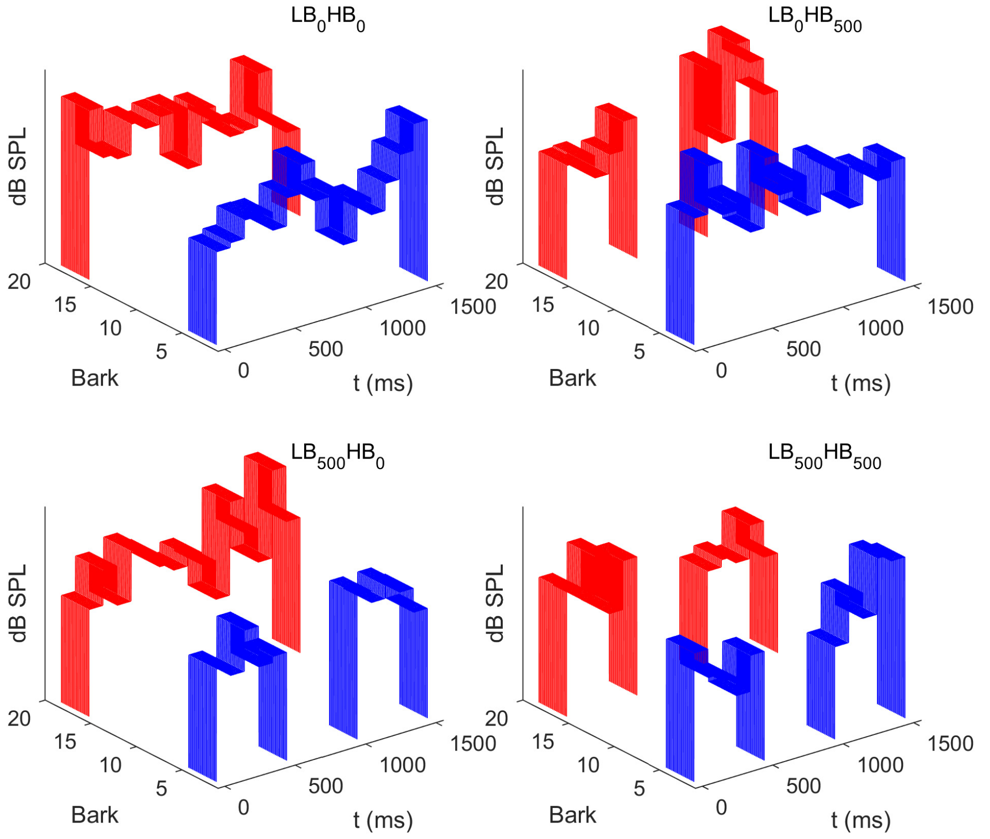

FIGURE 1 | Schematic spectrograms of the level fluctuating sounds presented in the four different conditions of Experiment 1. Each sound comprised two frequency bands (blue: lower band, LB, 2–5 Bark; red: higher band, HB, 16–19 Bark). Independent random level fluctuations were imposed on each of the 100-ms segments. In the example displayed here the same distribution mean was used for the higher and the lower band. In contrast, in the experiment, these means differed between the two bands as the HB was loudness matched to the LB (see section "Procedure"). Depending on the condition, a 500-ms silent gap was inserted in none of the frequency bands (LB_0HB_0), in only the higher band (LB_0HB_{500}), in only the lower band ($LB_{500}HB_0$), or in both bands ($LB_{500}HB_{500}$).

each trial, participants judged the overall loudness (i.e., the loudness of both frequency bands across the entire stimulus duration, encompassing potential silent temporal gaps) by deciding whether the presented sound had been louder or softer in comparison to previous trials within the same experimental block. Thus, a one-interval, *absolute identification task* (Braida and Durlach, 1972) with a virtual standard (e.g., Nachmias, 2006) was used.

The minimum silent interval between trials was 1500 ms. The next trial never started before the response to the preceding trial had been given. Trial-by-trial feedback was provided during the first seven trials of each block so that the participants could easily adopt a decision criterion for the new experimental condition. Those trials were not considered for the data analysis. A summarizing feedback was provided each time 50 trials were completed. It contained the number of correct and false answers, percent correct and the number of μ_H and μ_L trials as well as the number of "louder" and "softer" responses. Note that a response was classified as "correct" if the response ("louder"/"softer") matched the mean of the distribution that the stimulus' segment levels were drawn from (μ_H/μ_L).

To obtain a sufficient number of observations for the weight estimation, we presented 80 trials per temporal segment (cf. Oberfeld et al., 2018). As there were four different conditions in which the number of the temporal segments varied between a total of 20 (condition $LB_{500}HB_{500}$), 25 (conditions $LB_{500}HB_0$ and LB_0HB_{500}) and 30 segments (condition LB_0HB_0), we presented 1600, 2000, and 2400 trials per condition, amounting to a total of 8000 trials per listener.

Sessions

Each listener participated in nine experimental sessions, each containing 1000 trials of the loudness judgment task (300 for condition LB_0HB_0, 250 for $LB_{500}HB_0$, 250 for LB_0HB_{500}, and 200 for $LB_{500}HB_{500}$). Additionally, there was an initial session in which audiometric thresholds were measured, loudness matches were obtained, and practice blocks of the loudness judgment task were presented for all of the four conditions. The practice blocks were excluded from the data analysis. Within each session, sounds of the same condition were arranged into blocks with the above mentioned trial numbers. The order of conditions was chosen randomly. At the end of the final session, a second set of loudness matches (i.e., loudness matching of

HB to LB; see **Supplementary Material** "Loudness matching" for details) was obtained from each listener. The duration of each session was approximately 60 min, including a mandatory pause of about 5 min.

Data Analysis

The perceptual weights representing the importance of the 10–15 temporal segments of both bands for the decision in the loudness judgment task were estimated from the trial-by-trial data via multiple logistic regression. The decision model assumed that the listener compares a weighted sum of the segment levels of both bands to a fixed decision criterion, and responds that the sound was of the "louder" type if the weighted sum exceeds the criterion (a detailed description of the decision model is provided by Oberfeld and Plank, 2011). If the weighted sum is smaller than the criterion, then it is assumed that the listener classifies the sound as "softer." In the data analysis, the binary responses ("louder" or "softer") served as the dependent variable. The predictors (i.e., the 20, 25, or 30 segment levels) were entered simultaneously. The regression coefficients were taken as the decision weight estimates. Because the segment levels were drawn independently for each frequency band, this allowed for the detection of possible interactions between the bands in the observed temporal weights, especially in situations in which one band did contain a gap while the other one was contiguous.

A separate logistic regression model was fitted for each combination of listener and condition. The model included an intercept term so that potential biases toward one of the two responses were accounted for. The percentages of "softer" and "louder" responses as well as the SDT decision criterion *c* and the sensitivity in terms of d' is shown in **Table 1** for each listener in Experiment 1. A value of $c = 0$ represents unbiased responses. In general, the responses of the participants did not show strong response biases. As stated in the Methods section, we presented seven trials with trial-by-trial feedback at the beginning of each experimental block, so that the participants could easily adopt a decision criterion for the new experimental condition. Those trials were not considered for the data analysis. We assume that the decision criterion remained relatively stable across the remaining trials of the block. Still, it is of course possible that the listeners used information from preceding trials when forming their decision (Stewart et al., 2005), resulting in potential small shifts in the response criterion from trial to trial. Such a variability in the response criterion would reduce the goodness of fit of the logistic regression models that assumed a fixed response criterion. However, since the *relative* contributions of the different segments to the decision were of interest, rather than the absolute magnitude of the regression coefficients, this was of no significance for the research question of the present paper.

To focus the analyses on the *relative* contributions of the different segments to the decision, the regression coefficients for each frequency band were normalized so that the mean of the absolute values of the first five and the final five segments was 1.0. Thus, for each frequency band, exactly 10 segments contributed to the computation of the normalization factor in both the conditions with and without a gap, and the five middle weights in the conditions without a gap were not included in the normalization. This was done in order to avoid that the additional five middle segments presented in conditions without a gap lead to a different scaling of the weights compared to the conditions with a gap. The normalization per frequency band was done to compare the weights assigned to a specific band between the different conditions, independent of the weights assigned to the other band. We also conducted all analyses reported within this study for a normalization of the weights based on the mean of the absolute values of the first five segments for each frequency band. This kind of normalization was suggested by a reviewer and led to almost the same pattern of results as the normalization which was used within this study (see **Supplementary Material** "alternative normalization" for detailed plots and analyses).

TABLE 1 | Average percentages of "softer" and "louder" responses as well as the SDT decision criterion *c* and the sensitivity in terms of d' for each listener in Experiment 1.

Listener	% "louder"	% "softer"	Mean of c	SD of *c*	Mean of *d′*	SD of *d′*
1	0.54	0.46	−0.12	0.26	1.00	0.20
2	0.44	0.56	0.18	0.16	0.84	0.28
3	0.56	0.44	−0.17	0.23	0.77	0.27
4	0.47	0.53	0.10	0.23	1.15	0.22
5	0.59	0.41	−0.25	0.25	0.95	0.28
6	0.48	0.52	0.04	0.21	1.12	0.21
7	0.48	0.52	0.07	0.19	1.45	0.17
8	0.50	0.50	−0.01	0.13	1.27	0.18

Due to the sampling of all segment levels from either the distribution with higher or the distribution with lower mean, the segments levels were weakly correlated. Across all experiments reported in this paper, the maximum pairwise Pearson correlation between two segments levels was $r = 0.12$ (average *r* across listeners = 0.08). Multiple logistic regression analyses do not require the predictors to be uncorrelated. According to the Gauss–Markov theorem (Gauss, 1821), the estimated regression parameters from a (generalized) linear model are still unbiased when the predictors are correlated. We checked the validity of this assumption by fitting separate multiple logistic regression models to trials with segment levels sampled from the distribution with higher or the distribution with lower mean, for each combination of listener and condition. The averages of the normalized segment weights across the two level distributions per listener and condition were virtually identical (adjusted $R^2 \geq 0.975$) to the normalized weights estimated by fitting the logistic models to the trials from both level distributions simultaneously.

A summary measure of the predictive power of a logistic regression model is the area under the Receiver Operating Characteristic (ROC) curve (for details see Dittrich and Oberfeld, 2009). Areas of 0.5 and 1.0 correspond to chance performance and perfect performance of the model, respectively. Across the 32 (eight listeners, four conditions) fitted logistic regression models, the area under the ROC curve ranged between 0.70 and 0.88 ($M = 0.80$, $SD = 0.05$), indicating on average reasonably good predictive power (Hosmer and Lemeshow, 2000).

The individual normalized temporal weights were analyzed with repeated-measures analyses of variance (rmANOVAs) using a univariate approach with Huynh–Feldt correction for the degrees of freedom (Huynh and Feldt, 1976). The correction factor $\tilde{\varepsilon}$ is reported, and partial η^2 is reported as measure of association strength. An α-level of 0.05 was used for all analyses. If not stated otherwise, calculations were done with R 3.6.1 and R Studio 1.2.1335.

Results

The average sensitivity in terms of d' is shown in **Table 2** for each of the four conditions. There was a significant effect of condition on d', $F(3,21) = 5.19$, $\tilde{\varepsilon} = 0.701$, $p = 0.019$, $\eta_p^2 = 0.426$, with slightly higher mean sensitivity when both bands contained a gap ($LB_{500}HB_{500}$), and slightly lower sensitivity when none of the bands contained a gap (LB_0HB_0).

Figure 2 shows the mean normalized temporal weights assigned to the two frequency bands. Filled circles and open squares represent conditions where the plotted band did or did not contain a gap, respectively. For each of the plotted lines, the weights are averaged across the spectral context, that is, across the two conditions where the other frequency band either did or did not contain a gap. For both frequency bands, the patterns of the mean weights in both conditions (with and without a gap) showed a clear primacy effect at the beginning of the sound, in the sense that the weight on the first segment was higher than the weights on the following segments.

When a band contained a gap, the weight assigned to the first segment after the gap was higher compared to the condition in which the band did not contain a gap. Note that, in addition to this reoccurrence of the primacy effect after the silent gap, the primacy effect at the beginning of the sound was reduced when the band contained a silent gap. To investigate whether descriptive differences in the patterns of temporal weights can be explained by the stimulus properties in this condition (e.g., the frequency band that is concerned, whether the band contains a gap, or whether the other band contains a gap) one always has to compare the temporal weights for a given condition to a suitable control condition (e.g., HB without a gap vs. HB with a gap). This is necessary because, for example, even without a gap a difference between the segments weights is expected for the segments following the gap region as the weights tend to decline as a function of segment number/temporal onset even for later segments. The normalized temporal weights were analyzed with an rmANOVA with the within-subjects factors segment number (1–10 when the band contained a gap, 1–5 and 11–15 when the band did not contain a gap), target frequency band (LB, HB), target gap (no gap, 500-ms gap), and context (other band with 500-ms gap, other band without a gap). The rmANOVA showed a significant main effect of segment number, $F(9,63) = 40.06$, $\tilde{\varepsilon} = 0.434$, $p < 0.001$, $\eta_p^2 = 0.851$, highlighting the non-uniform temporal weighting patterns. The target gap × segment number interaction was also significant, $F(9,63) = 5.24$, $\tilde{\varepsilon} = 0.933$, $p < 0.001$, $\eta_p^2 = 0.429$, indicating that the pattern of the weights of the segments differed depending on whether a band was presented with or without a gap. This supports the observation that for the bands that contained a gap (filled circles in **Figure 2**), the weights assigned to the first segments following the gap were higher than the weights assigned to the same temporal positions when a band did not contain a gap (open squares in both panels of **Figure 2**). Thus, as expected, we observed a significant reoccurrence of the primacy effect.

TABLE 2 | Mean sensitivity (d') in the four different conditions of Experiment 1.

Condition	Mean of d'	SD of d'
LB_0HB_0	0.99	0.25
LB_0HB_{500}	1.09	0.25
$LB_{500}HB_0$	1.08	0.21
$LB_{500}HB_{500}$	1.17	0.13

N = 8.

The primary aim of our experiment was to test whether the temporal weights for a given frequency band are unaffected by the presence or absence of a gap in the other frequency band, and thus are frequency specific. Each panel in **Figure 3** shows the normalized weights for one band and depending on whether or not the plotted band did or did not contain a silent gap. The two lines shown in each panel represent the two spectral context conditions, that is, the presence or absence of a gap in the other frequency band. To answer the question whether the weights in one band are affected by the presence (or absence) of a gap in the other band, one has to compare the two lines in each panel of **Figure 3**. For example, the filled symbols in the left upper panel represent the weights observed for the higher band without gap, in the condition where the lower band also did not contain a gap (LB_0HB_0; see **Figure 1**). The open symbols represent the weights observed for the HB without gap, but this time in the condition where the LB contained a 500-ms gap ($LB_{500}HB_0$). Except for the segment with onset at 600 ms after sound onset, the weights in the two conditions were very similar. Thus, the temporal weights assigned to the higher band without gap were hardly affected by the temporal structure (with or without gap) of the other band. The same trend can be observed for the lower band without gap (left lower panel), for the lower band with gap (right lower panel), and, to a limited extent, also for the higher band with gap (right upper panel).

A first indicator that the temporal weights were frequency specific is that in the rmANOVA reported above, there were no significant interactions of the factor context (presence or absence of a gap in the other band) with segment number [$F(9,63) = 1.15$, $\tilde{\varepsilon} = 0.945$, $p = 0.347$, $\eta_p^2 = 0.141$], segment number and target gap [$F(9,63) = 1.58$, $\tilde{\varepsilon} = 0.952$, $p = 0.144$, $\eta_p^2 = 0.185$], or segment number, target gap and target frequency band [$F(9,63) = 1.22$, $\tilde{\varepsilon} = 1$, $p = 0.297$, $\eta_p^2 = 0.149$]. Thus, the temporal weights in a given band were not strongly affected by the presence or absence of gap in the other band. However, as discussed in the introduction, there are both arguments for expecting frequency-specific as well as for expecting frequency-unspecific temporal weights. For this reason, we conducted separate Bayesian rmANOVAs that quantify the relative evidence for both variants for the weights displayed in each panel of **Figure 3**, using the software JASP (JASP Team, 2019). These analyses encompass all potential effects that might occur if the weights were somehow dependent

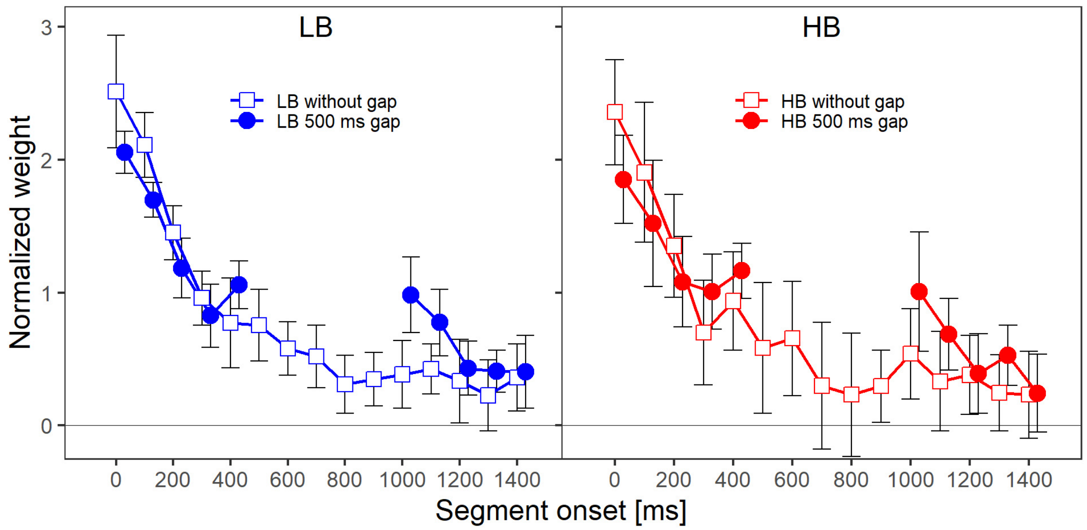

FIGURE 2 | Mean normalized weights as a function of segment onset for Experiment 1, averaged across spectral context. The two panels show the weights for the two frequency bands, lower band (LB, **left panel**) and higher band (HB, **right panel**). The frequency band is also indicated by the colors that were introduced in **Figure 1** (LB: blue, HB: red). The different symbols and separate lines within each panel indicate whether the band did or did not contain a gap (open squares: without gap, filled circles: 500-ms gap). Error bars show 95% confidence intervals (CIs). Note that for better visibility, the two lines are shifted slightly against each other along the *x*-axis.

between bands, as the analysis looks for any effect of context on the pattern of weights. In these analyses, we focused on the segments within and adjacent to the region of the gap as possible differences in the weighting patterns that would allow to differentiate between the two hypothesis were most likely to happen there. For each of these Bayesian rmANOVAs, the within-subjects factors were segment number (5–11 for bands without a gap, 5–6 for bands that contained a gap) and context (other band without a gap, other band with 500-ms gap). To quantify the influence of spectral context on the segment weights, we compared the posterior likelihood of the complete model that contained both main factors (segment number and context) and their interaction (segment number × context), to the posterior likelihood of a reduced model that included only the main factors segment number and context. The reduced model assumes no effect of spectral context on the temporal weights (that is, no segment × context interaction; H_0), while the full model assumes that the spectral context affects the weights (segment × context interaction; H_1). The scale parameter of the Cauchy prior distribution was set to commonly used values, i.e., $r = 0.5$ for fixed effects and $r = 1.0$ for random effects (for details on multivariate priors for Bayes factors see Rouder et al., 2012). We computed Bayes factors defined as the ratio between the posterior probability that the data occurred under H_0 (model without the segment × context interaction) and the posterior probability that the data occurred under H_1 (model including the segment × context interaction). Values of this Bayes factor (termed BF_{01} in the following) greater than 1.0 represent evidence in favor of the reduced model. For all four Bayesian rmANOVAs, the BF_{01} values were in favor of the reduced model not containing the interaction term, ranging from 2.08 (panel B) to 4.90 (panel C). This means that, for example, the patterns of weights displayed in panel B were 2.08 times more likely to occur under the null hypothesis of no segment × context interaction compared to the alternative hypothesis. According to the categories of Jeffreys (1961), there was thus anecdotal to moderate evidence for the null hypothesis that within each panel the segment weights within and around the region of the gap, depended only on the segments' temporal position, but not on the context (i.e., on whether the other frequency band was presented with or without a gap). To assess the robustness of the Bayes factors, we changed the width of the prior distribution for fixed effects over a range from 0.15 to 1.5. The resulting BF_{01} values are plotted in **Figure 4**. The changes in prior width did not affect the direction of the stated results. However, the size of the factors showed substantial variation ranging from 1.22 (prior width $r = 0.15$, panel A) to 89.11 (prior width $r = 1.5$, panel C). Taken together, the direction of the results indicates that the weights for both bands were hardly affected by the presence or absence of a gap in the other band.

In summary, Experiment 1 provides two main findings. First, it confirms previous data showing a reoccurrence of the primacy effect on the second sound part of a frequency band when this band contained a gap (Fischenich et al., 2020). In conditions where a band contained a gap, the weight assigned to the first segment of that band following the gap was higher than the weight assigned to the same segment when the band did not contain a gap. Even more important for the present study, the second finding was that the weights assigned to a given frequency band were virtually unaffected by its spectral context – that is, by whether the other band did or did not contain a gap. This observation was supported by Bayesian rmANOVAs, which consistently showed Bayes factors in favor of a reduced model not

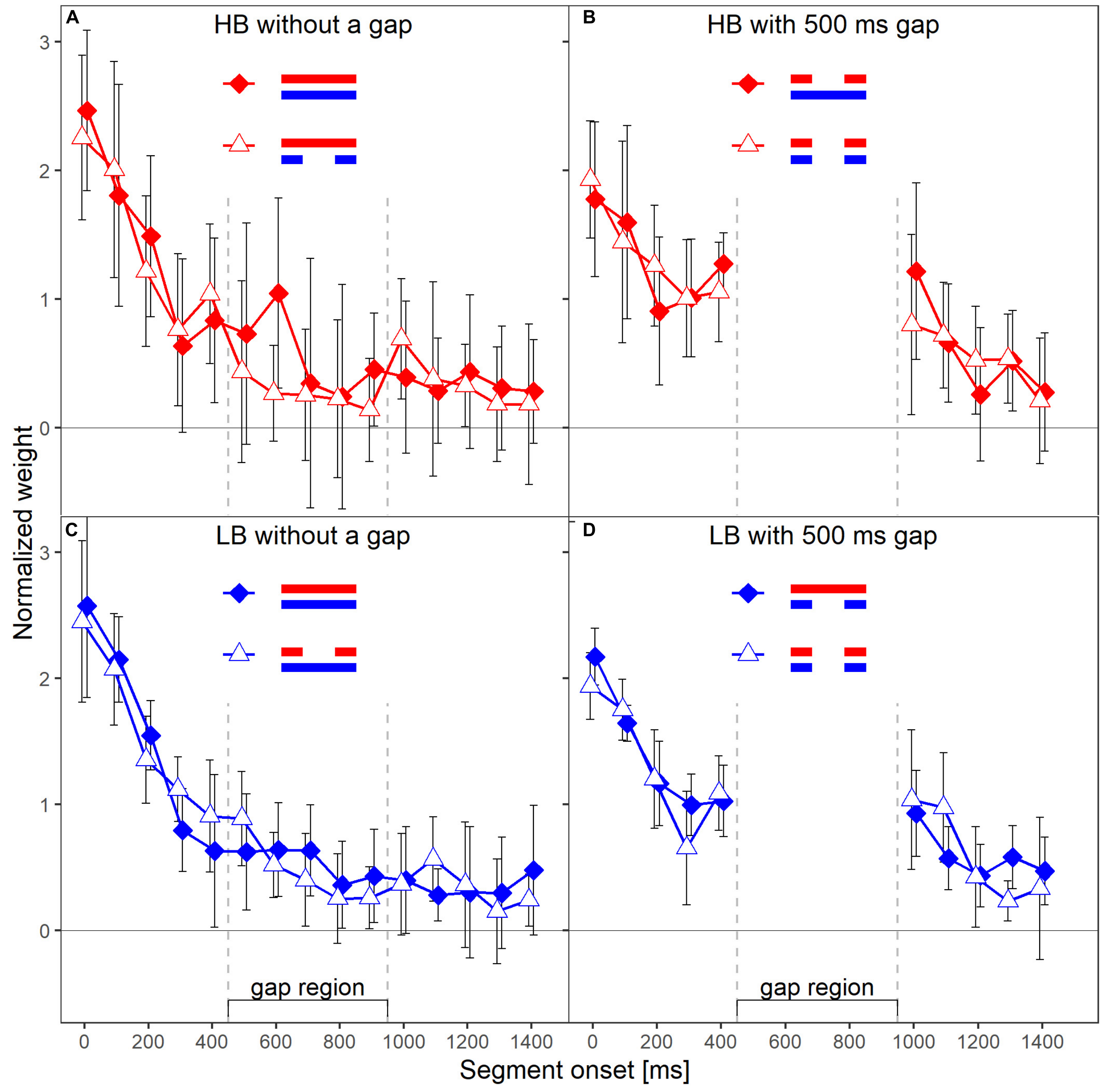

FIGURE 3 | Mean normalized temporal weights as a function of segment onset for Experiment 1. Upper panels **(A,B)** show the weights for the HB, lower panels **(C,D)** show the weights for the LB. The frequency band is also indicated by color, red = HB, blue = LB. Panels in the left column show the weights in the conditions without a gap in the analyzed band, panels on the right show the weights in the conditions with a gap. In each panel, the two different lines indicate the two different context conditions. Solid diamonds show the weights in the conditions in which the other band did not contain a gap, open triangles show the weights in the conditions in which the other band contained a gap. Error bars show 95% confidence intervals (CIs). Note that for better visibility, the two lines are shifted slightly against each other along the *x*-axis.

containing the segment × context interaction. The results from Experiment 1 thus indicate that temporal weights in loudness judgments are frequency specific.

EXPERIMENT 2

Experiment 1 showed a significant reoccurrence of the primacy effect for bands that contained a silent gap of 500-ms duration, and that the temporal weights were frequency specific. Experiment 2 was conducted to replicate these findings in an independent group of participants. Also, we presented a slightly longer gap duration of 700 ms compared to the 500 ms in Experiment 1. The reoccurrence of the primacy effect after a silent gap was reported to be more pronounced at longer gap durations (Fischenich et al., 2020). As a consequence, the presence or absence of the 700-ms gap was expected to cause a stronger change in the weights on the "context band" than for a 500-ms gap, and thus to provide a stronger test of our hypothesis that the temporal weights assigned to a given

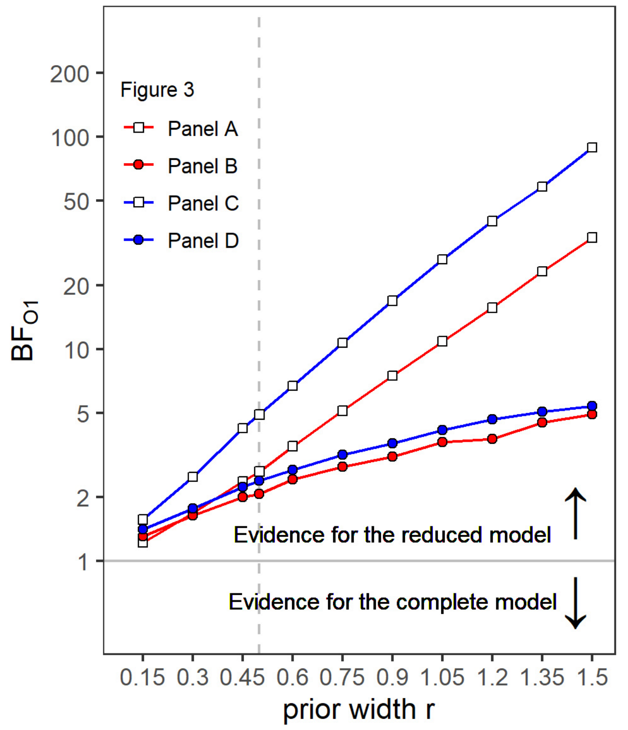

FIGURE 4 | Bayes factors (BF_{01}) as a function of prior width for the four different panels of **Figure 3**. Values of BF_{01} greater than 1.0 indicate a higher posterior probability for the reduced model not containing a segment × context interaction, compared to the complete model including a segment × context interaction.

frequency band are independent of the temporal weights assigned to a remote frequency band. Apart from the longer gap duration, the stimuli, apparatus and procedure were identical to those used in Experiment 1.

Method

Listeners

Eight normal hearing listeners (five female, three male, age 21–32 years) participated in this experiment. None of them had participated in Experiment 1. Hearing thresholds were measured by Békésy audiometry with pulsed 270-ms pure tones. All participants showed thresholds less than or equal to 15 dB HL bilaterally in the frequency range between 125 Hz and 8 kHz. All participants were students from the Johannes Gutenberg-Universität Mainz and received partial course credit for their participation.

Stimuli, Apparatus, Procedure, and Data Analysis

The apparatus was the same as used in Experiment 1. Except for the 700-ms gap duration, the stimuli were also identical to those presented in the previous experiment. The procedure and the data analysis were identical to Experiment 1. The average level difference between HB and LB at equal loudness was -3.27 dB (*SD* = 5.51 dB) and thus slightly larger than in Experiment 1. The individual mean sound pressure level differences between HB and LB at equal loudness are shown in **Supplementary Table 2**. As in Experiment 1, the loudness matches were stable across time (see **Supplementary Material** "Loudness matching" for more Information).

Table 3 shows the percentages of "softer" and "louder" responses as well as the SDT decision criterion c and the sensitivity in terms of d′ for each listener in Experiment 2.

Across the 32 fitted logistic regression models for all combinations of condition and listeners (eight listeners, four conditions), the area under the ROC curve ranged between 0.65 and 0.84 (*M* = 0.77, *SD* = 0.05), and was thus comparable to the values in Experiment 1.

Results

We report the same analyses as in Experiment 1. The average sensitivity in terms of d' is shown in **Table 4** for each of the four conditions of Experiment 2. There was a significant effect of condition on d', $F(3,21) = 6.09$, $\tilde{\varepsilon} = 0.658$, $p = 0.013$, $\eta_p^2 = 0.465$, with slightly higher mean sensitivity when both bands contained a gap (condition $LB_{700}HB_{700}$).

Figure 5 shows the mean normalized temporal weights assigned to the two frequency bands. Filled circles and open squares represent conditions where the plotted band did or did not contain a gap, respectively. For each of the plotted lines, the weights are averaged across the spectral context, that is, across the two conditions where the other frequency band either did or did not contain a gap.

As in Experiment 1, for both frequency bands, the patterns of the mean weights in both conditions (with and without a gap) showed a clear primacy effect at the beginning of the sound, in the sense that the weight on the first segment was higher than the weights on the following segments. Furthermore, when a band

TABLE 3 | Average percentages of "softer" and "louder" responses as well as the SDT decision criterion *c* and the sensitivity in terms of *d*′ for each listener in Experiment 2.

Listener	% "louder"	% "softer"	Mean of *c*	SD of *c*	Mean of *d*′	SD of *d*′
1	0.59	0.41	−0.28	0.19	1.15	0.17
2	0.47	0.53	0.10	0.27	1.02	0.28
3	0.51	0.49	−0.03	0.16	0.71	0.28
4	0.51	0.49	−0.04	0.11	1.02	0.25
5	0.40	0.60	0.31	0.18	1.15	0.21
6	0.59	0.41	−0.24	0.18	0.52	0.23
7	0.63	0.37	−0.36	0.20	0.76	0.24
8	0.61	0.39	−0.32	0.30	0.79	0.18

TABLE 4 | Mean sensitivity (*d*′) in the four different conditions of Experiment 2.

Condition	Mean of *d*′	SD of *d*′
LB_0HB_0	0.87	0.24
LB_0HB_{700}	0.85	0.27
$LB_{700}HB_0$	0.88	0.2
$LB_{700}HB_{700}$	1.00	0.23

N = 8.

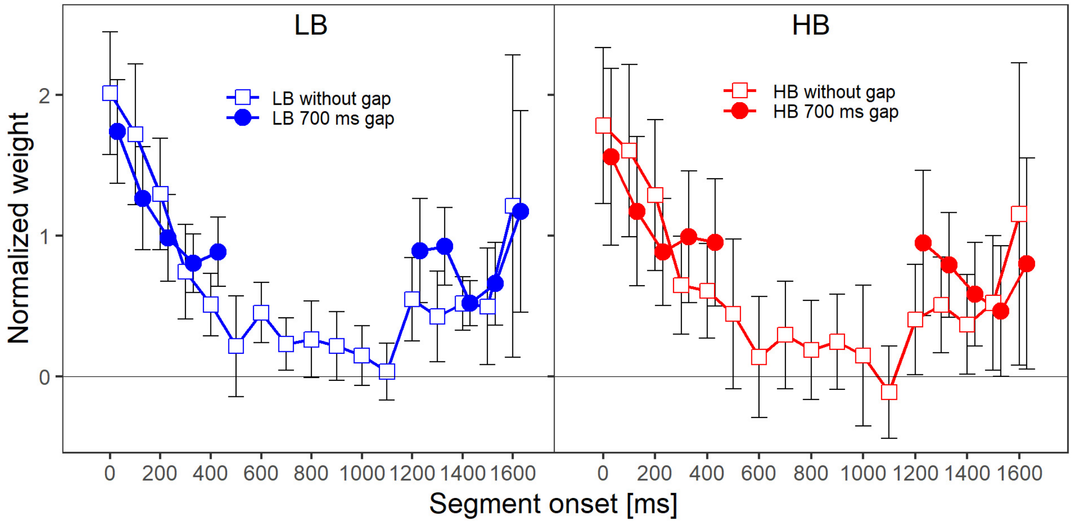

FIGURE 5 | Mean normalized weights as a function of segment onset for Experiment 2, averaged across spectral context. The two panels show the weights for the two frequency bands, LB (lower band, **left panel**) and HB (higher band, **right panel**). Frequency band is also indicated by color (LB: blue, HB: red). The different symbols and separate lines within each panel indicate whether the band did or did not contain a gap (open squares: without gap, filled circles: 700-ms gap). Error bars show 95% confidence intervals (CIs). Note that for better visibility, the two lines are shifted slightly against each other along the *x*-axis.

contained a gap, the weight assigned to the first segment after the gap was higher compared to the condition in which the band did not contain a gap. An rmANOVA with the within-subjects factors segment number (1–5 and 13–17 for bands without a gap, 1–10 for bands that contained a gap), frequency band (lower, higher), target gap (no gap, 700-ms gap) and spectral context (no gap in other band, 700-ms gap in other band) was conducted. The main effect of segment number was not significant, $F(9,63) = 2.98$, $\tilde{\varepsilon} = 0.177$, $p = 0.10$, $\eta_p^2 = 0.298$. This was likely caused by the response pattern of two listeners, who showed an almost exclusive weight on the last segment in almost all conditions for both bands (i.e., a strong recency effect; Oberfeld et al., 2018), while all remaining listeners showed a clear primacy effect. When these two listeners with strong recency effects were removed from the data analysis, the main effect of segment number was significant and comparable to the effect observed in Experiment 1, $F(9,45) = 34.26$, $\tilde{\varepsilon} = 0.524$, $p < 0.001$, $\eta_p^2 = 0.873$.

For the rmANOVA including the data from all participants, there was a significant segment number × target gap interaction, $F(9,63) = 3.25$, $\tilde{\varepsilon} = 0.775$, $p = 0.007$, $\eta_p^2 = 0.317$, indicating that the pattern of temporal weights differed depending on whether a band contained a gap or was presented without a gap. Thus, as in Experiment 1, we observed a significant reoccurrence of the primacy effect.

Each panel in **Figure 6** shows the normalized weights for one band and depending on whether or not the other band contained a silent gap. As in Experiment 1, to investigate the frequency-specificity of the weights, one has to compare the two lines in each panel of **Figure 6** that represent the two spectral context conditions (other band presented with our without a gap). For the lower band (lower panels), the two patterns of weights displayed in each panel are very similar. Thus, the weights assigned to the lower band were virtually unaffected by the presence or absence of a gap in the other band for both HB and LB. For the higher band (upper panels), the weights obtained in the two different context conditions showed differences for a few segments. However, for most segments, the weights were similar across the two context conditions.

Like in Experiment 1, the rmANOVA did not show significant interactions of the factor context with segment number [$F(9,63) = 1.81$, $\tilde{\varepsilon} = 1$, $p = 0.084$, $\eta_p^2 = 0.206$], segment number and target gap [$F(9,63) = 1.16$, $\tilde{\varepsilon} = 0.685$, $p = 0.343$, $\eta_p^2 = 0.143$] or segment number, target gap and target frequency band [$F(9,63) = 1.13$, $\tilde{\varepsilon} = 1$, $p = 0.297$, $\eta_p^2 = 0.139$). Separate Bayesian rmANOVAs were conducted per panel (that is, per combination of target band and target gap conditions) with the within-subjects factor segment number (5–13 for bands without a gap, 5 and 6 for bands that contained a gap) and context (other band without a gap, other band with 700-ms gap). The complete model that contained both main factors (segment number and context) and their interaction (segment number × context) was compared to a reduced model that included only the main factors segment number and context. Three of the resulting BF_{01} values were in favor of the reduced model, ranging from 1.37 (panel D) to 20.95 (panel C). Only the BF_{01} value of 0.37 for panel A was in favor of the complete model, showing according to Jeffreys (1961) categories anecdotal evidence for an effect of context. The robustness of the Bayes factors to changes in prior width is shown in **Figure 7**. Changes in prior width did only affect the direction of the stated results for panel A, where with increasing prior width, the results were also in favor of the reduced model. For panel C, the size of the factors showed substantial variation ranging from 2.67 to 1923.95. In general, the results thus indicate that the weights for both bands were hardly affected by the presence or absence of a gap in the other band.

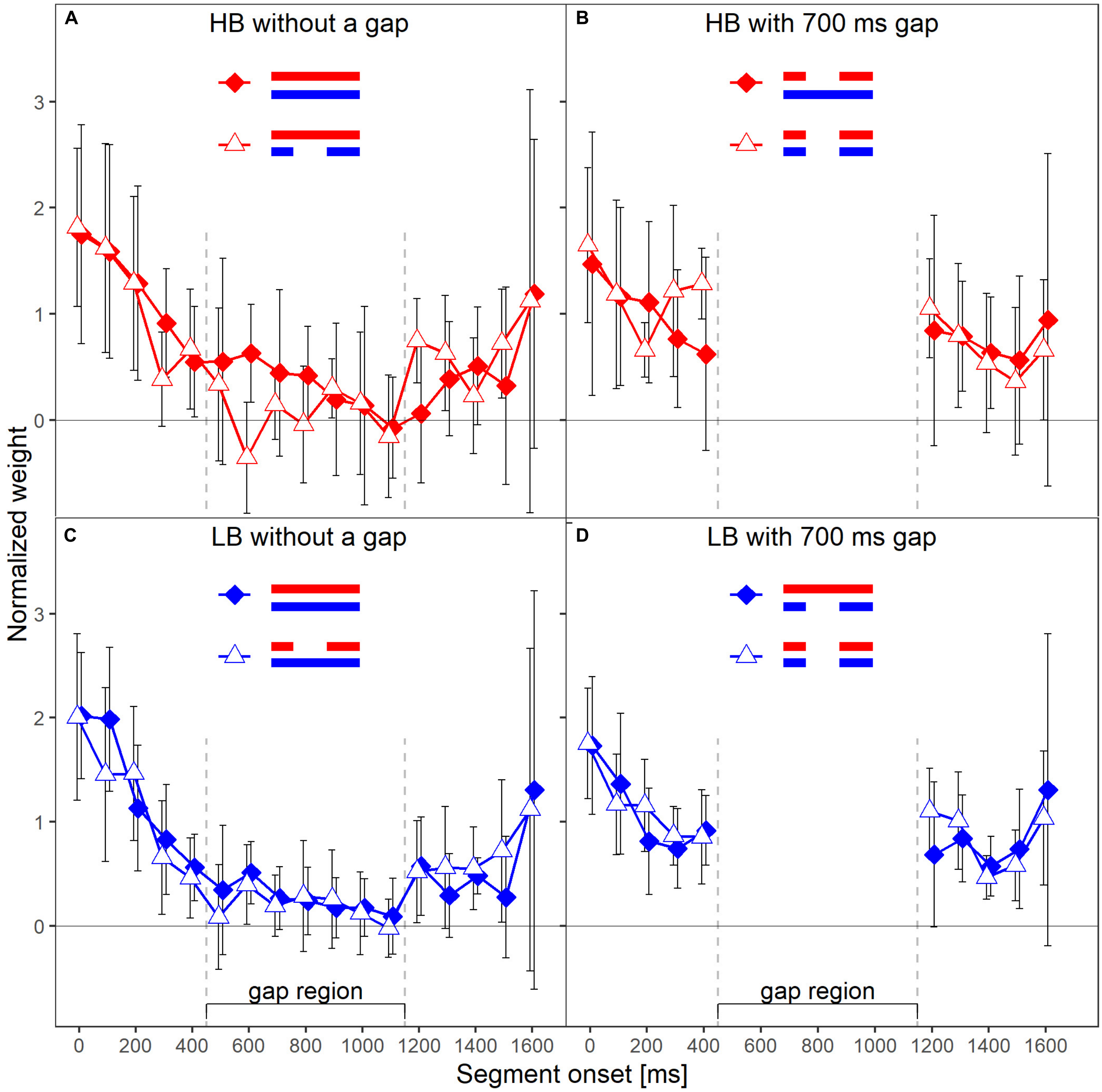

FIGURE 6 | Mean normalized weights as a function of segment onset for Experiment 2. Upper panels **(A,B)** show the weights for the higher frequency band HB, lower panels **(C,D)** show the weights for the lower band LB. The frequency band is also indicated by symbol and line color, red = HB, blue = LB. Panels in the left column show the weights in the conditions without a gap in the plotted band, panels on the right show the weights in the conditions with a gap. In each panel, the two different lines indicate the two different context conditions. Solid diamonds show the weights in the conditions in which the other band did not contain a gap, open triangles show the weights in the conditions in which the other band contained a gap. Error bars show 95% confidence intervals (CIs). Note that for better visibility, the two lines are shifted slightly against each other along the *x*-axis.

Thus, both of the main findings from Experiment 1 were confirmed in a different group of listeners and presenting a longer gap duration. There was a significant reoccurrence of the primacy effect on the second sound part when a frequency band contained a gap. For the majority of the analyses the weights assigned to a given band were largely unaffected by its spectral context, that is, by whether or not the other band contained a gap. The pattern of results thus confirms the conclusion from Experiment 1 that the temporal weights in loudness judgments are frequency specific.

DISCUSSION

The present study examined whether the temporal weights assigned to different frequency bands when listeners judge the overall loudness of a time-varying sound are frequency specific. In two experiments conducted in independent groups of listeners, temporal loudness weights were measured for stimuli consisting of two frequency bands. We introduced silent gaps in neither, only one, or both bands. According to previous research

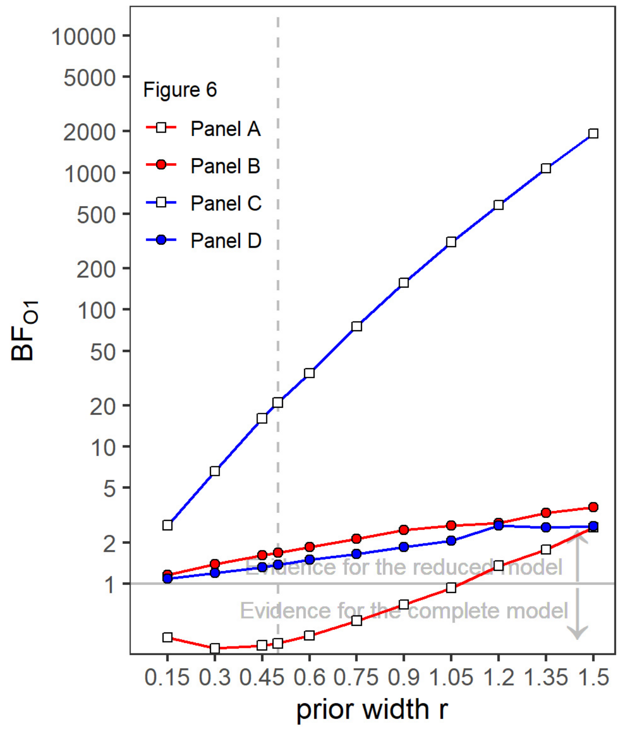

FIGURE 7 | Bayes factors (BF_{01}) as a function of prior width for the four different panels of **Figure 6**. Values of BF_{01} greater than 1.0 indicate a higher posterior probability for the reduced model not containing a segment × context interaction, compared to the complete model including a segment × context interaction.

(Fischenich et al., 2020), silent gaps result in a reoccurrence of the primacy effect after the silent gap. The temporal weights for conditions where only one of the bands contained a silent gap were compared to the weights observed when both bands were contiguous (no gap) or when both bands contained a gap. For all conditions in both experiments, primacy effects at the onset of the sounds were observed, in the sense that the first segments of a sound received higher weights compared to the following segments. This is compatible with previous data (e.g., Pedersen and Ellermeier, 2008; Rennies and Verhey, 2009; Fischenich et al., 2019).

In Experiment 2, two listeners consistently showed strong recency effects rather than a primacy effect for both bands and in all gap conditions. In previous studies, recency effects appeared from time to time in some conditions and for some listeners (condition without feedback, Pedersen and Ellermeier, 2008; in Experiment 3 and 4, Oberfeld and Plank, 2011; for five segment sounds with durations of 2.5 s and above, Oberfeld et al., 2018; sounds in background noise SL 7.5 dB, Fischenich et al., 2019). In general, they are less frequent and less pronounced than the primacy effect. The primacy effect has been observed very consistently across a large number of studies (for a review see Oberfeld et al., 2018). However, inter-individual differences in perceptual weights tend to be rather large, showing various kinds of patterns (e.g., Lutfi et al., 2011). This is even more pronounced for recency effects (Oberfeld and Plank, 2011; Oberfeld et al., 2018).

Bands that contained a gap showed higher weights on the first segments following the gap, compared to the weights assigned to segments at the same temporal position when the band did not contain a gap. This difference in the weighting patterns was statistically significant in both experiments. Thus, the results confirm the finding that the primacy effect reoccurs after a silent gap of a certain duration within a sound (Fischenich et al., 2020).

The main aim of the present study was to answer the question of whether the temporal loudness weights are applied independently for each frequency band contained in the stimulus, or to both bands simultaneously. Across the two experiments, the general patterns of the temporal weights assigned to the target band were hardly affected by the spectral context (i.e., presence or absence of a silent gap in the other frequency band). However, descriptively the weights in the gap region were sometimes smaller when the other band contained a gap compared to when it did not contain a gap (see **Figures 3A**, **6A**), indicating a context effect. If suppression of the HB by the LB and a resulting increase in loudness of the HB during the gap in the LB had played a role, the opposite pattern – higher weights on the HB weights during the gap region when the LB contained a gap – should have resulted. A potential explanation for these descriptive trends could rather be that loudness dominance takes place *after* spectral integration and therefore parts of the sound where both bands were present received higher weights. However, under this assumption, one should expect the weights in the continuous band to show a much stronger decline when the other band contains a gap. A reduction in sound pressure level by 10 dB has been shown to result in almost zero weights (e.g., Oberfeld and Plank, 2011). Because the two bands were loudness-matched in our experiments, we can assume that the total loudness during the gap in one band was approximately half of the total loudness when both bands were present. Thus, the effect of the gap on total loudness can be expected to be similar to the effect of a 10-dB level reduction within a single band, which also corresponds to a loudness reduction by approximately a factor of two. In addition, the loudness dominance effect would also have resulted in greater differences between all of the weights after the gap compared to the weights within the gap (see **Figure 5** in Fischenich et al., 2020). In addition, in **Figure 3C**, the weight on the first segment of the LB within the gap region in the HB was higher (rather than lower) when the other band did contain a gap, compared to when it did not contain a gap. This illustrates the variability in the data, as some descriptive data were compatible with an effect of spectral context, but the data also showed descriptive weight differences comparable in size that are incompatible with the assumption.

Another example for a descriptive pattern in the data that could be taken as an effect of spectral context is that in the continuous HB (without gap), the weight difference between the last segment within the gap region of the other band and the first segment after the gap region of the other band was higher when the other band contained a gap, compared to when the other band did not contain a gap. Interestingly, this pattern was present only for the continuous HB, but not for the continuous LB, in both experiments (see **Figures 3A,C**, **6A,C**). If one assumes that the gap in the other band caused an additional onset effect also in the ongoing band, it is difficult to argue why this was the case only for the HB, but not for the LB. Also, if one assumes higher loudness of the HB during the gap in the LB due to suppression, the first HB segment following the gap region should have been perceived as softer than the last HB segment in the gap region, due to suppression by the LB that was again present for this segment. In such a case, it is difficult to understand why then the weight on the first HB segment following the gap region was *higher* rather than lower when the LB contained a gap.

Apart from these relatively small effects of spectral context in a small subset of the weights, the more systematic and encompassing statistical evaluation of the size of the context effect, which was provided by the Bayes factors (BF_{01}) that compared the posterior likelihood for a model with an effect of context (i.e., assuming that the gap in the other band has a systematic effect on the weights for the target band), and a model without this effect of context, showed evidence for an absence of an effect of spectral context on the temporal weights for a given frequency band, for most of the conditions. The results thus indicate that the temporal weights in loudness judgments are, by and large, frequency specific.

In the context of loudness models, this finding suggests a weighting on the basis of a time-varying specific loudness, i.e., the loudness time function at the output of each frequency channel (auditory filter). The debate on whether temporal integration precedes spectral integration was already present when Zwicker (1977) proposed his original loudness model for time-varying sounds. Zwicker argued on the basis of results indicating spectral loudness summation for non-simultaneously presented frequency components (Zwicker, 1969; see also Heeren et al., 2011) that spectral integration should precede temporal integration. Thus, the dynamic loudness model (DLM) (Zwicker, 1977) and models based on it (Chalupper and Fastl, 2002) assume that spectral integration precedes temporal integration. The same order of the processing stages was assumed in the time-varying loudness model (TVL-model) proposed by Glasberg and Moore (2002). However, in the most recent versions of this model (Moore et al., 2016, 2018), the short-term specific loudness is calculated per frequency channel before spectral summation takes place. The assumption that temporal processing precedes spectral integration is compatible with the present finding of frequency-specific temporal loudness weights. It is also compatible with neurophysiological data showing entrainment to channel-specific instantaneous loudness in cortical MEG components up to about 100 ms (Thwaites et al., 2017). One should keep in mind, however, that the attack-decay type of temporal integration assumed by the TVL-model does not predict a primacy effect, as demonstrated by simulation results in Fischenich et al. (2019).

A possible explanation for the observed reoccurrence of the primacy effect within a frequency band when the band contained a gap might be that the re-onset of the band containing the gap might in principle capture the attention (Oberfeld and Plank, 2011). Such an attentional capture could cause a primacy effect on the post-gap part of the band containing the gap (if the weights are assigned per band), and even also on the band that did not contain a gap (if the weights are assigned across frequency). However, our previous work did not provide compelling support for such an attention-orienting explanation of the primacy effect. A reduction of the perceived abruptness of the onset effect by presenting the target sound in continuous background noise (Fischenich et al., 2019), or by imposing a gradual fade-in in level at the sound onset (Oberfeld and Plank, 2011), did not remove the primacy effect pattern.

Fischenich et al. (2020) discussed three possible explanations for the primacy effect and its reoccurrence after a silent gap. The first explanation, originally proposed by Oberfeld and Plank (2011), is based on the response characteristics of neurons in the AN, which tend to show a peak in the firing rate at the sound onset (Nomoto et al., 1964; Rhode and Smith, 1985). The inter-stimulus-interval that was reported to be necessary to see a reoccurrence of the initial peak in the firing rate of some types of nerve fibers (Relkin and Doucet, 1991) is roughly in line with the necessary interval to see a significant reoccurrence of the primacy effect (Fischenich et al., 2020). Because the inner hair cells that innervate the AN fibers are frequency specific, the recovery of the firing rate is also frequency specific (Harris and Dallos, 1979). The explanation of the primacy effect in temporal loudness weights based on the response characteristics of the AN fibers is thus compatible with the result of frequency-specific weights in the present study. However, the inter-individual differences in weighting patterns with pronounced recency effects for two listeners in Experiment 2 argue against an explanation based on the response characteristic of the AN. If the weighting patterns were due to the initial peak in the firing rate of the AN fibers, cases in which individuals show a completely reversed weighting pattern with strong recency effects should not occur.

A second potential explanation of the primacy effect and its reoccurrence is based on research on masking effects on intensity discrimination, which shows that for masker-target intervals below 400 ms, intensity-difference-limens (DLs) are increased substantially (e.g., Zeng et al., 1991; Oberfeld, 2008b). A segment presented in the middle of a longer sound might be subject to forward masking by preceding segments, which would result in a primacy effect if listeners adopted a reasonable strategy of placing higher weights on temporal portions of a sound for which the intensity resolution is higher (Green, 1958; Oberfeld et al., 2013). The silent gap necessary for a significant reoccurrence of the primacy effect in Fischenich et al. (2020) was approximately in line with the time course of masking effects on DLs. The explanation of the primacy effects and its reoccurrence based on masking effects on intensity discrimination, are in line with frequency-specific weights, because no DL elevations were observed when the masker-signal frequency separation is large (Zeng and Turner, 1992). However, as discussed in detail in Fischenich et al. (2020), several additional assumptions are needed in order to explain the primacy effect in temporal loudness weights by masking effects on intensity resolution.

A third potential explanation of the primacy effect and its reoccurrence is provided by an *evidence integration approach* (e.g., Vickers, 1970). Evidence integration suggests that when making perceptual judgments, listeners accumulate evidence for each of the possible response alternatives in a random walk process. As discussed by Fischenich et al. (2020), models that simulate such an evidence accumulation process can produce temporal weighting patterns with either primacy or recency effects. If one assumes that a separate evidence integration process is in effect for each frequency band or auditory stream, then frequency-specific weights are predicted. Furthermore, if one assumes that after a gap of sufficient duration within a band, a separate evidence integration process is carried out for both temporal parts of the band (the part before and the part after the gap), then evidence integration can also account for the reoccurrence of the primacy effect within a band.

It should be noted that while all three of the potential explanations of the frequency specificity of the temporal weighting patterns account for some aspects of the observed results, each of them has some clear limitations (for a discussion see Fischenich et al., 2020). It is currently not possible to decide which of the alternative mechanisms is the most likely explanation of the observed temporal loudness weights.

In addition, in an absolute identification task as the one presented in the experiments of this study, the decision of a participant might depend not only on the segment levels presented on the current trial, but also on the sounds presented on preceding trials (e.g., Stewart et al., 2005). It would be interesting to investigate such potential sequential effects in future research.

To summarize, in two experiments, the present study investigated whether the temporal weights assigned to different frequency bands when listeners judge the overall loudness of a time-varying sound are frequency specific. The results of both experiments indicated that temporal loudness weights are approximately frequency specific. While the frequency specificity of the weights is in accordance with several potential explanations of the primacy effect in loudness judgments, further research is needed to investigate the underlying mechanisms of the primacy effect as well as of its recovery during silent gaps.

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/qh3zt.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Institute of Psychology of the Johannes Gutenberg-Universität Mainz (reference number 2016-JGU-psychEK-002). The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

AF and DO supervised the data acquisition and data curation, performed the statistical analyses, and created the graphical illustrations and tables. DO and JV administered and supervised the project. DO was responsible for the funding acquisition. All the authors participated in writing the original draft and contributed to the development of the method and software, manuscript revision, and read and approved the submitted version.

FUNDING

This work was supported by a grant from Deutsche Forschungsgemeinschaft (DFG; www.dfg.de) to DO (OB 346/6-1) and JV (VE 373/2-1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We are grateful to Celine Gebbert, Jonas Krämer, Christoph Keßler, Anna-Lena Spies, and Sinah Neuwardt for their assistance in data collection.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.588571/full#supplementary-material

Supplementary Data Sheet 1 | Supplementary Material Loudness matching.

Supplementary Data Sheet 2 | Supplementary Material Alternative Normalization.

REFERENCES

Berg, B. G. (1990). Observer efficiency and weights in a multiple observation task. *J. Acoust. Soc. Am.* 88, 149–158. doi: 10.1121/1.399962

Braida, L. D., and Durlach, N. I. (1972). Intensity perception: II. Resolution in one-interval paradigms. *J. Acoust. Soc. Am.* 51, 483–502. doi: 10.1121/1.1912868

Chalupper, J., and Fastl, H. (2002). Dynamic loudness model (DLM) for normal and hearing-impaired listeners. *Acta Acust. U. Acust.* 88, 378–386.

Dittrich, K., and Oberfeld, D. (2009). A comparison of the temporal weighting of annoyance and loudness. *J. Acoust. Soc. Am.* 126, 3168–3178. doi: 10.1121/1.3238233

Ellermeier, W., and Schrödl, S. (2000). "Temporal weights in loudness summation," in *Fechner Day 2000. Proceedings of the 16th Annual Meeting of the International Society for Psychophysics*, ed. C. Bonnet (Strasbourg: Université Louis Pasteur), 169–173.

Ernst, S. M. A., Rennies, J., Kollmeier, B., and Verhey, J. L. (2010). Suppression and comodulation masking release in normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 128, 300–309. doi: 10.1121/1.3397582

Fischenich, A., Hots, J., Verhey, J., and Oberfeld, D. (2019). Temporal weights in loudness: investigation of the effects of background noise and sound level. *PLoS One* 14:e0223075. doi: 10.1371/journal.pone.0223075

Fischenich, A., Hots, J., Verhey, J. L., and Oberfeld, D. (2020). The effect of silent gaps on temporal weights in loudness judgments. *Hear. Res.* 395:108028. doi: 10.1016/j.heares.2020.108028

Gauss, C. F. (1821). Theoria combinationis observationum erroribus minimis obnoxiae. *Commentationes Soc. Regiae Sci. Gottingensis Recent.* 33–90.

Glasberg, B. R., and Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.* 50, 331–342.

Green, D. M. (1958). Detection of multiple component signals in noise. *J. Acoust. Soc. Am.* 30, 904–911. doi: 10.1121/1.1909400

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

Hartmann, W. M., and Pumplin, J. (1988). Noise power fluctuations and the masking of sine signals. *J. Acoust. Soc. Am.* 83, 2277–2289. doi: 10.1121/1.396358

Heeren, W., Rennies, J., and Verhey, J. L. (2011). Spectral loudness summation of nonsimultaneous tone pulses. *J. Acoust. Soc. Am.* 130, 3905–3915. doi: 10.1121/1.3652866

Hosmer, D. W., and Lemeshow, S. (2000). *Applied Logistic Regression*. New York, NY: Wiley.

Hots, J., Verhey, J., and Oberfeld, D. (2018). “Der einfluss von signalpausen auf die zeitliche gewichtung bei der Lautheitswahrnehmung,” in *Proceedings of the DAGA 2018*, (München).

Houtgast, T. (1974). *Lateral Suppression in Hearing*. Free Unversity. Doctoral dissertation.

Huynh, H., and Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *J. Educ. Stat.* 1, 69–82. doi: 10.3102/10769986001001069

IEC 318 (1970). *An IEC Artificial Ear, of the Wide Band Type, for the Calibration of Earphones Used in Audiometry*. Geneva: International Electrotechnical Commission.

ISO 389-8 (2017). “Acoustics - Reference zero for the calibration of audiometric equipment,” in *Proceedings of the Part 8: Reference Equivalent Threshold Sound Pressure Levels for Pure Tones and Circumaural Earphones*, (Geneva).

JASP Team (2019). *JASP (Version 0.11.1) [Computer software]*. Available online at: https://jasp-stats.org/ (accessed February 27, 2020).

Jeffreys, H. (1961). *Theory of Probability*. Oxford: Clarendon Press.

Kiang, N. Y. S., Watanabe, T., Thomas, E. C., and Clark, L. F. (1965). *Discharge Patterns of Single Fibers in the Cat’s Auditory Nerve*. Cambridge, MA: M.I.T. Press.

Kohlrausch, A., Fassel, R., Van Der Heijden, M., Kortekaas, R., Van De Par, S., Oxenham, A. J., et al. (1997). Detection of tones in low-noise noise: further evidence for the role of envelope fluctuations. *Acustica* 83, 659–669.

Lutfi, R. A., and Jesteadt, W. (2006). Molecular analysis of the effect of relative tone level on multitone pattern discrimination. *J. Acoust. Soc. Am.* 120, 3853–3860. doi: 10.1121/1.2361184

Lutfi, R. A., Liu, C. J., and Stoelinga, C. N. J. (2011). Auditory discrimination of force of impact. *J. Acoust. Soc. Am.* 129, 2104–2111. doi: 10.1121/1.3543969

Moore, B. C. J., Glasberg, B. R., Varathanathan, A., and Schlittenlacher, J. (2016). A loudness model for time-varying sounds incorporating binaural inhibition. *Trends Hear.* 20, 1–16.

Moore, B. C. J., Jervis, M., Harries, L., and Schlittenlacher, J. (2018). Testing and refining a loudness model for time-varying sounds incorporating binaural inhibition. *J. Acoust. Soc. Am.* 143, 1504–1513. doi: 10.1121/1.5027246

Nachmias, J. (2006). The role of virtual standards in visual discrimination. *Vision Res.* 46, 2456–2464. doi: 10.1016/j.visres.2006.01.029

Namba, S., Kuwano, S., and Kato, T. (1976). Loudness of sound with intensity increment. *Jpn. Psychol. Res.* 18, 63–72. doi: 10.4992/psycholres1954.18.63

Nomoto, M., Katsuki, Y., and Suga, N. (1964). Discharge pattern and inhibition of primary auditory nerve fibers in the monkey. *J. Neurophysiol.* 27, 768–787. doi: 10.1152/jn.1964.27.5.768

Oberfeld, D. (2008a). Does a rhythmic context have an effect on perceptual weights in auditory intensity processing? *Can. J. Exp. Psychol. Revue Can. Psychol. Exp.* 62, 24–32. doi: 10.1037/1196-1961.62.1.24

Oberfeld, D. (2008b). The mid-difference hump in forward-masked intensity discrimination. *J. Acoust. Soc. Am.* 123, 1571–1581. doi: 10.1121/1.2837284

Oberfeld, D. (2008c). Temporal weighting in loudness judgments of time-varying sounds containing a gradual change in level. *J. Acoust. Soc. Am.* 123:3307. doi: 10.1121/1.2933740

Oberfeld, D. (2015). Are temporal loudness weights under top-down control? Effects of trial-by-trial feedback. *Acta Acust. U. Acust.* 101, 1105–1115. doi: 10.3813/aaa.918904

Oberfeld, D., Heeren, W., Rennies, J., and Verhey, J. (2012). Spectro-temporal weighting of loudness. *PLoS One* 7:e50184. doi: 10.1371/journal.pone.0050184

Oberfeld, D., Hots, J., and Verhey, J. L. (2018). Temporal weights in the perception of sound intensity: effects of sound duration and number of temporal segments. *J. Acoust. Soc. Am.* 143, 943–953. doi: 10.1121/1.5023686

Oberfeld, D., Kuta, M., and Jesteadt, W. (2013). Factors limiting performance in a multitone intensity-discrimination task: disentangling non-optimal decision weights and increased internal noise. *PLoS One* 8:e79830. doi: 10.1371/journal.pone.0079830

Oberfeld, D., and Plank, T. (2011). The temporal weighting of loudness: effects of the level profile. *Atten. Percept. Psychophys.* 73, 189–208. doi: 10.3758/s13414-010-0011-8

Pedersen, B., and Ellermeier, W. (2008). Temporal weights in the level discrimination of time-varying sounds. *J. Acoust. Soc. Am.* 123, 963–972. doi: 10.1121/1.2822883

Plank, T. (2005). *Auditive Unterscheidung Von Zeitlichen Lautheitsprofilen (Auditory Discrimination of Temporal Loudness Profiles)*. Universität Regensburg. PhD thesis.

Ponsot, E., Susini, P., Saint Pierre, G., and Meunier, S. (2013). Temporal loudness weights for sounds with increasing and decreasing intensity profiles. *J. Acoust. Soc. Am.* 134, EL321–EL326.

Relkin, E. M., and Doucet, J. R. (1991). Recovery from prior stimulation. I: relationship to spontaneous firing rates of primary auditory neurons. *Hear. Res.* 55, 215–222. doi: 10.1016/0378-5955(91)90106-j

Rennies, J., and Verhey, J. L. (2009). Temporal weighting in loudness of broadband and narrowband signals. *J. Acoust. Soc. Am.* 126, 951–954. doi: 10.1121/1.3192348

Rhode, W. S., and Smith, P. H. (1985). Characteristics of tone-pip response patterns in relationship to spontaneousrate in cat auditory nerve fibers. *Hear. Res.* 18, 159–168. doi: 10.1016/0378-5955(85)90008-5

Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes factors for ANOVA designs. *J. Math. Psychol.* 56, 356–374. doi: 10.1016/j.jmp.2012.08.001

Stewart, N., Brown, G. D. A., and Chater, N. (2005). Absolute identification by relative judgment. *Psychol. Rev.* 112, 881–911. doi: 10.1037/0033-295x.112.4.881

Thwaites, A., Schlittenlacher, J., Nimmo-Smith, I., Marslen-Wilson, W. D., and Moore, B. C. (2017). Tonotopic representation of loudness in the human cortex. *Hear. Res.* 344, 244–254. doi: 10.1016/j.heares.2016.11.015

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics* 13, 37–58. doi: 10.1080/00140137008931117

Zeng, F. G., and Turner, C. W. (1992). Intensity discrimination in forward masking. *J. Acoust. Soc. Am.* 92, 782–787. doi: 10.1121/1.403947

Zeng, F. G., Turner, C. W., and Relkin, E. M. (1991). Recovery from prior stimulation II: effects upon intensity discrimination. *Hear. Res.* 55, 223–230. doi: 10.1016/0378-5955(91)90107-k

Zwicker, E. (1969). Influence of temporal structure of tones on addition of partial loudnesses. *Acustica* 21:16. doi: 10.1016/j.heares.2006.08.007

Zwicker, E. (1977). Procedure for calculating loudness of temporally variable sounds. *J. Acoust. Soc. Am.* 62, 675–682. doi: 10.1121/1.381580

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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 26 March 2021
doi: 10.3389/fpsyg.2021.623670

The "Missing 6 dB" Revisited: Influence of Room Acoustics and Binaural Parameters on the Loudness Mismatch Between Headphones and Loudspeakers

Florian Denk[1†], Michael Kohnen[2], Josep Llorca-Bofí[2], Michael Vorländer[2] and Birger Kollmeier[1]*

[1] *Medizinische Physik and Cluster of Excellence "Hearing4all", Universität Oldenburg, Oldenburg, Germany,* [2] *Institute of Technical Acoustics, RWTH Aachen University, Aachen, Germany*

OPEN ACCESS

Edited by:
Sabine Meunier,
Ecole Centrale Marseille, France

Reviewed by:
Vincent Koehl,
Université de Bretagne Occidentale, France
Ville Sivonen,
Helsinki University Central Hospital, Finland

****Correspondence:***
Birger Kollmeier
birger.kollmeier@uni-oldenburg.de

†Present address:
Florian Denk,
German Institute of Hearing Aids, Lübeck, Germany

Specialty section:
This article was submitted to Perception Science, a section of the journal Frontiers in Psychology

Received: *30 October 2020*
Accepted: *05 March 2021*
Published: *26 March 2021*

Citation:
Denk F, Kohnen M, Llorca-Bofí J, Vorländer M and Kollmeier B (2021) The "Missing 6 dB" Revisited: Influence of Room Acoustics and Binaural Parameters on the Loudness Mismatch Between Headphones and Loudspeakers. Front. Psychol. 12:623670. doi: 10.3389/fpsyg.2021.623670

Generations of researchers observed a mismatch between headphone and loudspeaker presentation: the sound pressure level at the eardrum generated by a headphone has to be about 6 dB higher compared to the level created by a loudspeaker that elicits the same loudness. While it has been shown that this effect vanishes if the same waveforms are generated at the eardrum in a blind comparison, the origin of the mismatch is still unclear. We present new data on the issue that systematically characterize this mismatch under variation of the stimulus frequency, presentation room, and binaural parameters of the headphone presentation. Subjects adjusted the playback level of a headphone presentation to equal loudness as loudspeaker presentation, and the levels at the eardrum were determined through appropriate transfer function measurements. Identical experiments were conducted at Oldenburg and Aachen with 40 normal-hearing subjects including 14 that passed through both sites. Our data verify a mismatch between loudspeaker and binaural headphone presentation, especially at low frequencies. This mismatch depends on the room acoustics, and on the interaural coherence in both presentation modes. It vanishes for high frequencies and broadband signals if individual differences in the sound transfer to the eardrums are accounted for. Moreover, small acoustic and non-acoustic differences in an anechoic reference environment (Oldenburg vs. Aachen) exert a large effect on the recorded loudness mismatch, whereas not such a large effect of the respective room is observed across moderately reverberant rooms at both sites. Hence, the non-conclusive findings from the literature appear to be related to the experienced disparity between headphone and loudspeaker presentation, where even small differences in (anechoic) room acoustics significantly change the response behavior of the subjects. Moreover, individual factors like loudness summation appear to be only loosely connected to the observed mismatch, i.e., no direct prediction is possible from individual binaural loudness summation to the observed mismatch. These findings – even though not completely explainable by the yet limited amount of parameter variations performed in this study – have consequences for the comparability of experiments using loudspeakers with conditions employing headphones or other ear-level hearing devices.

Keywords: psychoacoustics, headphone calibration, binaural loudness summation, room acoustics, cross site comparison

INTRODUCTION

While listening with ear-level devices, such as headphones, earphones, or hearing aids, it is often reasonable to assume that the presented acoustic signal is perceived with the same loudness as when presented *via* a loudspeaker, if the same acoustic signal is produced at the subject's eardrum at the same sound pressure level in both conditions. This "matching assumption" is important, e.g., for free-field equalization of headphones, for virtual reality applications, for hearing device fitting, or for protecting the earphone user from hazardous high sound pressure levels (Munson and Wiener, 1952; Killion, 1978; Rudmose, 1982; Fastl et al., 1985; Keidser et al., 2000). However, there is considerable evidence in the literature (see below) about a mismatch between headphone and loudspeaker presentation violating the "matching assumption" for yet unclear reasons. This contrasts with findings from more recent research (Völk and Fastl, 2011; Brinkmann et al., 2017) indicating that virtually no mismatch occurs if the individual sound filtering properties are adequately taken into account (i.e., using individual head related transfer functions, HRTFs, and headphone related transfer functions, HpTFs), thus ensuring that the same waveforms are created at the eardrums in both presentation modes. However, the reason why these studies provide contradicting findings and how the mismatch between headphone and loudspeaker listening might depend on the different experimental parameters employed in the various studies in the literature is yet unclear. The current study therefore attempts to pinpoint the origin of the mismatch by systematically investigating the influence of room acoustics, binaural parameters, and the stimulus on the reported mismatch, as well as potential lab-specific effects.

Beranek (1949) already reported that headphones require a 6–10 dB higher level at the eardrums to provide the same loudness impression as a loudspeaker in free field. This was confirmed by Munson and Wiener (1952) who reported a "6 dB mismatch" at low frequencies for diotic headphone presentation, which they explained by different perceived positions of the source. Further confirmation of the "missing 6 dB" was reported by Robinson and Dadson (1956) and Theile (1986). Rudmose (1982), however, reported to have resolved the "case of the missing 6 dB" by attributing its existence to transducer distortions, and the procedures employed including appropriate training of the subjects and structure-borne sound transmission from the electroacoustic transducers to the subject's body. The positioning of the loudspeaker was also acknowledged as an important factor, which was confirmed by Keidser et al. (2000) who found a mismatch of 8 dB for sounds around 500 Hz and no such difference around 3 kHz.

The observations outlined above were made under anechoic conditions, with diotic headphone presentation and a direct comparison between headphone and loudspeaker presentation, where the headphone was put on and off by the subject. Contrary, experimental designs using individual dynamic binaural synthesis, where headphones remained in place during loudspeaker playback, such that the subject was not informed which source they were listening to, achieve an authentic headphone presentation where no mismatch appeared (Völk and Fastl, 2011; Brinkmann et al., 2017). In a similarly blinded comparison, Bonnet et al. (2018) reported that an occlusion of the ear during stimulation by an external sound source did not result in a loudness mismatch to stimulation of the unoccluded ear with the same external sound source. Very recently, Meunier et al. (2020) compared loudness growth functions for headphone and loudspeaker presentation without a direct comparison of both sources, and also found no loudness mismatch. None of the experiments summarized above focused on the role of binaural hearing and interaural disparity for the mismatch. Their possible importance for the mismatch is highlighted by experiments performed by Edmonds and Culling (2009) who found a distinct influence of the interaural coherence (IC) of headphone stimuli in loudness judgment. Also, findings from Rudmose (1982) and Zahorik and Wightman (2001) using stimuli with varying distance of loudspeaker indicate that the differences in interaural coherence or the reverberant sound field influence loudness judgments. Hence, the binaural listening mode and the interaural coherence – which is usually also connected to the apparent source width (Rudmose, 1982; Zahorik and Wightman, 2001; Sivonen and Ellermeier, 2006) – appears to play an important role in the differential judgment of loudspeaker vs. headphone presentation. However, the specific influence of binaural reproduction parameters or the room on the perceived mismatch between headphone and loudspeaker presentation has not yet been assessed in a systematic way.

Another factor that might play a role in the reported loudness mismatch and the inconsistent study results is the interindividual variability in loudness perception. It is of considerable size if binaural hearing and binaural summation of loudness comes into play: Oetting et al. (2016) reported individual differences in categorical loudness scaling for the combined effect of loudness summation across both ears and across frequency that ranged up to 20 dB in effect size. Even though it is still unclear how to model these effects in current loudness models (e.g., Pieper et al., 2016), this high interindividual variability in binaural loudness summation might contribute to interindividual variability in the loudness mismatch between headphone and loudspeaker stimulation when broadband signals and an altered interaural coherence is involved.

The aim of the current study therefore is to systemically investigate the influence of a number of relevant parameters on the apparent mismatch in order to pinpoint its origin and the reason for non-consistent findings in the literature. Moreover, a thorough understanding of the influence of different parameters on the mismatch between headphone and loudspeaker presentation should be useful for avoiding this mismatch in designing modern ear-level communication systems such as, e.g., hearables or assistive listening devices. This paper focuses on the effect of room acoustics and interaural coherence on the mismatch while open-back headphones are used. Note that the influence of different kind of headphones on the mismatch is beyond the scope of the current study and will be examined in a companion paper by Kohnen et al. (in preparation).[1]

The study was designed to address the following hypotheses that are based on possible explanations for the differences across studies reported above:

> *H1*: *The same results with respect to the mismatch should be achieved across different labs if the same set of subjects and comparable conditions are used*. For testing this hypothesis, we performed a comparative study across two sites [Aachen (AC) and Oldenburg (OL)], employing the respective large anechoic room at each site and a group of subjects that performed the same experiments at both sites in addition to separate subjects at both sites. We extended this comparison across sites by including one additional moderately reverberant room at each site (termed as "non-anechoic" in the following, see below).
>
> *H2*: *The binaural presentation mode (diotic versus binaural headphone playback with different values of the interaural coherence) has a significant influence on the mismatch*. Hence, we used monaural as well as bilateral headphone presentation, the latter with diotic or dichotic playback. To systematically vary the reverberation time and, hence, the effective IC in the non-anechoic room as well as binaural headphone presentation, we performed the loudness matching experiments in four different rooms: The anechoic rooms in Oldenburg and Aachen, a sound-insulated lab room with little reverberation (OL earpiecelab, T30 = 0.4 s) and a medium-sized room without any specific acoustical treatment (AC tea kitchen) exhibiting a reverberation time T30 of approx. 0.6 s (see **Table 1**).
>
> *H2a*: *No mismatch between headphone and loudspeaker presentation in a non-anechoic room can be observed if the interaural coherence during headphone presentation is matched to the respective room*. To test this hypothesis, the "IC matched" condition was additionally tested throughout the experiments listed above.
>
> *H2b*: *The apparent source width is strongly connected to the mismatch*. To test this hypothesis, the apparent source width in the different experimental headphone playback conditions in comparison to the apparent source width of the target loudspeaker was evaluated and compared to the mismatch results.
>
> *H3*: *The interindividual spread in the mismatch across different conditions is related to the individual variability in binaural loudness summation or other individual binaural processing characteristics (like, e.g., the binaural benefit in a spatial speech recognition task)*. To test this hypothesis, we performed additional audiological evaluations with a subset of the subjects employed here.

MATERIALS AND METHODS

Subjects matched the perceived loudness of a headphone presentation to that of a loudspeaker presentation of the same stimulus, and the levels at the eardrum for equal loudness were compared. No equalization of the loudspeaker or headphone was applied during stimulus presentation. The loudness matching experiment was performed in four rooms distributed over two sites, three headphones, for four signals, and four headphone presentation modes (section Stimuli and Rooms). In this paper, only the results obtained with open-coupling headphones (HD 650, Sennheiser, Wedemark, Germany) are presented. The HD650 was chosen here due to its widespread use and due to low repositioning variation compared to the other headphones tested (Beyerdynamic DT770 Pro, Etymotic ER4, for further details see[1]), which does not depend on a tight fit on the ear due to the open-back design. Subjects underwent four experimental sessions at each site, including one session for auditory screening and characterization (section Subjects and Characterization), one for measurements of individual ear-related transfer functions (section Sound Levels at Eardrum), and two for the loudness matching and apparent source width experiments (section Procedure and Apparatus) that were separated between the two room conditions. All possible conditions in each room (Stimulus x Headphone Presentation Mode) were performed in random order. A part of the subjects conducted the experiments at both sites to assess possible lab-specific effects and reveal potential errors more easily. **Table 1** shows a summary of all conditions.

Procedure and Apparatus

The loudness matching experiment was implemented as a 1-up-1-down alternative forced choice paradigm (Levitt, 1971; Kollmeier et al., 1988) implemented in the AFC toolbox (Ewert, 2013). At all times, the subjects were aware whether the sound was presented from the headphones or the loudspeaker, and they saw their surroundings including the

[1] Kohnen, M., Denk, F., Llorca-Bofí, J., Kollmeier, B., and Vorländer, M. "Cross-site investigation on head-related and headphone transfer function measurements: Implications on loudness balancing," to be submitted to Acta Acustica.

TABLE 1 | Keys and description for each condition.

Room	**OL_anechoic**	**AC_anechoic**	**OL_earpiecelab**	**AC_teakitchen**
	Oldenburg Virtual Reality lab	Aachen hemianechoic chamber	Oldenburg shoebox-shaped sound isolated lab room T_{30} = 0.395 s	Aachen, non-shoebox shaped room, former tea kitchen T_{30} = 0.574 s
Signal	**tbn250**	**tbn1000**	**tbn4000**	**uen17**
	Third-octave-band noise, center frequency 250 Hz	Third-octave-band noise, center frequency 1,000 Hz	Third-octave-band noise, center frequency 4,000 Hz	Broadband Unified Excitation Noise, same energy in 17 auditory filters between 20 Hz and 4 kHz
Headphone Presentation Mode	**Monaural**	**Diotic**	**IC matched**	**Uncorrelated**
	Presentation on left ear only	Same sound on both ears	Interaural coherence matched to room	Independent sound samples at both ears

Each row shows the possibilities of the factor denoted in the left column. See main text for more details.

loudspeaker. For each condition, the loudness matching experiment began with loudspeaker presentation of the stimulus. The subjects then put on the headphones and started the headphone presentation by pressing a button on a foot switch. They then indicated whether the presentation on the headphone or the loudspeaker was perceived as louder, and the headphone playback level was adapted accordingly. A foot switch with three buttons ("Continue," "Headphone was Louder," and "Loudspeaker was Louder") allowed the subjects to quick response to instructions presented on a screen positioned on the floor in front of them, while they had the hands free for handling the headphones. The presentation order was alternated between trials, such that repositioning of the headphone was reduced to a minimum. The stepsize of the headphone playback level was reduced from the initial 10 to 5 dB and 1 dB after the first and second upper reversals, respectively. The initial headphone playback level was always chosen such that the loudspeaker was perceived as louder, which – in combination with the large initial stepsize – worked as a "bracketing" of the assumed level of equal loudness, thus reducing any bias produced by the selection of the start level. The median value of the three upper and lower reversals of the headphone playback level during the measurement phase was stored as the resulting equal-loudness level. A typical matching process needed between 10 and 20 comparisons until convergence was reached, which took approx. 1–2 min for each condition and approx. 60–80 min for a full session. Pauses were allowed after each condition and after one-third and two-thirds of the whole experiments were completed. Frozen stimuli were used, i.e., the same waveforms (except level adjustments for the headphone) were presented on each iteration. Sound pressure levels at the eardrum with loudspeaker and headphone presentation at equal loudness were calculated *post hoc* using individually measured transfer functions as described in section Sound Levels at Eardrum.

The loudspeaker was a Genelec 8030 active studio monitor that was mounted in view direction and head height (1.25 m) of the seated subjects at 2.25 m distance. The subjects were instructed to point their heads toward the loudspeaker at least during loudspeaker presentation. The loudspeaker presentation level was set to 65 Phon as per (ISO 226, 2003) for a pure tone at the center frequency of each stimulus (see section Stimuli and Rooms, 1 kHz for the broadband stimulus) to present all stimuli at roughly similar loudness. The loudspeaker presentation level was calibrated using a ½" free-field microphone (46AF, G.R.A.S., Holte, Denmark) pointed at the loudspeaker and mounted at the position of the subject's head. Both the loudspeaker and the headphone were connected to a laptop using an ADI-2 Pro FS sound interface (RME, Haimhausen, Germany) through its line and high-power headphone outputs, respectively.

Also, an experiment assessing the apparent source width of the headphone presentation with respect to the loudspeaker presentation was conducted. To this end, we adapted a graphical user interface originally designed for sound quality assessment (Völker et al., 2018). The interface was shown on a touch screen and consisted of a rating panel with a horizontal scale for the apparent source width ratings and buttons representing the different conditions. The loudspeaker playback served as the reference and could be started by pressing the appropriate button, which was fixed at the center of the panel. Pressing of the three other buttons started playback of the same stimulus over headphones with different interaural coherence (see section Stimuli and Rooms) at levels that were previously determined as equally loud as the loudspeaker playback. Monaural headphone presentation was not included in this experiment. The buttons could be positioned in the panel *via* drag and drop to indicate the apparent source width as compared to the loudspeaker presentation. The panel was labeled with a numerical scale ranging from −50 to 50, supplemented by descriptions (much smaller, smaller, larger, and much larger positioned at −40, −20, 20, and 40, respectively). Thus, negative values here indicate a smaller, 0 an equal, and positive values indicate a larger apparent source width in headphone presentation. A separate run of the interface was started for each of the four signals (see section Stimuli and Rooms). The experiment was only conducted in the non-anechoic rooms, in the same session, and directly after the loudness matching experiments were finished and lasted another approx. 10 min.

Stimuli and Rooms

Four different signals were used. Three signals were one-third-octave-band noises with center frequencies at 250 Hz, 1 kHz, and 4 kHz (referred to as tbn250, tbn1000, and tbn4000 in the following). The fourth was a broadband noise with equal energy in each of 17 critical frequency bands as defined by Zwicker (1961) in a frequency range between approx. 250 and 4 kHz, i.e., the same lower and upper boundary frequency as the narrowband noises. The signals were chosen to capture frequency regions with a high (250 Hz), intermediate (1 kHz), and low (4 kHz) ability of the human auditory system to integrate the temporal fine structure across the two ears (Moore, 2012). Also, differences between narrow-band sounds that fall within one auditory filter and broadband sounds can be characterized. In contrast to many other studies on loudness, the temporal envelope of the one-third-octave-band stimuli was not flattened (Kohlrausch et al., 1997) to facilitate manipulations of the interaural coherence in headphone presentation. The signals were 1 s in duration including 20 ms long rise and fall ramps. All level calculations excluded the ramps and possible reverberant tails.

Four different headphone presentation modes were employed for the headphone presentation:

- Monaural: presentation on left ear only.
- Diotic: same signal presented on both ears, interaural coherence = 1.
- IC matched: Interaural coherence matched to loudspeaker presentation.
- Uncorrelated: Independent noise samples presented on both ears, interaural coherence = 0.

Binaural stimuli with arbitrary interaural coherence were created by adding two independent noise samples with appropriate weights (symmetric generator method, Hartmann and Cho, 2011). The signal presented on the loudspeaker was always identical to the signal presented to the left ear over the headphone. The interaural coherence with loudspeaker presentation was determined using a KEMAR 45BB-12 mannequin with anthropometric pinnae and low-noise ear simulators (G.R.A.S., Holte, Denmark). The interaural coherences for third-octave band and the uen17 stimuli and rooms including observed standard deviations across several positions in a 20 cm radius around the reference position of the head are shown in **Figure 1**.

The experiments were conducted in an anechoic chamber and one office-like non-anechoic room at both sites in Oldenburg (OL) and Aachen (AC). At Oldenburg, a full anechoic chamber sized 8.6 m × 5.8 m × 5.5 m with 0.6 m foam wedge absorbers and a setup of 94 loudspeakers was used (OL_anechoic). While the loudspeakers generate mild reflections in the mid frequency range (Denk et al., 2018b), the reverberation time is still below 60 ms above 100 Hz. The non-anechoic room in Oldenburg was an isolated lab within a room with a shoe box shape (5.15 m × 3.85 m × 3.5 m) and a T_{30} reverberation time of 0.395 s (OL_earpiecelab). At Aachen, a hemianechoic chamber with a rigid floor of size 11 m × 5.97 m × 4.5 m and 0.8 m wedge length was used (AC_anechoic). The reflection from the floor was additionally attenuated through a 0.5-m foam wedge absorber layer laid out on the floor between the subject and the loudspeaker. The non-anechoic room in Aachen is the institute's old tea kitchen, which is non-shoebox (higher ceiling at approx. 1/3 of the ground area) with a ground area of approx. 2.7 m × 5 m, and average height of approx. 3 m, and a T_{30} reverberation time of 0.540 s (AC_teakitchen). In both non-anechoic rooms, the subjects and the loudspeaker were positioned asymmetrically to decorrelate the signals at both ears. The distance from the loudspeaker was at least a factor of four larger than the reverberation radii (OL_earpiecelab: 0.5 m, AC_teakitchen: 0.32 m, using Sabine's formula), i.e., the level of the reverberant sound field dominates at the position of the subjects. Room acoustic parameters were determined using the loudspeaker used in the loudness matching experiments and a free-field microphone (46AF, G.R.A.S., Holte, Denmark) positioned at the location of the subjects' head.

In the OL anechoic chamber, one of the installed Genelec 8030 loudspeakers was connected to the experimental laptop. This loudspeaker was mounted on a traverse system that was ultimately mounted on the supporting steel beam structure at the ceiling of the chamber. In all other rooms, the loudspeaker was mounted on a microphone stand on the floor.

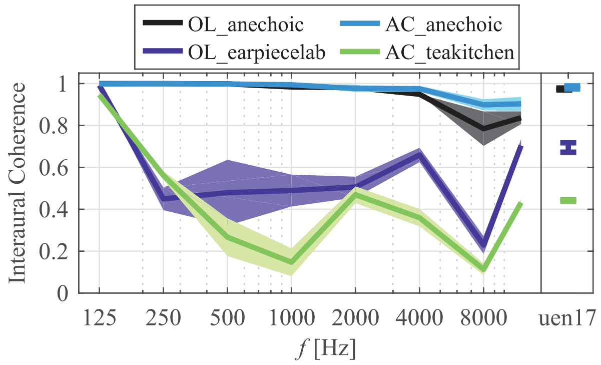

FIGURE 1 | Measured interaural coherence values of the rooms, measured in third octave bands, and for the broadband uen17 stimulus using a KEMAR. Thick lines incdicate average values and shaded areas indicate standard deviations across several positions around the reference head position.

Subjects and Characterization

Forty normal-hearing subjects (27.6 ± 7.2 years of age, half male and female, including three authors) participated in the study. Fourteen subjects (gender-balanced) went through the measurements at both sites, and additional 13 subjects were measured at each site, amounting to a total of 27 subjects measured at each site. The 14 subjects that went through the identical experiments at both sites allowed for a direct comparison of results and served to reveal any lab-specific differences.

Pure-tone audiometry with extended high frequencies was performed using an automated method (Bisitz and Silzle, 2011) to verify that the subjects had normal hearing. Subjects were excluded if their threshold exceeded 20 dB HL at one single audiometric frequency up to 8 kHz, or 35 dB HL at 12.5 or 16 kHz. For subjects participating in Oldenburg, further auditory characterization was conducted. This included the assessment of monaural and binaural loudness growth functions for the stimuli of the present study using the adaptive categorical loudness scaling method (ACALOS; Brand and Hohmann, 2002). Note that in the loundess growth function experiment, the narrowband stimuli had an optimized temporal envelope with minimal temporal level variations but the same spectrum ("low-noise noise"; Kohlrausch et al., 1997). Also, the SNR at 50% speech intelligibility (SRT50) was determined for a frontal speech source and noise at the front or the right, both with the left ear only and binaurally using the Oldenburg sentence test (Wagener et al., 1999). The subjects conducted all measurements autonomously using the Oldenburger Measurement Applications (Hoertech, 2019) with appropriate GUIs and using HDA300 audiometric headphones (Sennheiser, Wedemark, Germany).

Sound Levels at Eardrum

The sound pressure levels at the eardrum of the subjects were calculated *post hoc* using individually measured transfer functions. That is, the levels during headphone presentation were calculated by convolving the headphone stimulus (voltage at a level that produced equal loudness as free-field presentation) with individual HpTFs. The levels at eardrum during loudspeaker presentation were computed by convolving the loudspeaker stimulus (pressure waveforms at free field, known by calibration) with individual HRTFs. The transfer function-based calculation has the benefit that the same transfer function can be used for multiple conditions and sessions. Also, in transfer functions, it is easier to recognize faulty measurements (e.g., spurious notches due to placement too far away from the eardrum) and eliminates those from further calculations than in direct measurements of narrow-band sound pressures at the eardrum. We verified the transfer function-based approach against direct measurements of the stimuli in all rooms using the KEMAR.

The transfer functions to the eardrum were measured using probe tube microphones (ER7C, Etymotic Research, Elk Grove Village, IL, United States). The probe tubes were inserted into the ear canal until the subject reported contact with the eardrum, and then pulled back by a minimal amount and fixed at the check using medical tape. Comparatively, long probe tubes of 76 mm length (Type 76109MBB, Precision Cast Plastic Parts, Redding, CA, United States) were used, such that it was possible to place the body of the probe microphone outside of the headphone cushion to avoid leaks. Transfer function measurements were conducted in the anechoic chambers at each site. The transfer functions of the 14 subjects participating at both sites were measured at both sites, and for level calculations the transfer functions measured at the site of the appropriate room were utilized.

The HpTF was measured eight times using exponential sweeps including repositioning of the headphone to account for known variabilities (Kulkarni and Colburn, 2000; Müller and Massarani, 2001). In the frequency range of interest here (0.25–4 kHz), the typical standard deviation lies around 3 dB between subjects and 1 dB within one subject. The within-subject variations are in the same range as reported by Völk (2014) for 50 repetitions, showing that the eight repetitions employed here are sufficient to capture the variations that also occur during the listening tests when the subjects put the headphones off and on. The stored stimulus waveform that was presented during the psychoacoustic experiment was convolved with each instance of the HpTF, the RMS calculated for each ear and HpTF instance separately, the RMS values averaged and then transformed to dB SPL. The random variations of the HpTF included in the listening test, which contribute to the overall uncertainty of the results, are thus included in the level calculation procedure. For the monaural presentation mode, only the ear, where sound was presented, was regarded. HpTFs measured at both sites for the 14 cross-site subjects are shown in **Figure 2**, and a good correspondence between sites especially up to 4 kHz demonstrates a high data quality. Note that different headphones bought in one batch were used at both sites.

The transformation from free field to the eardrum of the subject for a specific incidence direction is defined by the

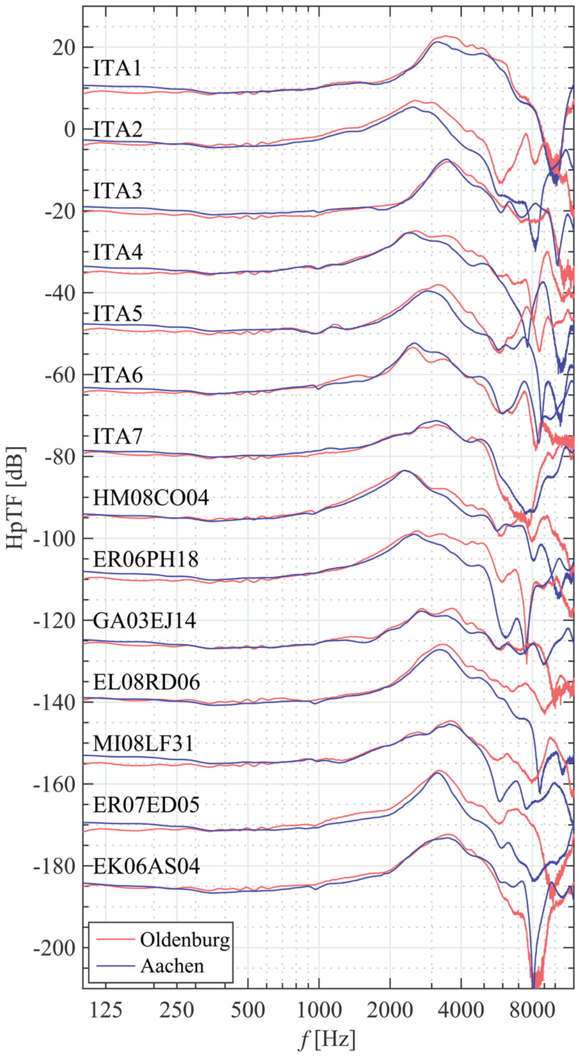

FIGURE 2 | HpTFs (power spectrum averages across eight repetitions, right ear) of the subjects that went through measurements at both sites. A good correspondence below 5 kHz verifies the validity of measurements at both sites in the frequency range of interest. Individual curves have been shifted in increments of −15 dB with respect to the top one for better display.

HRTF. HRTFs were measured for each subject in 87 and 3072 directions in Oldenburg and Aachen, respectively, using the techniques described in Denk et al. (2018a) and Richter (2019) in the same session as the HpTFs without repositioning of the probe tube. HRTFs for frontal incidence of the subjects that went through measurements at both sites are shown in **Figure 3**, and again a good correspondence demonstrates a generally high data quality. One subject participating only in Aachen had to be excluded due to a faulty HRTF measurement that could not be repeated due to the Corona pandemic.

For loudspeaker presentation, the level at free field at the location of the subject's head is known by calibration. In case of the anechoic chambers, a stimulus at eardrum and its corresponding level can thus be calculated by convolving the loudspeaker stimulus with the HRTF for frontal incidence. In the non-anechoic rooms, sound is reflected from the walls, the ceiling, and the floor, such that the sound field includes incidence from many other than the frontal incidence direction. This room-specific effect was approximated by a weighted average of the magnitudes of individual HRTFs for free- and diffuse-field incidences, representing the direct sound from the loudspeaker and the diffuse room reverberation. The individual diffuse-field HRTF was approximated by power spectrum averaging a subset of HRTFs uniformly distributed in space (Denk et al., 2018a). The weight between free- and diffuse-field incidence was adapted to each room (including anechoic chambers) to match KEMAR measurements of the stimuli level at eardrum in this room. This weight was used for each subject to compute a "room-matched HRTF" from individual free-and diffuse field HRTFs. This simple model matched the measured data with an accuracy of ±1 dB for the frequencies of interest, except for the AC_teakitchen. In this room, a prominent early reflection limited the accuracy of this model. The room-matched HRTF for this room was thus extended by an additional heuristic correction, which comprised the difference between estimated and measured levels in KEMAR.[2] The measured levels in KEMAR together with KEMAR's frontal- and diffuse-field HRTF and the weighted average are shown in **Figure 4**.

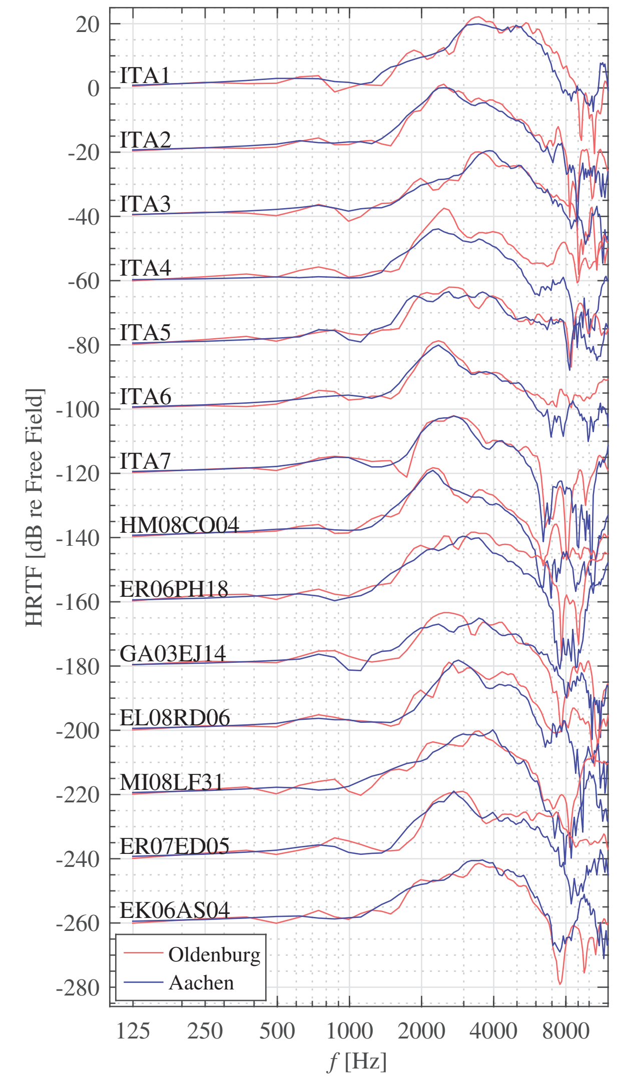

FIGURE 3 | HRTFs for frontal incidence in the left ear of the subjects that went through measurements at both sites. A good correspondence below 8 kHz verifies the validity of measurements at both sites in the frequency range of interest. Individual curves have been shifted in increments of −20 dB with respect to the top one for better display.

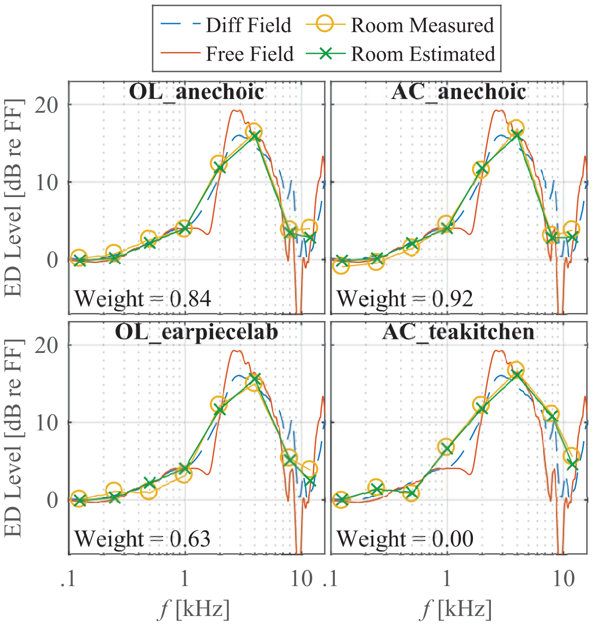

FIGURE 4 | Third-octave noise levels at eardrum (ED) with respect to free field measured in KEMAR (circles). Free- and diffuse-field responses are shown as solid red and dashed blue lines, respectively; estimated levels are free field obtained from KEMAR HRTFs and room-specific weighting factors are depicted as green crosses.

[2] Additional corrections in AC_teakitchen: 250 Hz: +0.8 dB; 1 kHz: +2.5 dB; 4 kHz: +1.5 dB.

RESULTS

Level Mismatch at Equal Loudness

Figure 5 shows the observed mismatch (headphone level minus loudspeaker level at eardrum at equal loudness in each subject) separately for each room, stimulus, and headphone presentation mode. For each condition, i.e., the combination of room, stimulus, and headphone presentation mode, the statistical significance of the difference from a mean of zero was assessed by *t*-tests including a Bonferroni correction for 64 paired comparisons. Statistically significant differences ($p < 0.05$) are marked by stars below the appropriate error bar in **Figure 5**. Aside from the conditions with monaural headphone presentation that are further assessed in section Binaural Parameters and Level Mismatch, a significant mismatch in the range of 3–6 dB is generally observed for the tbn250 stimulus. For the tbn1000, a significant mismatch is noted in all rooms with diotic headphone presentation, and with all headphone presentation modes in the AC anechoic chamber. For the tbn4000 stimulus, a significant mismatch is only observed in the AC anechoic chamber and room-matched and uncorrelated headphone presentation, although a trend towards a mismatch is also visible for this stimulus and diotic headphone presentation in both AC rooms. For the broadband uen17 stimulus and either binaural headphone presentation mode, no mismatch is observed.

Factors influencing the mismatch were further analyzed by means of a three-way ANOVA with the factors Room, Stimulus, and Headphone Presentation Mode.[3] Significant effects were revealed for all factors [Room: $F(3, 1{,}630) = 25.1$), $p < 0.001$;

[3] Note that this analysis grouped the cross-site subjects that participated in all rooms with the other subjects participating only at both rooms of one site. Although data independence between factors is not strictly given, we expect no effects on the statistical outcomes. This is supported by the observation that equivalent outcomes were obtained by running a repeated-measures ANOVA on the results with cross-site subjects only.

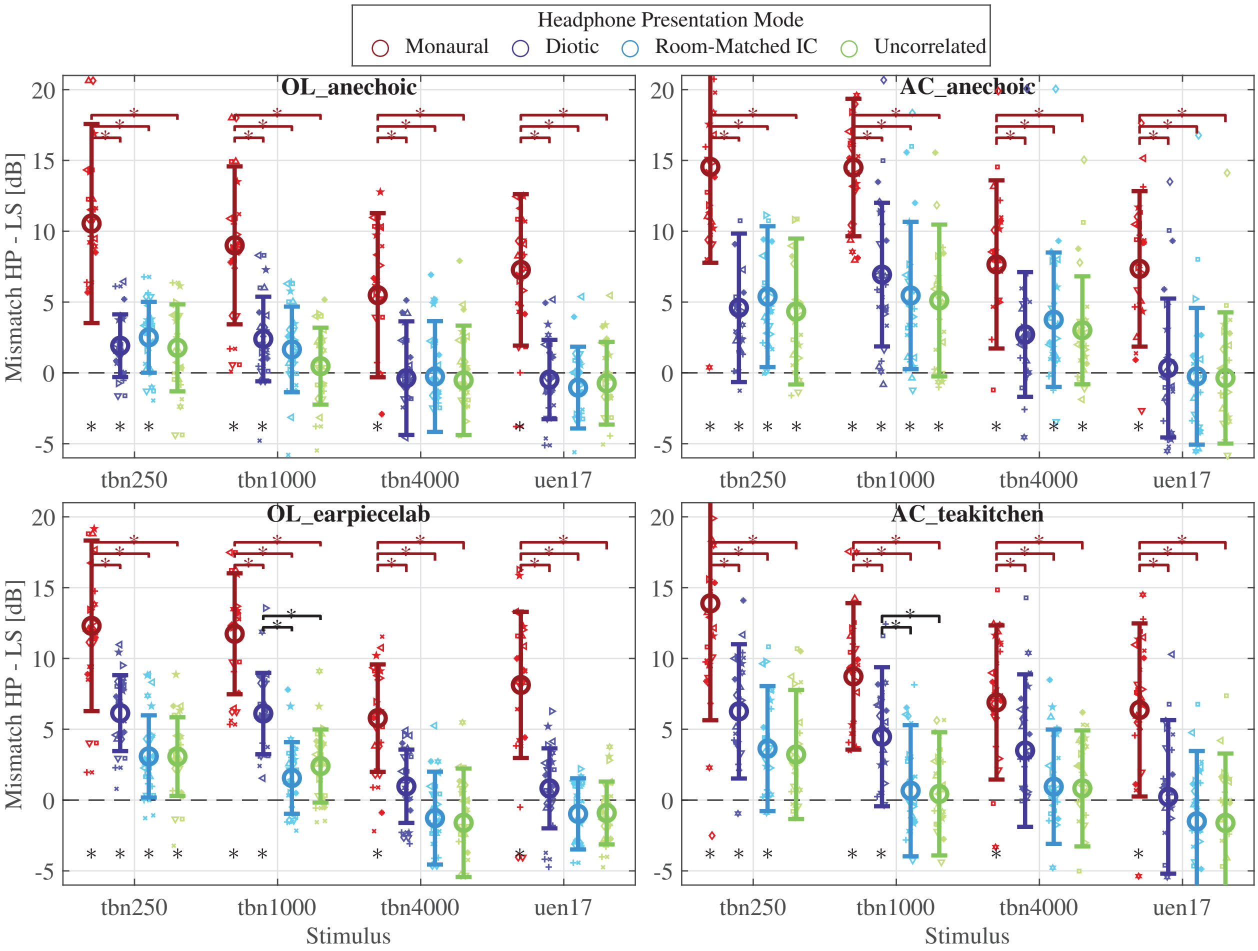

FIGURE 5 | Mismatch of levels at eardrum at equal loudness, headphone level minus loudspeaker level. Results for each room are shown in individual panels, the position on the *x*-axis denotes the stimulus, and the color denotes the Headphone Presentation Mode. Small symbols denote individual subjects, large symbols denote the mean, and error bars denote the standard deviation across subjects for each condition. A star at the bottom denotes a statistically significant mismatch for this condition, and stars above brackets indicate a significantly different mismatch between conditions connected by the bracket.

Stimulus: F(3, 1,630) = 103.0), $p < 0.001$; Headphone Presentation Mode: F(3, 1,630) = 322.9, $p < 0.001$], as well as all possible 2-way interactions [Stimulus × Headphone Presentation Mode: F(9, 1,630) = 2.6, $p < 0.001$; Stimulus × Room: F(3, 1,630) = 6.3, $p < 0.001$; Room × Headphone Presentation Mode: F(3, 1,630) = 2.7, $p = 0.003$]. The three-way interaction term was not significant [F(27, 1,630) = 0.358, $p = 0.99$].

As revealed by the ANOVA explicitly visible in the marginal means of the rooms as shown in **Figure 6**, the mismatch differs between rooms. These differences were assessed by means of a *post hoc* test on the marginal distributions for all subjects including a Bonferroni correction for six paired comparisons. An appropriate evaluation of the cross-site subjects' data that is shown for comparison in **Figure 6** yielded equivalent statistical results. On the one hand, significant differences between both anechoic chambers [$\Delta = 2.62 \pm 0.30$ dB (mean difference ± standard error), $p < 0.001$] with higher mismatch values in the Aachen chamber are noted. On the other hand, no significant difference is seen between the non-anechoic rooms at both locations ($\Delta = -0.33 \pm 0.31$ dB, $p = 1$). At Oldenburg, a larger mismatch is seen in the non-anechoic rooms than in the anechoic chamber (OL: $\Delta = 1.12 \pm 0.30$ dB, $p = 0.001$), while in Aachen the mismatch values are larger in the anchoic chamber ($\Delta = 1.76 \pm 0.31$ dB, $p < 0.001$). The mismatch was generally larger in the AC anechoic chamber as compared to the OL non-anechoic room ($\Delta = 1.42 \pm 0.31$ dB, $p < 0.001$), while the mismatch was smaller in the OL anechoic room as compared to the Aachen non-anechoic room ($\Delta = -0.86 \pm 0.31$ dB, $p = 0.03$).

Differences in the mismatch between stimuli are rather consistent between Headphone Presentation Modes in each room but differ between rooms. In both Oldenburg labs, the observed mismatch is very similar between the tbn250 and tbn1000, and larger in these two stimuli than with the tbn4000 or broadband uen17, where no mismatch is evident (except for monaural headphone presentation). In the AC_anechoic chamber, mismatches were slightly larger with the tbn1000 stimulus than with the others, while in the AC_teakitchen, only a minor dependence on the stimulus is seen. Common to all rooms is that no mismatch is observed with the broadband uen17 stimulus presented binaurally. Significant differences between marginal means of the stimuli were observed in all possible comparisons.

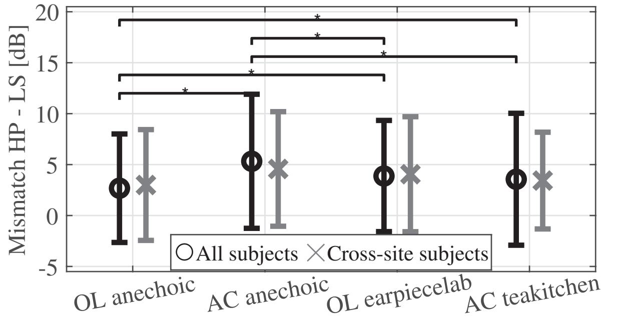

FIGURE 6 | Marginal means of mismatch observed in all rooms, mean, and standard deviation shown for all subjects (black symbols), and the subset of subjects that participated at both sites (cross-site subjects, gray symbols).

Differences between Headphone Presentation Modes were assessed within each combination of Stimulus and Room (as grouped in **Figure 5**, stars above bracket between conditions indicates $p < 0.05$) by pairwise t-tests with Bonferroni correction. First, little surprisingly there is a significant difference between monaural vs. all binaural headphone presentation modes. Second, in the non-anechoic rooms (OL_earpiecelab and AC_teakitchen), there is a tendency that the mismatch is larger in diotic vs. room-matched or uncorrelated headphone presentation, irrespective of the stimulus. However, this trend only reaches significance for the tbn1000 stimulus. This influence of the interaural coherence is exclusively seen in the non-anechoic rooms, i.e., where the interaural coherence is also considerably different from 1 with loudspeaker presentation (cf. **Figure 1**). Third, no considerable trends or significant differences are seen between uncorrelated and room-matched headphone presentation. Further evaluations regarding the influence of interaural coherence of the headphone presentation is given in section Binaural Parameters and Level Mismatch.

Binaural Parameters and Level Mismatch

In **Figure 5**, it is evident that especially in the non-anechoic rooms, a reduction of the interaural coherence in binaural headphone reproduction, on average, reduces the mismatch with respect to diotic presentation. **Figure 7** shows the individual correspondence of the mismatch with diotic and room-matched IC headphone presentation, separated for the different rooms and stimuli. High and significant correlations are seen between the mismatch results with both headphone presentation modes within the subjects. In the anechoic chambers, where the IC is very close to 1 (cf. **Figure 1**), thus diotic and room-matched headphone presentation are very similar, the results are centered around the diagonal and highly correlated, i.e., the mismatch was repeatable. In the non-anechoic rooms (OL_earpiecelab and AC_teakitchen), the distributions have an offset to the top of the diagonal, i.e., also for the individual level, the mismatch is generally larger with diotic presentation. The high correlation shows that, while the general size of the mismatch seems to be a rather individual quantity, the reduction of mismatch by adaptation of the interaural coherence to the room seems to be a factor that is consistent across subjects.

No links of the reduction of mismatch between headphone presentation modes to individual abilities to integrate across ears were found. Correlation analysis of the mismatch differences with the benefit of adding the worse ear in a spatially separated Speech-in-Noise task or difference between monaural and diotic categorical loudness growth functions (cf. section Subjects and Characterization) did not reveal any dependences on the individual level.

The equal-loudness levels are approx. 5–9 dB larger with monaural vs. binaural headphone playback, which obviously

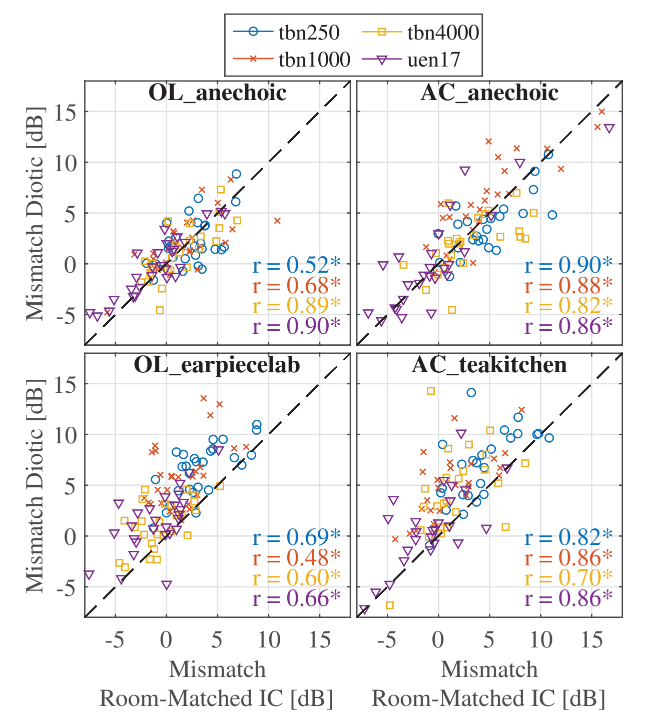

FIGURE 7 | Individual mismatch to loudspeaker presentation with room-matched (*x*-axis) or diotic (*y*-axis) headphone presentation. Each symbol denotes the result for one subject, and different stimuli are indicated by colors and symbols. The Pearson correlation coefficient for each stimulus is given by the values in the lower right corner of each panel, and a star behind the value denotes a statistically significant correlation.

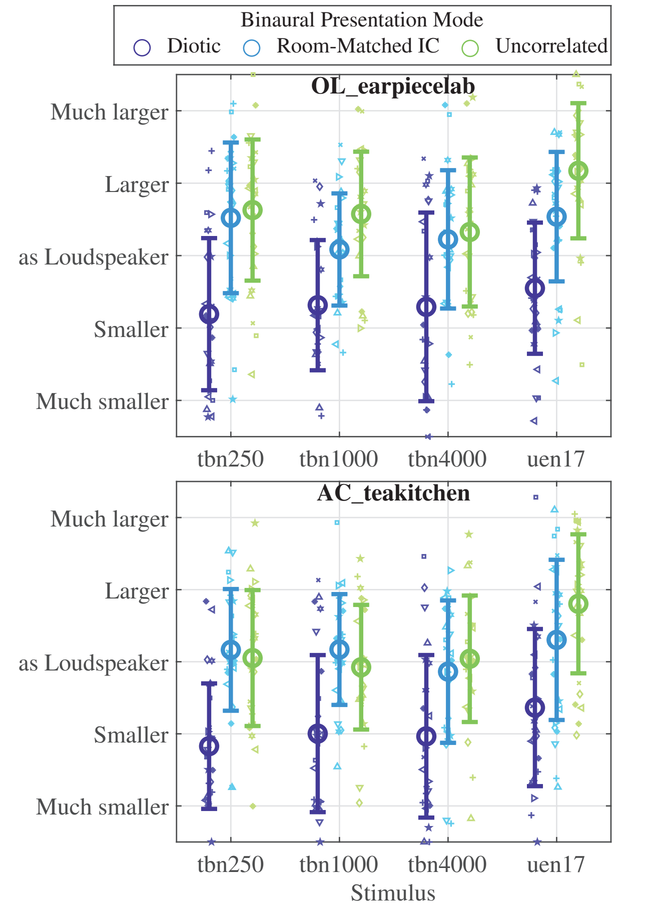

FIGURE 8 | Apparent Source Width ratings for the stimuli presented over headphones with different modes (colors), separated across stimuli and the two non-anechoic rooms. Small symbols denote subjects' ratings, and large symbols and error bars denote the average and standard deviation, respectively.

relates to the well-known effect of binaural loudness summation (Marks, 1978; Edmonds and Culling, 2009; Oetting et al., 2016). Similar to the effect of the interaural coherence, no correlation between the difference between monaural and binaural results from **Figure 5**, and difference in monaural and diotic loudness growth functions was seen. However, it should be noted that the equal-loudness level difference between monaural and diotic presentation seen here is larger than the typically reported effect of binaural loudness summation in headphone experiments, which lies around 3–6 dB (Edmonds and Culling, 2009).

Apparent Source Width and the Level Mismatch

Figure 8 shows the apparent source width ratings for the headphone presentation in the two non-anechoic rooms. The ratings are very similar between rooms and stimuli. The diotic stimulus presentation was generally perceived as smaller than the loudspeaker and headphone presentation with the room-matched or zero interaural coherence. For the narrowband stimuli (tbn250, tbn100, and tbn4000), the apparent source width was rated very similarly between the room-matched and uncorrelated conditions. At the same time, the apparent source width of uncorrelated and room-matched headphone presentation was rated very similar to that of the loudspeaker in the AC_teakitchen but a bit larger than the loudspeaker in the OL_earpiecelab. Only for the broadband uen17 noise, the uncorrelated headphone playback was perceived as larger than the room-matched playback. Reduction of the interaural coherence below that of the loudspeaker presentation thus led to an apparent source width that is larger than both the loudspeaker and the room-matched headphone presentation.

The apparent source width ratings are well in line with the mismatch between loudspeaker and headphone presentation: on average, headphone stimuli that were perceived as smaller also elicited a higher mismatch (diotic vs. room-matched, cf. section Level Mismatch at Equal Loudness). However, while the influence of the interaural coherence on the mismatch is smaller at high frequencies (tbn4000) or for the broadband noise (uen17), no such dependence is seen for the apparent source width ratings. No significant correlations between individual judgments of apparent source width and the mismatch were noted, however, this may be caused by the large variance of the apparent source width judgments.

DISCUSSION

Strengths and Limitations of the Current Study

The current study provides a rich dataset of loudness matching experiments with up to 40 subjects, four different rooms across two lab sites, four different binaural conditions, and four different signals that is unparalleled so far in the literature. While the investigation included three different headphone models, in the present work, only the results for the open-coupling Sennheiser HD650 are shown. Without pre-empting on the companion paper (see Footnote 1), it should be stated here that the main outcomes of the present work are no different for the other headphones.

The fact that individual HRTFs and HpTFs were recorded for each subject provided the possibility for an estimation of the mismatch in each condition that takes into account individual sound transfer characteristics of the ears both for the headphones and the loudspeaker. Contrary to the headphones and anechoic chambers, in the non-anechoic rooms, the transfer function comprises not only the measured direct transfer path between loudspeaker and eardrum, but also numerous reflections from different incidence directions and delays. This so-called binaural room transfer function was not directly measured, but modeled as a superposition of direct sound and reverberant field, where the weights of both components were determined for KEMAR and used for all subjects. While this approximation of the complex transfer behavior includes the effect of individual ear properties, it is still possible that errors in the estimated level at eardrum are introduced due to an oversimplification of the sound field. The additional heuristic correction necessary in the non-anechoic room in Aachen (cf. section Sound Levels at Eardrum), which was derived from differences between the originally estimated and measured levels for this room, gives a first estimate of the introduced accuracies. By doubling this correction, we estimate a worst-case error due to this approximation of around 3 dB. However, there is no reason why this inaccuracy should not be evenly distributed across subjects. Therefore, we assume that this estimation may lead to an increased uncertainty of the levels at eardrum for loudspeaker presentation in the non-anechoic rooms, but not to a change of the average mismatches observed.

In spite of the post-hoc compensation of individual transmission effects, sound presentation did not include any individual HpTF-compensation across frequency, but used the inherent free-field equalization of the headphones employed here. While for the three narrowband stimuli, it can be assumed that this approximation of the desired frequency response suffices to match the stimulus spectrum using headphones to that of the loudspeaker presentation, this is not the case for the broadband stimulus uen17, where coloration differences might interfere with the loudness matching task between loudspeaker and headphone presentation. However, this broadband stimulus provided the least mismatch across all conditions (cf. **Figure 5**), indicating that the spectral approximations during the measurement procedure do not interfere with the interpretability of the data. Nevertheless, future experiments should also perform the individual equalization of the headphones already during the measurements with broadband stimuli to test any potential influence of coloration artifacts and connected spatial cues on the loudness mismatch.

Occurrence and Size of the Mismatch: Diotic Headphone Presentation

With diotic headphone presentation, a significant mismatch of 3–7 dB higher level at eardrum with headphone as compared to loudspeaker presentation was consistently seen for narrow-band sounds at frequencies lower than 4 kHz. The mismatch occurred both in anechoic and non-anechoic conditions and was in each room very similar in size for the stimuli at 250 and 1,000 Hz. Our data hence confirm previous studies, e.g., Munson and Wiener (1952) and Keidser et al. (2000), indicating that for low frequencies up to 1 kHz a significant mismatch exists, albeit slightly smaller than the 6–8 dB reported previously. For the 4 kHz stimulus, no significant mismatch was observed in either room, although there is a tendency toward a mismatch of approx. 3 dB in both AC rooms, which is discussed below. For broadband stimuli, our results show very clearly that there is no mismatch, specifically confirming results by Brinkmann et al. (2017) who used binaural synthesis instead of diotic headphone playback.

For diotic headphone presentation and frequencies below 4 kHz, the occurring mismatch is smaller in the OL anechoic lab as compared to the other rooms (approx. 3 vs. 6 dB). While at OL, the mismatch is larger in the non-anechoic room, at AC, the mismatch values were similar in anechoic and non-anechoic conditions. Between the anechoic chambers at both sites, we see a striking and statistically significant difference of approx. 3 dB for all narrowband stimuli (incl. 4 kHz) and diotic headphone playback. These differences are also significant for our subset of 14 subjects who performed the experiment at both labs. Faulty calibration of equipment as a source of the difference between sites can be mostly ruled out due to the consistently non-existent mismatch with the broadband stimulus, and given the good correspondence between sites in the non-anechoic rooms. However, small differences in the experimental setup were unavoidable between the anechoic rooms in Oldenburg and Aachen (cf. section Stimuli and Rooms). Room acoustical consequences of these small differences included a smaller interaural coherence in the OL anechoic room (**Figure 1**), and a potential floor reflection in the AC anechoic room despite laying out absorbers on the floor. Vibration (Rudmose, 1982) might have a lower influence in the OL anechoic chamber due to the loudspeaker mounting on traverse system as opposed to a stand on the floor in the other rooms, although we do not expect vibration to reach a significant level in general. It should be stressed that all these acoustic factors would cause frequency-specific effects, while the observed difference in mismatch is very consistent across narrowband stimuli. Another possible explanation for the difference between anechoic rooms are non-acoustic differences such as the general impression of the room and visual cues like seeing one vs. many loudspeakers. While the potential influence of visual cues on loudness judgments is well-known, the presence of a total of 94 spatially separated loudspeakers in the OL anechoic lab is a feature of the experimental room that cannot be easily changed. The influence of such non-acoustic factors on the loudness mismatch should therefore be assessed in future experiments, e.g., by blindfolding subjects of providing different visual cues on a head-mounted display.

Contrary to the anechoic rooms, the mismatch results between the two non-anechoic rooms are quite consistent between both sites. These rooms were visually rather similar (single loudspeaker mounted in empty room) but acoustically different (T30 = 0.4 s vs. 0.57 s), although one may argue that the perceived difference between these rooms may be smaller than deviations from anechoic properties in one of the anechoic chambers. Altogether, we discard our hypothesis H1 and must conclude that small differences in the setup of the especially anechoic test conditions may lead to a considerable difference in obtained mismatch between diotic headphone and loudspeaker presentation. This might also explain the inconsistent reports from the literature about the (non-) observation of this mismatch since the experiments were all performed in somewhat different room conditions (Munson and Wiener, 1952; Rudmose, 1982; Fastl et al., 1985; Völk et al., 2011; Bonnet et al., 2018).

Influence of the Headphone Presentation Mode and Apparent Source Width

The interaural coherence in headphone presentation (diotic vs. room-matched/uncorrelated IC) exhibits a significant influence on the obtained loudness mismatch in the non-anechoic rooms, but virtually no difference in the anechoic chambers (cf. **Figure 5**). Also, no difference is evident between room-matched and uncorrelated headphone presentation. In the present data, the trend toward a difference in mismatch between diotic and room-matched/uncorrelated presentation in the non-anechoic rooms amounts up to 5 dB and is visible for all stimuli including the broadband sound. However, it only reaches significance for the 1,000 Hz narrowband stimulus (cf. **Figure 5**). The trend to smaller mismatches with uncorrelated headphone presentation in the non-anechoic rooms is consistent with results of Edmonds and Culling (2009), who reported lower levels in uncorrelated vs. diotic headphone presentation at equal loudness. The effect in their data was slightly smaller (up to 3 dB in size) and declined toward high frequencies and large bandwidth similarly to our data. However, given their data, it is quite surprising that the interaural coherence of the headphone presentation does not influence the mismatch to the loudspeaker – if the diotic/room-matched headphone playback (IC≈1 in the anechoic chambers) would have been directly compared with uncorrelated playback, a lower level at equal loudness would have been expected with the uncorrelated presentation. In conclusions, hypothesis H2 (influence of binaural presentation mode) can be supported for non-anechoic, but not for anechoic environments. Hypothesis H2a (matching the IC eliminates mismatch) has to be rejected: A mismatch was still significant with room-matched interaural coherence in all conditions where it was significant with diotic presentation, albeit reduced in non-anechoic conditions.

A closer look into the individual variations of mismatch in the diotic vs. room-matched IC conditions (section Binaural Parameters and Level Mismatch, **Figure 7**) indicated a high correlation across subjects in both conditions, i.e., individuals exhibiting a high mismatch in the diotic condition most often also show a high mismatch in the room-matched IC condition. This provides further evidence that the individually reported mismatch is an individual treat, where the exact distribution of the internal spatial impression as controlled by the IC only exerts a small influence. Other factors (e.g., the individual's ability to utilize binaural cues for better speech recognition under spatial talker-interferer conditions, cf. section Binaural Parameters and Level Mismatch) do not appear to have a stronger loading on the individually reported mismatch, thus making a prediction of this individual treat difficult. In other words, matching the IC during headphone presentation consistently reduces the size of the mismatch, while the general size of the mismatch is individual and determined by other factors that we could not identify in the present study in spite of an extensive auditory characterization of the subjects. Hypothesis H3 (individual markers of binaural hearing influences mismatch) thus has to be rejected.

To test hypothesis H2b, i.e., the influence of apparent source width on the mismatch, the relation between apparent source width and IC was analyzed in section Apparent Source Width and the Level Mismatch (**Figure 8**) for the non-anechoic rooms. With the broadband stimulus, the IC of the headphone presentation hardly affects the mismatch, but very clearly the average apparent source width rating. With the narrowband stimuli, the average apparent source width judgments are very consistent with the mismatch results (diotic vs. room-matched IC) – a "smaller" apparent source width as compared to the loudspeaker was associated with an increase of the mismatch by 3–5 dB (cf. **Figures 5, 8**). As for the mismatch, virtually no difference between average source width ratings was seen between the room-matched and uncorrelated conditions. On the individual level, no significant correlation between rated apparent source width and the loudness mismatch were observed. This can probably be attributed to the large variance in the apparent source width data, which may be caused by the rather hard task of comparing the perceived source width of a loudspeaker presentation occupying a certain part of auditory space around the loudspeaker with a headphone presentation that is most probably perceived as distributed somewhere in and around the head. The subjects may have had different internal interpretations of the apparent source width that could lead to much different results in the present experiment, e.g., the estimated absolute size of the source or its angular extent around the head. Also, the rather short stimuli of 1 s may have increased the difficulty of getting a feeling for the spatial characteristics of the different presentation modes.

Altogether, the present data support the hypotheses H2b that the mismatch can be reduced by adapting the interaural coherence during headphone to that with loudspeaker presentation, which also led to similar apparent source width judgments with loudspeaker and headphone presentation. This holds especially for narrowband stimuli in non-anechoic rooms, where diotic headphone presentation elicited a significant mismatch in most cases. With broadband stimuli, appropriate but weaker trends were also visible. Our data generally show that the mismatch is smaller with broadband stimuli, as is the influence of binaural parameters in headphone reproduction on the mismatch in general. We interpret the results as strong indicators of an influence of spatial perception on the mismatch. It cannot be finally concluded from the present data that a difference in spatial perception such as apparent source width, is the cause for a mismatch. However, in previous studies more spatially accurate headphone reproduction methods could avoid mismatch completely (Völk and Fastl, 2011; Brinkmann et al., 2017). While the apparent source width (cf. Rudmose, 1982) is one perceptual attribute of a plausible spatial perception, the present results show that eliciting the same apparent source width in headphone and loudspeaker presentation does not completely avoid the occurrence of a mismatch, especially when considering that the perception of source widths may differ fundamentally between loudspeaker and (unexternalized) headphone presentation. Similarly, the difference in spatial perception is even more different with monaural headphone presentation – which probably explains the difference to the mismatch seen with diotic headphone presentation that exceeded the common size of binaural loudness summation. In addition to the apparent source width, further perceptual attributes like the perceived externalization and distance, location, visual, and other multi-modal cues probably have to be adjusted correctly such that the mismatch disappears in a direct comparison, if the spatial perception is the dominant cause. Future studies should therefore examine the influence of more perceived spatial parameters on the loudness mismatch between headphone and loudspeaker presentation.

Implications for Headphone Studies and Hearing Aid Fitting

The results presented in this study indicate that

1. A substantial level difference at equal loudness up to 15 dB exists for monaural presentation at ear-level vs. loudspeaker presentation to both ears in basically all conditions.
2. The interaural coherence in binaural ear-level presentation (and corresponding apparent source width) has a moderate influence of up to 5 dB on the mismatch in non-anechoic rooms. This effect vanishes in anechoic environments.
3. Small acoustic and/or visual changes in an anechoic reference environment (OL anechoic vs. AC anechoic) exert a moderate effect up to 5 dB on the recorded loudness mismatch, whereas not such a large effect of the respective reference room employed is observed across listening rooms with some reverberation (OL earpiecelab vs. AC teakitchen).

These findings – even though not completely explainable by the yet limited amount of parameter variations performed in this study – have already notable consequences whenever an implication for experiments in the free field has to be drawn from a condition with ear-level hearing devices or vice versa.

For hearing aid fitting, for example, diagnostic and prescriptive measurements (including loudness judgments) are most often performed independently for both ears using headphones, whereas the verification of the fit is performed for loudspeaker-like sources listened binaurally. Hence, the expected value of the loudness difference for monaural vs. binaural presentation and the frequency dependence of the mismatch across different IC conditions might provide a level correction value for the prescriptive "first fit" settings of the hearing device. However, the large variability in the mismatch across normal-hearing subjects and across the two anechoic rooms in this study would lead to the recommendation to be careful about using anechoic rooms for hearing aid fitting. Moreover, extensive fine-tuning should be performed with the hearing-impaired user of the hearing device, who might even show a much higher variability in binaural loudness summation especially for broadband sounds (Oetting et al., 2016).

For headphone studies, virtual acoustic reality is often aimed for by presenting sound signals *via* headphones that should reflect as closely as possible the individual's perception (including loudness perception) in the free field. In applications of augmented reality, sounds from the free field and from ear devices are combined in order to enhance the free-field sound with added virtual sound. It is obvious that loudness perception from those two parts shall be matched. In order to minimize any loudness mismatch, narrowband stimuli should be used with the appropriate interaural coherence and special care has to be administered if non-anechoic conditions are employed. The present results further imply that in general a correct spatial perception of virtual sound sources is required to establish the same loudness at equal level.

CONCLUSION

- The loudness comparisons in headphones and loudspeaker presentation in various environments employed here were combined with individual recordings of the HRTF and HpTF. This allowed for a careful and individual post-hoc quantification of the level mismatch at the eardrum across conditions that exhibit the same loudness.
- A substantial mismatch exists with a high variability across conditions and subjects which is strongly influenced by the presentation mode (monaural vs. binaural headphone presentation with a varying interaural coherence) and by the room acoustic conditions for the loudspeaker presentation. Remarkably, even differences between the anechoic rooms across sites using the same set of subjects were detected that may be due to small, but yet not explainable differences in room acoustics or non-acoustic factors. Such differences across sites did not occur for the tested non-anechoic rooms. Hence, the non-conclusive findings from the literature appear to be related to the experienced disparity between headphone and loudspeaker presentation, where even small differences in (anechoic) room acoustics significantly change the perception and response behavior of the subjects.

- The difference between monaural and binaural presentation during headphone comparisons yields an effect of 10 dB that goes beyond usual values for binaural loudness summation, while another difference of up to 5 dB occurs between diotic, dichotic, and room-matched interaural coherence during headphone presentation. A room-matched interaural coherence reduces the mismatch with respect to diotic presentation in non-anechoic rooms, but does not completely eliminate it.
- Individual factors like loudness summation appear to be only loosely connected to the observed mismatch, i.e., no direct prediction of the mismatch is possible from individual binaural loudness summation.
- Apparent source width coincides well with the differences in IC across diotic, room-matched, and dichotic conditions that do, however, not predict the loudness mismatch in a satisfactory way for broadband stimuli. Hence, other possible perceptual factors like, e.g., perceived distance, size, location; visual, and other multi-modal cues should be considered in future studies.
- Further experiments will have to gain a more detailed understanding by avoiding some of the shortcomings of the current study, i.e., individual binaural synthesis to produce the correct spatial image already during headphone presentation, and a better control of non-acoustic factors like visual cues provided during the experimental conditions.

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 at: https://doi.org/10.5281/zenodo.4153118 (HRTF and HpTF data) and https://doi.org/10.5281/zenodo.4153154 (equal-loudness levels at eardrum, further psychoacoustical results).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethik-Kommission an der Medizinischen Fakultät der RWTH Aachen, EK 351/17. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FD, MK, JL-B, MV, and BK designed the general experiment. FD implemented the psychoacoustic experiments and the HRTF/HpTF measurements in Oldenburg. MK implemented the HRTF/HpTF measurements in Aachen. FD and JL-B conducted the measurements. FD and MK evaluated the data and prepared the figures. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Deutsche Forschungsgemeinschaft (DFG) – Projektnummer 352015383 – SFB 1330 HAPPAA A4.

ACKNOWLEDGMENTS

We thank Meike Renken and Frank Keller-Drees for conducting the probe tube measurements, Bernhard Eurich for assistance during the measurements in Oldenburg, and all subjects for participating. Manan Lamba is acknowledged for a critical review of the manuscript. We also thank Sabine Meunier for providing valuable comprehensive hints, as well as the two reviewers for a thorough review of previous versions of the manuscript.

REFERENCES

Beranek, L. L. (1949). *Acoustic measurements*. New York: John Wiley & Sons.

Bisitz, T., and Silzle, A. (2011). "Automated Pure-tone Audiometry Software Tool with Extended Frequency Range." in *Proceedings of the 130th AES Convention*; May 2011; London, UK.

Bonnet, F., Nélisse, H., and Voix, J. (2018). Effects of ear canal occlusion on hearing sensitivity: a loudness experiment. *J. Acoust. Soc. Am.* 143, 3574–3582. doi: 10.1121/1.5041267

Brand, T., and Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *J. Acoust. Soc. Am.* 112, 1597–1604. doi: 10.1121/1.1502902

Brinkmann, F., Lindau, A., and Weinzierl, S. (2017). On the authenticity of individual dynamic binaural synthesis. *J. Acoust. Soc. Am.* 142, 1784–1795. doi: 10.1121/1.5005606

Denk, F., Ernst, S. M. A., Ewert, S. D., and Kollmeier, B. (2018a). Adapting hearing devices to the individual ear acoustics: database and target response correction functions for various device styles. *Trends Hear.* 22, 1–19. doi: 10.1177/2331216518779313

Denk, F., Kollmeier, B., and Ewert, S. (2018b). Removing reflections in semianechoic impulse responses by frequency-dependent truncation. *J. Audio Eng. Soc.* 66, 146–153. doi: 10.17743/jaes.2018.0002

Edmonds, B. A., and Culling, J. F. (2009). Interaural correlation and the binaural summation of loudness. *J. Acoust. Soc. Am.* 125, 3865–3870. doi: 10.1121/1.3120412

Ewert, S. D. (2013). "AFC - A modular framework for running psychoacoustic experiments and computational perception models." in *Fortschritte der Akustik - DAGA*; March 2013; Meran.

Fastl, H., Schmid, W., Theile, G., and Zwicker, E. (1985). "Schallpegel im Gehörgang für gleichlaute Schalle aus Kopfhörern und Lautsprechern [Sound levels in the ear canal for equally loud sounds from headphones and loudspeakers]." in *Fortschritte der Akustik - DAGA*; March 1985, 471–474.

Hartmann, W. M., and Cho, Y. J. (2011). Generating partially correlated noise—a comparison of methods. *J. Acoust. Soc. Am.* 130, 292–301. doi: 10.1121/1.3596475

Hoertech (2019). Oldenburger Measurement Application R&D. Available at: https://www.hoertech.de/en/devices/oldenburger-measuring-programs.html (Accessed March 16, 2021).

ISO 226 (2003). Acoustics — Normal equal-loudness-level contours. Available at: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/03/42/34222.html (Accessed April 7, 2020).

Keidser, G., Katsch, R., Dillon, H., and Grant, F. (2000). Relative loudness perception of low and high frequency sounds in the open and occluded ear. *J. Acoust. Soc. Am.* 107, 3351–3357. doi: 10.1121/1.429406

Killion, M. C. (1978). Revised estimate of minimum audible pressure: where is the "missing 6 dB"? *J. Acoust. Soc. Am.* 63, 1501–1508.

Kohlrausch, A. G., Fassel, R., Van Der Heijden, M., Kortekaas, R., Van De Par, S., Oxenham, A. J., et al. (1997). Detection of tones in low-noise noise:

further evidence for the role of envelope fluctuations. *Acta Acustica united with Acustica* 83, 659–669.

Kollmeier, B., Gilkey, R. H., and Sieben, U. K. (1988). Adaptive staircase techniques in psychoacoustics: a comparison of human data and a mathematical model. *J. Acoust. Soc. Am.* 83, 1852–1862. doi: 10.1121/1.396521

Kulkarni, A., and Colburn, H. S. (2000). Variability in the characterization of the headphone transfer-function. *J. Acoust. Soc. Am.* 107, 1071–1074. doi: 10.1121/1.428571

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

Marks, L. E. (1978). Binaural summation of the loudness of pure tones. *J. Acoust. Soc. Am.* 64, 107–113. doi: 10.1121/1.381976

Meunier, S., Chatron, J., and Bonetto-Lemaire, T. (2020). "The Missing 6 dB: Perceptual or Procedural Effect?" in *Proceedings of the Forum Acusticum*; December 2020; Lyon, France, 3039–3043.

Moore, B. C. J. (2012). *An introduction to the psychology of hearing*. Bingley, UK: Emerald Publishing Group.

Müller, S., and Massarani, P. (2001). Transfer-function measurement with sweeps. *J. Audio Eng. Soc.* 49, 443–471.

Munson, W. A., and Wiener, F. M. (1952). In search of the missing 6 dB. *J. Acoust. Soc. Am.* 24, 498–501. doi: 10.1121/1.1906927

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., and Ewert, S. D. (2016). Spectral and binaural loudness summation for hearing-impaired listeners. *Hear. Res.* 335, 179–192. doi: 10.1016/j.heares.2016.03.010

Pieper, I., Mauermann, M., Kollmeier, B., and Ewert, S. D. (2016). Physiological motivated transmission-lines as front end for loudness models. *J. Acoust. Soc. Am.* 139, 2896–2910. doi: 10.1121/1.4949540

Richter, J.-G. (2019). Fast Measurement of Individual Head-related Transfer Functions. Available at: http://publications.rwth-aachen.de/record/760168 (Accessed March 16, 2021).

Robinson, D. W., and Dadson, R. W. (1956). A re-determination of the equal-loudness relations for pure tones. *Br. J. Appl. Phys.* 7:166. doi: 10.1088/0508-3443/7/5/302

Rudmose, W. (1982). The case of the missing 6 dB. *J. Acoust. Soc. Am.* 71, 650–659. doi: 10.1121/1.387540

Sivonen, V. P., and Ellermeier, W. (2006). Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. *J. Acoust. Soc. Am.* 119, 2965–2980. doi: 10.1121/1.2184268

Theile, G. (1986). On the standardization of the frequency response of high-quality studio headphones. *J. Audio Eng. Soc.* 34, 956–969.

Völk, F. (2014). Inter- and intra-individual variability in the blocked auditory canal transfer functions of three circum-aural headphones. *J. Audio Eng. Soc.* 62, 315–323. doi: 10.17743/jaes.2014.0021

Völk, F., Dunstmair, A., Riesenweber, T., and Fastl, H. (2011). "Bedingungen für gleichlaute Schalle aus Kopfhörern und Lautsprechern." in *Fortschritte der Akustik - DAGA*; March 2011; Düsseldorf, 825–826.

Völk, F., and Fastl, H. (2011). "Locating the missing 6 dB by loudness calibration of binaural synthesis" in *Audio engineering society convention. Vol. 131*; October 2011. New York, NY, USA: Audio Engineering Society, 1–12.

Völker, C., Bisitz, T., Huber, R., Kollmeier, B., and Ernst, S. M. A. (2018). Modifications of the multi stimulus test with hidden reference and anchor (MUSHRA) for use in audiology. *Int. J. Audiol.* 57, S92–S104. doi: 10.1080/14992027.2016.1220680

Wagener, K., Kühner, V., and Kollmeier, B. (1999). Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. *Z. Für Audiol.* 38, 4–16.

Zahorik, P., and Wightman, F. L. (2001). Loudness constancy with varying sound source distance. *Nat. Neurosci.* 4, 78–83. doi: 10.1038/82931

Zwicker, E. (1961). Subdivision of the audible frequency range into critical bands (Frequenzgruppen). *J. Acoust. Soc. Am.* 33:248. doi: 10.1121/1.1908630

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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 11 May 2021
doi: 10.3389/fpsyg.2021.635557

Continuous Magnitude Production of Loudness

Josef Schlittenlacher[1]* *and Wolfgang Ellermeier*[2]

[1] *Manchester Centre for Audiology and Deafness, Division of Human Communication, Development and Hearing, School of Health Sciences, University of Manchester, Manchester, United Kingdom,* [2] *Applied Cognitive Psychology Unit, Department of Psychology, Technische Universität Darmstadt, Darmstadt, Germany*

Continuous magnitude estimation and continuous cross-modality matching with line length can efficiently track the momentary loudness of time-varying sounds in behavioural experiments. These methods are known to be prone to systematic biases but may be checked for consistency using their counterpart, magnitude production. Thus, in Experiment 1, we performed such an evaluation for time-varying sounds. Twenty participants produced continuous cross-modality matches to assess the momentary loudness of fourteen songs by continuously adjusting the length of a line. In Experiment 2, the resulting temporal line length profile for each excerpt was played back like a video together with the given song and participants were asked to continuously adjust the volume to match the momentary line length. The recorded temporal line length profile, however, was manipulated for segments with durations between 7 to 12 s by eight factors between 0.5 and 2, corresponding to expected differences in adjusted level of −10, −6, −3, −1, 1, 3, 6, and 10 dB according to Stevens's power law for loudness. The average adjustments 5 s after the onset of the change were −3.3, −2.4, −1.0, −0.2, 0.2, 1.4, 2.4, and 4.4 dB. Smaller adjustments than predicted by the power law are in line with magnitude-production results by Stevens and co-workers due to "regression effects." Continuous cross-modality matches of line length turned out to be consistent with current loudness models, and by passing the consistency check with cross-modal productions, demonstrate that the method is suited to track the momentary loudness of time-varying sounds.

Keywords: loudness, time-varying, methods, cross-modality matching, line length, magnitude production

OPEN ACCESS

Edited by:
Sabine Meunier,
Aix-Marseille University – Ecole Centrale Marseille, France

Reviewed by:
Robert Schlauch,
University of Minnesota Twin Cities, United States
Walt Jesteadt,
Boys Town National Research Hospital, United States

****Correspondence:***
Josef Schlittenlacher
josef.schlittenlacher@manchester.ac.uk

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

Received: *30 November 2020*
Accepted: *13 April 2021*
Published: *11 May 2021*

Citation:
Schlittenlacher J and Ellermeier W (2021) Continuous Magnitude Production of Loudness. Front. Psychol. 12:635557. doi: 10.3389/fpsyg.2021.635557

INTRODUCTION

There are numerous methods for the subjective evaluation of auditory stimuli for a variety of purposes. Building upon Fechner's (1860) seminal work describing the three classical methods of threshold measurement, and proposing a rationale for psychophysical scale construction based on just-noticeable differences, transformed up-down methods (Levitt, 1971) have become the gold standard both for determining discriminability, and for adjusting two stimuli to equal sensation. Transformed up-down methods are subject to fewer biases than the classical methods because the task for the participant is rather simple. When evaluating loudness, the question is typically "Which of the two sounds was louder?," and the level of the target stimulus is adjusted before the next presentation of the pair.

Disadvantages of this method are that it needs a reference, that it can only be applied to measure thresholds or points of subjective equality, and that it is time-consuming because determining a point of subjective equality requires several trials.

In contrast, magnitude estimation (Stevens, 1956, 1957, 1975) does not require a reference, can easily cover a large range of stimulus intensities, and yields one estimate of a psychophysical scale value per trial. However, it is prone to biases because the task of scaling is left to the participant (Luce and Mo, 1965; Luce and Krumhansl, 1988). Some of these biases have been extensively studied in the framework of direct magnitude scaling (e.g., Stevens and Galanter, 1957; Teghtsoonian and Teghtsoonian, 1978; Poulton, 1979; DeCarlo and Cross, 1990). Others have been conceptualized within the framework of axiomatic measurement (Narens, 1996; Ellermeier and Faulhammer, 2000; Luce, 2002; Zimmer, 2005) or even Bayesian inference (Petzschner et al., 2015).

A basic check for the consistency of direct scaling outcomes has frequently been to perform magnitude production, which can be seen as the inverse procedure of magnitude estimation (Reynolds and Stevens, 1960): Instead of rating the magnitude of a stimulus, the stimulus is adjusted to match a given estimate. Magnitude production typically yields larger exponents than magnitude estimation, i.e., a smaller level change is needed to e.g., double loudness than magnitude estimates would suggest. Stevens and Greenbaum (1966) explained this phenomenon by "regression effects," which occur whenever two continua are matched in both directions because participants compress the range of the variable that they adjust.

A further opportunity to verify the consistency of scaling procedures is provided by the method of cross-modality matching. A very straightforward case is matching a given sensation with a line length to be produced: Instead of assigning a number to the magnitude of the stimulus, the length of a line is adjusted to match the subjective magnitude (Stevens and Galanter, 1957; Stevens and Guirao, 1963). Stevens and Greenbaum (1966) highlight the similarities between the matching and scaling methodologies by interpreting magnitude estimation as an "instance of the general method of cross-modality matching" (p. 441) to the number continuum.

Cross-modality matching with line length has also been used for continuous judgment of loudness (see Kuwano, 1996, and Kuwano and Namba, 2011, for an overview). Continuous judgment allows us to obtain estimates for the momentary loudness of time-varying sounds, where trial-based methods can only give estimates of the overall loudness of the segments that were presented. Continuous judgment may also be used with the goal to maximize the number of estimates that are obtained per experiment time, somewhat similar to Békésy tracking for obtaining thresholds (von Békésy, 1947).

Continuous judgment of auditory sensations, most commonly loudness, was first done using categories (Namba and Kuwano, 1980; Kuwano and Namba, 1985), with the participants pressing the button for the current category on a response box. Alternative methods used the position of a slider (Fastl, 1991) or cross-modality matching with a muscular force by employing a lever with force feedback (Susini et al., 2002). Several studies used continuous cross-modality matching with line length to track momentary loudness or similar auditory magnitudes, where typically the length of a line that is displayed on a computer screen can be modified by moving the mouse (e.g., Namba and Kuwano, 1990; Kuwano et al., 2003; Kuwano et al., 2014, 2017; Schlittenlacher et al., 2017).

To our knowledge, the methodology of continuous judgment lacks thorough investigation and consistency checks like the ones that have been performed for conventional magnitude estimation or cross-modality matching. To evaluate the consistency of continuous judgment, in Experiment 1, we had participants make continuous cross-modality matches of line length in response to temporally varying loudness patterns of musical songs. In Experiment 2, we inverted the procedure by having participants generate continuous magnitude productions of loudness in response to lines dynamically changing in length. We also analysed the temporal portion on which momentary line length matches are based. In contrast to Kuwano and Namba (1985), we did not use temporal windows with hard cutoffs but varied the exponential time constant in a loudness model (Moore et al., 2018) to find the highest correlation with the momentary line length matches and to evaluate the choice made by the loudness model.

MATERIALS AND METHODS

All participants completed two experiments involving cross-modality matches between loudness and line length. In the first experiment, they continuously adjusted the length of a line to match their impression of loudness. In the second experiment, they performed the reverse operation, i.e., they made magnitude productions of loudness by continuously adjusting sound levels so that their loudness matched dynamically changing line lengths that were displayed simultaneously with the sounds.

Participants

Twenty listeners, eight females and twelve males, participated in the experiments. They were aged 18 to 50 years, with a median age of 23 years. All of them participated in both experiments. Their hearing sensitivity was better than 20 dB HL at each frequency between 125 and 8,000 Hz at both ears. They participated voluntarily without compensation after having given informed consent.

Apparatus

The auditory stimuli were stored as wav files, D/A converted by an RME Hammerfall DSP Multiface II audio interface (Haimhausen, Germany) and presented via Sennheiser HDA 200 headphones (Wedemark, Germany). The participants sat in a double-walled sound-proof booth (IAC, Chandler's Ford, Hampshire, United Kingdom).

Calibration was done according to Richter (2003): The sound pressure level of a 1-kHz tone was measured in a Bruel & Kjaer 4153 coupler with DB-0843 adapter plate. To obtain a free-field level rather than coupler measurement, the difference of −3.5 dB between coupler sensitivity and free-field sensitivity (Table 3.1.3 in Richter, 2003) was added.

The participants used a computer mouse for making their responses. In experiment 1, movement of the mouse to the left or right changed the length of a line that was displayed horizontally on a screen. In experiment 2, the mouse was used to press buttons on the screen. The horizontal screen resolution was 1,280 pixels (px). Line lengths and button presses were recorded using the internal clock of Microsoft Windows XP, which has a rate of 16 ms. For statistical analyses and further processing, line lengths or adjusted levels between the timestamps were upsampled to a rate of 1 ms by linear interpolation.

Stimuli

The stimuli were fourteen excerpts of musical pieces with durations between 142 and 251 s. Their combined duration amounted to 45 min. Seven of the excerpts were from the rock genre and seven were selected from classical music. The distinction between genres was made to have stimuli with little variations in loudness over time (i.e., rock music excerpts) and other ones having a large dynamic range (i.e., the classical music samples). For each excerpt, the two tracks of the stereo file were merged for diotic presentation. The levels were adjusted so that the seven songs in each genre had overall calculated loudness levels (DIN 45631/A1, 2010, N5) of 70, 74, 78, 82, 86, 90, and 94 phon, respectively.

Procedure of Experiment 1

The participants were instructed to continuously adjust the length of a line to match the momentary loudness of the musical excerpt while it was being played: "Please adjust the length of the line by moving the mouse so that it matches your impression of loudness at any time." When participants asked for clarification of "at any time" (German: "zu jeder Zeit"), they were told that it was up to them to define "at any time," and they could form that opinion during three practice trials. The line was depicted horizontally, starting on the left of the screen, having a height of 2 px, and a maximum length of 1,260 px. At the start of a trial, its length was set to 10 px so that a line was clearly visible. The length of the line could be adjusted by moving the mouse. After a song finished, there was a silent interval of 3 s after which the participants were asked to adjust the length of the line to the perceived overall loudness of the sound that they had just heard. After this they could take a break or start the next song.

Before commencing with the fourteen songs, the participants went through a short practice consisting of three stimuli which were 20-s long segments of music with a calculated overall loudness of 70, 80, and 90 phon, respectively. After this practice, participants were told that these sounds represented the loudness range to be expected during the main experiment, so that they could "recalibrate" their line length. No reference line length was given. Participants were allowed to repeat the practice.

Procedure of Experiment 2

Experiment 2 took place right after Experiment 1. The participants were encouraged to take a break for as long as they wanted.

For Experiment 2, the participants were asked to continuously adjust the loudness of the sound to match the line length that they saw on the screen (displayed as in Experiment 1): "Please use the + and – buttons to adjust the loudness so that it matches the length of the line at any time." The lines were shown like in a video while the songs were being played. The level could be adjusted by using plus and minus buttons, each of which changed the level by 1 dB per click. The participants saw their individual line length sequences that they had produced during Experiment 1, with some critical manipulations, as specified in the next paragraph. That way, they did not experience a perfect covariation between line length and loudness, but rather had to react to make them match.

There were eight manipulations per sound, and each manipulation increased or decreased the line length for a segment of between 7 and 12 s duration. The magnitude of the manipulations corresponded to −10, −6, −3, −1, +1, +3, +6, and +10 dB according to Stevens's power law, i.e., the line length was multiplied by 0.5, 0.66, 0.81, 0.93 1.07, 1.23, 1.51, or 2.0, respectively. This implies exponents of 0.6 for sound pressure level and 1 for line length. These "line length gains" as we might call them were constant factors by which the time-varying line lengths were multiplied for the duration of the manipulation. They were introduced smoothly with linear rise and fall times of 500 ms before fully reaching the respective factor, i.e., the factor changed smoothly between 1 and the target factor.

An individual latency constant was derived from Experiment 1 and subtracted from the temporal position in the musical track, in order to subjectively align the line lengths displayed with the temporal segments of the songs they referred to. This latency constant was determined to be the offset that resulted in the highest correlation between adjusted line length and calculated momentary loudness (DIN 45631/A1). We assume that this latency covers the reaction time to changes in loudness and the time that is needed to handle the mouse. This was done for each participant and each sound.

In summary, the participants listened to a stimulus whose loudness varied over time (the music) and saw a line varying in length accordingly (the one that they produced in Experiment 1). This is different from traditional magnitude production where the participants make adjustments to a stationary stimulus. During eight intervals in each song, however, the line length displayed was manipulated and the participants were supposed to adjust the loudness after onset and offset of these manipulations.

RESULTS

Before comparing the cross-modality matching results of Experiment 1 to calculated loudness, we look at the magnitude productions made in Experiment 2.

Results of Experiment 2: Matching Sound Levels to Line Lengths

Table 1 shows the adjustments in sound level that were made 5 s after the onset of a line-length gain compared to the level 2 s before it started. 5 s were chosen because we think that this is

TABLE 1 | Adjustment in level in response to the onset of the line length manipulations given in the first column.

Line length factor	Mean change [dB]	SD [dB]	*t*-value	*p*-value
0.5	−3.3	2.9	−17.4	< 0.001
0.66	−2.4	2.9	−13.2	< 0.001
0.81	−1.0	2.4	−6.5	< 0.001
0.93	−0.2	3.1	−1.1	0.3
1.07	0.2	2.2	1.3	0.2
1.23	1.4	3.3	6.5	< 0.001
1.51	2.4	3.7	10.6	< 0.001
2	4.4	5.1	14.2	< 0.001

Means and standard deviations across subjects and stimuli, and one-sample t-tests comparing to zero change.

TABLE 2 | Adjustment in level to the offset of line length manipulations in the first column.

Line length factor	Mean change [dB]	SD [dB]	*t*-value	*p*-value
0.5	2.8	2.9	15.2	< 0.001
0.66	1.9	3.1	10	< 0.001
0.81	1.4	2.8	8.2	< 0.001
0.93	0.4	2.4	2.7	< 0.01
1.07	−0.3	1.8	−2.7	< 0.01
1.23	−1.2	3.1	−6.2	< 0.001
1.51	−2.7	3.4	−12.6	< 0.001
2	−3.9	3.9	−16.6	< 0.001

Means and standard deviations across subjects and stimuli, and t-tests comparing to zero.

long enough to account for any delay in a participant's reaction, and still within the 7 to 12-s window of the manipulation. These changes range from −3.3 dB for halving the line length to +4.4 dB for doubling it, which is considerably less than what would be expected from Stevens's power law. However, the changes in level made in response to the artificial line-length gains significantly differ from zero except for the two smallest manipulations in line lengths (factors of 0.93 and 1.07), according to *t*-tests which were calculated independently for each gain factor (last two columns of **Table 1**).

The opposite pattern in level adjustment would be expected after the offset of the manipulations in line length, i.e., after cancelling the artificial line-length gains and returning to the baseline pattern produced in Experiment 1. **Table 2** shows the adjustments that were made 5 s after the end of a manipulation in line length compared to the level 2 s before the end of a manipulation. They range from +2.8 to −3.9 dB, and all of them differ statistically significantly from zero. The sum of the mean values in **Tables 1**, **2** ranges from −0.5 to 0.5 dB and is 0.0 dB on average, which indicates that on average, the level adjustment to the onset of a manipulation was reversed after its offset.

In contrast to a classical magnitude production experiment, where one production is made per trial, participants may "fall asleep," lose track, and not make any adjustment. **Table 3** lists the percentages of adjustments in the correct direction (an increase of at least 1 dB when the line length increased or a decrease of at least 1 dB when the line length decreased), no change in level, or a change in the wrong direction. The largest line length gains led change of level in the correct direction in 78% of all cases. The two smallest gain factors produced no change in 40 or 45%, respectively, and 33% changes in the correct direction. For all of the gain factor manipulations, more changes were made in the expected direction than in the opposite.

Figure 1 shows the distributions of the adjustments in level 5 s after onset in 1-dB wide bins for the four largest gain factors (0.5, 0.66, 1.51, and 2.0), which were summarized in **Table 1**. For all four of them, only a small fraction reaches or exceeds the adjustment that would be expected according to Stevens's power law, i.e., of −10, −6, 6, and 10 dB, respectively. Each of them shows a peak at 0 dB, i.e., when no adjustment was made (as in **Table 3**). All distributions drop sharply on the "wrong" side of 0 dB.

One may speculate whether the continuous magnitude productions differ for the two music genres since rock songs have more uniform levels than classical music. **Table 4** shows the adjustments to the manipulations in sound level separately for each music genre (otherwise the same as the means in **Tables 1**, **2**). The level adjustments are rather similar for the two genres, except for the line length factor of 2 where the adjustments for rock music were about 1 dB larger than those for classical music. To test this discrepancy for statistical significance, we performed a three-way within-subjects analysis of variance with factors line length factor (−0.5 to 2), genre (rock, classic) and direction (onset, offset). The main effect for the line length factor was highly significant, $F(7{,}133) = 118$, $p < 0.001$. The main effect for genre was not statistically significant, $F(1{,}19) = 3.7$, $p = 0.07$, neither was the main effect for direction, $F(1{,}19) = 0.01$, $p = 0.92$. Most critically for the observed difference, the interaction between the line length factor and genre was statistically significant, $F(7{,}133) = 2.4$, $p < 0.05$. The interactions between line length factor and direction, $F(7{,}133) = 3.4$, $p < 0.01$, and between genre and direction, $F(1{,}19) = 18$, $p < 0.001$ were also statistically significant. The three-way interaction was not statistically significant, $F(7{,}133) = 1.7$, $p = 0.11$.

Results of Experiment 1: Continuous Matching of Line Length to Sound Levels

Experiment 1 was analysed to compare the line lengths produced via cross-modality matching with loudness calculations based

TABLE 3 | Type of level adjustment in response to the stimulus manipulations given in the first column.

Line length factor	Correct direction [%]	No change [%]	Wrong direction [%]
0.5	78	18	4
0.66	69	22	9
0.81	48	38	14
0.93	33	40	26
1.07	33	45	21
1.23	59	27	13
1.51	68	21	10
2	78	13	10

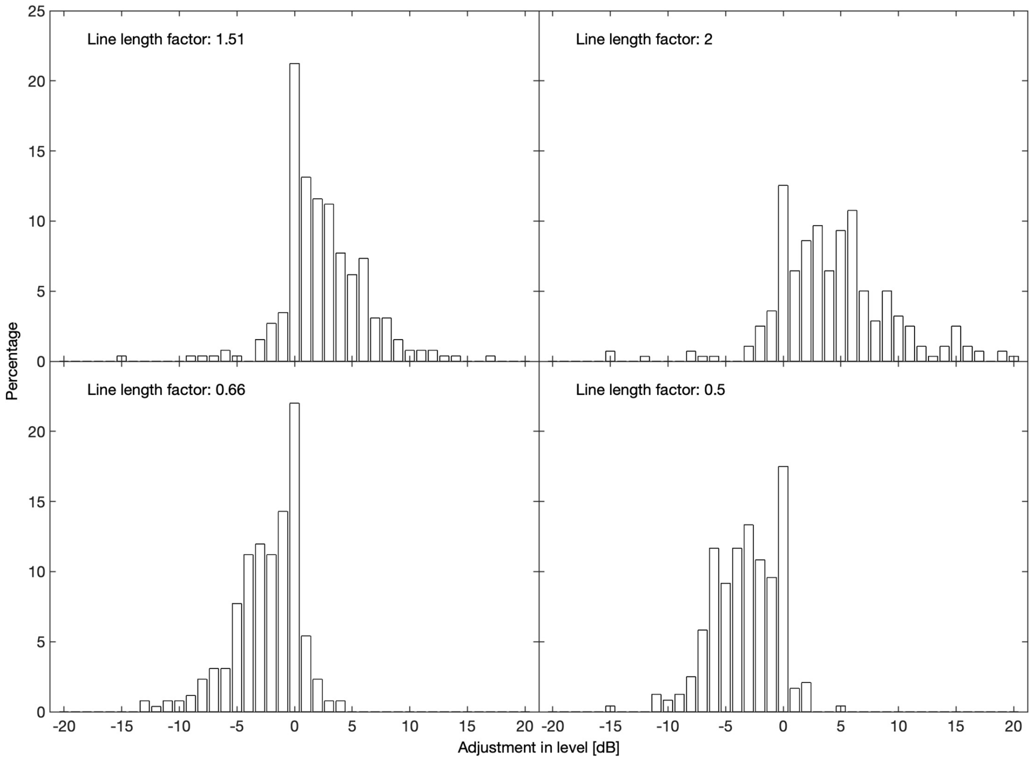

FIGURE 1 | Distributions of level adjustments 5 s after the onset of a manipulation in line length for factors in line length of 1.51 (upper left), 2 (upper right), 0.66 (lower left), and 0.5 (lower right).

on the model of Moore et al. (2018). This model produces three estimates of time-varying loudness: (1) Instantaneous loudness, which is not available to conscious perception and based on a single momentary spectrum; (2) Short-term loudness, which represents the loudness of short segments such as a syllable and calculated from instantaneous loudness using exponential time constants for attack and release in the order of a few ten milliseconds; and (3) Long-term loudness, which represents the loudness of longer segments such as a word or a sentence and is obtained from short-term loudness via exponential time constants, 100 ms for attack and 750 ms for release.

TABLE 4 | Adjustment in level [dB] to the onset and offset of the line length manipulations by music genre.

	Classic music		Rock music	
Line length factor	Onset	Offset	Onset	Offset
0.5	−3.5	2.4	−3.1	3.1
0.66	−2.6	1.4	−2.3	2.3
0.81	−1.0	1.4	−1.0	1.5
0.93	−0.4	0.1	0.0	0.7
1.07	0.1	−0.3	0.3	−0.3
1.23	0.9	−1.1	1.9	−1.4
1.51	2.6	−2.6	2.3	−2.8
2	3.5	−3.6	5.2	−4.2

*Same as second columns (means) of **Tables 1**, **2** but separately for each genre.*

Figure 2 shows mean logarithmic line length as a function of calculated long-term loudness level (thick black line). Error bars represent the standard deviation across points in time that fell within a 1-phone wide bin of calculated loudness level after logarithmic line lengths were averaged across subjects for each point in time. Note, that only 20 s of the total stimulus time of 45 min had loudness levels lower than 40 phon and it is probably difficult to discriminate very short line lengths, explaining the noisy function evident at these low levels, while each 1-phon-wide bin above 65 phon represents 30 to 140 s. The participants seem to have chosen a short line of about 10 px in length independently of loudness level to represent loudnesses below 40 phon. Above 40 phon, mean line length correlates highly with calculated long-term loudness, $r(58) = 0.99$, $p < 0.001$. The correlation between line length and calculated long-term loudness without averaging across time per phon bin, i.e., using the raw data points, is still high, $r(2690225) = 0.89$, $p < 0.001$. To compute this correlation, line length was upsampled to a resolution of 1 ms to match the sample rate of calculated long-term loudness. The fact that the relationship (above 40 phons) is nearly linear in log-log coordinates is evidence for an excellent fit to a power function. The dashed line shows it to imply a 13 px line-length increment for each loudness increase by 1 sone.

An important question in continuous psychophysical scaling is which temporal portions of the sound impact a momentary

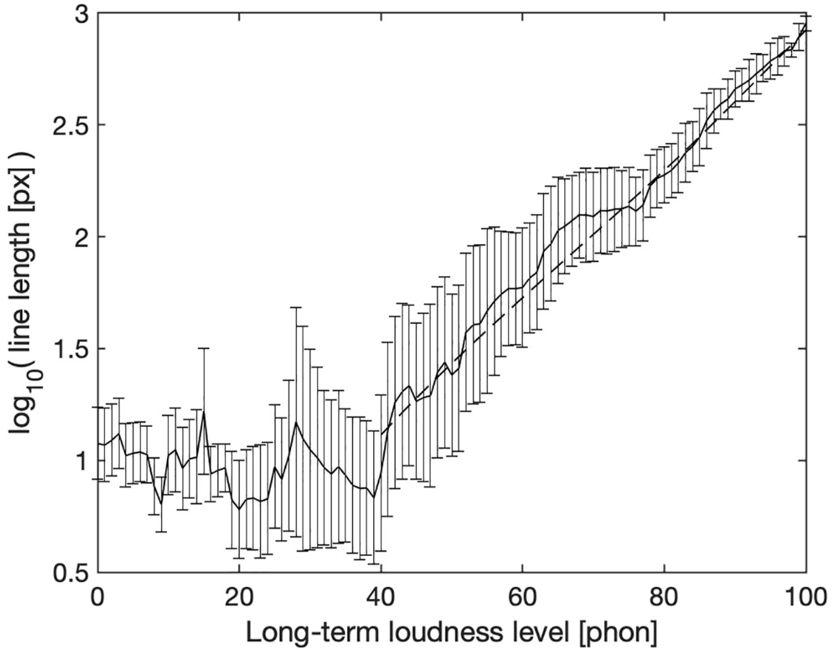

FIGURE 2 | Logarithmic line length as a function of momentary long-term loudness level in 1-phon wide bins. The solid line shows averages across participants and temporal segments (and thus stimuli), error bars ± 1 standard deviation across time. The dashed line shows a correspondence of 1 sone to 13 pixels.

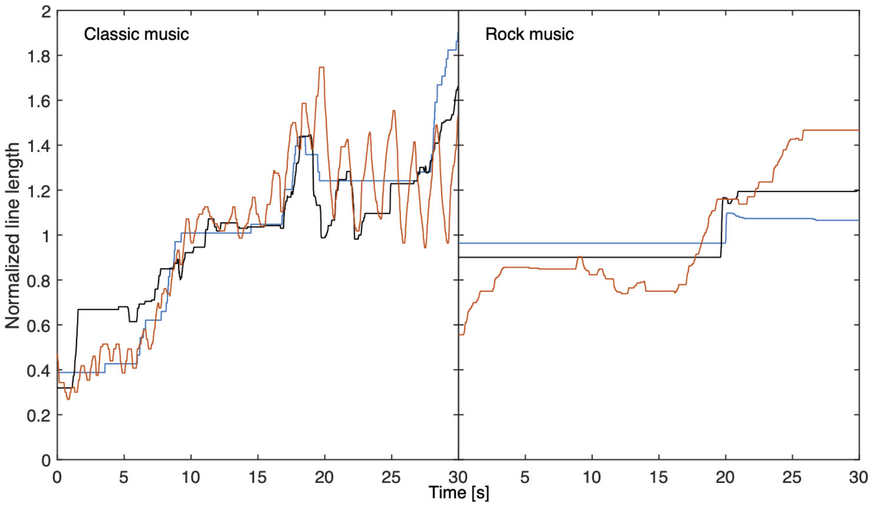

FIGURE 3 | Normalized matched line lengths for three example participants for a 30-s segment of a classical piece (left) and of a rock song (right). The line length was normalized for the 30-s segment.

judgment. **Figure 3** shows normalized line length matches to 30-s excerpts of a classical piece and a rock song for three illustrative participants who apparently had different strategies for making continuous line-length adjustments. The loudness model of Moore et al. (2018) uses an exponential time constant of 750 ms for long-term loudness based on time-varying binaural stimuli. The present data can also be used to estimate this time constant, although this estimate may be limited by the ability to move the mouse. For this purpose, we calculated correlation coefficients between adjusted line length and long-term loudness for each song and participant, and varied the release time constant of long-term loudness between 0 and 3,000 ms, while keeping all other time constants as suggested by the model. The latency, a delay for producing a corresponding line length, was varied between 0 and 3,000 ms. The time constant and latency that yielded the highest correlation coefficient for each song and participant were taken as the "true" values.

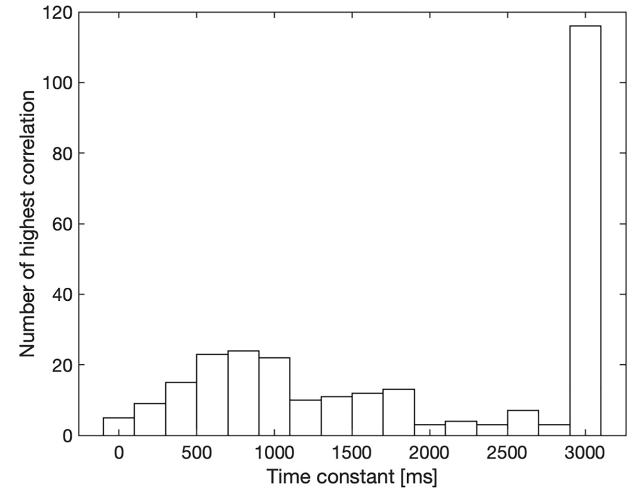

FIGURE 4 | Release time constants for long-term loudness that yielded the highest correlation between line length and calculated long-term loudness for each participant and sound. The number of occurrences within 200-ms wide bins is shown on the ordinate. The search ranged from 0 to 3,000 ms.

The mean latency turned out to be 826 ms. The time constants estimated for long-term loudness are shown in **Figure 4**. 41% of the stimuli (classic: 31%, rock: 50%) yielded the maximum time constant of 3,000 ms, indicating that an even longer integration time may have produced a higher correlation and participants only moved the line to considerable changes in loudness. Interestingly, the distribution in **Figure 4** shows a local maximum close to the model's time constant of 750 ms. Fitting the distribution with two Gaussians, and not taking into account the time constants of 3,000 ms or longer, yielded means of 710 and 1,730 ms (averaged across 100 runs for the fit, ranges of the means: 700 to 720 ms and 1,690 to 1,760 ms).

DISCUSSION

In two laboratory experiments it was shown that two instances of cross-modality matching, (1) continuous matching of line length to time-varying loudness, and (2) its inverse, continuous matching of loudness to temporally varying line lengths yielded meaningful results in terms of (a) validity of responses to stimulus changes, (b) psychophysical functions, and (c) the time constants involved: (a) Participants followed the direction of the experimentally manipulated line length changes in the magnitude-production task despite those manipulations being embedded in long musical excerpts that already varied over time. (b) The exponent of the psychophysical function for continuous matching of line length agreed with predictions of a loudness model since on average, line length as a function of long-term loudness exhibited a simple linear relation between pixels and sone (dashed line in **Figure 2**), and was steeper in the magnitude production task due to a regression effect that was known to exist to a lesser extent for stationary stimuli. (c) The time constants exhibited a local mode at a value that was also found using a different approach based on binaural effects (Moore et al., 2018). This had not been demonstrated to that extent for stimuli continuously varying in magnitude over time.

Some peculiarities of the present results, however, deserve discussion. The exponent of loudness as a function of sound pressure is typically steeper for magnitude production than it is for magnitude estimation (Reynolds and Stevens, 1960), i.e., a difference of less than 10 dB is required to double loudness. Stevens and co-workers found exponents of 0.7, corresponding to 9 dB being required to double loudness, in magnitude-production tasks (Stevens and Guirao, 1962; Stevens and Greenbaum, 1966; **Figure 5**); Hellman (1981) reported an exponent of 0.81 (7 dB to double loudness) for a 1-kHz pure tone. Teghtsoonian and Teghtsoonian (1978) found that the exponent depended on the range of magnitudes that is presented. For a range of 0.5 log units, which corresponds to a factor of 3.2 (close to the maximal ratio of manipulations used in the present magnitude production experiment), they reported an exponent of 1.1 (i.e., 5 dB to double loudness).

The results of the present magnitude production task (Experiment 2) suggest a difference of 3 or 4 dB to match the loudness after doubling or halving line length. This is considerably less than the 10 dB that Stevens's power law suggests, and also less than in all other studies cited. However, the range of our manipulations from 0.5 to 2 (a factor of 4) was rather narrow, for which Teghtsoonian and Teghtsoonian (1978) found results more similar to ours. Furthermore, the modes at 0 dB (i.e., no adjustment made, see **Figure 1**) indicate that the participants sometimes failed to track a stimulus change in the continuous task. About 50% of the adjustments to factors of 0.5 and 2 in line length had absolute values between 2 and 6 dB, confirming a mean sound level change of about 4 dB for line length changes suggesting a doubling or halving of subjective magnitude. Another contribution to the regression effect may be that participants were reluctant to change the stimulus level in the magnitude production task to the extent called for by the altered line lengths because they would be producing sounds unlike those they had heard in the estimation task.

The results of the line-length task agreed well with calculated long-term loudness (**Figure 2**), which suggests that they reproduced the exponent that underlies the loudness model. The loudness model predicts a doubling of loudness for an increase of 10 dB for a 1-kHz tone above 40 dB SPL. For other sounds, the amount that is needed to double loudness is slightly different, but similar. For example, a pink noise that spans from 50 to 20,000 Hz and has an overall level of 40 dB SPL needs an increase of 9 dB to double its loudness. In contrast to this, the 3 to 4 dB that were necessary to double loudness in the magnitude production task are considerably less.

The possible difference between genres deserves attention, too. There was no statistically significant main effect of genre in the magnitude production task. This was to be expected since the line length manipulations were balanced in both directions and thus the grand means are close to 0 dB for both genres. However, the interaction between line length factor and genre was statistically significant. This could suggest that the slightly higher absolute values for rock music, in particular for a line length factor of 2, were not due to chance. We want to emphasize that we did not formulate a specific hypothesis prior to this analysis between genres, which is why it should be considered exploratory.

Figure 2 shows a good correspondence between the continuously tracked line length (Experiment 1) and calculated momentary loudness, a line that relates 1 sone to 13 pixels approximates the averages well. Standard deviations decrease on a logarithmic scale of line length with increasing calculated loudness level, suggesting that the participants judged the louder parts reliably, which are the most important ones to inform judgments of overall loudness (DIN 45631/A1, 2010; Schlittenlacher et al., 2014, 2017; Moore et al., 2018). The good agreement between the line lengths of Experiment 1 and calculated momentary loudness in linear units (pixel and sone, dashed line in **Figure 2**) is at odds Stevens's (1975) suggestion to average the exponents across estimation and production experiments to obtain a "balanced" estimate: The present results suggest that predictions of the loudness model agree with subjective evaluations in a line-length task.

To our knowledge the study of Kuwano and Namba (1985) has been the only one to date that analysed the time interval that is used to inform a momentary judgment. They presented a 20-min long recording of road traffic range during which A-weighted sound pressure level varied between about 50 and 90 dB(A), and depended mainly on the presence or absence of vehicles. They correlated the momentary judgment by category with the equivalent A-weighted sound pressure level and found the highest correlation for an integration time interval of 2.5 s. The analysis of the present paper did not use a time window but an exponential time constant, which is expected to produce somewhat shorter durations for the best match. Thus, the means of 710 and 1730 ms of the Gaussian mixture that represents 60% of the stimuli in the present study are broadly in line with the results of Kuwano and Namba.

The loudness model of Moore et al. (2018) uses an exponential time constant of 750 ms to calculate long-term loudness. This time constant was derived from time-varying synthetic stimuli that differed across the two ears. **Figure 4** provides some support for this time constant: Time constants around 700 ms yielded the highest correlation between calculated long-term loudness and momentary line length more often than others. However, for many songs and participants a rather long time constant of 3 s or more produced the highest correlation. In these cases, the participants may have seen the line length to reflect the current setting of a volume control in which one would tolerate regular fluctuations in loudness or different loudness for different instruments. Furthermore, they may have been reluctant to follow the marginal changes in loudness of rock songs that are typically compressed to a small dynamic range. The long total duration of stimuli, 45 min, though with breaks, may have contributed to this effect. This kind of bias may also occur to a lesser extent in noise studies, where participants focus on a single noise source and not a band or orchestra. The long duration of a music piece compared to echoic memory in combination with the fact that adjustments in line length took time may be a further explanation for the long time constants found.

The mean latency of 826 ms is in the range of values found in the literature for cross-modality matching: Kuwano and Namba (1985) reported 1.0 s, Susini et al. (2002) 0.9 and 1.1 s for their two experiments, and Schlittenlacher et al. (2017) 495 ms.

Taken together, the results of Experiments 1 and 2 suggest that cross-modality matching of line length is a suitable method to assess momentary loudness. Its counterpart, continuous magnitude production of loudness in response to varying line length stimulation, largely agreed with the literature, though the level changes that were produced for a given change in magnitude were on the lower end of the expected range.

DATA AVAILABILITY STATEMENT

The data analyzed in this study are subject to the following licenses/restrictions: The data are owned by TU Darmstadt. Requests to access these datasets should be directed to JS, josef.schlittenlacher@manchester.ac.uk.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JS implemented the software to run the experiments, supervised the data collection, and analysed the data. JS and WE interpreted the results and wrote and revised the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

JS was supported by the NIHR Manchester Biomedical Research Centre and by the Wellcome Trust University of Manchester TPA (209741/Z/17/Z).

ACKNOWLEDGMENTS

We thank Özlem Çelebi, Katharina Staab, Angelo Schleussner, and Alisa Samel for collection the data. Results of Experiment 1 on overall loudness were published in Schlittenlacher et al. (2014). We thank SM, RS, and WJ for helpful comments and contribution to the discussion.

REFERENCES

DeCarlo, L. T., and Cross, D. V. (1990). Sequential effects in magnitude scaling: models and theory. *J. Exp. Psychol.* 119, 375–396. doi: 10.1037/0096-3445.119.4.375

DIN 45631/A1 (2010). *Calculation of Loudness Level and Loudness From the Sound Spectrum–Zwicker Method–Amendment 1: Calculation of the Loudness of Time-Variant Sound*. Berlin: Beuth.

Ellermeier, W., and Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Percept. Psychophys.* 62, 1505–1511. doi: 10.3758/bf03212151

Fastl, H. (1991). "Evaluation and measurement of perceived average loudness," in *Proceedings of the Results of the Fifth Oldenburg Symposium on Psychological Acoustics*, Oldenburg, 205–216.

Fechner, G. T. (1860). *Elemente der Psychophysik (Elements of psychophysics)*. Leipzig: Breitkopf und Härtel.

Hellman, R. P. (1981). Stability of individual loudness functions obtained by magnitude estimation and production. *Percept. Psychophys.* 29, 63–70. doi: 10.3758/bf03198841

Kuwano, S. (1996). "Continuous judgment of temporally fluctuating sounds," in *Recent Trends in Hearing Research*, eds H. Fastl, S. Kuwano, and A. Schick (Oldenburg: BIS), 193–214.

Kuwano, S., Fastl, H., and Namba, S. (2017). "Loudness of traffic noises using the method of continuous judgment by line length," in *Proceedings of the Internoise 2017*, Hong Kong, 153–160.

Kuwano, S., and Namba, S. (1985). Continuous judgment of level-fluctuating sounds and the relationship between overall loudness and instantaneous loudness. *Psychol. Res.* 47, 27–37. doi: 10.1007/bf00309216

Kuwano, S., and Namba, S. (2011). *Loudness in the Laboratory, Part II: Non-Steady-State Sounds. In: Loudness*. New York, NY: Springer, 145–168.

Kuwano, S., Namba, S., Fastl, H., and Putner, J. (2014). "Continuous judgment of sound quality of electric home appliances," in *Proceedings of Internoise 2014*, Melbourne, VIC, 1835–1842.

Kuwano, S., Namba, S., Kato, T., and Hellbrück, J. (2003). Memory of the loudness of sounds in relation to overall impression. *Acoust. Sci. Technol.* 24, 194–196. doi: 10.1250/ast.24.194

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

Luce, R. D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychol. Rev.* 109, 520–532. doi: 10.1037/0033-295X.109.3.520

Luce, R. D., and Krumhansl, C. L. (1988). "Measurement, scaling, and psychophysics," in *Stevens' Handbook of Experimental Psychology*, 2nd Edn, eds R. C. Atkinson, R. J. Herrnstein, G. Lindzey, and R. D. Luce (Hoboken, NJ: Wiley), 3–74.

Luce, R. D., and Mo, S. S. (1965). Magnitude estimation of heaviness and loudness by individual subjects: a test of a probabilistic response theory. *Br. J. Math. Stat. Psychol.* 18, 159–174. doi: 10.1111/j.2044-8317.1965.tb00339.x

Moore, B. C., Jervis, M., Harries, L., and Schlittenlacher, J. (2018). Testing and refining a loudness model for time-varying sounds incorporating binaural inhibition. *J. Acoust. Soc. Am.* 143, 1504–1513. doi: 10.1121/1.5027246

Namba, S., and Kuwano, S. (1980). The relation between overall noisiness and instantaneous judgment of noise and the effect of background noise level on noisiness. *J. Acoust. Soc. Jpn. (E)* 1, 99–106. doi: 10.1250/ast.1.99

Namba, S., and Kuwano, S. (1990). Continuous multi-dimensional assessment of musical performance. *J. Acoust. Soc. Jpn. (E)* 11, 43–52. doi: 10.1250/ast.11.43

Narens, L. (1996). A theory of ratio magnitude estimation. *J. Math. Psychol.* 40, 109–129. doi: 10.1006/jmps.1996.0011

Petzschner, F. H., Glasauer, S., and Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends Cogn. Sci.* 19, 285–293. doi: 10.1016/j.tics.2015.03.002

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychol. Bull.* 86, 777–803. doi: 10.1037/0033-2909.86.4.777

Reynolds, G. S., and Stevens, S. S. (1960). Binaural summation of loudness. *J. Acoust. Soc. Am.* 32, 1337–1344. doi: 10.1121/1.1907903

Richter, U. (2003). Characteristic data of different kinds of earphones used in the extended high frequency range for pure-tone audiometry. *PTB-Berichte Mechanik und Akustik* 72, 1–24.

Schlittenlacher, J., Hashimoto, T., Kuwano, S., and Namba, S. (2017). Overall judgment of loudness of time-varying sounds. *J. Acoust. Soc. Am.* 142, 1841–1847. doi: 10.1121/1.5003797

Schlittenlacher, J., Samel, A., Schleussner, A., Rost, K., Çelebi, Ö, and Ellermeier, W. (2014). "Instantaneous and overall loudness of music," in *Proceedings of the Forum Acusticum 2014*, Krakow.

Stevens, S. S. (1956). The direct estimation of sensory magnitudes: loudness. *Am. J. Psychol.* 69, 1–25. doi: 10.2307/1418112

Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.* 64, 153–181. doi: 10.1037/h0046162

Stevens, S. S. (1975). *Psychophysics: Introduction to its Perceptual, Neural and Social Prospects*. New York, NY: Wiley.

Stevens, S. S., and Galanter, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *J. Exp. Psychol.* 54, 377–411. doi: 10.1037/h0043680

Stevens, S. S., and Greenbaum, H. B. (1966). Regression effect in psychophysical judgment. *Percept. Psychophys.* 1, 439–446. doi: 10.3758/bf03207424

Stevens, S. S., and Guirao, M. (1962). Loudness, reciprocality and partition scales. *J. Acoust. Soc. Am.* 34, 1466–1471. doi: 10.1121/1.1918370

Stevens, S. S., and Guirao, M. (1963). Subjective scaling of length and area and the matching of length to loudness and brightness. *J. Exp. Psychol.* 66, 177–186. doi: 10.1037/h0044984

Susini, P., McAdams, S., and Smith, B. K. (2002). Global and continuous loudness estimation of time-varying levels. *Acta Acust. United. Acust.* 88, 536–548.

Teghtsoonian, R., and Teghtsoonian, M. (1978). Range and regression effects in magnitude scaling. *Percept. Psychophys.* 24, 305–314. doi: 10.3758/bf03204247

von Békésy, G. (1947). A new audiometer. *Acta Otolaryngol.* 35, 411–422. doi: 10.3109/00016484709123756

Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Percept. Psychophys.* 67, 569–579. doi: 10.3758/bf03193515

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.

frontiers in Psychology

ORIGINAL RESEARCH
published: 22 June 2021
doi: 10.3389/fpsyg.2021.634943

Toward an Individual Binaural Loudness Model for Hearing Aid Fitting and Development

*Iko Pieper, Manfred Mauermann, Birger Kollmeier and Stephan D. Ewert**

Medizinische Physik and Cluster of Excellence Hearing4All, Universität Oldenburg, Oldenburg, Germany

The individual loudness perception of a patient plays an important role in hearing aid satisfaction and use in daily life. Hearing aid fitting and development might benefit from individualized loudness models (ILMs), enabling better adaptation of the processing to individual needs. The central question is whether additional parameters are required for ILMs beyond non-linear cochlear gain loss and linear attenuation common to existing loudness models for the hearing impaired (HI). Here, loudness perception in eight normal hearing (NH) and eight HI listeners was measured in conditions ranging from monaural narrowband to binaural broadband, to systematically assess spectral and binaural loudness summation and their interdependence. A binaural summation stage was devised with empirical monaural loudness judgments serving as input. While NH showed binaural inhibition in line with the literature, binaural summation and its inter-subject variability were increased in HI, indicating the necessity for individualized binaural summation. Toward ILMs, a recent monaural loudness model was extended with the suggested binaural stage, and the number and type of additional parameters required to describe and to predict individual loudness were assessed. In addition to one parameter for the individual amount of binaural summation, a bandwidth-dependent monaural parameter was required to successfully account for individual spectral summation.

Keywords: loudness summation, hearing aid, hearing impairment, binaural inhibition, binaural summation, binaural loudness summation, loudness function

Edited by:
Sabine Meunier,
Laboratoire de Mécanique et d'Acoustique—CNRS—Aix-Marseille University—Ecole Centrale Marseille, France

Reviewed by:
Jeremy Marozeau,
Technical University of Denmark, Denmark
Daniel M. Rasetshwane,
Boys Town National Research Hospital, United States

****Correspondence:***
Stephan D. Ewert
stephan.ewert@uni-oldenburg.de

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

Received: *29 November 2020*
Accepted: *27 May 2021*
Published: *22 June 2021*

Citation:
Pieper I, Mauermann M, Kollmeier B and Ewert SD (2021) Toward an Individual Binaural Loudness Model for Hearing Aid Fitting and Development. Front. Psychol. 12:634943. doi: 10.3389/fpsyg.2021.634943

INTRODUCTION

Being "too loud" is the most frequent descriptor for fitting problems with hearing aids (Jenstad et al., 2003), and current hearing aid fitting procedures take loudness into consideration (e.g., Moore and Glasberg, 1998; Byrne et al., 2001; Keidser et al., 2012). For instance, when deriving the widely used fitting formula NAL-NL1, loudness models were used to ensure that speech stimuli are not perceived louder by aided hearing impaired (HI) listeners than by normal hearing (NH) listeners (Byrne et al., 2001). Nevertheless, gains prescribed by NAL-NL1 or similar fitting procedures were still too high for many HI listeners (Keidser et al., 2012). This indicates that the loudness of HI listeners was underestimated by the loudness model and the prescribed gains were reduced in NAL-NL2 (Keidser et al., 2012).

Loudness perception differs significantly across individuals with similar audiometric hearing loss (Moore, 2000) and, to some extent, for NH listeners (e.g., Pieper et al., 2018). This suggests that loudness models with parameters based on averaged data, and with individualization of parameters for HI listeners inferred solely from their audiogram (e.g., Moore et al., 1999), might not be sufficient to predict individual loudness perception (Oetting et al., 2013; Pieper et al., 2018). Accordingly, if such models are involved in the first fitting of a hearing aid, subsequent manual adjustments are likely required. In order to improve individualized loudness predictions, existing parameters of loudness models need to be considered for individualization or additional parameter-controlled stages need to be introduced. Pieper et al. (2018) extended the physiologically motivated loudness model for average NH listeners for individualized loudness predictions of NH and HI listeners. In addition to typical assumptions like an expansive component (or reduced compression component) related to cochlear gain loss and an attenuation component (e.g., Launer, 1995; Derleth et al., 2001; Chalupper and Fastl, 2002; Moore and Glasberg, 2004; Chen et al., 2011a), they suggested a frequency-dependent post gain, potentially reflecting central gain mechanisms (for review, see Brotherton et al., 2015) to improve individual loudness predictions. Although the post gain improved the ability to fit the extended loudness model to individual loudness data for narrowband stimuli, predictions for broadband stimuli were not improved. Furthermore, their model was only applied to monaural stimuli. However, in realistic environments, sounds are typically perceived binaurally in addition to showing broadband properties, as observed for, e.g., speech and environmental noise.

With bilaterally aided HI listeners, it has been shown that binaurally presented broadband stimuli are perceived louder by aided HI listeners than by NH listeners at high levels, i.e., the uncomfortable level perceived as "too loud" is reached at lower levels (Oetting et al., 2018; van Beurden et al., 2020). For monaural presentation, this effect is smaller (Strelcyk et al., 2012; Oetting et al., 2016; Ewert and Oetting, 2018). Taken together, these findings suggest that binaural loudness summation can be affected by hearing loss and depend on the bandwidth of the stimulus. Parameters of binaural loudness summation might be related to physiological processes like the middle ear-muscle (MEM) reflex (Møller, 1962) or the medial olivocochlear (MOC) reflex (Berlin et al., 1993, 1995; Norman and Thornton, 1993; Guinan, 2006). Binaural loudness summation might as well be influenced by later stages of the central auditory system: Binaural inhibition was found in the inferior colliculus (see Li and Yue, 2002, for an overview) probably mediated in part by auditory neuronal stages prior to the inferior colliculus, such as the lateral superior olive (Finlayson and Caspary, 1991) or the dorsal nucleus of the lateral lemniscus (Faingold et al., 1993).

Most studies on binaural loudness summation use the loudness ratio between binaural loudness N_B and monaural loudness N_M in sones to quantify the amount of binaural loudness summation. The ratio N_B/N_M had been assumed to be close to 2 based on the assumption that binaural loudness is the sum of the monaural loudness in sones (e.g., Hellman and Zwislocki, 1963; ANSI S3.4, 2007), while more recent studies have suggested $N_B/N_M < 2$, i.e., binaural loudness in sones is less than twice the monaural loudness in sones (Zwicker and Zwicker, 1991; Sivonen and Ellermeier, 2006; Whilby et al., 2006; Epstein and Florentine, 2009). Current loudness models include a binaural inhibition stage to account for these findings (Moore et al., 2014, 2016): If assuming $N_B/N_M = 1.5$, a wide variety of averaged NH loudness data can be successfully predicted (Moore et al., 2016). For HI listeners, binaural level differences for equal loudness (BLDELs) were (slightly) underestimated (Moore et al., 2014), indicating that the assumed ratio of $N_B/N_M = 1.5$ might be too low to account for binaural loudness summation in HI individuals. Ewert and Oetting (2018) found higher ratios for HI listeners ($\frac{N_B}{N_M} = 2.1 \pm 0.5$, mean ± standard deviation) than for NH listeners ($\frac{N_B}{N_M} = 1.7 \pm 0.4$) using broadband stimuli. In combination, these results suggest that particularly in HI, individual differences in binaural loudness summation might exist.

The goal of this study is to develop a binaural loudness model that can be individualized for NH and HI listeners. Hearing aid fitting and development might benefit from individualized loudness models, enabling better adaptation of the processing to the individual needs, including model-based control of hearing aid signal processing to optimize loudness perception for arbitrary stimuli. Hereby, the critical question is how many and which parameters are required in addition to the commonly used cochlear gain or outer hair cell (OHC) loss and attenuation or inner hair cell (IHC) loss component, to allow for both the ability of the model to account for and to predict individual binaural loudness data in NH and HI listeners. Additional parameters should have a psychoacoustical or physiological motivation resulting in a structured functional model. In light of applicability for hearing aid development and fitting, loudness is modeled in four basic conditions covering the variety in bandwidth and binaurality occurring in natural sounds: (i) monaural narrowband, (ii) binaural narrowband, (iii) monaural broadband, and (iv) binaural broadband. For the model development and evaluation of this study, monaural and binaural loudness data were collected, focusing on narrowband stimuli with different center frequencies in order to be able to access the frequency dependency of binaural loudness summation. Additional loudness data were available from an earlier study of Oetting et al. (2016).

In a first experiment, a simplified binaural summation stage was devised and tested in a data-driven approach in which the binaural stage was applied directly to the measured monaural loudness data for the two ears of individual listeners. By this, binaural loudness summation can be assessed without relying on accurate loudness predictions for monaural stimulus presentation. The simplified binaural stage has a single parameter that controls the overall binaural gain. However, using monaural loudness data as input, this approach assumes that the binaural summation itself cannot be frequency- or bandwidth-dependent. Thus, in a second experiment, the monaural loudness model of Pieper et al. (2018) was extended with an augmented version of the above binaural summation stage where the binaural gain depends on the modeled internal excitation pattern after basilar membrane (BM) processing, which in turn depends on the bandwidth and level of the stimulus. Hereby, the modeled excitation pattern is influenced by individual properties of the peripheral auditory system, such as an individual OHC and IHC loss, and a central gain (Pieper et al., 2018). In order to improve monaural loudness predictions over those of Pieper et al. (2018), a bandwidth-dependent central gain is introduced into the monaural paths of the model prior to the binaural summation stage. Taken together, in addition to the frequency-dependent peripheral components OHC and IHC loss, commonly contained in HI loudness models, and the frequency-dependent post gain introduced in Pieper et al. (2018), four further frequency-independent parameters were introduced, which control a bandwidth-dependent monaural gain in each ear (two parameters), an overall binaural gain (one parameter), and the bandwidth dependency of the binaural gain (one parameter). The extended loudness model was then used to determine which of these individual parameters are required to describe loudness perception in the four basic conditions mentioned above. The improvement of the goodness of fit for each of the parameters was estimated in a hierarchic manner.

Suggestions are devised on which parameters and measurements are required to provide an individual loudness model applicable for hearing aid fitting and aided performance prediction.

MODEL EXTENSIONS AND MODIFICATIONS

The suggested binaural loudness model is based on the monaural loudness model of Pieper et al. (2018). **Figure 1** shows the block diagram of the model. The colored parts contain model parameters used for individualization, with blue parts reflecting model extensions of this study. Here and in the following, subscripts L and R denote constants and variables of the left and right ear, respectively. Subscript B denotes constants and variables of the binaural summation stage. The model follows a signal processing chain structure where each stage receives input only from the previous stage.

The stimulus first passes through a fixed filter representing the transfer function from the sound source to the eardrums for free-field conditions. For frontal incident, the filter meets the ANSI S3.4 (2007) standard, in line with existing loudness models, e.g., Moore et al. (1997) and Chen et al. (2011b). If azimuth shifts from the frontal incidence were simulated, the same amplitude and phase shifts as for the stimuli were applied (see Section Apparatus, Procedure, and Stimuli).

The correction filter, attenuating low and amplifying high frequencies (see Pieper et al., 2016, 2018 for details), is applied to obtain a frequency-dependent absolute threshold according to ISO 389-7 (2005).

The middle ear transfer function is realized with a fixed finite impulse response filter that was fitted closely to the data of Puria (2003).

A physiologically plausible transmission-line model (TLM, e.g., Verhulst et al., 2018) of the cochlear simulates basilar membrane motion. The BM is divided in $N = 1,000$ equidistant segments n. The cochlear gain of the BM can be reduced to account for OHC loss (indicated in red in **Figure 1**), typically referred to as compression loss component in the literature, accounting for steepening of the loudness function (loudness recruitment) as well as widening of auditory filter bandwidth. The TLM provides the segment velocities at the time steps m at a sampling frequency of 100 kHz. The absolute values (denoted $|v_{L,n,m}|$ for the left ear and $|v_{R,n,m}|$ for the right ear) are used as the input of the temporal integration stage.

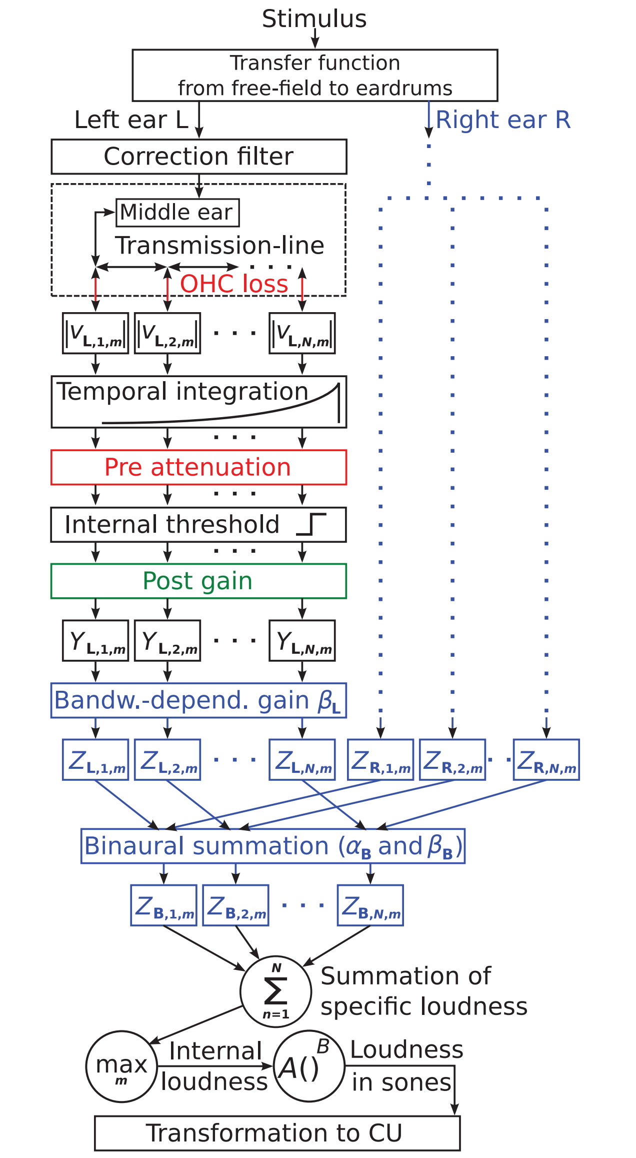

FIGURE 1 | Block diagram of the individual loudness model based on Pieper et al. (2018). Model parts colored in red and green contain frequency-dependent parameters to account for individual hearing loss. Red indicates attenuation and green indicates amplification. Here, the suggested model extensions for further individualization are colored in blue: the parameters β_L and β_R control a monaural bandwidth-dependent central gain for the left and right ears, respectively. α_B controls the overall amount of binaural summation, and β_B controls the binaural summation depending on the bandwidth of the input signals $Z_{L,n,m} + Z_{R,n,m}$.

Temporal integration is performed with a first-order low pass filter (time constant $\tau = 25$ ms). Subsequently, the sampling frequency is reduced to 200 Hz.

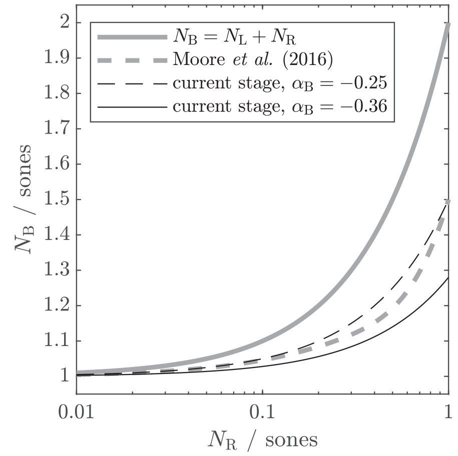

FIGURE 2 | Different theoretical assumptions of binaural summation, depicted as input-output functions. The output is the binaural loudness in sones N_B. The input is the monaural loudness for the right ear $N_{R,}$ while the monaural loudness for the left ear N_L is kept constant at 1 sone. In the suggested binaural stage, the individual amount of binaural summation can be controlled *via* a parameter α_B. Gray solid line: the classical assumption that binaural loudness is the summed monaural loudness in sones. Gray dashed line: binaural stage of the loudness model of Moore et al. (2016). Black lines: simplified version ($\beta_B = 0$, N_L and N_R as input) of the current binaural stage with $\alpha_B = -0.25$ (dashed; comparable to Moore et al., 2016) and $\alpha_B = -0.36$ (solid; obtained from the first experiment of this study).

IHC loss reflecting damage or loss of IHCs, often referred to as attenuation component in the literature, is implemented as linear attenuation prior to a constant internal threshold (referred to as pre attenuation, indicated red in **Figure 1**). The pre attenuation might be interpreted as a reduction of the summed spike rate of an adjacent IHC population, e.g., due to a reduction in the number of intact IHCs, attached synapses, or stereocilia (Pieper et al., 2018). The attenuation component shifts the entire loudness function to higher levels.

All signal parts below the internal threshold are set to 0, simulating the absolute hearing threshold. Thus, attenuation related to OHC and IHC loss prior to the internal threshold (shown in red) effectively increases the hearing threshold. The internal threshold might be interpreted as a specific summed spike rate, which has to be overcome in order to evoke responses in higher processing stages (Pieper et al., 2018).

The subsequent post gain (shown in green in **Figure 1**) is linear amplification applied to the signal part above the internal threshold, assumed to reflect effects of central gain (e.g., Heinz et al., 2005; Zeng, 2013). For HI listeners, a post gain exactly opposite to the pre attenuation counteracts the effect of the pre attenuation for high levels. This leads to the same uncomfortable level as in NH and steepening of the loudness function above the hearing threshold (see **Figure 2** in Pieper et al., 2018). If the post gain is viewed as a part of the IHC loss component and depends on the pre attenuation, the current implementation of IHC loss as well as that of OHC loss are both functionally comparable to the respective components of other HI loudness models (e.g., Launer, 1995; Derleth et al., 2001; Chalupper and Fastl, 2002; Moore and Glasberg, 2004; Chen et al., 2011a). However, Pieper et al. (2018) demonstrated that the post gain is required to be a free parameter for both HI and NH listeners to account for individual differences in the steepness of loudness functions for narrowband stimuli.

In Pieper et al. (2018), the output of the post gain stage for each BM segment is denoted as $Y_{n,m}$ and is summed over the BM segments n at every time step m to yield the time-dependent internal loudness I_m for a single ear (summation of specific loudness). In the current binaural model, $Y_{n,m}$ is calculated separately for each ear and is denoted as $Y_{L,n,m}$ for the left ear and $Y_{R,n,m}$ for the right ear.

At the output of the post gain stage, the model is extended (blue parts in **Figure 1**) by a monaural bandwidth-dependent central gain and a binaural summation stage. These extensions introduce four additional parameters that are considered for individualization:

i) Two monaural parameters, β_L, and $\beta_{R,}$ for the left and right ears to adjust the monaural bandwidth-dependent central gain individually (see Equation 1),
ii) One parameter that controls the overall amount of binaural inhibition α_B, and
iii) One parameter that controls the bandwidth dependency of binaural inhibition β_B.

In the following, equations that are applied in both ears separately are described only for the left side.

Monaural Extension

In Pieper et al. (2018), the individually adjusted post gain did not improve the individual loudness predictions for broadband stimuli. No peripheral parameters (such as outer and middle ear transfer function, OHC and IHC loss, thresholds of BM compression) were identified, which quantitatively explain the remaining individual variations in spectral loudness summation. In principle, the physical BM properties could be individually altered in the TLM to change the auditory filter bandwidth and therefore affect the modeled spectral loudness summation. However, since auditory filters are already widened in HI models because of the OHC loss, and excitation patterns for narrowband stimuli already cover a large portion of the BM, further substantial changes in spectral loudness summation are not expected by such modifications (see, e.g., Zwicker and Scharf, 1965). Pieper et al. (2018) supposed that the medial olivocochlear reflex (Guinan, 2006) or more central mechanisms might be involved. These mechanisms might be altered as a consequence of hearing impairment and might therefore differ across ears. Therefore, as a first functional approach, an additional monaural bandwidth-dependent central gain $\left[1 + \beta_L \cdot W_{L,m}\right]$ is introduced here (see Equation 1), which is multiplied with the output of the post gain stage $Y_{L,n,m}$ in all segments of the BM at every time step m to obtain the final output of the monaural stage $Z_{L,n,m}$ (see **Figure 1**). The gain can be individualized with the constant parameter β_L. The bandwidth estimator $W_{L,m}$ is calculated as a function of $Y_{L,n,m}$ by dividing the average of $Y_{L,n,m}$ across segments ($\overline{Y}_{L,m} = \frac{1}{N}\sum_{n=1}^{N} Y_{L,n,m}$) by the mean of the absolute differences between $Y_{L,n,m}$ and $\overline{Y}_{L,m}$

(Equation 2).

$$Z_{L,n,m} = \left[1 + \beta_L \cdot W_{L,m}\right] \cdot Y_{L,n,m}, \tag{1}$$

$$W_{L,m} = \frac{\overline{Y}_{L,m}}{\frac{1}{N}\sum_{n=1}^{N}\left|Y_{L,n,m} - \overline{Y}_{L,m}\right|} - \frac{1}{2} \cdot \frac{1}{1-\frac{1}{N}}. \tag{2}$$

The second term $-\frac{1}{2} \cdot \frac{1}{1-\frac{1}{N}} \approx -0.5$ ensures that $W_{L,m} = 0$ at the hearing threshold for a narrowband signal, for which $Y_{L,n,m} > 0$ occurs at a single segment only. Above hearing threshold, the excitation pattern broadens with level, particularly for narrowband stimuli, resulting in bandwidth estimations $W_{L,m}$ higher than 0. As a consequence, the bandwidth-dependent gain will alter the model output not only for broadband stimuli but for narrowband stimuli as well. However, as $W_{L,m}$ grows exponentially with the width of $Y_{L,n,m}$ (in contrast to the linear growth of other bandwidth estimators used in, e.g., Rennies et al., 2009; Oetting et al., 2018), good separation between narrowband and broadband signals is maintained despite the broadening of excitation pattern for narrowband stimuli. The same is independently introduced in the right ear, resulting in the two monaural parameters β_L and β_R, respectively.

Binaural Stage

The binaural stage sums $Z_{L,n,m}$ and $Z_{R,n,m}$ present at the monaural paths of each segment. The sum is multiplied with a binaural gain to obtain the output $Z_{B,n,m}$:

$$Z_{B,n,m} = \left[1 + \alpha_B V_{B,n,m} + \beta_B V_{B,n,m} W_{B,m}\right] \cdot \left[Z_{L,n,m} + Z_{R,n,m}\right], \tag{3}$$

where $V_{B,n,m}$ denotes the binaural difference of $Z_{L,n,m}$ and $Z_{R,n,m}$:

$$V_{B,n,m} = 1 - \frac{|Z_{L,n,m} - Z_{R,n,m}|}{Z_{L,n,m} + Z_{R,n,m}}. \tag{4}$$

Equation 4 is a simplified version of the equation that is used in Oetting et al. (2018) to estimate the binaural loudness difference. Here, $V_{B,n,m}$ equals 0 for monaural conditions (with either $Z_{L,n,m} = 0$ or $Z_{R,n,m} = 0$), in which case the stage does not alter loudness. $V_{B,n,m}$ equals 1 if the signals $Z_{L,n,m}$ and $Z_{R,n,m}$ are identical in the monaural paths (diotic stimuli), and is between 0 and 1 if a signal is present in both monaural paths.

Two constant parameters α_B and β_B are used to individualize the binaural stage. α_B alters the gain as a function of the binaural difference $V_{B,n,m}$. Binaural inhibition is modeled if $\alpha_B < 0$, as the gain is lower (and smaller than 1) the higher $V_{B,n,m}$ is, i.e., the more equal $Z_{L,n,m}$ and $Z_{R,n,m}$ are. β_B alters the gain as a function of the binaural difference and the binaural bandwidth estimator $W_{B,m}$. For $W_{B,m}$, the same bandwidth estimation as for the monaural stage (Equation 2) is applied where $Y_{L,n,m}$ is replaced by the binaural sum $Z_{L,n,m} + Z_{R,n,m}$. If β_B is set to 0 and if the signal is identical in the monaural paths ($V_{B,n,m} = 1$), α_B directly reflects the amount by which $Z_{L,n,m} + Z_{R,n,m}$ is altered.

Finally, summation of specific loudness is performed to derive the time-dependent internal binaural loudness:

$$I_m = \frac{1}{N}\sum_{n=1}^{N} Z_{B,n,m} \tag{5}$$

and transformed to loudness in sones by a power-law function and subsequently to loudness in categorical units CU by a non-linear transformation as described in the **Appendix**.

In order to test the binaural stage independently from the monaural stages of the loudness model, a slightly modified version of the suggested binaural stage was first used in a data-driven approach. This approach aims to predict binaural loudness from the monaural loudness measurements. For this, the empirically derived loudness in sones for monaural stimulus presentation N_L and N_R was used as input to the binaural stage, replacing the signals $Z_{L,n,m}$ and $Z_{R,n,m}$ of the monaural model paths. Given that only a single input value per ear exists and the bandwidth estimation $W_{B,m}$ is unknown in this case, the binaural bandwidth-dependent gain in Equation 4 is deactivated by setting β_B to 0.

Figure 2 shows the input–output function of this simplified binaural stage in comparison to other assumptions for binaural loudness summation from the literature. The binaural loudness estimate N_B is shown as a function of N_R ranging from 0.01 to 1 sone, with N_L kept constant at 1 sone. If α_B is set to 0, the current binaural stage follows the classical assumption that the binaural loudness in sones N_B is simply the sum of the monaural loudness in sones (gray solid line, Hellman and Zwislocki, 1963; ANSI S3.4, 2007). If α_B is set to -0.25 (black dashed line), the input–output function is comparable to that of the binaural stage of Moore et al. (2016, gray dashed line), which accounts for a wide variety of averaged NH loudness data. If N_R is much lower than N_L (e.g., $N_R = 0.1$, $N_L = 1$), the contribution of N_R to N_B is still further reduced by the binaural inhibition, resulting in $N_B = 1.05$, i.e., for large loudness differences between ears, the softer ear hardly contributes to binaural loudness.

METHODS

Listeners

Eight NH listeners and eight HI listeners with slight-to-moderate sensorineural hearing loss (SNHL) participated in the study. The NH listeners had audiometric thresholds of 15 dB HL or better at the test frequencies 0.25, 0.5, 1, 2, 4, and 6 kHz. The mean audiometric thresholds and standard deviation of the HI group were 23 ± 11, 33 ± 12, 40 ± 11, 48 ± 18, 61 ± 12, and 59 ± 17 dB HL at the six test frequencies, respectively.

Apparatus, Procedure, and Stimuli

Adaptive categorical loudness scaling (Brand and Hohmann, 2002) was performed to obtain loudness estimates for narrowband low-noise noise (LNN) stimuli with center frequencies of 0.25, 0.5, 1, 2, 4, and 6 kHz and a bandwidth of one-third octave. In comparison to other loudness measurement procedures, loudness scaling offers an easy and fast method applicable in a clinical context. The listener judges loudness on a scale with 11 labeled and unlabeled categories. Labeled categories are “no heard” (0 CU), “very soft” (5 CU), “soft” (15 CU), “medium” (25 CU), “loud” (35 CU), “very loud” (45 CU), and “too loud” (50 CU). In between the categories “very soft” and “very loud,” the categories alternate between labeled and unlabeled categories (10, 20, 30, and 40 CU). In addition to the widely used narrowband stimuli, broadband stimuli were presented to a subset of four NH and four HI listeners. The broadband stimuli, referred to as international female (IF) speech noise, were stationary speech-shaped noise generated from the international speech test signal (Holube et al., 2010). The spectral shape of the signal is the same as the (international) long-term average speech spectrum for females (Byrne et al., 1994). IF noise stimuli were presented unaided or aided. For the aided condition, the monaural narrowband loudness compensation, as described in Oetting et al. (2016), was employed: The spectrum of the stimulus is divided in adjacent frequency bands, and the monaural loudness of the HI listener is restored to average NH loudness in each band. For HI listeners, this resulted in higher amplifications of frequency bands with lower power, resembling the non-linear, level-dependent gain in a hearing aid. Exactly the same procedure was applied for the individual NH listeners (with considerably smaller adjustments than required for HI) and is also referred to as "aided" for NH.

For each listener, the stimuli were presented monaurally in the left, in the right ear, and diotically *via* headphones (Sennheiser HDA200). The listeners were seated in a sound attenuating booth, and responses were collected from a touchscreen connected to a personal computer. Signal generation and experimental control were performed in MATLAB using the AFC package (Ewert, 2013). The headphones were free-field equalized. For the diotic presentation, this means a simulated frontal incident of the sound waves. For the IF noise stimuli, additional azimuth shifts of the incident to the left by 60° and to the right by 60° were simulated. For this, frequency-dependent interaural level and phase differences were applied derived from the interaural differences in the head-related binaural impulse responses for ±60° re frontal incidence of the database from Kayser et al. (2009).

Headphone calibration and equalization ensured level differences between the left and right ears in the binaural conditions of <3 dB for all tested frequencies. In the following, we refer to the outcome of the respective measurements performed in this study as Dataset 1.

As a second set of data (referred to as Dataset 2), categorical loudness scaling data of eight NH and 10 HI listeners of Oetting et al. (2016) were used in this study[1]. The monaural data for the left ear were the same as those used in Pieper et al. (2018). Data for the same narrowband LNN stimuli as used in this study were available but only for monaural presentation. However, data for a narrowband uniform exciting noise (UEN1, Fastl and Zwicker, 2007) with a center frequency of 1,370 Hz and a bandwidth of 210 Hz were available for both monaural aided and binaural aided conditions. For narrowband stimuli, aided conditions imply that not the same level but the same (monaural) loudness was presented to each ear. Broadband stimuli were the same IF noise stimuli used in this study. For NH listeners, only data for unaided conditions are available.

Estimation of Loudness Functions

The fitting procedure "BTUX," as recommended by Oetting et al. (2014), was performed to derive individual loudness functions as well as the hearing thresholds from the raw data of Datasets 1 and 2. BTUX fits a loudness function proposed by Brand and Hohmann (2002) that consists of two straight lines connected with a Bezier curve to the loudness data. First, the hearing threshold level is estimated from all raw data points and assumed to correspond to loudness of 2.5 CU between categories 0 CU ("not heard") and 5 CU ("very soft"). The function is then fitted to all data points above threshold with the prescribed threshold level at 2.5 CU using the least-squares method in the direction of the level (X-direction). If <5 data points are available between 35 and 50 CU, the slope of the upper straight line is set to a fixed value. The hearing threshold estimations of BTUX were used in the following experiment II.

The reference NH functions used for the narrowband loudness compensation (aided conditions) were the same as those shown in **Table 2** in Oetting et al. (2016). These functions are the average loudness functions across nine NH listeners.

Experiment I: Binaural Loudness Summation and Data-Driven Binaural Stage

In the first experiment, the binaural summation ratios R were calculated for the individual loudness functions from Datasets 1 and 2. R was defined as:

$$R = \frac{2N_B}{N_L + N_R}, \tag{6}$$

where N_B denotes the binaural loudness in sones and N_L and N_R the monaural loudness for the left and right ear, respectively. $R > 2$ indicates that the binaural loudness N_B is higher than the sum of the monaural loudness values. This is referred to as binaural excitation in the following. $R < 2$ indicates binaural inhibition (e.g., Moore et al., 2016). Given that loudness had been measured in CU, the CU values were transformed to sone values with the five-parameter cubic function as suggested by (Heeren et al., 2013). Contrary to the procedure commonly used in the literature, where loudness categories are calculated for fixed sound pressure levels, loudness ratios were calculated for the given loudness categories of the binaural (diotic) condition. This allows the comparison of loudness ratios across listeners for, e.g., medium loudness (25 CU) or the "very loud" category (45 CU) close to the uncomfortable level. To obtain the ratio R, the level at which the individual binaural loudness function yields the desired CU value was determined. For that level, the respective CU values of the individual monaural loudness functions were obtained. These binaural and monaural CU values at an equal level were then transformed to sone values.

The value of R is comparable to the binaural summation ratios given in earlier studies if the same loudness was present in both ears, i.e., N_L equals N_R. For the diotic data of this study, loudness can be unequal in both ears, in particular in the case of unaided asymmetric hearing loss, so that the individual values of R are expected to be closer to 2 compared with those

[1] Listeners were the same as in Pieper et al. (2018). As in Pieper et al. (2018), a ninth NH listener was excluded from the original dataset because the listener's uncomfortable levels were identified as outliers at three of the six test frequencies.

found in the literature. Thus, the ratio R (Equation 6) does not directly indicate binaural inhibition or excitation if unequal loudness occurs across the ears. In order to enable comparisons with the literature and across listeners and conditions in the case of unequal loudness across ears, the simplified binaural summation stage (operating on the monaural empirical loudness data in sones, see model extensions above) can be used to derive "corrected" ratios as would have been observed for equal loudness in both ears (see below).

In order to quantify the possible benefit of individualizing the binaural stage, the simplified binaural summation stage (which was fitted to the individual binaural summation ratios for each listener) was compared with a binaural stage that only considers the average (non-individualized) binaural gain derived for NH. After transformation of the stage output in sones back to CUs, the error between CUs inferred with the individualized and non-individualized stage and measured loudness functions was calculated and compared.

The individualized implementation of the binaural stage was realized by allowing for individual values of the binaural gain α_B. Individual values were derived by fitting the loudness ratios calculated from the model outputs $\hat{R}_{s,c}$ to the empirically derived loudness ratios $R_{s,c}$ for the LNN stimuli, using α_B as the fit parameter. The error function that was minimized in the fit was:

$$\sum_{s=1}^{6}\sum_{c=15}^{50}\left(\lg\left(\hat{R}_{s,c}\right)-\lg\left(R_{s,c}\right)\right)^2, \tag{7}$$

where $s = 1, 2, \ldots, 6$ is the number of the six LNN stimuli, and $c = 15, 20, \ldots, 50$ is the loudness category of the empirical loudness function for diotic conditions before transformation to loudness in sones. 5 and 10 CU were excluded in experiment I to ensure that the monaural loudness was always above the hearing threshold. The non-individualized stage used the mean value across the individual values for α_B of the NH listeners.

The amount of binaural inhibition can be assessed more directly if the summation ratios are derived for equal monaural loudness in both ears. Since the hearing thresholds of the HI listeners are usually less symmetric than for the NH listeners, unequal loudness in both ears is to be expected in particular for unaided HI listeners if measurement conditions are diotic. Unequal loudness generally reduces the effect of binaural inhibition or excitation, leading to ratios of R closer to 2 compared with the case where loudness is equal in both ears[2]. Thus, in addition to the calculations of R for a diotic condition as given above (Equation 6), binaural summation ratios were estimated for equal loudness in both ears: substitution of the binaural loudness N_B in Equation 6 with the simplified version of Equation 3 (internal loudness $Z_{n,m}$ replaced by loudness in sones N, $\beta_B = 0$, see model extensions above) and inserting equal loudness $N_L = N_R$ yields the "corrected" binaural summation ratio for assumed equal loudness in both ears:

$$R = 2\cdot(1+\alpha_B)\,. \tag{8}$$

To obtain the binaural summation ratio for assuming equal loudness for a certain stimulus, the fit procedure described above was applied to the stimulus in isolation to determine α_B in Equation 8. It should be noted that the binaural summation ratios resulting from this procedure are assumed to be independent of the loudness category.

In the fits, α_B had a lower limit of $\alpha_B = -0.5$. This constraint ensured that binaural loudness is not lower than monaural loudness in the stage predictions as well as in the ratios for equal loudness ($R \geq 1$ in Equation 8).

Experiment II: Individual Parameters to Describe Monaural and Binaural Loudness

In the second experiment, the complete loudness model was individualized for each listener. In contrast to the isolated simplified binaural stage in experiment I, the loudness model accounts for the auditory preprocessing of the stimuli before they enter the binaural stage. Auditory preprocessing includes the frequency-place transformation on the BM, and thus spectral and bandwidth properties of the stimuli are available to the binaural summation stage and their effect on binaural loudness summation can be assessed.

Similar to experiment I, different model versions were tested for their ability to account for the empirical loudness data. Each version added an additional free parameter. In order to determine the individual parameter values, the individualized models were fitted to measured data for appropriate selected measurement conditions, e.g., monaural broadband data were added to the selection once the bandwidth-dependent individual monaural gain was enabled. The remaining loudness data were then predicted with the individualized models. The non-linear correlation coefficient (ncc), the root mean squared error (rmse), and the bias (bias) were used as performance measures. These measures are based on the level differences between modeled and measured loudness functions at a certain CU. The non-linear correlation coefficient ncc was calculated as:

$$ncc = 1 - \frac{\sum_s \sum_{c=5}^{50}\left(L_{s,c}-\hat{L}_{s,c}\right)^2}{\sum_s \sum_{c=5}^{50}\left(L_{s,c}-\overline{L}\right)^2}. \tag{9}$$

$\overline{L}$ denotes the mean of the empirically derived levels $L_{s,c}$ across all stimuli s and 10 categories $c = 5, 10, \ldots, 50$. $\hat{L}_{s,c}$ are the respective model predictions. The category 0 CU was excluded as the model output is 0 CU for all levels below the hearing threshold. If all predicted levels match the empirically derived levels, i.e., $\hat{L}_{s,c}$ equals $L_{s,c}$ for all s and c, $ncc = 1$ is obtained. If ncc equals 0, the predictions are as good as with $\overline{L}$ as predictor. Additionally, the adjusted ncc' was calculated when using all stimuli s to account for the number of individualized parameters p, i.e., the degrees of freedom of the model:

$$ncc' = 1-(1-ncc)\,\frac{n-1}{n-p-1}, \tag{10}$$

where $n = 10\cdot s$ is the number of observations. The adjusted ncc' was not calculated for a single stimulus where the number of parameters is higher than the number of observations.

[2] The extreme case where the signal is only perceived in one, for example the left, ear ($N_L = N_B$ and $N_R = 0$) Equation 6 will result in a ratio $R = 2$.

The root mean square error rmse estimates the average deviation in dB between model and data:

$$rmse = \sqrt{\frac{1}{10}\sum_{c=5}^{50}(L_c - \hat{L}_c)^2}. \qquad (11)$$

The bias was calculated to identify systematic offsets in dB:

$$bias = \frac{1}{10}\sum_{c=5}^{50}(L_c - \hat{L}_c). \qquad (12)$$

Positive bias values indicate that the predicted loudness function is on average shifted to higher levels compared with the empirically derived loudness function (loudness is on average underestimated). rmse and bias were calculated for each stimulus in isolation.

The extension of the model with a binaural summation stage made it necessary to refine fixed parameters of the final transformations to loudness in sones and CU in the model (see **Appendix**). The procedure performed to refine those parameters also resulted in a non-individualized binaural stage modeling the average NH inhibition of the NH listeners in Datasets 1 and 2.

The following four model versions were considered, which incorporate a successively increasing number of free parameters in the above-described monaural and binaural stages:

1) Binaural stage with average NH binaural inhibition:
    - Model version 1 is the loudness model of Pieper et al. (2018), modified to account for average NH binaural inhibition. As in Pieper et al. (2018), the individual OHC and IHC losses were derived from the hearing threshold, and the lower slope of the loudness functions for monaurally presented narrowband LNN stimuli at frequencies 0.25, 0.5, 1, 2, 4, and 6 kHz. The cochlear gains of the TLM were set to account for OHC loss, and the pre attenuations were set to account for the IHC loss. The monaural post gains were fitted to the loudness functions for the monaural LNN stimuli. These individualization steps were performed for each ear separately. No monaural or binaural bandwidth dependencies were assumed, i.e., $\beta_L = \beta_R = \beta_B = 0$. The above-mentioned non-individualized binaural stage was used to account for average NH inhibition.
2) Addition of individualized bandwidth-dependent gain in the monaural paths:
    - The bandwidth-dependent gain in the monaural paths was individualized by adding β_L and β_R to the set of free parameters and by adding the monaural loudness data for the aided IF noise stimuli with frontal incidence to the targeted empirical data[3]. For the NH data of Dataset 2, no aided conditions were available; thus, the unaided IF noise stimuli were used. As new data were added to the fit, the weightings in the error function needed to be reconsidered[4]. The same binaural stage as in model version 1 was used.
3) Individualized gain of the binaural stage, independent of bandwidth:
    - Similar to experiment I, α_B was used as the free parameter to fit the model to the diotic/binaural narrowband loudness functions. For the listeners measured in this study, the data from Dataset 1 were used for the six diotic LNN stimuli. From Dataset 2, only one binaural narrowband stimulus was available per group: aided UEN1 for HI and unaided UEN1 for NH.
4) Individualized bandwidth-dependent gain of the binaural stage:
    - Here β_B was considered as a free parameter in addition to α_B in the binaural stage, and the loudness function for the aided binaural IF noise stimulus was added to the targeted empirical data. In order to ensure equal weighting of the narrowband and broadband data, the error for the IF noise stimulus was weighted by the number of narrowband stimuli used, which is 6 for Dataset 1 and 1 for Dataset 2. Again, for the NH listeners of Dataset 2, only unaided conditions were available and used.

It should be noted that adjustments of model parameters to the data are referred to as "fitting." The modeled loudness data are only referred to as model "prediction" if the empirical data for the same stimulus were not used in the process of fitting model parameters.

RESULTS

Experiment I: Binaural Loudness Summation and Data-Driven Binaural Stage

The data collected here characterize binaural loudness summation for narrowband stimuli with relatively fine frequency resolution for a wide range of frequencies (0.25, 0.5, 1, 2, 4, and 6 kHz) and for the whole level range from hearing threshold to or close to an uncomfortable level, as covered by the loudness functions. In addition, binaural loudness summation for broadband stimuli (IF noise aided and unaided) was assessed. The raw loudness data and the loudness functions are provided in the **Supplementary Material**. Here, **Figures 3**, **4** show the inferred loudness ratios R in sones/sones from the data for the NH and HI listeners, respectively. The empirically derived ratios of the NH listeners are usually lower than 2, indicating binaural inhibition in all the NH listeners (solid lines and x symbols of lightened colors). The values are similar across the

[3] The aided broadband condition was chosen because pretests revealed better overall predictions than for using the unaided condition.

[4] Good convergence of the fits was achieved if the broadband IF noise data was weighted half as high as the narrowband LNN data and the weighting of the adjacent post gain differences was increased. Weighting of LNN = 1 (x6 signals), weighting of IFN = 3, weighting of post gain differences: 1 (Pieper et al., 2018: weighting of LNN = 1 (x6 signals), weighting of IFN = 0, weighting of post gain differences: 1/5).

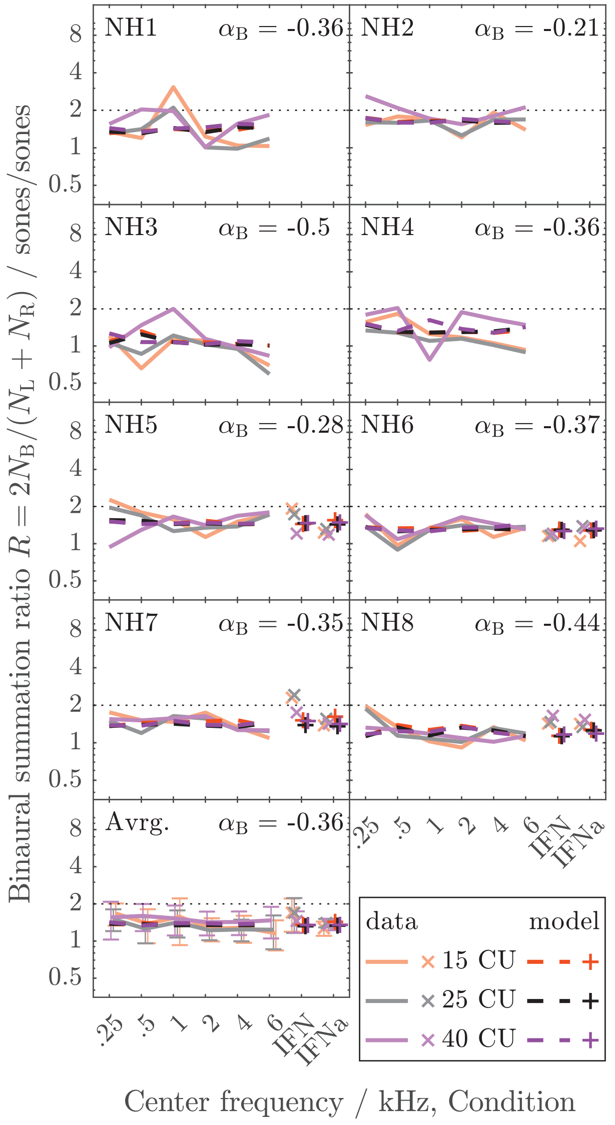

FIGURE 3 | Loudness ratios $R = 2N_B/(N_L + N_R)$ between loudness for diotic presentation N_B in sones and summed monaural loudness $N_L + N_R$ in sones for the NH listeners of Dataset 1. The lines indicate diotically presented narrowband LNN stimuli with different center frequencies (0.25, 0.5, 1, 2, 4, and 6 kHz). Broadband IF noise stimuli with frontal incidence for unaided (IFN) and aided (IFNa) conditions are indicated with symbols. The colors indicate the loudness in CU for the diotic condition (orange: 15 CU, "soft," gray/black: 25 CU, "medium," purple: 40 CU, "loud–very loud"). The empirical data are shown as solid lines (narrowband stimuli) and x symbols (broadband stimuli) in lightened colors. Dashed lines and + symbols indicate the respective model calculations of the individualized binaural stage (the individual values for α_B are given for each listener). The bottom panel shows the averaged data across the NH listeners. Standard deviations are indicated with error bars.

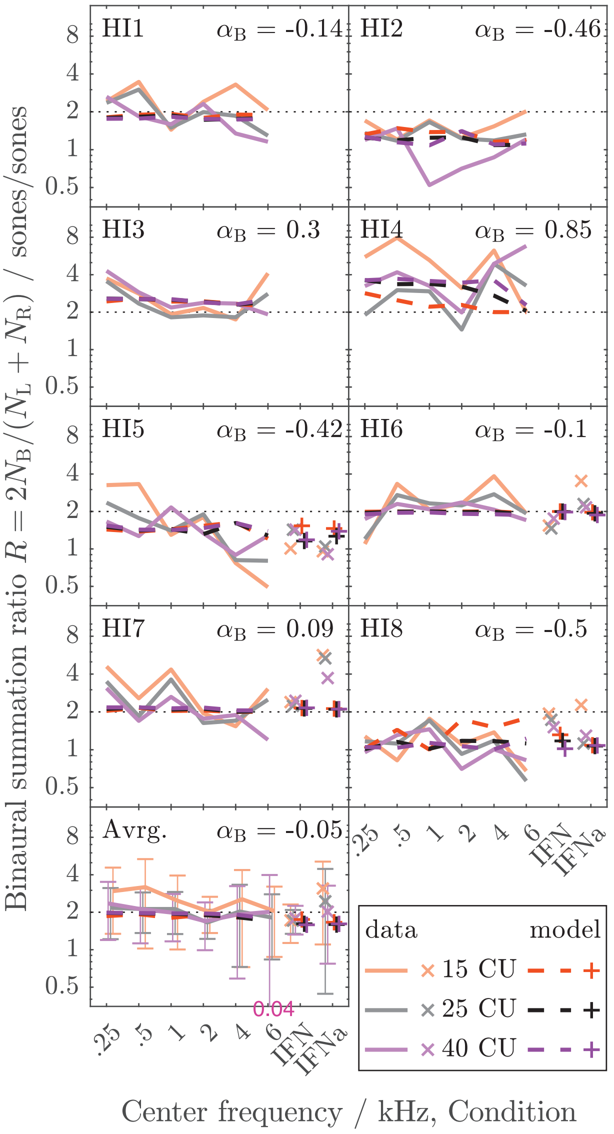

FIGURE 4 | As **Figure 3** but for the HI listeners of Dataset 1.

NH listeners with the exception of NH3, whose data show basically no binaural summation (R close to 1). Nevertheless, as for the other NH listeners, the ratios for NH3 show no dependency of R on the loudness region ("soft," "medium," "loud–very loud" at 15, 25, and 40 CU, respectively), indicated by the loudness of the diotically presented stimulus (orange: 15 CU, gray: 25 CU, purple: 40 CU), and therefore R also does not depend on the stimulus level, as previously found by Marozeau et al. (2006). In some of the NH listeners, some unsystematic variation of the ratios with the stimulus frequency is observed. In contrast to the NH listeners in **Figure 3**, the ratios for the HI listeners in **Figure 4** vary considerably across listeners and within listeners across stimuli. The ratios for HI3, HI4, and HI7 are predominantly higher than 2, indicating binaural excitation. Occasionally, quite high ratios are observed for few frequencies and mostly low levels (and low loudness categories) close to the hearing threshold (maximum ratio $R = 7.9$ at 0.5 kHz, 15 CU for HI4). On average, across the listeners (bottom panels), the ratios of NH and HI decrease slightly with increasing frequency.

For the NH listeners (**Figure 3**), the binaural model stage (dashed lines and + symbols) can be closely fitted to the ratios across all frequencies and loudness regions *via* parameter α_B. For the HI listeners (**Figure 4**), the fit of the binaural model stage (dashed lines and + symbols) shows deviations to the empirically derived loudness ratios (solid lines and x symbols), particularly in the low loudness region (orange lines and symbols). However, because low loudness categories cover a high loudness range in sones if sones are plotted on a logarithmic scale (see **Figure 3** in Heeren et al., 2013), ratios inferred after transformation to loudness in sones at low loudness categories are most sensitive to inconsistencies in the response of the listener and any biases in the method[5]. This is particularly true for the steep loudness functions found in HI listeners at low loudness categories (see, e.g., Oetting et al., 2016). Conversely, in the low loudness region, high deviations of the loudness ratios translate into relatively low deviations for the modeled binaural loudness in CUs.

Figure 5 shows the error of the binaural summation stage if the modeled binaural loudness in sones is transformed back to CUs (HI listeners only). Dashed lines and + symbols indicate the errors for the fits of the binaural stage *via* α_B, i.e., the individualized binaural stage for which the modeled loudness ratios are shown as dashed lines in **Figure 4**. Solid lines and x symbols indicate errors if α_B was fixed to the average value $\alpha_B = -0.36$ of the NH listeners, i.e., for the non-individualized binaural stage using average NH binaural inhibition. Errors in CUs are indeed low for the low loudness region. The absolute errors are lower than 5 CU for (binaural) loudness values of 15 CU (orange lines and symbols). The errors increase with loudness (gray: 25 CU, purple: 40 CU), as reflected by the mean values of the absolute errors across listeners shown in the bottom panel of **Figure 5**.

Individualized binaural summation reduces the errors of the modeled loudness for individual HI listeners at the high loudness region. This is shown by a comparison of the errors at 40 CU for the individualized binaural stage (purple dashed lines and + symbols) and the errors for the average NH binaural inhibition (purple solid lines and × symbols): For HI3 and HI4, the individualization reduces these errors by approximately 10 CU or two loudness categories. For HI7, these errors are reduced by more than a loudness category (5 CU) for the narrowband LNN stimuli with low center frequencies and for the broadband IFN stimuli. From the subgroup for which broadband data are available (HI 5–8), the two listeners HI6 and HI7 show decreased binaural inhibition for the narrowband stimuli (i.e., the fit to the narrowband data resulted in parameters $\alpha_B = -0.1$ and $\alpha_B = 0.09$, respectively). For these listeners, the accuracy of the binaural broadband predictions is increased (+ symbols in **Figure 5**) compared with no individualization ($\alpha_B = -0.36$, × symbols at "IFNa" label in **Figure 5**). However, ratios inferred with the binaural model stage are similar for all stimuli within a listener. Thus, the stage does not account for, e.g., the frequency dependencies of binaural summation ratios in HI2 (compare purple solid with a dashed line in **Figure 4** and see purple dashed line in **Figure 5**) or the increased binaural summation for the aided broadband stimulus (IFNa) in HI7 (compare x with + symbols at "IFNa" label in **Figure 4**).

The model fits are mostly determined by high loudness categories and almost not affected by low loudness categories, as indicated by the differences in error between fitted (dashed lines and + symbols) and unfitted (solid lines and x symbols)

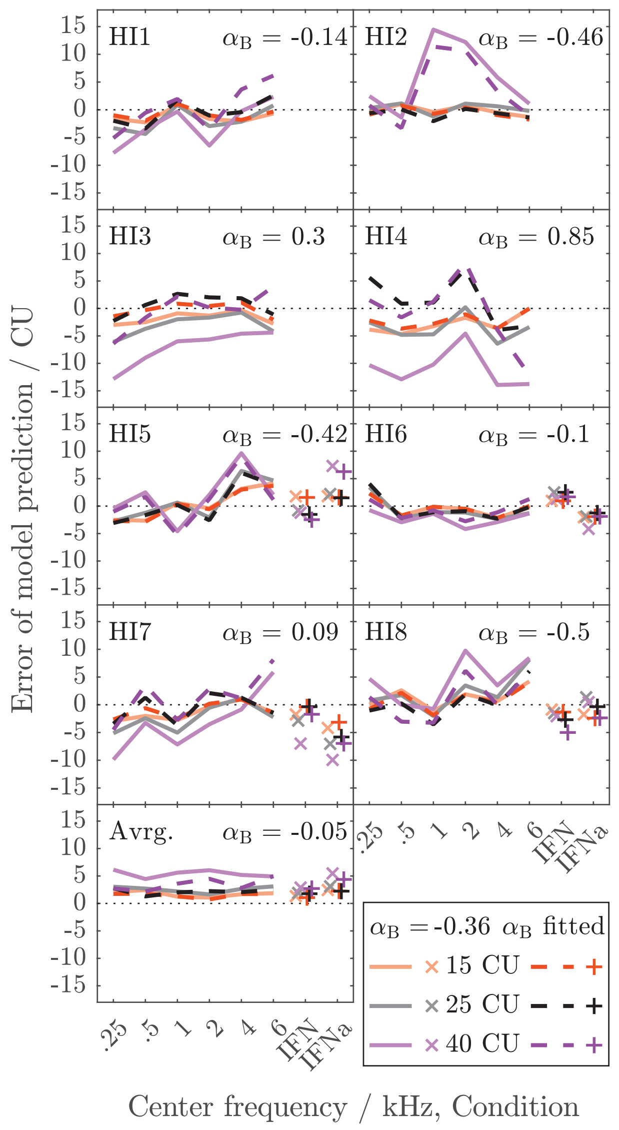

FIGURE 5 | Deviations between model calculations of binaural loudness in CUs and empirical binaural loudness data in CUs for the HI listeners of Dataset 1. As in **Figure 4**, the color indicates the loudness of the diotic signal. Dashed lines and x symbols are for the individualized binaural model stage (see **Figure 4**). Solid lines and + symbols of lightened color are for the model using average NH binaural inhibition (**Figure 3**, mean of $\alpha_B = -0.36$). The bottom panel shows the averaged *absolute* errors across listeners.

[5] Biases may be possible due to the choice of the loudness function that is used to fit the raw CU data. The function used is based on the assumption that low CUs are linearly related to the stimuli levels (Brand and Hohmann, 2002). Another possible cause for biases may be small inaccuracies in the transformation from CU to sones.

TABLE 1 | Binaural summation ratios *R* averaged (mean ± standard deviation) across listeners of Dataset 1 (column 4: NH, column 5: HI) and across categories (15–50 CU).

Stimulus type	Freq./kHz	Aided	Binaural summ. ratio *R*		Variance *p*	Mean *p*
			NH	HI		
ORIGINAL DIOTIC MEASUREMENT CONDITION						
Narrowband LNN (eight listeners)	0.25	No	1.56 ± 0.30	2.40 ± 1.09	**0.002**	0.071
	0.5	No	1.43 ± 0.29	2.29 ± 1.07	**0.045**	0.061
	1	No	1.47 ± 0.31	2.10 ± 0.82	0.054	0.073
	2	No	1.30 ± 0.21	1.73 ± 0.53	**0.030**	0.061
	4	No	1.32 ± 0.20	2.07 ± 1.38	**0.046**	0.175
	6	No	1.31 ± 0.30	2.01 ± 1.43	0.072	0.217
	ANOVA	No	1.40 ± 0.28	2.10 ± 1.06	**0.008**	0.054
Broadband IFN (four listeners)		No	1.54 ± 0.33	1.73 ± 0.44	0.605	0.529
		Yes	1.34 ± 0.10	2.33 ± 1.57	0.056	0.298
	ANOVA		1.44 ± 0.25	2.03 ± 1.12	0.092	0.294
ASSUMED EQUAL LOUDNESS IN BOTH EARS						
Narrowband LNN (eight listeners)	0.25	No	1.46 ± 0.32	2.35 ± 1.20	**0.001**	0.078
	0.5	No	1.35 ± 0.30	2.24 ± 1.13	**0.023**	0.063
	1	No	1.39 ± 0.30	1.97 ± 0.95	**0.024**	0.140
	2	No	1.21 ± 0.19	1.76 ± 0.63	**0.018**	**0.045**
	4	No	1.22 ± 0.23	2.01 ± 1.55	0.069	0.196
	6	No	1.22 ± 0.27	2.65 ± 3.80	0.055	0.326
	ANOVA	No	1.31 ± 0.27	2.16 ± 1.79	**0.008**	0.087
Broadband IFN (four listeners)		No	1.49 ± 0.32	1.55 ± 0.63	0.322	0.883
		Yes	1.27 ± 0.09	2.69 ± 2.32	**0.045**	0.308
	ANOVA		1.38 ± 0.25	2.12 ± 1.69	0.050	0.347

The upper half of the table shows the results for the diotic measurement condition. The lower half shows the ratios for assumed equal loudness in both ears (see Method section of experiment I). The first column shows the stimulus type. Stimuli were unaided narrowband LNN with center frequencies given in column 2 (eight listeners) or broadband IFN for unaided and aided conditions (four listeners). Column 6 shows the p-value of Levene's test for equal variances between groups. Column 7 shows the p-value for equal means of Welch's t-test. The rows labeled ANOVA in column 2 show the respective mean values for the NH and HI groups (columns 3 and 4). Here, the p-values for the main effect of listener group (NH or HI) for the absolute deviation statistic of the ratios according to Levene and for the ratios are provided in columns 6 and 7. Significant differences ($p < 0.05$) are shown in bold.

models in **Figure 5**. These differences are highest for high loudness categories (purple) and almost nonexistent for low loudness categories (orange). This is beneficial for the ratios for "assumed equal loudness between ears" addressed below, which are based on the model fits, because, as mentioned above, the ratios inferred for low loudness categories are most sensitive to inconsistencies in the response of the listener and any biases in the method.

Tables 1, **2** show the binaural summation ratios for each stimulus in isolation averaged across categories (15 to 50 CU) and NH or HI listeners. **Table 1** shows the resulting ratios (mean ± standard deviation across listeners) for the data collected in this study and discussed above (Dataset 1) and **Table 2** for the additional Dataset 2 provided by Oetting et al. (2016). In both tables, columns 4 and 5 list the ratios averaged across NH and HI listeners, respectively. As in **Figures 3**, **4**, the ratios listed in the upper half of the table are for diotic conditions.

In both datasets, the standard deviations across the HI listeners (column 5) are more than twice as high as across the NH listeners (column 4) for all diotic conditions, indicating a higher inter-subject variability of binaural summation for the HI listeners. For the narrowband stimuli in Dataset 1, this observation is confirmed by a significant main effect of listeners group ($p < 0.05$) performing a two-way mixed-design ANOVA applied to Levene's absolute deviation estimate for each condition (column 6 of the upper part of **Table 1** in the row labeled with ANOVA). For the individual narrowband LNN stimuli, Levene's test for equal variances between groups (NH and HI listeners) showed a significant difference in variances ($p < 0.05$) for most of the center frequencies (0.25, 0.5, 2, and 4 kHz) as well as for the broadband unaided IFN stimulus in Dataset 2 (column 6 in **Table 2**). Here and in the following, no correction for multiple comparisons was applied, as they were performed following a significant main effect of the ANOVA omnibus test.

In both datasets, the average binaural summation ratios are higher for the HI listeners than for the NH listeners, again for all diotic conditions: However, a significant main effect of the listeners group ($p < 0.05$) was not found performing a two-way mixed-design ANOVA (column seven in the row labeled with ANOVA). Significant differences in the average ratios ($p < 0.05$) between the groups were found for unaided IFN in Dataset 2 (Welch's *t*-test, column seven).

The lower halves of both **Tables 1**, **2** list the ratios for assumed equal loudness in both ears as described in the Method section

TABLE 2 | Binaural summation ratios averaged across listeners of Dataset 2.

Stimulus type	Freq./kHz	Aided	Binaural summ. ratio *R*		Variance *p*	Mean *p*
			NH (eight listeners)	HI (10 listeners)		
ORIGINAL DIOTIC MEASUREMENT CONDITION						
Narrowband UEN1	1.37	No	1.38 ± 0.16			
		Yes		1.72 ± 0.56		
Broadband IFN		No	1.93 ± 0.35	2.86 ± 1.01	**0.022**	**0.020**
		Yes		3.37 ± 2.06		
ASSUMED EQUAL LOUDNESS IN BOTH EARS						
Narrowband UEN1	1.37	No	1.28 ± 0.16			
		Yes		1.68 ± 0.69		
Broadband IFN		No	1.79 ± 0.35	2.97 ± 1.27	**0.014**	**0.018**
		Yes		3.65 ± 2.45		

Narrowband stimuli were UEN1 with a center frequency of 1.37 kHz and were presented unaided for the NH listeners (column 4) but aided for the HI listeners (column 5). As both measurement conditions should have led to near equal loudness at the ears (see Method section of experiment I), the values for assumed equal loudness in the ears (lower half of table) are similar to the original values (upper half of table). However, this does not hold for the broadband IFN stimuli. The broadband conditions are the same as those in ***Table 1****, but the results are based on a bigger listener pool (eight NH listeners and 10 HI listeners instead of four NH and four HI listeners). Column 6 shows the p-value of Levene's test for equal variances between groups. Column 7 shows the p-value of Welch's t-test. Significant differences (p < 0.05) are shown in bold.*

of experiment I. The ratios for assumed equal loudness in the ears averaged across the NH listeners of Dataset 1 differ by 5.6, 5.8, 5.3, 5.2, 6.7, and 10.7% (median across absolute percentages) from the ratios for the original diotic measurement conditions at the stimuli center frequencies of 0.25, 0.5, 1, 2, 4, and 6 kHz, respectively. For the HI listeners of Dataset 1, the respective values are 5.7, 5.2, 7.1, 7.8, 9.2, and 15.3 %, and therefore similar to the values of the NH listeners at low frequencies but increased for medium to high frequencies.

For the NH listeners, inter-subject variability of the ratios for assumed equal loudness in ears is similar to the ratios for diotic conditions (compare standard deviations in column 4 between upper and lower halves in both tables). Contrary to the NH listeners, inter-subject variability is increased for the HI listeners (compare standard deviations in column 45 between upper and lower halves in both tables). As for the upper part of **Table 1**, a significant main effect of the listeners group ($p < 0.05$) on variance was found performing a two-way mixed-design ANOVA applied to Levene's absolute deviation estimate (column six in the row labeled with ANOVA). For the individual condition, significantly different variances were found for LNN with low to medium center frequencies (0.25, 0.5, 1, and 2 kHz) and aided IFN in Dataset 1. No significant differences are observed for LNN with high center frequencies (4 and 6 kHz) and unaided IFN. On the contrary, in Dataset 2 where more listeners participated in the measurements for broadband conditions (8 NH and 10 HI listeners in Dataset 2 compared with the subset of 4 NH and 4 HI listeners in Dataset 1), a significant difference is found for unaided IFN.

The mean loudness ratios are generally higher for the HI listeners than for the NH listeners for all conditions in both datasets; but no significant main effect of the listener group was found performing a two-way mixed-design ANOVA applied to Dataset 1.

Except for the IFN where only four listeners participated in the measurements for Dataset 1, comparable measurement conditions yielded similar results between datasets. The ratios of the NH groups are similar for both datasets for narrowband noises with a center frequency of approximately 1 kHz: The ratio for assumed equal loudness in ears inferred from Dataset 1 for the LNN stimulus with a center frequency of 1 kHz is 1.39 ± 0.3. The ratio for assumed equal loudness in ears inferred from Dataset 2 for the UEN1 stimulus (center frequency: 1.37 kHz) is 1.28 ± 0.16. The respective ratios for the HI groups are both higher than for the NH groups: the ratio is 1.97 ± 0.95 for Dataset 1 and 1.68 ± 0.69 for Dataset 2.

Binaural summation ratios derived from Dataset 2 suggest an increase in the ratio with bandwidth in the NH group, as the ratio for assumed equal loudness in the ears for the (unaided) IFN stimulus is higher (1.79 ± 0.35) than for the UEN1 stimulus (1.28 ± 0.16). On the contrary, ratios derived from Dataset 1 show only a slight increase for unaided IFN (1.49 ± 0.32 compared with 1.39 ± 0.3 for 1 kHz LNN) and no increase for aided IFN (1.27 ± 0.09). For the HI group of Dataset 1, the ratios for assumed equal loudness in the ears suggest increased binaural summation for aided IFN (2.69 ± 2.32 compared with 1.97 ± 0.95 for 1 kHz LNN) but a decrease for unaided IFN (1.55 ± 0.63). Ratios derived from Dataset 2 suggest an increase for both aided and unaided IFNs (3.65 ± 2.45 and 2.97 ± 1.27, respectively, compared with 1.68 ± 0.69 for aided UEN1).

Taken together, based on the binaural loudness summation data for narrowband stimuli, it was shown that for NH no level dependency and for HI no systematic level dependency of binaural summation exist. For both the NH and HI listeners, binaural summation ratios slightly decreased with frequency if averaged across listeners. Some of the HI listeners showed loudness ratios >2, i.e., indicating super additivity (binaural excitation). Individualization of the amount of binaural

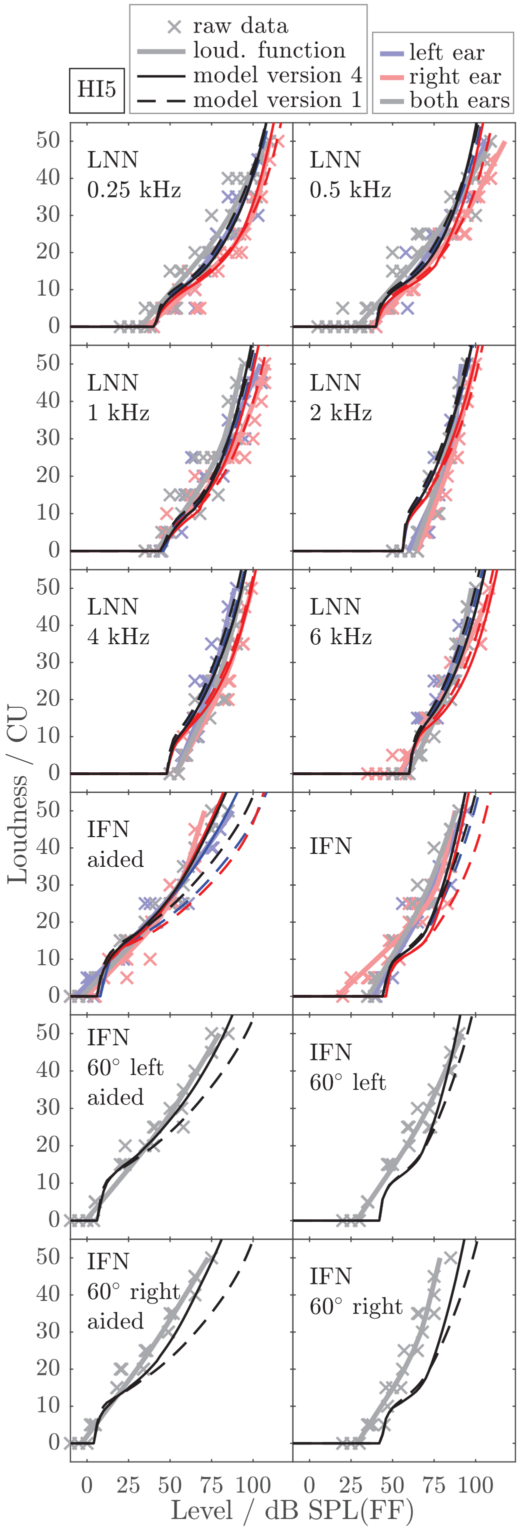

FIGURE 6 | Empirical loudness functions of listener HI5 (brightened thick solid lines) and modeled loudness (thin lines) for monaural conditions (left ear: blue, right ear: red) and diotic conditions (gray/black) with simulated frontal sound incidence or from ±60° in the horizontal plane as indicated in the panels.

(Continued)

FIGURE 6 | The raw measurement data are indicated by crosses. Thin solid lines are for model version 4, for which all monaural (β_L and β_R) and binaural parameters (α_B and β_B) have been individualized by fitting the model to the LNN and the aided IFN stimuli. The loudness of four IFN stimuli in the lower panels and the unaided IFN are model predictions. Thin dashed lines are for model version 1, i.e., without bandwidth-dependent monaural and binaural gains ($\beta_L = \beta_R = \beta_B = 0$) and with average NH binaural inhibition ($\alpha_B = -0.273$, see **Appendix**). For this model version, only the monaural LNN loudness data were used in the fitting procedure. The modeled loudness for the remaining binaural LNN stimuli and all IFN stimuli is a prediction.

summation in the model can reduce the error in fitting the data by 10 CU (two loudness categories) in the high loudness region (or high stimulus levels). The binaural stage allows the calculation of "corrected" binaural summation ratios for assuming equal loudness, enabling better comparability between conditions, listeners, and other studies. The inter-subject variability of the "corrected" ratios is higher for the HI listeners than for the NH listeners. Mean ratios were higher in all the conditions for HI; however, the effect was not significant in most conditions, likely because of the small sample size.

Experiment II: Individual Parameters to Describe Monaural and Binaural Loudness

To exemplarily show the effects of individualized parameters for modeling loudness, **Figures 6**, **7** show the empirically derived loudness functions (thick solid lines, lightened colors), the underlying raw data (crosses), and modeled loudness functions (thin lines, darkened colors) of listeners HI5 and HI7. The corresponding figures for the subgroup of listeners for which broadband conditions were measured in Dataset 1 are provided in the **Supplementary Material**. The dashed lines show model version 1, for which only the monaural post gains have been individualized. The solid lines are for model version 4, for which all monaural (post gains, β_L and β_R) and binaural parameters (α_B and β_B) have been individualized. Red and blue indicate monaural presentation to the left and right ears, respectively. Gray and black indicate diotic presentation.

The model output of model version 1 (dashed lines) closely fits the loudness functions of the monaural LNN stimuli, which have been targeted in the fit (upper six panels, center frequencies indicated inside the panels). Model predictions for the monaural broadband stimuli are inaccurate (compare thin dashed red and blue lines with thick solid red and blue lines in the panels indicated as IFN and IFN aided). This result is in line with Pieper et al. (2018) who have shown that the fit of the post gain to narrowband loudness data does not improve the model predictions for broadband loudness data.

In model version 4 (solid lines), the aided monaural IFN stimuli were added to the targeted stimuli. Consequently, the modeled loudness functions better account for these data. However, the fit slightly alters the modeled loudness functions for the monaural narrowband LNN stimuli. Particularly at low frequencies, this can result in decreased accuracy of the model for narrowband stimuli (see, e.g., **Figure 6**, red lines in

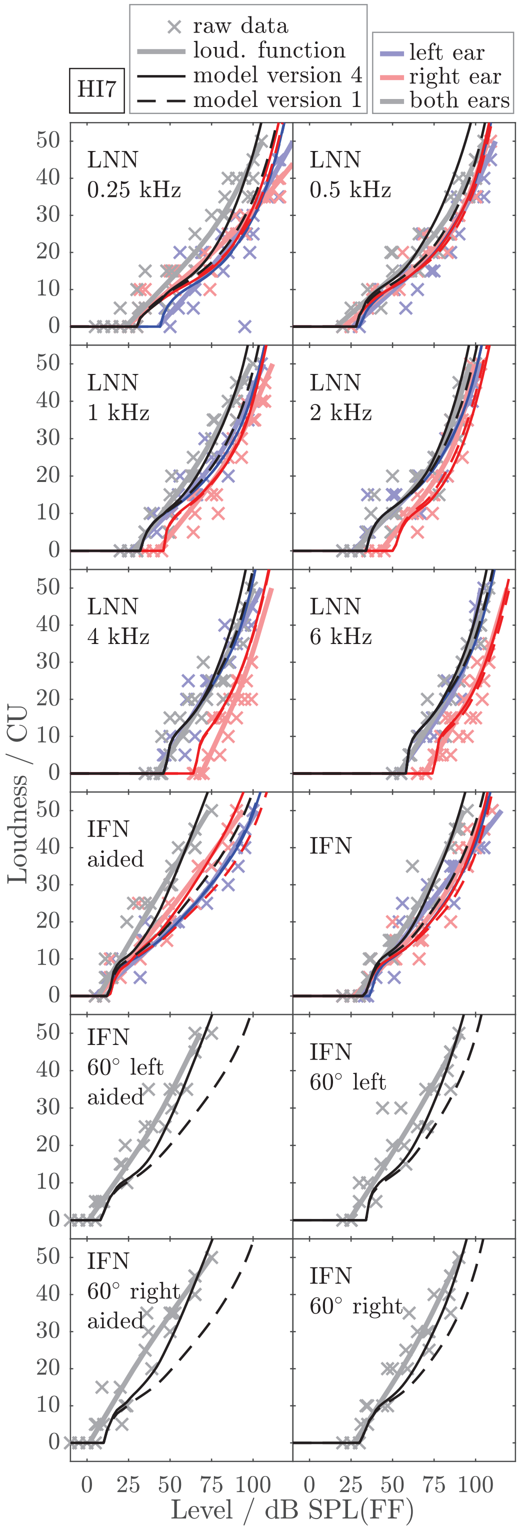

FIGURE 7 | As **Figure 6** but for listener HI7.

the panels indicated as 0.5 kHz and 1 kHz). Model version 4 improves the predictions for the unaided and aided monaural and binaural broadband stimuli, which were not involved in the fitting procedure (see panels indicated as IFN, IFN 60° left, IFN 60° left aided, IFN 60° right, and IFN 60° right aided in **Figures 6**, **7**), which also holds for the other two HI listeners (not shown).

To assess the performance of all the models and both data sets, **Figures 8**, **9** show the median (symbols) and the 25 and 75 percentiles (bars) of the performance measures (ncc, rmse, and bias) across the HI subgroup of Dataset 1 (**Figure 8**) and the HI listeners of Dataset 2 (**Figure 9**). While the figures show the performance measures for each stimulus in isolation, **Table 3** lists the adjusted ncc' values across all the stimuli for the HI listeners, and **Table 4** for the NH listeners.

For the HI listeners of both datasets, the monaural bandwidth-dependent gain in model versions 2–4 (disabled in model version 1, $\beta_L = \beta_R = 0$) improves the model predictions across listeners for monaural broadband conditions (IFN unaided and aided, model version 1: red circles, model versions 2–4: red squares in **Figures 8**, **9**). For the aided IFN stimuli, the rmse is reduced from 11.6 to 2.6 dB for the HI subgroup of Dataset 1 and from 8.8 to 5.3 dB for the HI group of Dataset 2. For the respective unaided IFN stimuli, which were not considered in the fits, the rmse is reduced from 8.8 to 6.1 dB for the HI subgroup of Dataset 1 and from 5.7 to 3.8 dB for the HI listeners of Dataset 2. Performance improvements are nearly as high for the respective NH listeners of both Datasets (not shown in figures). Median rmse values for aided IFN are reduced from 7.6 to 4 dB for the NH subgroup of Dataset 1. Median rmse values for unaided IFN are reduced from 6.8 to 5.5 dB for the NH subgroup of Dataset 1 and from 6.7 to 4.5 dB for the NH listeners of Dataset 2. The median of the adjusted ncc' values across the HI listeners is increased from 0.84 for model version 1 to 0.936 for model version 2 for the subgroup of Dataset 1 and from 0.933 to 0.95 for Dataset 2 (**Table 3**). The respective values for the subgroup of NH listeners in Dataset 1 are 0.955 and 0.973 for model versions 1 and 2, and 0.965 and 0.974 for Dataset 2. The higher values and the smaller benefit for Dataset 2 reflect that this dataset contains fewer broadband conditions than Dataset 1.

It has already been shown in experiment I that the individualization of the overall binaural gain α_B is necessary to describe the binaural data for certain listeners. In experiment II, benefits from individualized binaural summation are reflected in the performance measures for model version 3 in both datasets (black triangles in **Figures 8**, **9**). Compared with the performance measures of model version 2 (without individualized binaural summation; black squares), individualization in model version 3 leads to a slight increase in the median nccs and a slight decrease in the median rmses for most conditions, indicating small overall improvements for the HI listeners. The reduced percentile ranges of the ncc and rmse measures for all the conditions indicate improvements in the model predictions for certain HI listeners. The median adjusted ncc' is slightly increased from 0.936 to 0.944 for the HI subgroup of Dataset 1 and from 0.950 to 0.959 for the HI listeners of Dataset 2. The respective 25 percentile is increased from 0.919 to 0.937 for the subgroup of Dataset 1 and increased from 0.905 to 0.947 for Dataset 2, indicating improvements in the worst-performing individual models. For the NH listeners, individualized predictions of model version 3 are almost not improved over model version 2.

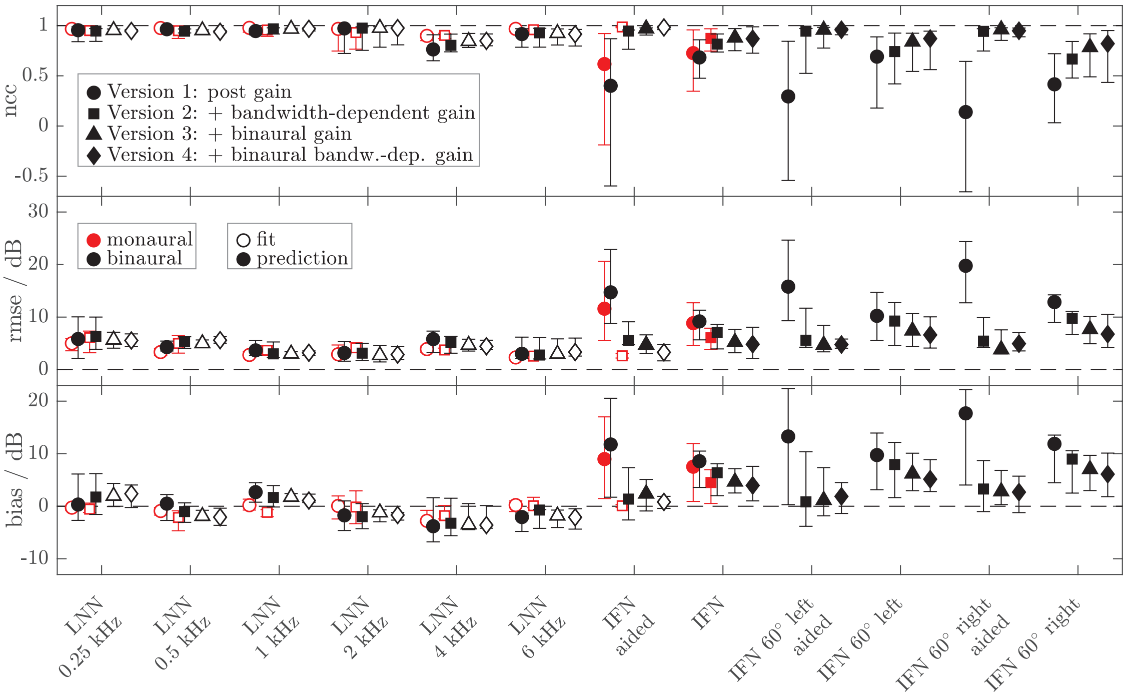

FIGURE 8 | Median (symbols) and 25 and 75 percentiles (bars) of the model performance measures (top panel: ncc, middle panel: rmse, bottom panel: bias) for monaural (red, mean value across left and right ear per listener) and diotic (black) conditions across the subgroup of four HI listeners for which loudness data of broadband IFN stimuli were collected in Dataset 1. The dashed lines indicate optimal performance. Four different model versions were tested as described in the Method section of experiment II (circles: version 1 with monaural post gain, squares: version 2 with additional monaural bandwidth-dependent gain, i.e., individualized parameters β_L and β_R, triangles: version 3 with additional overall binaural gain, i.e., individualized parameter α_B, diamonds: version 4 with additional binaural-bandwidth-dependent gain, i.e., individualized parameter β_B). Open symbols mark the conditions that were utilized in the model fits. Filled symbols indicate model predictions.

The individualized model version 4 has been successfully fitted to the aided IFN without many tradeoffs for the narrowband stimuli (open diamonds in **Figures 8**, **9**). However, the predictions of the HI data of Dataset 1 for all stimuli not used for the fitting procedure (closed diamonds) show no clear performance improvements in comparison with model version 3, whereas the predictions for the HI data of Dataset 2 are improved (unaided IFN only). For the HI group of Dataset 2, the median ncc for unaided IFN is increased from 0.89 for model version 3 to 0.96 for model version 4. The median rmse is reduced from 6.4 to 4.3 dB, and the median bias is reduced from 4 to 3 dB. The medians of the adjusted ncc' values are increased for both HI datasets (from 0.944 to 0.952 for the subgroup of Dataset 1 and from 0.959 to 0.964 for Dataset 2).

It has to be considered here that for version 4 the aided IFN data are added to the data used for fitting to allow a fit of the additional free parameter β_B. However, nearly the same improvements can be archived when using model version 3 (i.e., no parameter β_B) and including the same aided IFN data into the fitting procedure as well, referred to as model "version 3 modified." The last row in the lower half of **Table 3** shows the adjusted ncc' value of 0.961 for this modification for Dataset 2, which is almost as high as that for model version 4 (0.964). Likewise, the percentile values are comparable. The effect of this modification on the model performance was assessed for Dataset 1 as well. Again, aided IFN was added to the fitting data. Instead of the six binaural narrowband loudness functions originally used for fitting, only a single function (1 kHz) was used more comparable to the single UEN1 function (1.37 kHz) used for Dataset 2. Thus, comparable fitting data as for the above modified version 3 in Dataset 2 were used. Again, the resulting adjusted ncc' values are improved over the original model version 3 and are almost the same as those for model version 4 (see upper half in **Table 3**). Overall, although the additional parameter β_B improves the ability to fit the model to the loudness data, predictions of model version 4 are almost not improved over the predictions of model version 3 if the same underlying data are used for fitting. Using one binaural narrowband and one binaural broadband loudness function to determine α_B results in better performance than using the six binaural narrowband loudness functions.

The ncc' values only show a quite small increase for some of the models introducing binaural parameters. Given that the ncc' calculation includes more narrowband and monaural conditions than binaural and broadband conditions, differences between, e.g., models 2 and 3 might not be well-captured by the global ncc'.

To better illustrate the effect of the different model parameters, **Figure 10** shows example aided binaural IF noise loudness functions (solid gray lines) of listeners HI5 and HI7 of Dataset

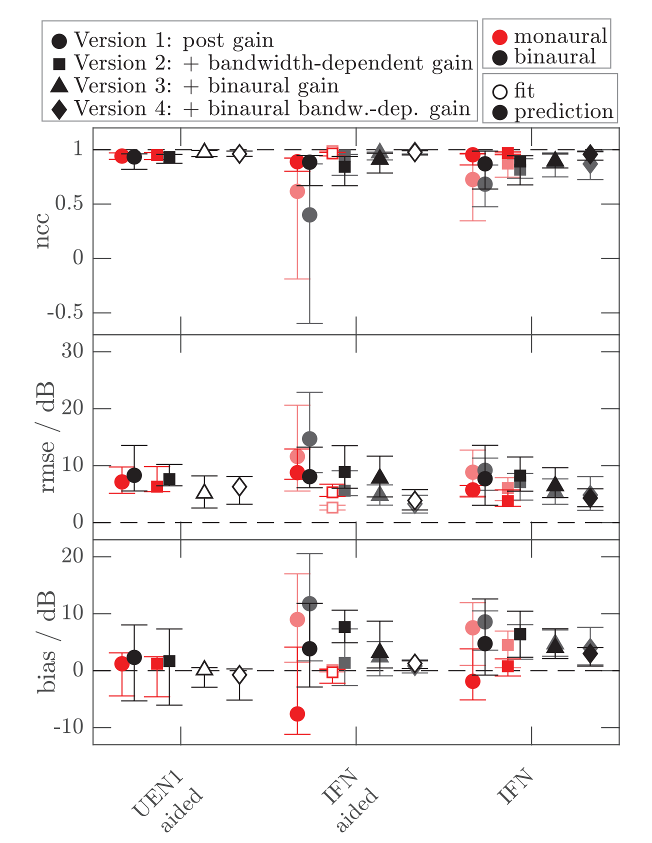

FIGURE 9 | As **Figure 8** but for the additional data of 10 HI listeners of Dataset 2. Performance measures from **Figure 8** are replotted in brightened colors if measurement conditions are comparable. In contrast to **Figure 8**, the performance measures for the monaural narrowband LNN stimuli are not shown. Binaural or diotic data are not available for the LNN stimuli. Instead, the loudness data of the aided UEN1 stimulus were used to fit the parameters of the binaural stage.

1 (also shown in **Figures 6**, **7**) and two additional listeners of Dataset 2 (HI04 and HI11) together with the respective predictions of all individualized model versions. HI7 and HI04 are examples where model 2 (dash-dotted) considerably underestimates loudness by two categories (10 CU) at a level that is perceived as "too loud" (50 CU). Similar or worse mismatches are obtained for four more HI listeners (not shown). Model versions 3 (dotted black) and 4, (solid), and modified model version 3 (dotted green) considerably reduce the deviations for HI7 and HI04 and two other HI listeners. Two of 14 listeners remain with errors slightly higher than 10 CU (not shown). Thus, for realistic conditions with binaural speech-shaped noise, for six of 14 individuals model version 3 and higher avoid a severe underestimation of loudness in conditions that would otherwise lead to overly loud sensations, which are particularly problematic in the context of hearing aids. This demonstrates that even if the benefit of additional model parameters might be small on average, it can be highly relevant for individuals.

DISCUSSION

Binaural Loudness Summation and Hearing Impairment

Both the modeling approaches in experiment II and the more data-driven approach in experiment I indicate decreased binaural inhibition (or even binaural excitation) in some of the HI listeners. This result appears to contradict the findings of van Beurden et al. (2018), who found similar binaural loudness summation between HI and NH listeners. The reason for this apparent contradiction is the differences in the methods on how to access and calculate binaural loudness summation. Whereas, in this study binaural loudness summation is calculated from the ratios between binaural and monaural loudness in sones at a given level (experiment I) or by fitting a single model parameter that alters the modeled binaural inhibition to the empirical loudness data (experiments I and II), van Beurden et al. (2018) calculated the level differences between monaural and binaural loudness functions at given loudness categories, but if the loudness ratios (in sones or CUs) are kept constant, the increase in the steepness of the loudness functions caused by the hearing impairment decreases the level differences between the functions (Moore et al., 2014). The fact that van Beurden et al. (2018) did not find such a decrease in these level differences in HI (in case of high hearing losses they even found a slight but not significant—increase) indicates that the loudness ratios at a given level must have been increased, which is in line with the observations in this study. On the contrary, Moore et al. (2014) found reduced level differences at frequencies where hearing loss was present. However, as mentioned in the introduction, a model that assumed average NH loudness ratios/inhibition predicted even lower-level differences and therefore underestimated binaural loudness summation in HI. To avoid conversion from CU to sones, other methods to directly assess loudness in sones, such as absolute magnitude estimation, appear suited. However, categorical loudness scaling has been shown to be well-applicable in the clinical context and for hearing aid fitting. Using absolute magnitude estimation, Marozeau and Florentine (2009) found increased inter-subject variability of the ratios in HI listeners, in line with the results of this study, but overall lower ratios and therefore no binaural excitation ($R > 2$). Their overall lower ratios can be explained with differences in the method: In this study, loudness in CU was transformed to loudness in sones using the transformation of Heeren et al. (2013). This transformation is based on the (sone-) loudness function in ANSI S3.4 (2007) resembling the loudness function in Hellman and Zwislocki (1961). Their loudness functions are considerably steeper than the respective loudness function derived with the method used by Marozeau and Florentine (see Epstein and Florentine, 2005). Rerunning the calculations in experiment I using the shallower loudness function yielded no binaural excitation, except for few stimuli for individual HI listeners. The increased inter-subject variability in the HI group, compared with that in the NH group, was still significant.

Based on this consideration, we hypothesize that the underlying physiological basis for *binaural excitation* ($R > 2$) could be: (1) super-additivity: neural excitation from both sides is

TABLE 3 | Adjusted ncc' (Equation 10) across all conditions.

Model version	Number of parameters	Number of observations used in fit	25 percentile	Median	75 percentile
SUBGROUP OF DATASET 1 (4 LISTENERS, 280 OBSERVATIONS PER LISTENER)					
1	36	120	0.765	0.840	0.904
2	38	140	0.919	0.936	0.949
3	39	200	0.937	0.944	0.956
4	40	210	0.935	0.952	0.963
3 modified	39	160	0.936	0.952	0.962
DATASET 2 (10 LISTENERS, 210 OBSERVATIONS PER LISTENER)					
1	36	120	0.901	0.933	0.958
2	38	140	0.905	0.950	0.960
3	39	150	0.947	0.959	0.964
4	40	160	0.946	0.964	0.971
3 modified	39	160	0.947	0.961	0.971

Shown are the 25 percentiles (column 4), the medians (column 5) and the 75 percentiles (column 6) across the HI models for the different model versions (column 1). Column 2 shows the total number of parameters for each model version. Hearing loss, distribution of OHC/IHC loss, and post gain at 6 frequencies per ear yield a total of 36 parameters for model version 1. Model version 2 introduces one parameter per ear (bandwidth-dependent gain), version 3 introduces the binaural gain, and version 4 introduces the bandwidth-dependent binaural gain. Column 3 shows the number of observations used to determine the parameter values in the model fit. Observations are 10 loudness categories for each stimulus (0 CU lies below the absolute threshold and is therefore excluded). The number of stimuli and, therefore, the number of observations differ across datasets. The model version denoted "3 modified" uses the same (Dataset 2) or comparable (Dataset 1) stimuli in the model fit as model version 4 in Dataset 2.

TABLE 4 | Same as **Table 3** but for NH listeners.

Model version	Number of parameters	Number of observations used in fit	25 percentile	Median	75 percentile
SUBGROUP OF DATASET 1 (FOUR LISTENERS, 280 OBSERVATIONS PER LISTENER)					
1	36	120	0.934	0.955	0.968
2	38	140	0.964	0.973	0.980
3	39	200	0.967	0.976	0.981
4	40	210	0.966	0.975	0.981
DATASET 2 (EIGHT LISTENERS, 180 OBSERVATIONS PER LISTENER)					
1	36	120	0.953	0.965	0.976
2	38	140	0.960	0.974	0.979
3	39	150	0.961	0.975	0.979
4	40	160	0.962	0.975	0.980

not added, but excitation from the contralateral side causes excess excitation on the ipsilateral side; and (2) an internal loudness representation with another slope than the sone scale used in this study in which case one would not observe $R = 2$ even if the binaural summation was purely additive.

Another limitation of this study might be the low sample size of eight NH and eight HI listeners for Dataset 1, and eight NH and 10 HI listeners for Dataset 2.

Individualization of Loudness Models

It has been shown that for some HI individuals, severe deviations from average loudness perception exist, likely causing problems in daily life and with hearing aids (Oetting et al., 2018; van Beurden et al., 2020). Even if only a subgroup of HI listeners is affected, loudness models that aim to support hearing aid fitting and development need to account for these listeners. Current HI loudness models fail in this regard (Pieper et al., 2018).

In Pieper et al. (2018), a monaural frequency-dependent post gain was introduced. The post gain allows fitting of the loudness model to individual narrowband loudness data but does not improve the model predictions for broadband loudness. In experiment II of this study, a bandwidth-dependent gain has been added to the monaural paths, controlled *via* parameters β_L, and β_R. The monaural bandwidth-dependent gain improves the ability of the loudness model to describe and predict monaural broadband data for both the NH and HI listeners.

The results of experiments I and II show that accounting for individual increased binaural loudness summation can decrease prediction errors of binaural loudness for narrowband and broadband stimuli. This holds in particular for a subgroup of the HI listeners and is almost independent of frequency and bandwidth.

In order to allow for individual binaural loudness summation, a binaural gain has been introduced that is controlled *via*

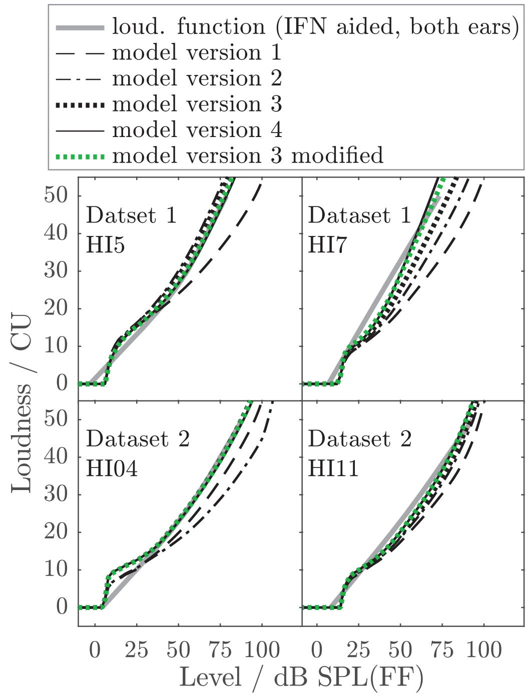

FIGURE 10 | Examples of empirical loudness functions for aided binaural IF noise (solid gray lines) of listeners HI05 and HI07 of Dataset 1 (also shown in **Figures 6**, **7**) and two listeners of Dataset 2 (HI04 and HI11), together with the respective predictions of all individualized model versions indicated in the legend.

parameter α_B. The binaural gain linearly attenuates (if $\alpha_B < 0$) or amplifies (if $\alpha_B > 0$) the signal, the more the signal amplitudes in the monaural paths are equal. Unlike the frequency-dependent post gain but like parameters β_L and β_R, α_B is a single parameter, independent of BM location and therefore independent of stimulus frequency. If modeling average NH inhibition (by setting α_B to the average value of -0.273 of the NH listeners), the predictions of individual binaural loudness data were inaccurate for the HI listeners compared with the more accurate predictions for the NH listeners. Allowing individual values α_B for the NH listeners slightly improved the ability to fit the binaural narrowband data but did not improve the predictions of binaural broadband data. For the HI listeners, predictions were improved and similar prediction accuracies as for the NH listeners were obtained.

A mechanism similar to the monaural bandwidth-dependent gain has been added to the binaural path, controlled *via* parameter β_B. The mechanism increases (for $\beta_B > 0$) or decreases (for $\beta_B < 0$) the binaural gain, the larger the bandwidths are and the more similar the signal amplitudes are in the monaural paths. This modification of the binaural gain did not further improve the predictions of the binaural broadband data, indicating that the individual amount of binaural inhibition would be almost independent of the bandwidth of the signal representation after monaural processing.

Therefore, it can be recommended that for the individualization of loudness models, the individual monaural spectral loudness summation should be addressed, independently of sensorineural hearing loss. If hearing loss is present, binaural loudness summation might be affected and therefore needs to be individualized as well. A single bandwidth-independent binaural gain (controlled here *via* parameter α_B), i.e., model version 3, might be sufficient for most applications.

In this study, a subset of the individual loudness data has been used to determine the parameter values of an individual model. The “cost” for the measurement time to obtain this subset might be too high for certain applications, in particular for clinical use. Most time consuming are the 12 monaural narrowband loudness functions that were used to obtain the OHC loss, IHC loss, and post gain. Approximations of these functions could be derived from hearing thresholds and UCL measurements to reduce measurement time. Two monaural broadband loudness functions were used to determine the values of β_L and β_R. Values of β_L and β_R were similar for most but not all listeners (not shown), so that the required measurement time cannot be reduced. Using one binaural narrowband loudness function and one binaural broadband loudness function to determine the individual value of α_B resulted in better model performance than using six binaural narrowband loudness functions.

Overall, to obtain a well-performing individual loudness model for an NH or HI listener, the measurement of 14 monaural loudness functions (12 narrowband, and two broadband) was required. For the HI listener, two additional binaural loudness functions (one narrowband, and one broadband) were required. Balancing “cost” and “value” of the measurements, a substantial reduction in measurement effort might be possible if the 12 monaural narrowband loudness functions are approximated based on more clinical data, such as the hearing threshold and uncomfortable level.

Frequency Dependency of Binaural Summation

The results of experiment I suggest high individual but only slight systematic frequency dependencies of binaural loudness summation averaged across listeners. Comparing the averaged results from NH and HI listeners, increased binaural loudness summation was found for the HI listeners (see **Tables 1**, **2**). This might suggest a connection between hearing loss and binaural loudness summation, which might also occur for frequency-dependent hearing loss within a listener. Given that hearing loss is typically increased at high frequencies, one could, thus, expect increased binaural summation at high frequencies. Contrary to this consideration, all except one HI listener (HI04) show decreasing summation ratios with increasing stimulus frequency (**Figure 4**). On average across listeners, decreasing summation ratios with increasing frequency was observed for both the NH and HI listeners. However, the binaural summation (parameter α_B) of the model was chosen to be independent of frequency. Consequently, on average, the binaural loudness summation is underestimated at low frequencies (compare positive bias values for binaural conditions with bias values close to zero for monaural conditions for LNN stimuli with low center frequencies in **Figure 8**) and overestimated at high frequencies (compare negative bias values for binaural conditions with bias values close to zero for monaural conditions for LNN stimuli with high center frequencies in **Figure 8**) in the model calculations of experiment II. Such underestimation at low frequencies was already observed in Moore et al. (2014) using a model that also assumed binaural inhibition to be independent of frequency. Thus, further small improvements can be expected for the predictions of narrowband signals if a slight decrease in binaural summation with increasing frequency would be considered. This can be realized in this model by a decrease in the value of α_B as a function of the center frequency of the TLM segment.

Relations to Hearing Impairment and Physiologic Mechanisms

In this model, the binaural stage is located subsequent to the simulation of basilar membrane movements and monaural central gain mechanisms (post gain and bandwidth-dependent central gain). Therefore, binaural inhibition is considered to be a more central effect. Before its introduction into binaural loudness models (Moore and Glasberg, 2007), the idea of central binaural inhibition mechanisms has been used in binaural auditory models for sound localization and binaural unmasking (Lindemann, 1986; Breebaart et al., 2001). Breebaart et al. (2001) argued that a subgroup of cells in the mammalian lateral superior olive and the inferior colliculus are excited by signals from one ear and inhibited by signals from the other ear. Since there is evidence for homeostatic adaptations of the neurons to reduced firing rates at the inputs in case of hearing impairment (Qiu et al., 2000; Kotak et al., 2005), reduced binaural inhibition might be a side effect of these adaptations. However, although central inhibition can explain the mentioned psychoacoustic observations, the link between neural stimulus encoding and the neural representations of percepts, including loudness, is not well-understood (Schreiner and Malone, 2015). Further candidates that can potentially influence binaural loudness summation are efferent reflexes like the MEM or the medial olivocochlear (MOC) reflex. The MOC reflex is directly affecting the cochlear gain (e.g., Berlin et al., 1993), whereas the MEM reflex causes a reduction of sound transmission by the middle ear of up to 10 dB for high stimulus levels at frequencies below 1 kHz (Rabinowitz, 1977). In both MEM reflex and MOC reflex, threshold and strength depend on stimulus level, frequency, and bandwidth as well as stimulus presentation (monaural or binaural). Both reflexes are feedback mechanisms that are controlled by post cochlear processes, and they are affected by damages to the auditory path prior to the central processing stages. For example, the characteristics of the MEM reflex threshold depend on the different peripheral compression in the NH and HI listeners (Müller-Wehlau et al., 2005). Thus, these effects might provide a direct link between the individual state of outer and inner hair cells and the individual amount of binaural inhibition. If the influence of the hair cell states on binaural inhibition is high in comparison to central binaural inhibition effects, a proper implementation of these feedback mechanisms could reduce the number of parameters that are required for the individualization of loudness models.

By reducing the cochlear gain of the TLM, the proposed loudness model accounts for reduced spectral loudness summation in the HI listeners. Nevertheless, subsequent to the TLM, significant bandwidth-dependent gain changes (controlled *via* parameters β_L and β_R) were necessary to describe the individual monaural broadband data (results of experiment II). These subsequent corrections were not only necessary for the HI but for the NH listeners as well, for which only small cochlear gain losses were expected and therefore simulated. Together with the finding that a similar mechanism applied to the binaural path of the model did not improve binaural loudness predictions of broadband stimuli, the model simulations suggest an additional mechanism besides the cochlear non-linearities that influence spectral loudness summation, which is not related to hearing loss, as already hypothesized in Pieper et al. (2018).

Implications for Loudness Models and Application in Hearing Aid Fitting and Development

This model analysis estimated the number and type of monaural and binaural parameters required to improve fitting to and prediction of individual loudness data. It is generally expected that an increasing number of free parameters increase the ability of any model to fit the data. The goal was, therefore, to systematically assess the benefit of successively adding perceptually motivated and physiologically plausible, effective model stages with respective parameters and to devise the minimum number required for individualization. We show that, in fact, additional stages beyond commonly considered peripheral processes, such as non-linear gain loss and linear attenuation, typically associated with OHC and IHC, loss are required to account for individual loudness data in NH and HI. At least a bandwidth-dependent retro-cochlear gain parameter, likely reflecting central gain, is required to individualize the amount of spectral summation, and a frequency- and bandwidth-independent binaural gain parameter is required to individualize binaural summation (inhibition or excitation). The authors deem it unlikely that individual loudness can be accounted for with any improved peripheral model without the need for the suggested or similar additional parameters.

Although a specific loudness model was used in this study, the findings can be generally applied to other loudness models and do not depend on the front end of the current model. Other loudness models (e.g., Chalupper and Fastl, 2002; Chen et al., 2011a; Moore et al., 2014) could be extended with the suggested or modified retro-cochlear stages, which introduce additional individual parameters. The front ends of these models offer the advantage of strongly reduced computational complexity compared with the current TLM front end.

The potential of individualized loudness predictions is in hearing aid fitting, where a fitting rule can contain the individualized model and improved hearing aid gains can be devised for different situations based on the respective prominent signal properties and the wearers individual loudness perception. Further potential is in hearing aid development, where loudness can be predicted for a certain set of prototypical HI with different loudness perceptions. Future potential can be expected in hearing aid algorithms with real-time updates of their processing based on integrated individual loudness predictions. Although there is still room for further improvement of individual loudness predictions, the relevance of the already achieved, at times seemingly small differences or improvements in dB, should not be underestimated. Due to the steep progression of HI loudness functions, in particular close to uncomfortable levels, according gain changes in hearing aids can easily make the difference between acceptable and uncomfortable loudness.

SUMMARY AND CONCLUSIONS

Loudness perception of the NH and HI listeners with sensorineural hearing loss was measured by categorical loudness scaling for narrowband and broadband stimuli, presented monaurally, and binaurally. To assess the individual amount of binaural summation, binaural loudness ratios were calculated, and a data-driven model approach was employed to account for binaural loudness based on the measured monaural loudness by individually fitting a single binaural summation parameter, α_B. Analysis of the loudness data showed a higher individual variability of binaural loudness summation for HI compared with the NH. While NH showed binaural inhibition in line with previous findings from the literature, the data of some of the HI listeners of this study suggest reduced binaural inhibition ($\alpha_B < 0$) or even super additive summation, i.e., binaural excitation ($\alpha_B > 0$).

In the second step, the monaural loudness model of Pieper et al. (2018) was extended by a functional binaural loudness summation stage (Equation 3). The stage sums the signals in the monaural paths and weights the result depending on the amplitude difference in the monaural paths and the value of the parameter α_B that controls the overall amount of binaural inhibition or excitation. Loudness model predictions for binaural stimulus presentation were improved for individual HI listeners if the individual amount of binaural inhibition/excitation was considered. The introduction of an additional parameter β_B that alters the amount of binaural inhibition/excitation depending on the bandwidth of the summed monaural signals did not substantially improve the model predictions. However, for the accuracy of the model predictions in both the NH and HI listeners, it was crucial to include bandwidth-dependent weightings of the signals in the monaural paths (Equation 1) that were controlled with parameters β_L and β_R for the left and right ears, respectively.

The following conclusions can be drawn:

1. Individual ratios of binaural loudness summation vary across the HI listeners and are sometimes increased compared with the NH listeners, indicating that binaural inhibition, as typically observed in NH, might be affected by sensorineural hearing loss.
2. The empirical data suggest a slight increase in binaural inhibition (or decrease in binaural excitation) with frequency for both the NH and HI listeners.
3. To correctly account for spectral loudness summation, individualized loudness models for NH and HI should include an individually adapted bandwidth-dependent retro cochlear gain stage in the monaural pathway.
4. Individualized loudness models for HI listeners should account for the individual amount of binaural inhibition/excitation.

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.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

IP, MM, and SE co-conceived the presented ideas. SE supervised the project. IP developed the test software and carried out the simulations and experiments. All the authors discussed the results and contributed to the manuscript.

FUNDING

This study was supported by the Deutsche Forschungsgemeinschaft (DFG, Cluster of Excellence EXC 2177 Hearing4all and project 352015383—SFB 1330 A2). Most of the simulations were performed on the HPC Cluster CARL funded by the DFG under INST 184/157-1 FUGG.

ACKNOWLEDGMENTS

We are grateful to Dirk Oetting for sharing experimental data and discussion, to Hongmei Hu for comments on earlier versions of the manuscript, and to all the participants in the experiments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.634943/full#supplementary-material

The loudness functions and the underlying raw loudness data of 8 NH and 8 HI listeners collected for the current study (Dataset 1) as well as model calculations are provided.

REFERENCES

ANSI S3.4 (2007). *Procedure for the Computation of Loudness of Steady Sounds. American National Standard Acoustical Society of America Accredited Standards Committee S3 Bioacoustics* (New York, NY).

Berlin, C. I., Hood, L. J., Hurley, A. E., Wen, H., and Kemp, D. T. (1995). Binaural noise suppresses linear click-evoked otoacoustic emissions more than ipsilateral or contralateral noise. *Hear. Res.* 87, 96–103. doi: 10.1016/0378-5955(95)00082-F

Berlin, C. I., Hood, L. J., Wen, H., Szabo, P., Cecola, R. P., Rigby, P., et al. (1993). Contralateral suppression of non-linear click-evoked otoacoustic emissions. *Hear. Res.* 71, 1–11. doi: 10.1016/0378-5955(93)90015-S

Brand, T., and Hohmann, V. (2002). An adaptive procedure for categorical loudness scaling. *J. Acoust. Soc. Am.* 112, 1597–1604. doi: 10.1121/1.1502902

Breebaart, J., Van de Par, S., and Kohlrausch, A. (2001). Binaural processing model based on contralateral inhibition. I. Model structure. *J. Acoust. Soc. Am.* 110, 1074–1088. doi: 10.1121/1.1383297

Brotherton, H., Plack, C. J., Maslin, M., Schaette, R., and Munro, K. J. (2015). Pump up the volume: could excessive neural gain explain tinnitus and hyperacusis? *Audiol. Neurotol.* 20, 273–282. doi: 10.1159/000430459

Byrne, D., Dillon, H., Ching, T., Katsch, R., and Keidser, G. (2001). NAL-NL1 procedure for fitting non-linear hearing aids: characteristics and comparisons with other procedures. *J. Am. Acad. Audiol.* 12, 37–51.

Byrne, D., Dillon, H., Tran, K., Arlinger, S., Wilbraham, K., Cox, R., et al. (1994). An international comparison of long-term average speech spectra. *J.Acoust. Soc. Am.* 96, 2108–2120. doi: 10.1121/1.410152

Chalupper, J., and Fastl, H. (2002). Dynamic loudness model. (DLM). for normal and hearing-impaired listeners. *Acta Acust. United Acust.* 88, 378–386.

Chen, Z., Hu, G., Glasberg, B. R., and Moore, B. C. J. (2011a). A new model for calculating auditory excitation patterns and loudness for cases of cochlear hearing loss. *Hear. Res.* 282, 69–80. doi: 10.1016/j.heares.2011.09.007

Chen, Z., Hu, G., Glasberg, B. R., and Moore, B. C. J. (2011b). A new method of calculating auditory excitation patterns and loudness for steady sounds. *Hear. Res.* 282, 204–215. doi: 10.1016/j.heares.2011.08.001

Derleth, R. P., Dau, T., and Kollmeier, B. (2001). Modeling temporal and compressive properties of the normal and impaired auditory system. *Hear. Res.* 159, 132–149. doi: 10.1016/S0378-5955(01)00322-7

Epstein, M., and Florentine, M. (2005). A test of the Equal-Loudness-Ratio hypothesis using cross-modality matching functions. *J. Acoust. Soc. Am.* 118, 907–913. doi: 10.1121/1.1954547

Epstein, M., and Florentine, M. (2009). Binaural loudness summation for speech and tones presented via earphones and loudspeakers. *Ear Hear.* 30, 234–237. doi: 10.1097/AUD.0b013e3181976993

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* (Merano), 1326–1329.

Ewert, S. D., and Oetting, D. (2018). Loudness summation of equal loud narrowband signals in normal-hearing and hearing-impaired listeners. *Int. J. Audiol.* 57:S71–S80. doi: 10.1080/14992027.2017.1380848

Faingold, C. L., Anderson, C. A. B., and Randall, M. E. (1993). Stimulation or blockade of the dorsal nucleus of the lateral lemniscus alters binaural and tonic inhibition in contralateral inferior colliculus neurons. *Hear. Res.* 69, 98–106. doi: 10.1016/0378-5955(93)90097-K

Fastl, H., and Zwicker, E. (2007). *Psychoacoustics: Facts and Models, Third ed.* Berlin: Springer. doi: 10.1007/978-3-540-68888-4

Finlayson, P. G., and Caspary, D. M. (1991). Low-frequency neurons in the lateral superior olive exhibit phase-sensitive binaural inhibition. *J. Neurophys.* 65, 598–605. doi: 10.1152/jn.1991.65.3.598

Guinan, J. J. Jr. (2006). Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans. *Ear Hear.* 27, 589–607. doi: 10.1097/01.aud.0000240507.83072.e7

Heeren, W., Hohmann, V., Appell, J. E., and Verhey, J. L. (2013). Relation between loudness in categorical units and loudness in phons and sones. *J. Acoust. Soc. Am.* 133, EL314–EL319. doi: 10.1121/1.4795217

Heinz, M. G., Issa, J. B., and Young, E. D. (2005). Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment. *J. Assoc. Res. Otolaryngol.* 6, 91–105. doi: 10.1007/s10162-004-5043-0

Hellman, R. P., and Zwislocki, J. (1961). Some factors affecting the estimation of loudness. *J. Acoust. Soc. Am.* 33, 687–694. doi: 10.1121/1.1908764

Hellman, R. P., and Zwislocki, J. (1963). Monaural loudness function at 1000 cps and interaural summation. *J. Acoust. Soc. Am.* 35, 856–865. doi: 10.1121/1.1918619

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

ISO 389-7 (2005). *Acoustics - Reference Zero for the Calibration of Audiometric Equipment. Part 7: Reference Threshold of Hearing Under Free-Field and Diffuse-Field Listening Conditions.* Geneva: International Organization for Standardization.

Jenstad, L. M., Van Tasell, D. J., and Ewert, C. (2003). Hearing aid troubleshooting based on patients' descriptions. *J. Am. Acad. Audiol.* 14, 347–360. doi: 10.1055/s-0040-1715754

Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B. (2009). Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses. *EURASIP J. Adv. Signal Process.* 2009:6. doi: 10.1155/2009/298605

Keidser, G., Dillon, H., Carter, L., and O'Brien, A. (2012). NAL-NL2 empirical adjustments. *Trends Amplif.* 16, 211–223. doi: 10.1177/1084713812468511

Kotak, V. C., Fujisawa, S., Lee, F. A., Karthikeyan, O., Aoki, C., and Sanes, D. H. (2005). Hearing loss raises excitability in the auditory cortex. *J. Neurosci.* 25, 3908–3918. doi: 10.1523/JNEUROSCI.5169-04.2005

Launer, S. (1995). *Loudness Perception in Listeners With Sensorineural Hearing Impairment*. Oldenburg: Unpublished Ph D thesis, Oldenburg University.

Li, L., and Yue, Q. (2002). Auditory gating processes and binaural inhibition in the inferior colliculus. *Hear. Res.* 168, 98–109. doi: 10.1016/S0378-5955(02)00356-8

Lindemann, W. (1986). Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. *J. Acoust. Soc. Am.* 80, 1608–1622. doi: 10.1121/1.394325

Marozeau, J., Epstein, M., Florentine, M., and Daley, B. (2006). A test of the binaural equal-loudness-ratio hypothesis for tones. *J. Acoust. Soc. Am.* 120, 3870–3877. doi: 10.1121/1.2363935

Marozeau, J., and Florentine, M. (2009). Testing the binaural equal-loudness-ratio hypothesis with hearing-impaired listeners. *J. Acoust. Soc. Am.* 126, 310–317. doi: 10.1121/1.3133703

Møller, A. R. (1962). Acoustic reflex in man. *J. Acoust. Soc. Am.* 34, 1524–1534. doi: 10.1121/1.1918384

Moore, B. C. J. (2000). Use of a loudness model for hearing aid fitting. IV. Fitting hearing aids with multi-channel compression so as to restore 'normal'loudness for speech at different levels. *Br. J. Audiol.* 34, 165–177. doi: 10.3109/03005364000000126

Moore, B. C. J., Gibbs, A., Onions, G., and Glasberg, B. R. (2014). Measurement and modeling of binaural loudness summation for hearing-impaired listeners. *J. Acoust. Soc. Am.* 136, 736–747. doi: 10.1121/1.4889868

Moore, B. C. J., and Glasberg, B. R. (1998). Use of a loudness model for hearing-aid fitting. I. Linear hearing aids. *Br. J. Audiol.* 32, 317–335. doi: 10.3109/03005364000000083

Moore, B. C. J., and Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hear. Res.* 188, 70–88. doi: 10.1016/S0378-5955(03)00347-2

Moore, B. C. J., and Glasberg, B. R. (2007). Modeling binaural loudness. *J. Acoust. Soc. Am.* 121, 1604–1612. doi: 10.1121/1.2431331

Moore, B. C. J., Glasberg, B. R., and Baer, T. (1997). A model for the prediction of thresholds, loudness, and partial loudness. *J. Audio Engin. Soc.* 45, 224–240.

Moore, B. C. J., Glasberg, B. R., and Stone, M. A. (1999). Use of a loudness model for hearing aid fitting: III. A general method for deriving initial fittings for hearing aids with multi-channel compression. *Br. J. Audiol.* 33, 241–258. doi: 10.3109/03005369909090105

Moore, B. C. J., Glasberg, B. R., Varathanathan, A., and Schlittenlacher, J. (2016). A loudness model for time-varying sounds incorporating binaural inhibition. *Trends Hear.* 20, 1–16. doi: 10.1177/2331216516682698

Müller-Wehlau, M., Mauermann, M., Dau, T., and Kollmeier, B. (2005). The effects of neural synchronization and peripheral compression on the acoustic-reflex threshold. *J. Acoust. Soc. Am.* 117, 3016–3027. doi: 10.1121/1.1867932

Norman, M., and Thornton, A. R. D. (1993). Frequency analysis of the contralateral suppression of evoked otoacoustic emissions by narrow-band noise. *Br. J. Audiol.* 27, 281–289. doi: 10.3109/03005369309076705

Oetting, D., Brand, T., and Ewert, S. D. (2014). Optimized loudness-function estimation for categorical loudness scaling data. *Hear. Res.* 316, 16–27. doi: 10.1016/j.heares.2014.07.003

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., and Ewert, S. D. (2016). Spectral and binaural loudness summation for hearing-impaired listeners. *Hear. Res.* 335, 179–192. doi: 10.1016/j.heares.2016.03.010

Oetting, D., Hohmann, V., Appell, J.-E., Kollmeier, B., and Ewert, S. D. (2018). Restoring perceived loudness for listeners with hearing loss. *Ear Hear.* 39, 644–678. doi: 10.1097/AUD.0000000000000521

Oetting, D., Hohmann, V., Ewert, S. D., and Appell J-E. (2013). "Model-based loudness compensation for broad- and narrow-band signals," *ISAAR-International Symposium on Auditory and Audiological Research Auditory Plasticity - Listening With the Brain* (Ballerup), 365–372.

Pieper, I., Mauermann, M., Kollmeier, B., and Ewert, S. D. (2016). Physiological motivated transmission-lines as front end for loudness models. *J. Acoust. Soc. Am.* 139, 2896–2910. doi: 10.1121/1.4949540

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

Puria, S. (2003). Measurements of human middle ear forward and reverse acoustics: implications for otoacoustic emissions. *J. Acoust. Soc. Am.* 113, 2773–2789. doi: 10.1121/1.1564018

Qiu, C., Salvi, R., Ding, D., and Burkard, R. (2000). Inner hair cell loss leads to enhanced response amplitudes in auditory cortex of unanesthetized chinchillas: evidence for increased system gain. *Hear. Res.* 139, 153–171. doi: 10.1016/S0378-5955(99)00171-9

Rabinowitz, W. M. (1977). *Acoustic-reflex effects on the input admittance and transfer characteristics of the human middle-ear*. (Ph.D. Thesis). Massachusetts Institute of Technology, United Stastes.

Rennies, J., Verhey, J. L., Chalupper, J., and Fastl, H. (2009). Modeling temporal effects of spectral loudness summation. *Acta Acust. United With Acust.* 95, 1112–1122. doi: 10.3813/AAA.918243

Schreiner, C. E., and Malone, B. J. (2015). Representation of loudness in the auditory cortex. *Handbook Clin. Neurol.* 129, 73–84. doi: 10.1016/B978-0-444-62630-1.00004-4

Sivonen, V. P., and Ellermeier, W. (2006). Directional loudness in an anechoic sound field, head-related transfer functions, and binaural summation. *J. Acoust. Soc. Am.* 119, 2965–2980. doi: 10.1121/1.2184268

Strelcyk, O., Nooraei, N., Kalluri, S., and Edwards, B. (2012). Restoration of loudness summation and differential loudness growth in hearing-impaired listeners. *J. Acoust. Soc. Am.* 132, 2557–2568. doi: 10.1121/1.4747018

van Beurden, M., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., and Dreschler, W. A. (2018). Potential consequences of spectral and binaural loudness summation for bilateral hearing aid fitting. *Trends Hear.* 22:5690. doi: 10.1177/2331216518805690

van Beurden, M., Boymans, M., van Geleuken, M., Oetting, D., Kollmeier, B., and Dreschler, W. A. (2020). Uni-and bilateral spectral loudness summation and binaural loudness summation with loudness matching and categorical loudness scaling. *Int. J. Aud.* 60, 1–9. doi: 10.1080/14992027.2020.1832263

Verhulst, S., Altoe, A., and Vasilkov, V. (2018). Computational modeling of the human auditory periphery: auditory-nerve responses, evoked potentials and hearing loss. *Hear. Res.* 360, 55–75. doi: 10.1016/j.heares.2017.12.018

Whilby, S., Florentine, M., Wagner, E., and Marozeau, J. (2006). Monaural and binaural loudness of 5-and 200-ms tones in normal and impaired hearing. *J. Acoust. Soc. Am.* 119, 3931–3939. doi: 10.1121/1.2193813

Zeng, F.-G. (2013). An active loudness model suggesting tinnitus as increased central noise and hyperacusis as increased non-linear gain. *Hear. Res.* 295, 172–179. doi: 10.1016/j.heares.2012.05.009

Zwicker, E., and Scharf, B. (1965). A model of loudness summation. *Psychol. Rev.* 72, 3–26. doi: 10.1037/h0021703

Zwicker, E., and Zwicker, U. T. (1991). Dependence of binaural loudness summation on interaural level differences, spectral distribution, and temporal distribution. *J. Acoust. Soc. Am.* 89, 756–764. doi: 10.1121/1.1894635

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.

APPENDIX

Binaural Model Stage for Normal Hearing and Refinement of Loudness Transformations

To obtain loudness in sones S_m (at time step m), a power law with the exponent B is applied to the internal binaural loudness I_m and the result scaled with the factor A (Pieper et al., 2016, 2018):

$$S_m = A \cdot I_m^B$$

Loudness in CU can then be obtained using the five-parameter cubic function of Heeren et al. (2013):

$$\begin{aligned} CU = \; & a_3 \lg\left(S/sone + b\right)^3 + a_2 \lg\left(S/sone + b\right)^2 \\ & + a_1 \lg\left(S/sone + b\right) + c \end{aligned}$$

In Pieper et al. (2018) the parameters of the transformations have been fitted to averaged NH binaural (A and B) and monaural (a_1, a_2, a_3 b, and c) narrowband loudness data at 1 kHz, resulting in the values: $A = 2.1 \cdot 10^6$, $B = 0.768$, $a_1 = 8.8$, $a_2 = 3.02$, $a_3 = 4.47$, $b = 0.092$, and $c = 13.3$. However, to fit the transformation from sones to CU, binaural loudness in sones was, for simplicity, divided by 2. Given that the current binaural summation stage differs from this assumption, these parameters needed to be refined. Furthermore, a binaural summation stage for averaged NH data was required for model versions 1 and 2 to access the benefit of its individualization in model version 3 and 4. For this, an iterative procedure was performed involving the fitting procedure of the loudness transformations and the individualization procedure of model version 3:

In the first iteration step, the parameter α_B of the binaural summation stage (Equation 3) was set to a value that reflects average NH inhibition. The transformations from internal loudness to loudness in sones and from sones to CU were fitted to average NH loudness data from other studies as mentioned above. In the first iteration α_B was set to -0.36 as inferred from experiment I.

In the second iteration step, the individual model version 3 was adjusted for all NH listeners of Datasets 1 and 2 as described in the method section of experiment II, using the loudness transformations from the first iteration step. The mean across the resulting individual values of α_B was then used for the binaural summation stage in the first step of the next iteration.

The iteration was repeated until the average value of α_B did not change by more than 1% from the previous iteration (here four iterations were sufficient). The resulting average value is $\alpha_B = -0.273$. The resulting parameter values for the transformations are $A = 1.44 \cdot 10^6$, $B = 0.795$, $a_1 = 6.23$, $a_2 = 2.22$, $a_3 = 3.709$, $b = 0.053$, and $c = 12.1$. These transformations were used for all model versions in experiment II.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 01 September 2021
doi: 10.3389/fnins.2021.634967

Manual Gestures Modulate Early Neural Responses in Loudness Perception

*Jiaqiu Sun[1,2], Ziqing Wang[2,3] and Xing Tian[1,2,3]**

[1] Division of Arts and Sciences, New York University Shanghai, Shanghai, China, [2] NYU-ECNU Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai, China, [3] Shanghai Key Laboratory of Brain Functional Genomics, Ministry of Education, School of Psychology and Cognitive Science, East China Normal University, Shanghai, China

How different sensory modalities interact to shape perception is a fundamental question in cognitive neuroscience. Previous studies in audiovisual interaction have focused on abstract levels such as categorical representation (e.g., McGurk effect). It is unclear whether the cross-modal modulation can extend to low-level perceptual attributes. This study used motional manual gestures to test whether and how the loudness perception can be modulated by visual-motion information. Specifically, we implemented a novel paradigm in which participants compared the loudness of two consecutive sounds whose intensity changes around the just noticeable difference (JND), with manual gestures concurrently presented with the second sound. In two behavioral experiments and two EEG experiments, we investigated our hypothesis that the visual-motor information in gestures would modulate loudness perception. Behavioral results showed that the gestural information biased the judgment of loudness. More importantly, the EEG results demonstrated that early auditory responses around 100 ms after sound onset (N100) were modulated by the gestures. These consistent results in four behavioral and EEG experiments suggest that visual-motor processing can integrate with auditory processing at an early perceptual stage to shape the perception of a low-level perceptual attribute such as loudness, at least under challenging listening conditions.

Keywords: multisensory integration, cross-modal modulation, audiovisual, manual gesture, motion perception, action, loudness perception

OPEN ACCESS

Edited by:
*Maaike Van Eeckhoutte,
Technical University of Denmark,
Denmark*

Reviewed by:
*Hamish Innes-Brown,
Eriksholm Research Centre, Denmark
Gerard Encina-Llamas,
Technical University of Denmark,
Denmark*

***Correspondence:**
*Xing Tian
xing.tian@nyu.edu*

Specialty section:
*This article was submitted to
Auditory Cognitive Neuroscience,
a section of the journal
Frontiers in Neuroscience*

Received: *29 November 2020*
Accepted: *06 August 2021*
Published: *01 September 2021*

Citation:
*Sun J, Wang Z and Tian X (2021)
Manual Gestures Modulate Early
Neural Responses in Loudness
Perception.
Front. Neurosci. 15:634967.
doi: 10.3389/fnins.2021.634967*

INTRODUCTION

Imagine that you are boasting about the size of the fish you caught last weekend to your friend. You would probably raise your voice volume when you say the word "big," and at the same time move your hands away from each other. The iconic gestures in this example not only represent the size of the fish visually but also parallel the volume of your voice. Let's go a bit further. Suppose that two utterances have the same intensity; if a gesture accompanies one but not the other sound, would you perceive one sound as quieter or louder than the other sound? In general, whether and how the informational contents in one modality penetrate the processing in another modality is a fundamental question for understanding the nature of human perception.

Multisensory integration has been extensively documented (Calvert et al., 2004; Ghazanfar and Schroeder, 2006; Stein and Stanford, 2008). In the domain of multisensory audiovisual interaction, most studies explored the cross-modal effects in ecologically valid connections. For example, the McGurk effect (McGurk and MacDonald, 1976) is established by naturally linked speech categorical representations in the visual and auditory domain (Möttönen et al., 2002; Besle et al., 2004; van Wassenhove et al., 2005, 2007; Arnal et al., 2009; Baart et al., 2014). The ventriloquist effect is based on a high probability in the natural world that the source of visual and auditory information comes from a common identity and location (Howard and Templeton, 1966; Alais and Burr, 2004; Bonath et al., 2007; Alais et al., 2010). However, the boundary and efficacy of cross-modality modulation effects have not been thoroughly explored. For example, it has been extensively demonstrated that gestures and language processing are linked (Krauss, 1998; Arbib et al., 2008; Goldin-Meadow and Alibali, 2013). Most studies revealed this cross-modal connection at higher levels such as semantic, lexical, and phonological levels. Studies on cross-modal interaction occurring for low-level perceptual attributes were relatively rare, such as loudness in auditory perception and distance in visual perception. Compared with other auditory perceptual attributes such as phonetic, phonological, and prosodic features of speech sound, the perceived loudness is at a lower level in the hierarchy of auditory and speech processing.

Recent studies of audiovisual integration using gestures can provide some hints. Gestures can influence auditory perception *via* the linked speech categorical representations at the semantic and phonological levels (Kelly et al., 2004; Özyürek et al., 2007; Willems et al., 2007). For example, gestures (either semantically matching or mismatching) interacted with the N1–P2 auditory responses of words (Kelly et al., 2004). Gestures such as beat (Hubbard et al., 2009) and clapping (Stekelenburg and Vroomen, 2012; van Laarhoven et al., 2017) also influence auditory processing *via* a spatial–temporal contingency. Such modulation resulted from the expected frequency of an acoustic event predicted by the gesture (van Laarhoven et al., 2017). Recently, the basis of cross-modal connections has extended to more basic features such as direction. For example, manual directional gestures can facilitate learning lexical tones in Mandarin Chinese (Zhen et al., 2019). All these results suggest that gestures and acoustic features may share overlapped or transformable representations that would enable across-modal integration for low-level perceptual attributes that do not necessarily link in two modalities. One more interesting phenomenon is that when human participants instinctively made gestures during listening to music, the position of gesture positively correlated with the intensity of the sound (Caramiaux et al., 2010). Will the universal dimension of magnitude, the lowest level perceptual attribute in the perception of all modalities, serves as a connection for multisensory integration in general and a basis for gestural effects on auditory perception in particular?

In this study, we investigated whether and how manual gestures can modulate loudness perception. We developed a new multimodal paradigm in which participants heard the same vowel/a/twice with manual gestures concurrently presented with the second sound. Participants judged the loudness change of the second sound relative to the first sound. We hypothesized that the visual-motion information of gestures would modulate the perceived loudness. To test this hypothesis, we first carried out two behavioral experiments (BE1 and BE2). In BE1, we probe the effects of natural motion gestures on the judgments of loudness changes. To distinguish which features (the distance or the motion) of the gestures influenced the judgments of loudness changes, we carried out BE2 using still images of gestures. We carried out two more EEG experiments (EE1 and EE2) to further investigate whether the effects were perceptual (rather than decisional) in nature by examining the temporal dynamics of the modulation effects. Specifically, we compared the early auditory event-related potential (ERP) responses between conditions of different gestures (EE1) and between trials of different loudness judgment to the same sound (EE2).

In the EEG experiments, we focused on the ERP N100 component that is an early cortical response reflecting (auditory) perceptual analysis (Roberts et al., 2000). The auditory N100 is a fronto-centrally distributed negative wave that is mainly generated in the (primary and associative) auditory cortex (Näätänen and Picton, 1987). Previous studies found that N100 amplitude correlates with perceived loudness. Schmidt et al. (2020) observed that the preceding tone (inducer tone) decreased the perceived loudness of the target tone. Tian et al. (2018) demonstrated that the preceding imagined speech lowered the loudness ratings of the target sound. Both studies showed that the contextual effects on changing loudness perception correlated with the magnitude changes in N1/P2 components in the responses to the target sound. In addition, Lu et al. (1992) observed that the decay rate of N100 amplitude correlated with the decay rate of the loudness perception. Our results suggested that certain visual-motion information in manual gestures modulated the early auditory neural responses (around 110 ms) that corresponded to changes in loudness perception at the just-noticeable difference (JND) threshold.

MATERIALS AND METHODS

Participants

Fifteen young adults (10 females; mean age, 22.1 years; range, 19–25 years) participated in BE1; 12 young adults (7 females; mean age, 22.0 years; range, 20–24 years) participated in BE2; 23 young adults (16 females; mean age, 22.0 years; range, 17–27 years) participated in EE1; 20 young adults (10 females; mean age, 21.8 years; range, 18–25 years) participated in EE2. There was no overlapping of the participants among all four experiments. All participants were native Chinese speakers. They all had normal hearing and normal or corrected-to-normal vision (this was listed in the requirement when recruiting participants, and we also verbally confirmed that). None of them had any neurological deficits (self-reported). They received monetary incentives for their participation. Written informed consents were obtained for all participants before the experiments. The local Research Ethics Committee at NYU Shanghai approved all protocols.

Stimuli and Trial Procedure

Behavioral Experiment 1 (BE1): The Effects of Motional Gestures on the Judgment of Loudness Changes

In BE1, we used natural motional gestures in 1920 × 1080-pixel movie clips with a frame rate of 25 fps. The movie clips were made by combining video recordings of natural gestures and an audio recording of syllable/a/in a male voice. The gestures were performed by a male in front of his torso in black clothes, with gray backgrounds (**Figure 1B**, the first row). At the beginning of the videos, two hands appeared apart at an intermediate distance approximately the same width as the shoulder. The still frame was presented for 1320 ms, followed by videos of three different conditions. The hands keep constant (CONST condition, with no movement) for the rest of the trial, or they moved horizontally toward each other (CLOSER condition) or away from each other (AWAY condition) for 600 ms. The motion was naturally smooth, and the moving distances in CLOSER and AWAY were the same.

The auditory stimuli were a 400-ms vowel/a/adjusted in different levels of intensity, delivered through Sennheiser HD 280 headphones. The sound was extracted from a recording

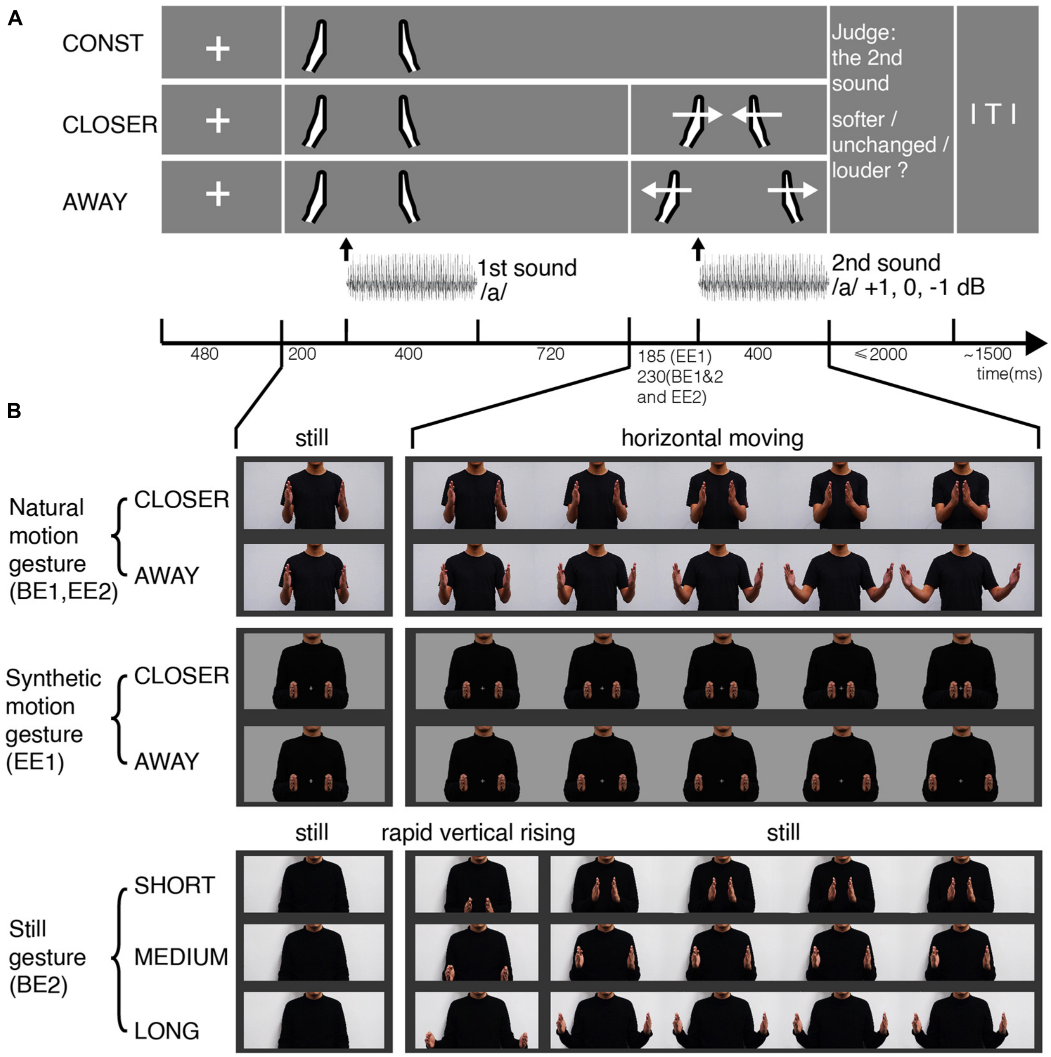

FIGURE 1 | Experimental procedures and stimuli for the four experiments. **(A)** Experimental procedures in BE1, EE1, and EE2 (not in BE2). Each row depicts one condition in which the gestures vary. Participants first saw the still image of two hands in the middle position relative to the body and heard the vowel/a/. Then they either saw the motional gestures (CLOSER, AWAY) or the same still image (CONST). Meanwhile, they heard the same vowel/a/that either remained unchanged or increased 1 dB or decreased 1 dB from the first-time presentation. Participants performed a loudness judgment task. **(B)** Frames of gesture video clips. A male torso was shown from waist to chin in the middle of the screen. The hands always started from the middle position as a still image, followed by moving horizontally either toward each other (CLOSER) or away (AWAY). In BE1 and EE2, we used gestures with natural motion (the first row), recorded using a video camera and edited in Adobe Premiere. In EE1, we used gestures with computer-synthetic motion (the second row), controlled by a set of displacement functions (three periods: constant acceleration, uniform motion, constant deceleration). In BE2, we used still images of gestures with three different distances (the last row). The gestures appeared from the initial image of the body with a rapid vertical rising in two video frames and remained still until the end of the second sound.

(sampling rate 44.1 kHz) of protracted pronunciation with steady intensity and then ramped in the first and last 20-ms duration. We only used one auditory token/a/because different tokens were hard to equalize to have the same loudness. Tokens with different perceived loudness would add confounding effects to the loudness-change judgment. We measured the sound intensity by using a sound meter (AWA5636, Aihua) with an acoustic coupler (ear simulator, AWA6160, Aihong) for circumaural headphones. We then calibrated the output sound intensity levels (see description below) by adjusting the sound wave files.

As shown in **Figure 1A**, at the beginning of each trial, a fixation cross was presented on the center of the screen for 480 ms, followed by the still image of two hands with an intermediate distance presented for 1320 ms. The first sound was played 200 ms after the onset of the still hands. The duration of the sound was 400 ms. Seven hundred twenty milliseconds after the sound offset, a video of gesture motion was presented in the CLOSER and AWAY conditions. In the CONST condition, the hands remained still. The stimulus onset asynchrony (SOA) between the onset of the gesture motion and the onset of the target sound (the second sound) was set at 230 ms in BE1. This sound latency was selected for three reasons. First, gestures usually precede articulations in natural conversations (Butterworth and Beattie, 1978; Morrel-Samuels and Krauss, 1992). Second, this amount of interval would be enough for the gesture to predictively modulate the auditory processing while still fall in the effective integration time window (van Wassenhove et al., 2007). Third, this interval would allow visual information to pre-modulate the auditory cortex or polymodal areas (Besle et al., 2008; Schroeder et al., 2008; Arnal et al., 2009; Talsma, 2015).

The intensity of the first sound was randomly selected from 55, 60, and 65 dB, while the intensity change of the second sound was randomly selected from −1, 0, and +1 dB relative to the first sound in that trial. We used the 1-dB step because the effect size of gestures' cross-modulation could be small. Intensity change of 1 dB is around JND of most people with normal hearing (Johnson et al., 1993). The participants were given a maximum of 2 s to judge whether the second sound was softer, the same, or louder than the first sound. The inter-trial intervals (ITIs) were 1500 ms with no jitter because the study focuses on the interaction between the gestures and perception of the second sound. Three gestures and three intensity changes in the second sound were fully crossed and yielded nine conditions. In this experiment, 648 trials with 72 trials for each condition were divided into six blocks. All conditions were evenly distributed across blocks and were randomly presented in each block. The experiment was programmed and presented by using a Python package, Psychopy.

Behavioral Experiment 2 (BE2): The Effects of Still Images of Gestures on the Judgment of Loudness Changes

In BE2, we replaced motional gestures with still images of gestures (**Figure 1B**, the last row). The initial image was the torso, followed by four different conditions. The hands may not appear (NO-GESTURE), or they appeared at different distances apart – SHORT (the final frame of the CLOSER clip), MEDIUM (same as CONST), and LONG (the final frame of the AWAY clip). We added two frames of transitional motion to make them appear naturally. The two hands quickly moved up vertically from outside the bottom edge of the frame to the height of the chest. The hands then remained still until the end of the second sound.

We used in BE2 similar trial procedures as in BE1. Three still gestures (SHORT, MEDIUM, and LONG) appeared at the same time point when the gestures in BE1 started to move. No gestures appeared in the control condition (NO-GESTURE). Four types of visual displays and three sound intensity changes yielded 12 conditions in BE2.

EEG Experiment 1 (EE1): Comparing the Early Auditory ERP Responses Between Conditions of Different Gestures

In EE1, to make sure that the two motional gestures would elicit a similar response, we synthesized the movement of gestures in Python (**Figure 1B**, the second row). First, we limited the movement ranges of the two hands within the torso boundary to avoid a sudden contrast change. Second, we used a set of displacement functions with three stages (constant acceleration, uniform motion, and constant deceleration) to make the synthesized motion as natural as possible. Thirdly, we increased the video frame rate to 80 fps for smoother motion. We also shrank the clip frame to 613 × 318 pixels. As a result, the torso was within a visual range of 3.4°, both lateral and vertical, from fixation. The maximum horizontal range of the gesture movements was 3.1° lateral from fixation.

The trial procedure of EE1 was the same as in BE1 (**Figure 1A**). More control conditions were included in EE1 to quantify the neural responses of modulation effects. Specifically, a total of 14 conditions (**Supplementary Table 1**) were divided into three categories: audiovisual (AV), auditory-only (A), and visual-only (V). In the AV category, six conditions were included: two motional gestures (CLOSER and AWAY) were fully crossed with three intensity changes (−1, 0, +1 dB relative to the first sound). In the A category, another three conditions were included: one still gesture (CONST) and one fixation-only (BLANK) were fully crossed with three intensity changes (−1, 0, +1 dB). In the V category, additional two conditions were included: the two motional gestures (CLOSER and AWAY) were presented without any sound. We included 48 trials for each condition in the AV and A categories and 72 trials for each condition in the V category. A total of 576 trials in AV and A conditions were mixed together and evenly divided into 12 sets. A total of 144 trials in V conditions were also divided into 12 sets. The whole experiment contained 12 blocks. Each block included two parts – the first part contained a mix of AV and A conditions, and the second part contained only V conditions. The stimuli in each part were randomly presented. The same gesture appeared in no more than two consecutive trials. After the main experiment, participants went through an intensity localizer block, in which they passively listened to a sequence of 140 1-kHz pure tones at an average interstimulus interval (ISI) of 1 s, jittered between 800 and 1200 ms. Tones with two levels of intensity (67 and 69 dB) were randomly presented, with each intensity level presented 70 times. The intensity localizer was aimed to check if the sound intensity level *per se* would induce different ERPs.

Visual stimuli were presented *via* a display screen of Dell S2417DG with a resolution of 1920 × 1080 and a refresh rate of 165 Hz. The graphic card was GeForce RTX 2060. We fixed the intensity of the first sound at 70 dB to simplify EEG experiments and increase power. The intensity change of the second sound was randomly selected from −1, 0, and +1 dB relative to the first sound in that trial. Sounds were delivered through plastic air tubes connected to foam earpieces (ER-3C Insert Earphones; Etymotic Research). The sound intensities were measured using a sound-level meter (AWA5636, Aihua) with the acoustic coupler for insert earphones (occluded ear simulator, AWA6162, Aihong). Further, we adjusted the SOA between the onset of gesture motion and the onset of the second sound to 185 ms because the audiovisual integration has a high probability of occurring in a time window of 0–200 ms and is likely skewed toward the later part of this window (van Wassenhove et al., 2007). The pure tones (sampling rate of 44.1 kHz; duration of 400 ms) in the intensity localizer were generated in Praat (Boersma and Weenink, 2021).

To control the timing of visual and auditory stimuli precisely, we recorded the onset timing of both the visual and auditory stimuli *via* StimTracker Duo (Cedrus) system (the trigger box). A light sensor was attached to the lower-left corner of the monitor and connected to the trigger box. The acoustic signals were split into the trigger box (another went to the earphones). The onset time of each physical stimulus was captured with a sampling frequency of 1 kHz, which provided a set of temporal markers to the physical stimuli measured in the timeline of EEG recordings. The actually measured distribution of SOAs in EE1 had a mean of 185.5 ms and an SD of 10.4 ms. We did not align the stimuli to the refresh rate of the screen. However, the refresh rate of the screen was more than double the corresponding video frame rate in both EEG experiments.

EEG Experiment 2 (EE2): Comparing the Early Auditory ERP Response Between Trials of Different Loudness Judgment to the Same Sound

In EE2, we examined the modulation effects of gestures by quantifying the neural responses in trials with different perceptual judgments to the same stimuli. We used the same gestures with natural motion (**Figure 1B**, the first row) as in BE1 because there was no need to control the visual responses to CLOSER and AWAY gestures in this experiment. We only compared between conditions within either gesture. Moreover, we used the same SOA (230 ms) between the onset of motional gesture and the onset of the second sound as in BE1. The reason to increase the SOA was to further separate the ERP responses to the second auditory stimulus from those driven by the motion gestures so that our questions can be better addressed. The neural responses take time to accumulate so that EEG signals can be recorded. For example, the early perceptual components in visual and auditory domains can take about 200 ms – the classic N1/P2 components. The actually measured distribution of SOAs in EE2 had a mean of 229.9 ms and an SD of 8.9 ms. Similar trial procedures as in BE1 were used with several modifications to yield enough trials of biased responses to the same auditory stimuli. First, we excluded the CONST conditions. Second, we fixed the intensity of the first sound at 68 dB. Third, we adjusted the proportions of intensity changes (−1, 0, and +1 dB) to a ratio of 1:5:1. Reasonable percentages of −1 and +1 dB intensity changes were included to convince the participants that the intensity did vary and to avoid any strategies. A large portion of trials was intensity unchanged (0 dB) so that enough trials would be obtained in situations of different loudness judgment to the same intensity. In total, six conditions were included in EE2 (2 motional gestures × 3 levels of intensity change). A total of 672 trials were divided into 12 blocks. The presentation order was randomized in each block. EE2 was carried out using a display screen of Dell E2214Hv with a resolution of 1920 × 1080 and a refresh rate of 60 Hz. The graphic card was AMD Radeon HD 5450. The video stimuli were presented with 25 fps.

Procedure

General Experimental Procedures for BE1, BE2, EE1, and EE2

BE1 and BE2 were carried out in a small room with participants sitting on a comfortable chair. Before each experiment, participants were given instructions on how to attend to the stimuli properly. They were asked to watch the hands, to pay attention to the sounds, and to make judgments based on the auditory stimuli. Importantly, they were explicitly told that the gestures and sound intensity changes were randomly paired. Participants went through a brief training to familiarize the changes in sound intensity. During training, they judged the loudness change of the second sound without the presence of the visual stimulus and with real-time feedback. After they passed the training, they went through a practice block to familiarize themselves with all the stimuli and tasks. We verbally confirmed that they could see the gestural motion easily and that they could hear the intensity changes in the practice block. During the experiments, participants were required to take a break for at least 1 min between two blocks.

EE1 and EE2 share the same procedure with BE1 and BE2 except for a few aspects. The two EEG experiments were carried out in an electromagnetically shielded and soundproof booth. The location of the chair was fixed to control the retinal angle of the visual stimuli. We asked the participants to fixate at the tiny cross displayed at the center between two hands, sit still, and avoid unnecessary head movement and eye blink during the trial.

EEG Data Acquisition

EEG signals were recorded with a 32-electrode active electrodes system (actiChamp system, Brain Products GmbH, Germany). Electrodes were placed on EasyCap, on which electrode holders were arranged according to the 10–20 international electrode system. Two additional electrooculogram (EOG) electrodes were used to monitor horizontal and vertical ocular movements, respectively. The ground electrode was placed at the forehead. Electrode impedances were kept below 10 kΩ. The data were continuously recorded in single DC mode, sampled at 1000 Hz and referenced online to the electrode Cz. The EEG data were acquired with Brain Vision PyCoder software and filtered online by the acquisition system using a low-pass filter (second order Butterworth) with a cutoff frequency of 200 Hz. A 50-Hz notch filter was applied to filter out AC noise online during EEG recordings. EEG data processing and analysis were conducted with customized Python codes, MNE-python (Gramfort et al., 2014), EasyEEG (Yang et al., 2018).

Data Analysis

Behavioral Data Analysis of BE1, BE2, EE1, and EE2

For the behavioral data in each experiment, we calculated a judgment score in each condition to characterize a participant's judgment preference. The score was obtained by averaging all the judgments (1 for choosing louder, 0 for unchanged, and −1 for softer) across trials. We also calculated in each condition the participants' accuracy – the ratio of the number of correct trials to the total number of trials. We applied repeated measure two-way analyses of variance (ANOVA) to the judgment scores and accuracy, respectively, with the factors of intensity change and gesture, followed by *post hoc* pairwise comparisons using t-tests with Bonferroni correction. We checked the normality of ANOVA residuals by visual inspection of the Q–Q plot and Shapiro–Wilk test. The residuals were approximately normal. We checked the sphericity assumption using Mauchly's Test of Sphericity. We applied the Greenhouse–Geisser correction when the sphericity assumption was violated. In EE1, we also calculated the accuracy of behavioral judgment for the BLANK condition in each intensity change. The accuracy without any gestural influence in the training session of each experiment and in the EE1 BLANK condition was compared with the 0.33 chance level by using a one-sample one-tailed t-test.

Furthermore, we calculated the bias ratios in each of the three intensity changes to index how the manual gestures influence the loudness judgment to different intensity changes. It is a summary statistic based on the confusion matrix (the percentage of choice responses with respect to the total trial in that condition) shown in **Supplementary Table 2**. For ±1 dB intensity change, there were two kinds of judgment biases. The response could be off the actual intensity change by 1 level (level-1 bias), such as responding "unchanged" when the second sound increased or decreased by 1 dB. The bias could also be off by 2 levels (level 2 bias), such as responding "louder" when the intensity change was −1 dB and vice versa. For 0 dB (no intensity change), only level-1 bias could be induced. For −1 and 0 dB intensity change, the judgment bias of louder percepts was obtained for each gesture: bias ratio = frequency of louder bias/frequency of all bias. For +1 dB intensity change, the judgment bias of softer percept was calculated for each gesture: bias ratio = frequency of softer bias/frequency of all bias. We applied planned paired t-tests to the bias ratios between different gestures in each level of intensity change.

EEG Data Analysis

For each participant, data were band-pass filtered (0.1–30 Hz, Kaiser windowed FIR filter) offline and re-referenced to the average potential of all the EEG electrodes. For EE1, epochs were extracted according to the conditions (AV conditions, −310 to 400 ms, time-locked to the onset of the second sound; A conditions, −100 to 400 ms, time-locked to the onset of the second sound; V conditions, −100 to 600 ms, time-locked to the onset of the motional gesture; intensity localizer, −100 to 400 ms, time-locked to the onset of the tones). All epochs were baseline-corrected using the 100-ms pre-stimulus data, except for the AV conditions, which were baseline-corrected using the 100-ms pre-motion data (−310 to −210 ms). For EE2, epochs were extracted from −350 to 400 ms time-locked to the onset of the second sound. All epochs were baseline-corrected using the pre-motion data from −350 to −250 ms.

To ensure data quality, epochs with peak-to-peak amplitude exceeding 100 μV were automatically excluded, and epochs with artifacts that resulted from eye blinks and other muscle movements were manually rejected. We identified eye blinks by visual inspection of the two EOG channels. We identified and removed muscle artifacts also by visual inspection. The remaining epochs were used to obtain the ERP in each condition. An average of 30.2 trials (SD = 7.2, out of 48 trials per condition) of AV conditions and an average of 45.9 trials (SD = 10.7, out of 72 trials per condition) of V conditions were included in EE1. An average of 52.6 trials (SD = 19.3) were included in EE2 (the total number of trials in each condition varied depending on loudness judgment). Two participants were excluded from EE1, and three participants were excluded from EE2 because they either produced many artifacts (more than 50%) or made almost identical behavior responses in all conditions (probably not following instructions nor paying attention to the task). Their behavioral data were also excluded from the analysis.

For EE1, the EEG epochs were averaged and created an ERP response in each condition. Instead of selecting sensors, we calculated a more conservative index, the global field power (Murray et al., 2008). GFP, calculated as the root mean square of data in all sensors, represents the amount of energy change in all sensors throughout the time. GFP provides more holistic and unbiased information (Murray et al., 2008). We applied a temporal cluster analysis (Maris and Oostenveld, 2007) to compare the GFP waveforms of two conditions. Specifically, we first calculated a paired t-statistics between the two conditions at each time point. Then, temporal clusters were formed with more than two adjacent time points where the corresponding p-value was above the threshold (0.05). We summed all the t-values within each temporal cluster as its summary empirical statistics. To form a distribution of the null hypothesis, we permutated the condition labels 10,000 times and collected the maximum cluster sum-t value in each permutation. Finally, the summary empirical statistics of each temporal cluster identified in the original data were tested in the permutation distribution of max-t values. The same temporal cluster analysis was separately applied in the AV conditions, V conditions, A conditions, and the intensity localizers in the absence of visual modulation.

For EE2, to examine how neural responses were modulated as a function of perception to the same auditory stimuli, only the data in the conditions of 0 dB (no intensity change) were used in EEG analysis. Data were divided into three groups based on participants' judgment in either gesture level (CLOSER and AWAY). The three groups were (1) trials of "softer" perceptive shift (choosing softer), (2) trials of "louder" perceptive shift (choosing louder), and (3) trials of no perceptive shift (choosing unchanged). We applied the ERP component analysis and temporal cluster analysis to the comparison between "softer" perceptive shift and no perceptive shift in the CLOSER condition, as well as to the comparison between "louder" perceptive shift and no perceptive shift in the AWAY condition. The exact N100 peak latency varied in individual participants. Therefore, in the ERP component analysis, the N100 was automatically located with an in-house algorithm (Wang et al., 2019) for each participant in a pre-determined time range (65–135 ms). We took an average of the amplitudes in a 20-ms window centered at the individual N100 peak as the N100 response magnitude. For all paired comparisons, the numbers of epochs in the pair of conditions were equalized with the function of "equalize_epoch_counts" included in the MNE toolbox. Basically, this function equalizes the number of trials in two conditions by selecting trials in the conditions that have more trials. The criterion of selection is that the selected trials would occur as close as possible in time to the trials in the other condition.

RESULTS

The analysis of the training data (**Supplementary Figure 1**) showed that participants were able to discriminate the intensity changes above the chance level (0.33) without gestural influence [BE1, $M = 0.47$, SD = 0.12, $t(14) = 4.34$, $p < 0.001$, $d_z = 1.09$; BE2, $M = 0.41$, SD = 0.09, $t(11) = 2.97$, $p = 0.006$, $d_z = 0.86$; EE1, $M = 0.47$, SD = 0.14, $t(20) = 4.21$, $p < 0.001$, $d_z = 0.92$; EE2, $M = 0.43$, SD = 0.09, $t(16) = 3.95$, $p < 0.001$, $d_z = 0.96$].

Behavioral Experiment 1 (BE1): The Effects of Motional Gestures on the Judgment of Loudness Changes

Response accuracy was differentially influenced by gestures across intensity changes (**Figure 2B**). The average accuracy was around 0.5 where the gesture direction and intensity change matched, and the accuracy was lower than that when the gesture direction and intensity change did not match. ANOVA revealed that the main effect of intensity change on accuracy was not significant [$F(1.16,28) = 3.98$, $p = 0.058$, $\eta_p^2 = 0.22$, $\varepsilon = 0.58$], and the main effect of gestures was not significant [$F(2,28) = 0.64$, $p = 0.535$, $\eta_p^2 = 0.04$, $\varepsilon = 0.76$] either. Crucially, there was an interaction between gesture and intensity change [$F(1.48,56) = 14.62$, $p < 0.001$, $\eta_p^2 = 0.41$, $\varepsilon = 0.31$]. Pairwise t-tests revealed that the accuracies of CONST ($M = 0.27$, SD = 0.14) and AWAY ($M = 0.23$, SD = 0.15) were lower than the accuracy of CLOSER ($M = 0.47$, SD = 0.20) in −1 dB intensity change [$t(14) = 4.20$, $p < 0.01$, $d_z = 1.21$; $t(14) = 3.72$, $p = 0.02$, $d_z = 1.40$]. The accuracies of CLOSER ($M = 0.42$, SD = 0.18) and AWAY ($M = 0.40$, SD = 0.19) were lower than the accuracy of CONST ($M = 0.61$, SD = 0.17) in 0 dB intensity change [$t(14) = 4.51$, $p = 0.004$, $d_z = 1.11$; $t(14) = 4.05$, $p = 0.01$, $d_z = 1.20$]. In addition, the accuracies of CLOSER ($M = 0.33$, SD = 0.17) and CONST ($M = 0.33$, SD = 0.14) were lower than the accuracy of AWAY ($M = 0.56$, SD = 0.16) in +1 dB intensity change [$t(14) = 3.54$, $p = 0.03$, $d_z = 1.46$; $t(14) = 4.05$, $p = 0.01$, $d_z = 1.57$]. That is, the highest accuracy was found where the gestural direction matched with the intensity change (CLOSER with intensity −1 dB, AWAY with intensity +1 dB, CONST with intensity unchanged). On the contrary, accuracy was much lower where the gestural direction and the intensity change did not match.

Both gesture and intensity change positively affected the judgment scores (**Figure 2A**). On average, the judgment score monotonically increased as a function of the intensity change (−1, 0, +1 dB) or the gesture (CLOSER, CONST, AWAY). ANOVA showed that the main effect of intensity change was significant [$F(1.16,28) = 50.61$, $p < 0.0001$, $\eta_p^2 = 0.78$, $\varepsilon = 0.58$]. More importantly, the main effect of gesture was also significant [$F(1.09,28) = 13.92$, $p = 0.002$, $\eta_p^2 = 0.450$, $\varepsilon = 0.54$]. However, the interaction was not significant [$F(4,56) = 1.19$, $p = 0.30$, $\eta_p^2 = 0.08$, $\varepsilon = 0.19$]. The pairwise t-tests revealed that the judgment scores under AWAY ($M = 0.31$, SD = 0.23) were higher than the judgment scores under CONST ($M = 0.04$, SD = 0.08) [$t(44) = 6.41$, $p < 0.0001$, $d_z = 1.13$]. The judgment scores under CONST ($M = 0.04$, SD = 0.08) was higher than the judgment scores under CLOSER ($M = –0.15$, SD = 0.26) [$t(44) = 5.46$, $p < 0.0001$, $d_z = 0.72$]. These results suggest that: (1) participants were able to detect intensity changes (not by pure guessing); (2) The moving directions of the gestures were in line with the bias they caused in intensity judgment.

The bias ratio characterized the direction and extent to which gestures biased the judgments (**Figure 2C**). When the intensity change was −1 dB, AWAY ($M = 0.47$, SD = 0.21) induced higher ratio of level-2 bias (judge louder) than CONST ($M = 0.18$, SD = 0.13) [$t(14) = 4.94$, $p < 0.001$, $d_z = 1.67$] and CLOSER ($M = 0.23$, SD = 0.13) [$t(14) = 3.83$, $p = 0.006$, $d_z = 1.38$]. When the intensity change was +1 dB, CLOSER ($M = 0.38$, SD = 0.22) also induced higher ratio of level-2 bias (judge softer) than CONST ($M = 0.17$, SD = 0.12) [$t(14) = 4.71$, $p = 0.001$, $d_z = 1.22$] and AWAY ($M = 0.26$, SD = 0.21) [$t(14) = 2.94$, $p = 0.03$, $d_z = 0.56$]. When the intensity change was 0 dB, AWAY ($M = 0.78$, SD = 0.16) produced higher ratio of louder bias than CONST ($M = 0.57$, SD = 0.12) [$t(14) = 4.55$, $p \leq 0.001$, $d_z = 1.49$] and CONST produced higher ratio of louder bias than CLOSER ($M = 0.41$, SD = 0.22) [$t(14) = 3.25$, $p = 0.097$]. The analyses using bias ratios further suggested that the gestures biased the judgments of loudness changes when they were inconsistent. BE1 results provided overall behavioral evidence supporting the hypothesis that gestures influence loudness perception.

Behavioral Experiment 2 (BE2): Still Gesture Images Modulated Judgment of Loudness Less Than Motional Gestures

The goal of BE2 was to examine whether the influence of gestures on loudness judgment was only due to the distance between hands in the gestures. Therefore, in BE2, we replaced motional gestures used in BE1 with still images of gestures. The judgment score increases as the intensity change goes from −1 dB to 0 and to +1 dB but is only moderately influenced by gestures

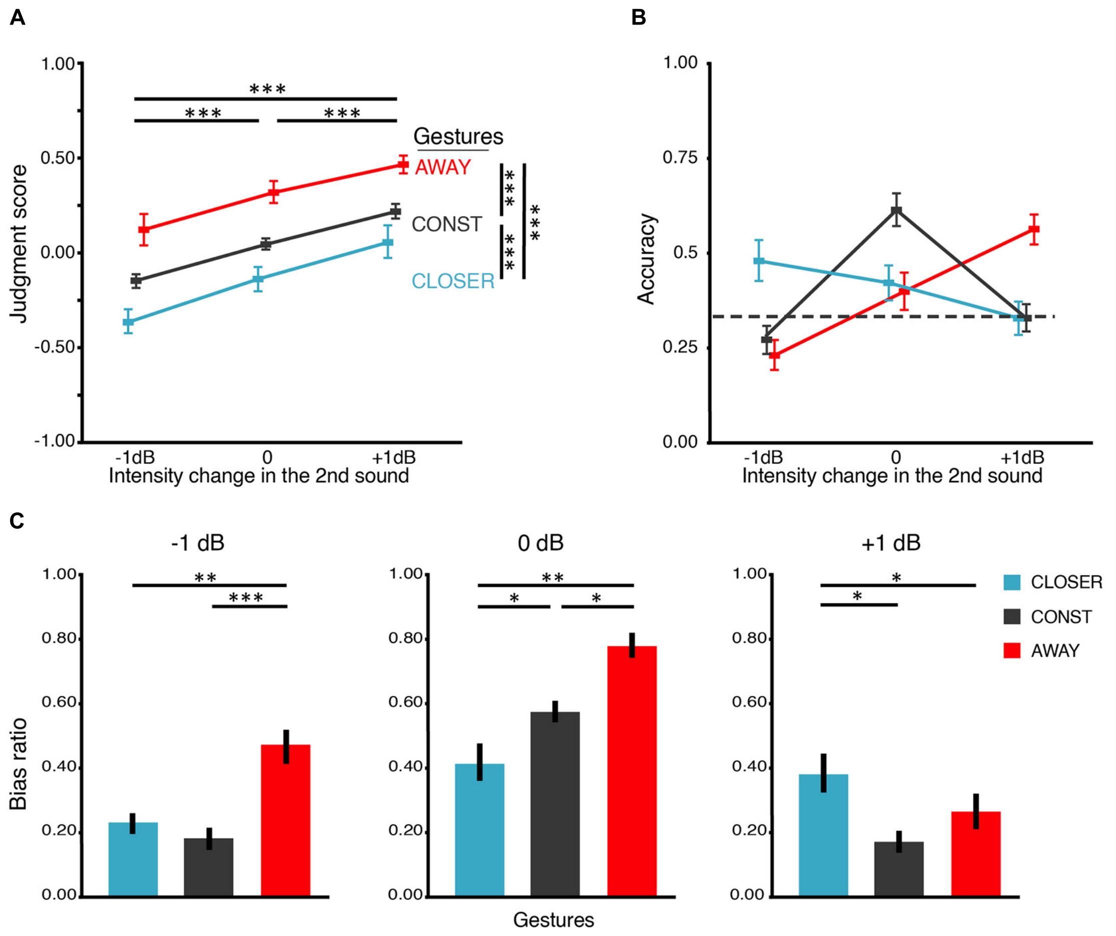

FIGURE 2 | Results of BE1. **(A)** Judgment score. BE1 investigated how motional gestures influenced loudness perception. The judgment score was obtained by averaging all the responses in which "1" for "louder," "0" for unchanged, and "–1" for softer. Therefore, the judgment score reflects the overall judgment tendency, where 0 stands for no change, positive for louder and negative for softer. Participants can correctly identify the intensity changes – the judgment scores increased as the intensity increased in all gesture conditions. Moreover, gesture modulated the loudness judgment – in all levels of intensity change. Judgment scores in the AWAY gesture condition were larger than those in CONST, and judgment scores in CONST were larger than those in CLOSER. **(B)** Accuracy of the behavioral judgments about intensity change. We obtained accuracy by calculating the portion of trials that participants correctly identified the intensity change. There is an interaction between the factors of gesture and intensity change. The interaction was driven by higher accuracies in conditions where the changes in intensity and gestures were consistent. The dashed line indicates the chance level (0.33). **(C)** Bias ratios. The bias ratio was calculated to index how the manual gestures influence the loudness judgment to different intensity changes. For –1 dB intensity change (left panel) and 0 dB intensity change (middle panel), the judgment bias of louder percepts was obtained for each gesture: bias ratio = frequency of louder bias/frequency of all bias. For +1 dB intensity change (right panel), the judgment bias of softer percept was calculated for each gesture: bias ratio = frequency of softer bias/frequency of all bias. The results indicate that the judgment was biased toward the "matched" gesture in all intensity changes. All error bars indicate ±one SEM. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$.

(**Figure 3A**). The statistical results suggest that both gesture and intensity change positively affected the judgment score. ANOVA showed that the main effects of both intensity change and gesture were significant [$F(1.28,22) = 114.20$, $p < 0.0001$, $\eta_p^2 = 0.91$, $\varepsilon = 0.64$; $F(1.34,33) = 8.35$, $p = 0.007$, $\eta_p^2 = 0.43$, $\varepsilon = 0.45$]. The interaction was not significant [$F(6,66) = 3.34$, $p = 0.071$, $\eta_p^2 = 0.23$, $\varepsilon = 0.26$]. For the judgment scores, pairwise t-tests (Bonferroni corrected) showed that LONG ($M = 0.15$, SD = 0.15) was higher than MEDIUM ($M = 0.04$, SD = 0.12) [$t(35) = 3.65$, $p < 0.005$, $d_z = 0.33$]; MEDIUM was higher than SHORT ($M = -0.09$, SD = 0.20) [$t(35) = 3.99$, $p = 0.002$, $d_z = 0.38$]; LONG was higher than NO-GESTURE ($M = -0.01$, SD = 0.10) [$t(35) = 5.61$, $p < 0.0001$, $d_z = 0.48$]. However, the judgment scores of SHORT and MEDIUM were not significantly different from NO-GESTURE [$t(35) = 2.20$, $p = 0.21$, $d_z = 0.23$; $t(35) = 2.67$, $p = 0.07$, $d_z = 0.17$]. Moreover, the judgment of loudness change was not modulated by the gestures in −1-dB intensity change: the judgment scores were not different between any pair of gestures. Finally, the differences of judgment scores across the three still gestural conditions (SHORT, MEDIUM, LONG) in BE2 (**Figure 3A**) were less than the differences of judgment scores across the three motion gestural conditions (CLOSER, CONST, AWAY) in BE1 (**Figure 2A**).

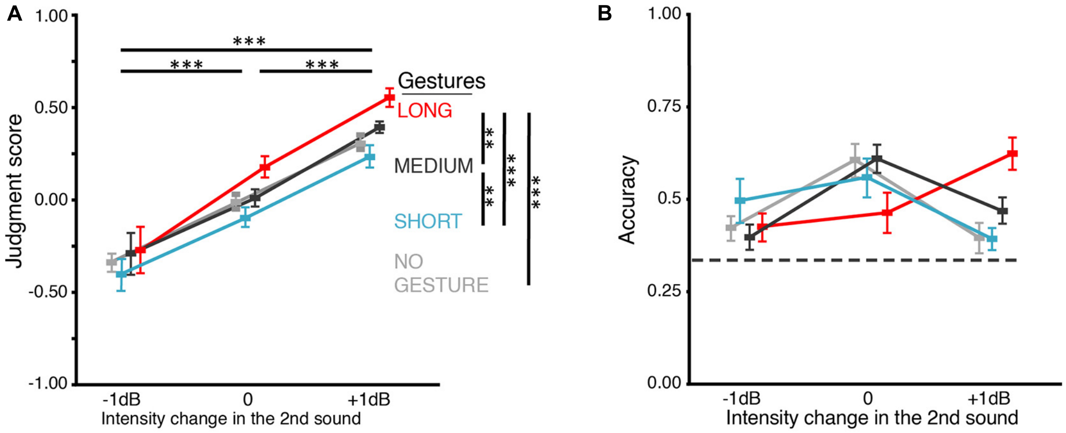

FIGURE 3 | Results of BE2. **(A)** Judgment score. BE2 investigated how the gestures in still images influenced loudness perception. Because for each of the three gestures, no gesture was shown before the hands showed up, we included a NO-GESTURE as the baseline condition. The loudness judgment was positively influenced by both the intensity change and still gesture. However, the difference was neither significant between MEDIUM and NO-GESTURE nor between SHORT and NO-GESTURE. The influence of still gestures on loudness judgment was smaller than motional gestures in BE1 (**Figure 2A**). In addition, the judgment scores were not different between any pair of gestures in –1-dB intensity change. **(B)** Accuracy of the behavioral judgments about intensity change. When intensity changes were –1 or 0 dB, no gesture differed in its influence on judgment accuracy, though the overall interaction between the factors of gesture and intensity change was significant. The dashed line indicates the chance level (0.33). All error bars indicate ±one SEM. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

Response accuracy further showed a difference in BE2, as compared to that in BE1. In BE2 (**Figure 3B**), although ANOVA still showed that gesture interacted with intensity change [$F(2.69,66) = 9.05$, $p < 0.0001$, $\eta_p^2 = 0.45$, $\varepsilon = 0.30$], the pairwise t-tests failed to show any difference between gestures in -1 and 0 dB intensity changes. When the intensity change was $+1$ dB, only LONG ($M = 0.62$, SD = 0.15) showed higher accuracy than NO-GESTURE ($M = 0.39$, SD = 0.15) [$t(11) = 7.9$, $p = 0.0001$, $d_z = 1.5$] and SHORT ($M = 0.39$, SD = 0.10) [$t(11) = 5.47$, $p = 0.004$, $d_z = 1.82$]. These results suggested that the visual-motor information in the motional gestures contributed more greatly to the modulation effects on loudness judgment than the final distance between two hands. We used motional gestures to investigate further the dynamics of the modulation effects in the following EEG experiments.

EEG Experiment 1 (EE1): Gestures Modulated Early Neural Responses in Loudness Perception

Behavioral Results

The behavioral results in EE1 replicated those in BE1. In EE1, both gesture and intensity change positively affected the judgment scores (**Figure 4A**). ANOVA showed that the main effects of both intensity change and gesture were significant [$F(1.23,40) = 139.55$, $p < 0.0001$, $\eta_p^2 = 88$, $\varepsilon = 0.62$; $F(2,40) = 44.26$, $p < 0.0001$, $\eta_p^2 = 0.69$, $\varepsilon = 0.83$]. The interaction was also significant [$F(4,80) = 12.82$, $p < 0.001$, $\eta_p^2 = 0.39$, $\varepsilon = 0.36$]. Pairwise t-tests (Bonferroni corrected) showed that the judgment scores under AWAY ($M = 0.30$, SD = 0.18) was higher than the judgment scores under CONST ($M = -0.04$, SD = 0.11) [$t(62) = 9.88$, $p < 0.0001$, $d_z = 0.76$] and CONST was higher than CLOSER ($M = -0.12$, SD = 0.18) [$t(62) = 2.53$, $p = 0.04$, $d_z = 0.18$]. Second, gesture interacted with intensity change in terms of their effects on accuracy [$F(4,84) = 54.94$, $p < 0.0001$, $\eta_p^2 = 0.74$, $\varepsilon = 0.43$] (**Figure 4B**).

The response accuracy for the BLANK condition (where only the fixation was shown, **Supplementary Figure 2**) was above the 0.33 chance level in each intensity change [-1 dB, $M = 0.51$, SD = 0.16, $t(20) = 5.14$, $p < 0.0001$, $d_z = 1.12$; 0 dB, $M = 0.69$, SD = 0.14, $t(20) = 11.97$, $p < 0.0001$, $d_z = 2.61$; $+1$ dB, $M = 0.45$, SD = 0.19, $t(20) = 2.86$, $p = 0.005$, $d_z = 0.62$]. These results confirmed that participants were able to discriminate all three intensity changes when no gesture was shown.

The pattern of the bias ratio (**Figure 4C**) in EE1 was similar to that in BE1. The judgment was biased toward the "matched" gesture in all intensity changes. CLOSER and AWAY biased the choice toward opposite directions in 0-dB intensity change: the louder bias ratio of AWAY ($M = -0.83$, SD = 0.13) was larger than the louder bias ratio of CLOSER ($M = 0.40$, SD = 0.22) [$t(20) = 8.97$, $p < 0.0001$, $d_z = 2.40$]. These results suggest that (1) participants were able to detect the real changes of intensity (not by pure guessing); (2) The influence of gestures on the judgments of loudness change positively correlated to the direction of movement in gestures, and (3) The gestures biased the judgments of loudness changes when the direction of change was not congruent across modalities.

EEG Results

If the gestures CLOSER and AWAY modulated loudness perception rather than decisional processes, the modulation effects should be observed in ERPs at early latencies (e.g., ~100 ms) rather than at late latencies. For the 0 dB intensity change, the GFP waveforms in both gesture conditions rose following the gesture motion onset and increased again about 50 ms after the second sound's onset (**Figure 4D**). An apparent

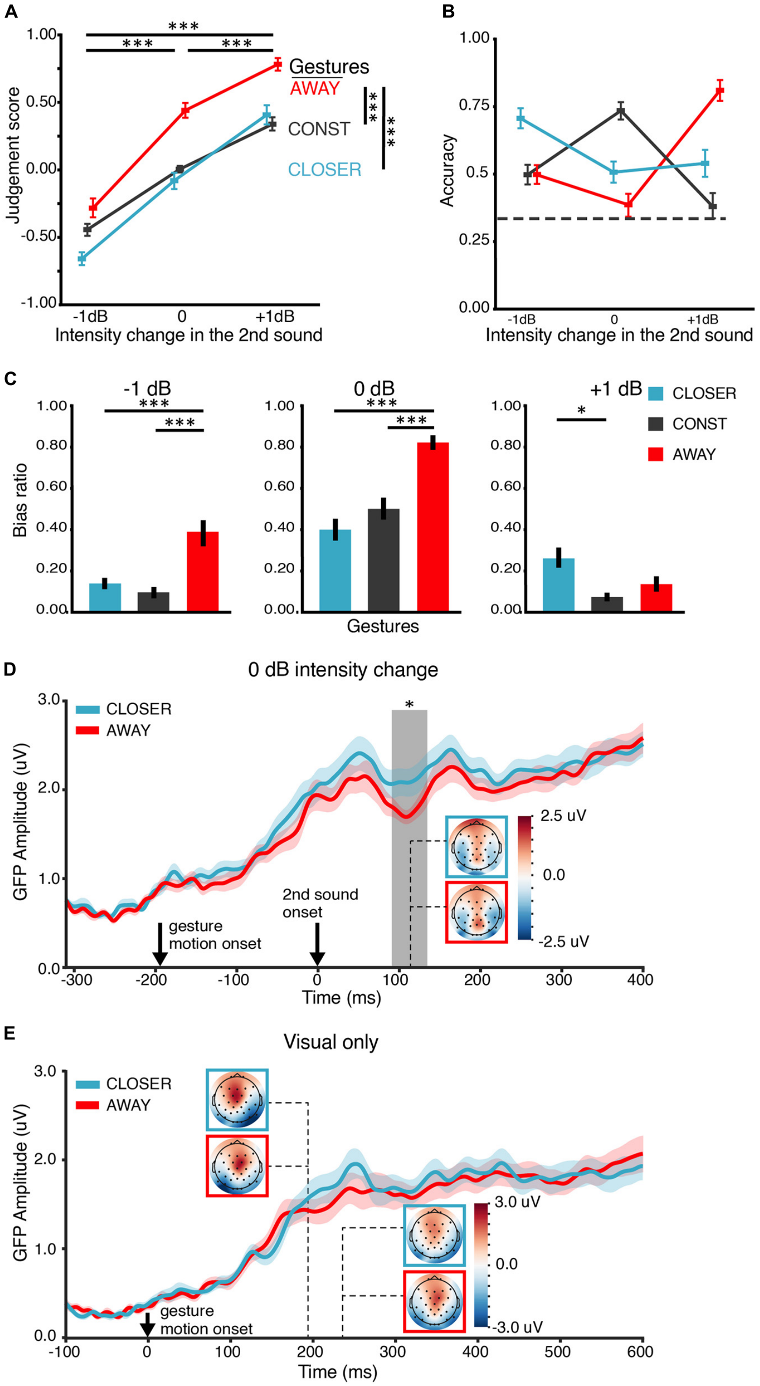

FIGURE 4 | Results of EE1. **(A)** Judgment score. Behaviorally, both gesture and intensity change positively affected the judgment of loudness changes. The main effects were similar to the results in BE1 in **Figure 2A**. **(B)** Accuracy of the behavioral judgment about intensity change. An interaction between the factors of gesture and intensity change was observed, consistent with the main results in BE1 in **Figure 2B**. The dashed line indicates the chance level (0.33). **(C)** Bias ratios. The judgment was biased toward the "matched" gesture 0-dB-intensity-change conditions. **(D)** ERP responses to the second sound in the 0-dB-intensity-change conditions were modulated by the gestures. Gesture CLOSER elicited a stronger ERP response than gesture AWAY at around 110 ms after the onset of the second sound. Solid lines in each plot indicate the grand mean global field power (GFP) waveform. The shades around the lines represent +, – one SEM (*n* = 21). Response topographies are shown in colored boxes with dashed lines pointing to their latencies. The colored boxes use the same color schemes as the waveform responses to indicate different conditions. The gray vertical rectangular shade indicates the temporal cluster (91–135 ms) in which the two GFPs were significantly different in the temporal cluster analysis. **(E)** ERP responses to video stimuli in the visual-only (V) conditions. The visual ERP responses to gesture CLOSER and AWAY were not significantly different, suggesting that different visual stimuli evoked similar early visual responses and the observed effects in panel **(D)** were not caused by different visual gesture stimuli. The depicting formats are the same as in panel **(D)**. The error bars indicate ±one SEM. **p* < 0.05; ***p* < 0.01; ****p* < 0.001.

diverge of the two waveforms was observed around 100-ms latency. CLOSER evoked a larger response than AWAY in a temporal cluster from 91 to 135 ms (p = 0.039). The effects of gestures on early auditory responses were not caused by visual responses because the two gestures elicited similar GFPs across the time in visual-only (V) conditions (**Figure 4E**). No significant cluster was found in the temporal cluster analysis. For the −1 and +1-dB intensity change, respectively, no significant difference was found in the GFP waveforms of the two gesture conditions.

It was somewhat surprising that CLOSER evoked larger auditory responses than AWAY for the 0-dB intensity change. This modulation pattern could arise from the interaction between gestures and a particular neural response profile to physical stimuli in the current experimental setting. Therefore, we further investigated the neural response profile to auditory stimuli with different levels of intensity without motional gestures. First, we examined the ERP responses in auditory-only (A) conditions in which different intensity changes were presented with the still image of the CONST gesture. No difference was found in the temporal cluster analysis (**Supplementary Figure 3A**). Also, we did not find any difference between the low-intensity and high-intensity conditions in the intensity localizer (**Supplementary Figure 3B**). These results suggest that the ERP difference we found in the AV conditions was specific to gestural modulation.

In summary, the behavioral results in EE1 were similar to those in BE1 and supported that audiovisual gesture information biased the judgments of loudness changes. More importantly, the modulation effects were observed in auditory neural responses at an early latency (around 110 ms after the second sound onset). The modulation pattern in the 0-dB-intensity-change conditions was somewhat surprising and specific to the gestures. Therefore, to replicate the results of EE1 and to provide further evidence about across-modal effects on loudness perception, we carried out EE2 in which the modulation effects of gesture were examined as a function of loudness perception to the same physical stimuli.

EEG Experiment 2 (EE2): Changes of Loudness Perception Were Reflected by the Modulation in Early Auditory Responses

Behavioral Results

The behavioral response in BE1, EE1, and EE2 followed the same pattern (**Supplementary Figure 4**). The trends of the judgment score and accuracy in EE2 (**Figures 5A,B**) were similar to those in BE1. We applied the same statistical analyses used in BE1 to the behavioral data of EE2. The behavioral results were similar to those in BE1 and EE1, although we removed the still gesture CONST and increased the number of 0-dB-intensity-change trials. ANOVA showed that the main effects of both intensity change and gesture on the judgment scores were significant [$F(1.04,32) = 21.82$, $p < 0.001$, $\eta_p^2 = 0.58$, $\varepsilon = 0.52$; $F(1,16) = 21.93$, $p < 0.001$, $\eta_p^2 = 0.58$, $\varepsilon = 1.00$] (**Figure 5A**). However, the interaction was not significant [$F(2,32) = 1.21$, $p = 0.31$, $\eta_p^2 = 0.07$, $\varepsilon = 0.98$]. These results suggested that participants could detect the actual intensity change, and their judgments of loudness changes positively correlated with the direction of movement in gestures. Moreover, gesture interacted with intensity change in terms of their effects on accuracy [$F(1.30,32) = 25.69$, $p < 0.0001$, $\eta_p^2 = 0.62$, $\varepsilon = 0.65$] (**Figure 5B**). Especially, the louder bias ratio of CLOSER (M = 0.26, SD = 0.18) was significantly lower than the louder bias ratio of AWAY (M = 0.75, SD = 18) in 0 dB intensity change [$t(16) = 3.88$, $p = 0.004$, $d_z = 0.61$] (**Figure 5C**). These results indicated that when the second sound was identical to the first sound, gesture CLOSER drove participants toward "softer" bias, whereas gesture AWAY drove participants toward "louder" bias.

EEG Results

We designed the EE2 to further investigate the relation between neural modulations and loudness perception changes caused by gestures. Specifically, we examined how gestures changed the auditory neural responses as a function of subjective biases in loudness perception to the same physical stimuli. Based on what we found in EE1, we expected that "softer" bias (induced by gesture CLOSER) would have stronger ERP responses at early latency (N100) than no bias. Indeed, the ERP time course of "softer" perceptive shifts had larger responses than no perceptive shifts shortly after 100-ms latency (**Figure 5D**). A cluster from 104 to 127 ms (p = 0.047) was found by the temporal cluster analysis. This was consistent with results in the paired t-test on N100 component response magnitude [$t(16) = 2.17$, $p = 0.045$, $M_{\text{diff}} = -0.34\ \mu\text{V}$, $d_z = 0.53$] (**Figure 5E**). Note that these two conditions ("softer" perceptive shifts and no perceptive shifts) were identical in all physical aspects. The only difference between them was in subjective judgment. These results suggested that the bias in loudness perception induced by gesture CLOSER was accompanied by an early perceptive modulation at around 100 ms after the onset of the second sound. However, we did not observe any significant differences between different loudness percepts in gesture AWAY: the ERPs of "louder" perceptive shifts and no perceptive shifts did not differ (**Figure 5F**). The response magnitudes of N100 component were not significantly different either [$t(16) = 0.17$, $p = 0.87$, $d_z = 0.04$] (**Figure 5G**).

In general, these results, together with EE1, supported that the perceived loudness changes induced by the gesture CLOSER were consistently reflected in neurological measures as increases in early auditory ERP responses.

DISCUSSION

Our results from two behavioral experiments and two EEG experiments consistently demonstrate that visual-motor information in gestures can modulate the perception of a low-level auditory perceptual attribute such as loudness at the JND threshold. In BE1 and BE2, we found that gestures affected the judgment of loudness in accordance with their moving directions. In addition to the final position of hands, the visual-motor information exhibited extra influence on the loudness perception. The behavioral results in two EEG experiments replicated BE1. More importantly, in EE1, we found that the early neural responses to the sound stimuli were differently modulated by two gestures. In EE2, we found that biased

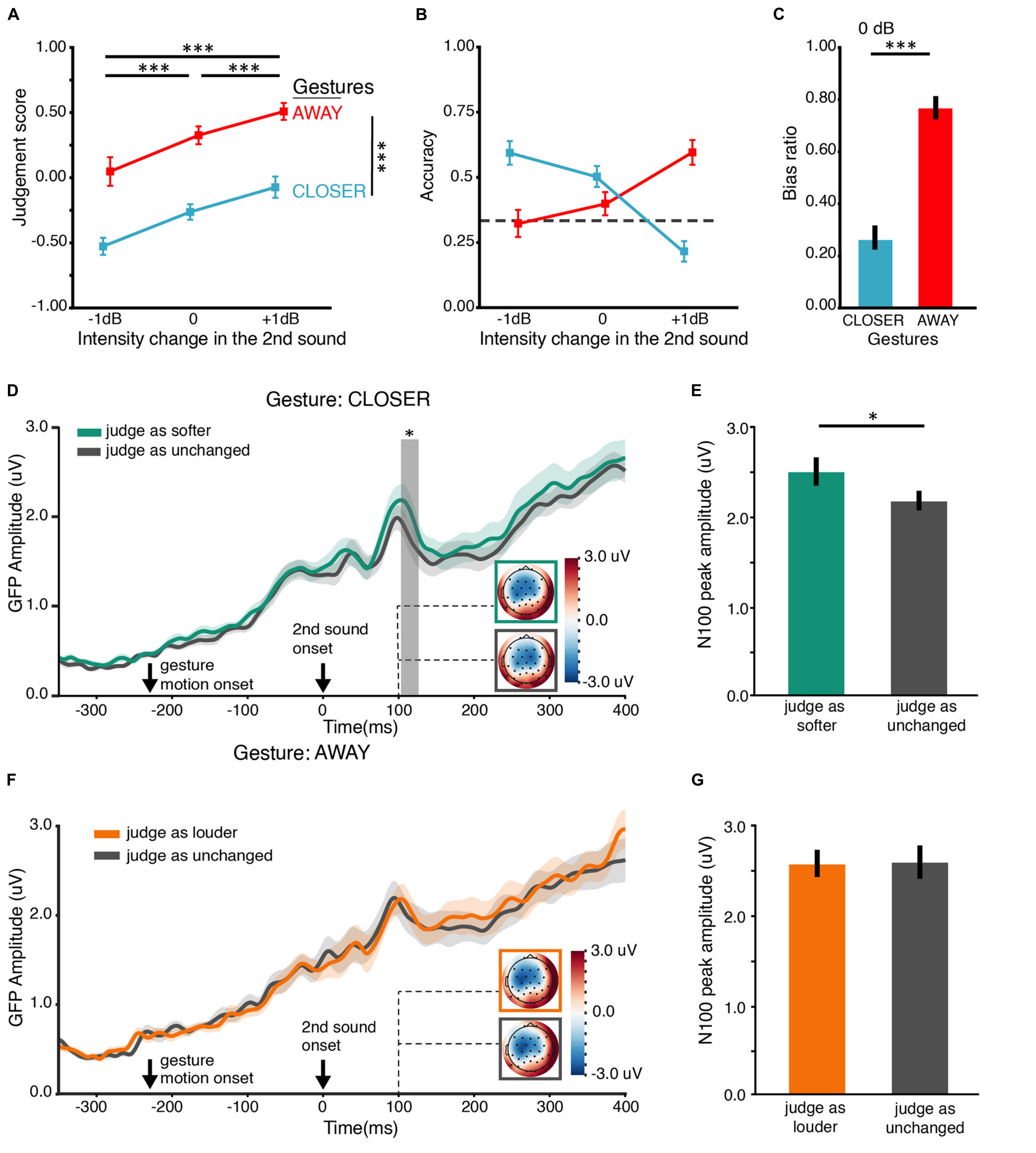

FIGURE 5 | Results of EE2. **(A)** Judgment score. Behaviorally, both gesture and intensity change positively affect the judgment of loudness changes, like the results in BE1 in **Figure 2A**. **(B)** Accuracy of the behavioral judgment about intensity change. An interaction between the factors of gesture and intensity change was observed, consistent with the results in BE1 (**Figure 2B**). The interaction was driven by the boost of accuracy by the CLOSER gesture in −1-dB intensity change and by the AWAY gesture in the +1-dB intensity change. **(C)** Bias ratios in the 0 dB intensity change conditions. The bias ratio was calculated as the judgment biased toward a louder percept: bias ratio = frequency of louder bias/frequency of all bias. AWAY biased participants toward choosing louder (higher bias ratio) while CLOSER biased participants toward choosing softer (lower bias ratio); the actual intensity did not change. **(D)** ERP responses to the second sound with 0-dB intensity change in the CLOSER conditions as a function of loudness perception. Stronger ERP responses were observed at around 115-ms latency when the second sound was perceived as "softer" (green) than was perceived as unchanged (gray), although the stimuli were the same sound of 0 dB intensity change. The solid lines indicate the grand mean global field power (GFP) waveforms. The shades around the solid lines represent ±one SEM. Response topographies are shown in colored boxes with dashed lines pointing to their latencies. The colored boxes use the same color schemes as the waveform responses to indicate different conditions. The gray vertical shades indicate the temporal cluster (104–207 ms) in which the two GFPs were significantly different in the temporal cluster analysis.

(Continued)

FIGURE 5 | Continued
(E) Response magnitude of N100 component in "softer" perceptive shifts (green) and no perceptive shifts (gray), obtained by temporally averaging a 20-ms time window centered at the individual early peak latencies (100 ms) observed in panel **(D)**. The response magnitude of N100 was larger in "softer" perceptive shifts than that in no perceptive shifts. **(F)** ERP responses to the second sound of 0 dB intensity change in the AWAY conditions as a function of loudness perception. No difference between ERP responses was observed. The depicting formats are the same as in panel **(D)**. **(G)** Response magnitude of N100 components in "louder" perceptive shifts (orange) and no perceptive shifts (gray), obtained by temporally averaging a 20-ms time window centered at the individual early peak latencies (100 ms) observed in panel **(F)**. "Louder" perceptive shifts had a similar response magnitude to no perceptive shifts in the early auditory response of N100. $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Error bars indicate ±one SEM.

judgments of loudness perception induced by gesture CLOSER showed larger N100 responses than unbiased judgments to the same stimuli. These consistent results collaboratively suggest that loudness perception can be modulated by the informational contents in other modalities that do not necessarily relate to auditory perception.

In BE1, we found that motional gestures modulated and interacted with the judgment of loudness change. Using the still CONST gesture as baseline conditions, we found that the AWAY gesture pushed participants' judgments toward louder across all intensities, whereas the CLOSER gesture had the opposite effect – it pulled the judgments toward softer (**Figure 2A**). This tendency was also observed in the accuracy of judgments (**Figure 2B**). Specifically, accuracy was highest when the intensity changes of −1, 0, and +1 dB were paired with CLOSER, CONST, and AWAY gestures, respectively. The bias ratio further characterized the direction and the extent to which gestures biased the judgment of loudness under specific intensity changes (**Figure 2C**). We found that gestural directions biased the judgment of loudness when they were inconsistent. Interestingly, for +1 dB intensity change, CLOSER biased the responses off two levels (choosing "softer") for about 40% of all the misjudgments made under that condition, and it was vice versa for −1 dB paired with AWAY. This effect was surprisingly big. One might argue that the 1 dB intensity change is hard to detect because it is close to the threshold, and participants might make their decision solely based on gestures. However, this was less likely given that the participants' accuracy in the training sessions was above chance level (**Supplementary Figure 1**). Moreover, we explicitly told participants that the paring between sounds and gestures was completely random so that they should judge the sound intensity change only by what they heard.

We further probed the modulation effects of still gestures in BE2 to dissociate the factors of the distance between hands from the moving trajectories of gestures. We found that although both types of gestural stimuli had similar overall effects, the still gestures (SHORT, LONG) had a weaker influence on the judgment of loudness change (**Figure 3A**) than their motional versions (CLOSER, AWAY in BE1). Specifically, still gestures did not induce any significant effects on the judgment accuracy in most of the intensity change conditions (**Figure 3B**). Our findings suggest that the visual-motor information of gestures modulated the judgment of loudness differently from the spatial location between hands. This was in line with an fMRI study (Calvert and Campbell, 2003) reporting that moving speaking faces activated the auditory cortex and STS greater than still speech face images did. More importantly, our findings further suggested that motional gestures affected the judgment of loudness change not just by effects like a psychological suggestion or priming. Otherwise, the motional gestures would have very similar effects to still gestures.

EE1 was designed and analyzed in a "stimulus" perspective to investigate the nature of the observed modulation effects. That is, we examined whether and how the two motional gestures, CLOSER and AWAY modulated the early auditory neural responses. We identified an early neural modulation effect at around 110 ms (**Figure 4D**). Surprisingly, the sound stimuli induced stronger responses at around 110-ms latency when CLOSER rather than AWAY gesture was presented. The observed effect was not due to differences in visual responses to gestures because CLOSER and AWAY elicited similar visual neural responses in the time range of interest (**Figure 4E**). We did not observe such a pattern of stronger early auditory responses to lower intensity sounds when participants saw a blank screen or a still image (**Supplementary Figure 3A**) throughout the trial. Therefore, this ERP pattern we found in the AV conditions was specific to gestural modulation.

To provide further and stronger evidence, we carried out EE2 that tackled the same question in EE1 but from a complementary "perception" angle. We compared neural responses to the same auditory stimuli of no intensity change across two instances but with different loudness judgments to the second sound. In this case, the physical stimuli in each comparison were identical. The only difference was the participants' loudness judgments. We found that the N100 ERP response was stronger when participants were biased by the CLOSER gesture to choose "second sound softer" than that when they were not perceptively biased. However, no significant neural modulation effect was found when participants were biased by AWAY gesture to choose "second sound louder." This modulation pattern was consistent with the observation in EE1. It is worth mentioning that both EE1 and EE2 replicated the behavioral results of BE1. Crucially, EE2 ruled out the possibility that the observed modulation effects were caused by task demand, context, and stimuli in the specific experimental procedures. The finding strongly suggested that the biased judgments of loudness induced by gesture CLOSER were perceptual in nature.

In EE2, we found significant effects of CLOSER but not AWAY gesture on modulating auditory neural responses. These surprising but consistent asymmetric results could root in the inherent properties of auditory perception. Asymmetry of loudness perception and neural responses has been reported in various auditory tasks, such as auditory habituation (Butler, 1968), loudness recalibration (Marks, 1994;

Mapes-Riordan and Yost, 1999), loudness adaptation (Canévet et al., 1985), and changing-loudness after effect (Reinhardt-Rutland, 2004). Interestingly, the changing-loudness after effect can also be induced if participants adapted to visual changing-depth, e.g., a box expanding or shrinking (Kitagawa and Ichihara, 2002). These studies suggest that asymmetry could indeed be a property in some forms of auditory perception. The modulation effects of gesture on loudness perception could also be asymmetrical so that the modulated responses associated with AWAY are smaller than the threshold that could be detected. Regardless of the asymmetry, the observations of modulation effects in early auditory responses support the hypothesis that visual-motor information in gestures can influence loudness perception.

The audiovisual paradigm we used introduces a challenge for the ERP analysis – the leading visual display induces visual responses that may temporally overlap with the subsequent auditory responses. In fact, we did not observe a clear auditory N1 component in EE1. This may be because the SOA of stimuli in different modalities was too short. Therefore, we used a longer SOA (230 ms) in EE2, which would minimize potential overlaps between the early auditory responses and visual responses to the preceding visual stimuli. The audiovisual integration likely occurs in a rather wide time window. So, the modulation effect would still be observed.

How loudness is represented in the brain is still unclear. Neuroimaging studies suggest that a full representation of perceived loudness completes at the cortical rather than the subcortical level (Röhl and Uppenkamp, 2012). According to electrophysiological studies (Thwaites et al., 2016) that analyzed the relations between EEG/MEG signals and loudness perception in different duration stimuli, the transformation of instantaneous loudness took place at 45- to 165-ms latency in Heschl's gyrus and dorsal lateral sulcus. The cortical loudness representation (short-term loudness) can form as early as 45 ms in Heschl's gyrus. Another transformation of the short-term loudness took place at 165- to 275-ms latency, such as at the length of a typical auditory word, in both dorsal lateral sulcus and superior temporal sulcus. We observed the modulation effect of gestures on loudness perception to simple vowels (/a/in our experiment) around 110 ms. The latency of the effect fell between the windows characterizing instantaneous loudness and the following possible transformation. Our results are consistent with previous literature about the dynamics of loudness perception and suggest that the perception of loudness might "superimpose" on auditory stimuli of different contents at different latencies.

Loudness perception is sensitive to context. Many studies have reported that loudness perception could be influenced by preceding sounds (Butler, 1968; Canévet et al., 1985; Näätänen and Picton, 1987; Lu et al., 1992; Marks, 1994; Mapes-Riordan and Yost, 1999; Näätänen and Winkler, 1999; Schmidt et al., 2020). At the neural population level, the dynamic range of auditory neurons of mammals could adapt to the intensity statistics of preceding sounds within a few seconds – a phenomenon called dynamic range adaptation (DRA). Evidence suggests that DRA first occurs in the auditory periphery (Wen et al., 2009) and develops along the auditory pathway, including the inferior colliculus (Dean et al., 2005) and the primary auditory cortex (Watkins and Barbour, 2008). Although different mechanisms have been proposed to account for various contextual effects, a common assumption is that loudness might be represented as relativity in the brain. In other words, what has been encoded is the change from a previous level instead of absolute magnitude. Our findings fit this view. We did not observe a simple relation among the sound intensity, loudness judgment, and ERP responses. In contrast, we observed larger N1 ERP responses to the sound with gesture CLOSER than with gesture AWAY in EE1. Moreover, the trials with "soft" bias evoked by CLOSER also showed a larger N1 component than trials with no bias in EE2. Such neural modulation effects were most likely reflecting the degree of loudness change. Moreover, we observed that the visual-motor information in gestures could influence auditory responses of loudness perception. This cross-modulation effect further suggests that the loudness perception is relative rather than directly linked to the absolute magnitude of physical, auditory stimuli.

Our results of cross-modulation on loudness perception are consistent with the framework of multisensory integration with some detailed exceptions. Audiovisual integration occurs in distributive cortices in various stages, with the most stable early effects around 100 ms after the sound onset (Talsma, 2015). The observed cross-modulation on loudness perception agrees with the timing of multisensory integration. Some theories assume the integration as unsupervised and bottom-up by combining information in two modalities based on spatial and temporal proximity (Alais and Burr, 2004; Baart et al., 2014). On the other hand, the multisensory integration could base on temporal predictions (van Wassenhove et al., 2005; Arnal et al., 2009) or predictions about features and categories (van Laarhoven et al., 2017). Moreover, iconic gestures and vocalization are innately connected in humans (Perlman and Lupyan, 2018). Our findings suggest that the early auditory responses reflect the modulation of gestures on loudness perception. However, the two gestures did not differ in their predictability or other aspects such as congruency or attention. The only difference was the moving direction that was remotely linked to loudness perception. Therefore, our results imply that factors other than predictability are likely to influence the amplitudes of early neural responses mediating loudness perception.

Manual gestures and speech have long been thought of as being integrated at the semantic and lexical level, with a few pieces of evidence suggesting a lower-level perceptual and productive connection. For example, observing motor acts of hand grasp modulated syllable pronunciation (Gentilucci, 2003). The lip aperture, voice peak amplitude and F0 frequency were greater when the observed hand grasp was directed to the large object. Our findings also suggest that gesture could modulate loudness perception, an attribute linking low-level features of speech perception. However, whether such modulation is due to a specific gesture-speech interaction or a general audiovisual interaction requires further investigation. It would be informative to probe the modulation effect by replacing the manual gestures with non-biological moving objects such as dots and bars, as well as extending to a wider range of featural differences.

The integrity of the gesture-speech system was often disrupted in various types of motor and psychopathological disorders, such as stuttering (Mayberry and Jaques, 2000), schizophrenia (Nagels et al., 2019), and autism spectrum (Silverman et al., 2010). Notably, tests based on sensory dominance and multisensory integration have been proposed as effective tools for the diagnosis of mild cognitive impairment in the elderly population (Murray et al., 2018). The findings in the current study may contribute to the development of new screening tools for psychopathological disorders involve the degradation of low-level multisensory processing.

The neural mechanisms that mediate multisensory integration, in particular the observed modulation effects of gestures on loudness perception, necessitate further investigation. Evidence suggests direct neural pathways between visual and auditory areas (Cappe and Barone, 2005). Silent movie clips of lip-movement activated auditory areas (Calvert et al., 1997; Calvert and Campbell, 2003; Besle et al., 2008). The earliest audiovisual cortical interaction can appear as early as 30 ms before the activation of polymodal areas (Besle et al., 2008). These studies indicate that cross-modal interaction could occur in a direct way between visual and auditory systems. Another possibility is that the interaction is mediated by the motor system. The motional gesture videos started 200 ms before the sound. The motion of gestures may induce corresponding motor representations that, in turn, transfer to sensory representations *via* the internal forward models (Tian and Poeppel, 2010, 2012). These sensory representations may share some common representational features that might be much easier to integrate with the processing of external auditory stimuli (Tian et al., 2018; Zhen et al., 2019). For loudness perception, the converted distance and speed information from the motor system may have an abstract representation for magnitude that can interact with the rate coding of loudness perception (Glasberg and Moore, 2002; Röhl and Uppenkamp, 2012; Thwaites et al., 2016).

In conclusion, we found that motional gestures influenced the judgment of loudness change at the JND threshold. Moreover, the cross-modal effects on loudness perception were temporally localized in the early auditory neural responses. The consistent results in four behavioral and EEG experiments suggest that gestures can modulate loudness perception. These findings provide evidence suggesting that visual-motor events can penetrate the processes of primary perceptual attributes in auditory perception.

DATA AVAILABILITY STATEMENT

The data will be available upon request to interested researchers.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by local Research Ethics Committee at NYU Shanghai. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

XT, JS, and ZW conceived the study. JS and XT wrote the manuscript. JS, ZW, and XT designed the experiments. JS and ZW carried out the experiments and pre-processed the EEG data. JS analyzed the data and made the tables and figures. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China 32071099, Natural Science Foundation of Shanghai 20ZR1472100, Program of Introducing Talents of Discipline to Universities, Base B16018, and NYU Shanghai Boost Fund.

ACKNOWLEDGMENTS

We thank Hao Zhu for his support in collecting and processing the EEG data, Zhuoshi Liu for her help in preprocessing part of our EEG data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.634967/full#supplementary-material

REFERENCES

Alais, D., and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262. doi: 10.1016/j.cub.2004.01.029

Alais, D., Newell, F., and Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing Perceiving* 23, 3–38. doi: 10.1163/187847510X488603

Arbib, M., Liebal, K., and Pika, S. (2008). Primate vocalization, gesture, and the evolution of human language. *Curr. Anthropol.* 49, 1053–1063. doi: 10.1086/593015

Arnal, L. H., Morillon, B., Kell, C. A., and Giraud, A.-L. (2009). Dual neural routing of visual facilitation in speech processing. *J. Neurosci.* 29, 13445–13453. doi: 10.1523/JNEUROSCI.3194-09.2009

Baart, M., Stekelenburg, J. J., and Vroomen, J. (2014). Electrophysiological evidence for speech-specific audiovisual integration. *Neuropsychologia* 53, 115–121. doi: 10.1016/j.neuropsychologia.2013.11.011

Besle, J., Fischer, C., Bidet-Caulet, A., Lecaignard, F., Bertrand, O., and Giard, M.-H. (2008). Visual activation and audiovisual interactions in the auditory cortex during speech perception: intracranial recordings in humans. *J. Neurosci.* 28, 14301–14310. doi: 10.1523/JNEUROSCI.2875-08.2008

Besle, J., Fort, A., Delpuech, C., and Giard, M.-H. (2004). Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234. doi: 10.1111/j.1460-9568.2004.03670.x

Boersma, P., and Weenink, D. (2021). *Praat: Doing Phonetics by Computer [Computer program]. Version 6.1.40.* Available online at: http://www.praat.org/ (accessed February 27, 2021).

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., et al. (2007). Neural basis of the ventriloquist illusion. *Curr. Biol.* 17, 1697–1703. doi: 10.1016/j.cub.2007.08.050

Butler, R. A. (1968). Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *J. Acoust. Soc. Am.* 44, 945–950. doi: 10.1121/1.1911233

Butterworth, B., and Beattie, G. (1978). "Gesture and silence as indicators of planning in speech," in *Proceedings of the Recent Advances in the Psychology of Language: Formal and Experimental Approaches NATO Conference Series*, eds R. N. Campbell and P. T. Smith (Boston, MA: Springer), 347–360. doi: 10.1007/978-1-4684-2532-1_19

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science* 276, 593–596. doi: 10.1126/science.276.5312.593

Calvert, G. A., and Campbell, R. (2003). Reading speech from still and moving faces: the neural substrates of visible speech. *J. Cogn. Neurosci.* 15, 57–70. doi: 10.1162/089892903321107828

Calvert, G. A., Spence, C., and Stein, B. E. (2004). *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.

Canévet, G., Scharf, B., and Botte, M.-C. (1985). Simple and induced loudness adaptation. *Audiology* 24, 430–436. doi: 10.3109/00206098509078362

Cappe, C., and Barone, P. (2005). Heteromodal connections supporting multisensory integration at low levels of cortical processing in the monkey. *Eur. J. Neurosci.* 22, 2886–2902. doi: 10.1111/j.1460-9568.2005.04462.x

Caramiaux, B., Bevilacqua, F., and Schnell, N. (2010). "Towards a gesture-sound cross-modal analysis," in *Gesture in Embodied Communication and Human-Computer Interaction*, eds S. Kopp and I. Wachsmuth (Berlin: Springer), 158–170. doi: 10.1007/978-3-642-12553-9_14

Dean, I., Harper, N. S., and McAlpine, D. (2005). Neural population coding of sound level adapts to stimulus statistics. *Nat. Neurosci.* 8, 1684–1689. doi: 10.1038/nn1541

Gentilucci, M. (2003). Grasp observation influences speech production. *Eur. J. Neurosci.* 17, 179–184. doi: 10.1046/j.1460-9568.2003.02438.x

Ghazanfar, A. A., and Schroeder, C. E. (2006). Is neocortex essentially multisensory? *Trends Cogn. Sci.* 10, 278–285. doi: 10.1016/j.tics.2006.04.008

Glasberg, B. R., and Moore, B. C. J. (2002). A model of loudness applicable to time-varying sounds. *J. Audio Eng. Soc.* 50:25.

Goldin-Meadow, S., and Alibali, M. W. (2013). Gesture's role in speaking, learning, and creating language. *Annu. Rev. Psychol.* 64, 257–283. doi: 10.1146/annurev-psych-113011-143802

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C., et al. (2014). MNE software for processing MEG and EEG data. *NeuroImage* 86, 446–460. doi: 10.1016/j.neuroimage.2013.10.027

Howard, I. P., and Templeton, W. B. (1966). *Human spatial orientation*. Oxford: John Wiley & Sons.

Hubbard, A. L., Wilson, S. M., Callan, D. E., and Dapretto, M. (2009). Giving speech a hand: gesture modulates activity in auditory cortex during speech perception. *Hum. Brain Mapp.* 30, 1028–1037. doi: 10.1002/hbm.20565

Johnson, J. H., Turner, C. W., Zwislocki, J. J., and Margolis, R. H. (1993). Just noticeable differences for intensity and their relation to loudness. *J. Acoust. Soc. Am.* 93, 983–991. doi: 10.1121/1.405404

Kelly, S. D., Kravitz, C., and Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain Lang.* 89, 253–260. doi: 10.1016/S0093-934X(03)00335-333

Kitagawa, N., and Ichihara, S. (2002). Hearing visual motion in depth. *Nature* 416, 172–174. doi: 10.1038/416172a

Krauss, R. M. (1998). Why do we gesture when we speak? *Curr. Dir. Psychol. Sci.* 7, 54–54. doi: 10.1111/1467-8721.ep13175642

Lu, Z. L., Williamson, S. J., and Kaufman, L. (1992). Behavioral lifetime of human auditory sensory memory predicted by physiological measures. *Science* 258, 1668–1670. doi: 10.1126/science.1455246

Mapes-Riordan, D., and Yost, W. A. (1999). Loudness recalibration as a function of level. *J. Acoust. Soc. Am.* 106, 3506–3511. doi: 10.1121/1.428203

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190. doi: 10.1016/j.jneumeth.2007.03.024

Marks, L. E. (1994). "Recalibrating" the auditory system: the perception of loudness. *J. Exp. Psychol. Hum. Percept. Perform.* 20, 382–396. doi: 10.1037/0096-1523.20.2.382

Mayberry, R. I., and Jaques, J. (2000). "Gesture production during stuttered speech: insights into the nature of gesture–speech integration," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 199–214. doi: 10.1017/CBO9780511620850.013

McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748. doi: 10.1038/264746a0

Morrel-Samuels, P., and Krauss, R. M. (1992). Word familiarity predicts temporal asynchrony of hand gestures and speech. *J. Exp. Psychol. Learn. Mem. Cogn.* 18:615. doi: 10.1037/0278-7393.18.3.615

Möttönen, R., Krause, C. M., Tiippana, K., and Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cogn. Brain Res.* 13, 417–425. doi: 10.1016/S0926-6410(02)00053-58

Murray, M. M., Brunet, D., and Michel, C. M. (2008). Topographic ERP analyses: a step-by-step tutorial review. *Brain Topogr.* 20, 249–264. doi: 10.1007/s10548-008-0054-55

Murray, M. M., Eardley, A. F., Edginton, T., Oyekan, R., Smyth, E., and Matusz, P. J. (2018). Sensory dominance and multisensory integration as screening tools in aging. *Sci. Rep.* 8:8901. doi: 10.1038/s41598-018-27288-27282

Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x

Näätänen, R., and Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* 125:826. doi: 10.1037/0033-2909.125.6.826

Nagels, A., Kircher, T., Grosvald, M., Steines, M., and Straube, B. (2019). Evidence for gesture-speech mismatch detection impairments in schizophrenia. *Psychiatry Res.* 273, 15–21. doi: 10.1016/j.psychres.2018.12.107

Özyürek, A., Willems, R. M., Kita, S., and Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *J. Cogn. Neurosci.* 19, 605–616. doi: 10.1162/jocn.2007.19.4.605

Perlman, M., and Lupyan, G. (2018). People can create iconic vocalizations to communicate various meanings to naïve listeners. *Sci. Rep.* 8:2634. doi: 10.1038/s41598-018-20961-20966

Reinhardt-Rutland, A. H. (2004). Perceptual asymmetries associated with changing-loudness aftereffects. *Percept. Psychophys.* 66, 963–969. doi: 10.3758/BF03194988

Roberts, T. P. L., Ferrari, P., Stufflebeam, S. M., and Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *J. Clin. Neurophysiol.* 17, 114–129.

Röhl, M., and Uppenkamp, S. (2012). Neural coding of sound intensity and loudness in the human auditory system. *J. Assoc. Res. Otolaryngol.* 13, 369–379. doi: 10.1007/s10162-012-0315-316

Schmidt, F. H., Mauermann, M., and Kollmeier, B. (2020). Neural representation of loudness: cortical evoked potentials in an induced loudness reduction experiment. *Trends Hear.* 24:2331216519900595. doi: 10.1177/2331216519900595

Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., and Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends Cogn. Sci.* 12, 106–113. doi: 10.1016/j.tics.2008.01.002

Silverman, L. B., Bennetto, L., Campana, E., and Tanenhaus, M. K. (2010). Speech-and-gesture integration in high functioning autism. *Cognition* 115, 380–393. doi: 10.1016/j.cognition.2010.01.002

Stein, B. E., and Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* 9, 255–266. doi: 10.1038/nrn2331

Stekelenburg, J. J., and Vroomen, J. (2012). Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. *Front. Integr. Neurosci.* 6:26. doi: 10.3389/fnint.2012.00026

Talsma, D. (2015). Predictive coding and multisensory integration: an attentional account of the multisensory mind. *Front. Integr. Neurosci.* 9:19. doi: 10.3389/fnint.2015.00019

Thwaites, A., Glasberg, B. R., Nimmo-Smith, I., Marslen-Wilson, W. D., and Moore, B. C. J. (2016). Representation of instantaneous and short-term loudness in the human cortex. *Front. Neurosci.* 10:183. doi: 10.3389/fnins.2016.00183

Tian, X., Ding, N., Teng, X., Bai, F., and Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nat. Hum. Behav.* 2, 225–234.

Tian, X., and Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Front. Psychol.* 1:166. doi: 10.3389/fpsyg.2010.00166

Tian, X., and Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. *Front. Hum. Neurosci.* 6:314. doi: 10.3389/fnhum.2012.00314

van Laarhoven, T., Stekelenburg, J. J., and Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: electrophysiological evidence from stimulus omissions. *Brain Res.* 1661, 79–87. doi: 10.1016/j.brainres.2017.02.014

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. *Proc. Natl. Acad. Sci. U S A.* 102, 1181–1186. doi: 10.1073/pnas.0408949102

van Wassenhove, V., Grant, K. W., and Poeppel, D. (2007). Temporal window of integration in auditory-visual speech perception. *Neuropsychologia* 45, 598–607. doi: 10.1016/j.neuropsychologia.2006.01.001

Wang, X., Zhu, H., and Tian, X. (2019). Revealing the temporal dynamics in non-invasive electrophysiological recordings with topography-based analyses. *bioRxiv [preprint]* doi: 10.1101/779546

Watkins, P. V., and Barbour, D. L. (2008). Specialized neuronal adaptation for preserving input sensitivity. *Nat. Neurosci.* 11, 1259–1261. doi: 10.1038/nn.2201

Wen, B., Wang, G. I., Dean, I., and Delgutte, B. (2009). Dynamic range adaptation to sound level statistics in the auditory nerve. *J. Neurosci.* 29, 13797–13808. doi: 10.1523/JNEUROSCI.5610-08.2009

Willems, R. M., Özyürek, A., and Hagoort, P. (2007). When language meets action: the neural integration of gesture and speech. *Cereb. Cortex* 17, 2322–2333. doi: 10.1093/cercor/bhl141

Yang, J., Zhu, H., and Tian, X. (2018). Group-Level multivariate analysis in EasyEEG toolbox: examining the temporal dynamics using topographic responses. *Front. Neurosci.* 12:468. doi: 10.3389/fnins.2018.00468

Zhen, A., Van Hedger, S., Heald, S., Goldin-Meadow, S., and Tian, X. (2019). Manual directional gestures facilitate cross-modal perceptual learning. *Cognition* 187, 178–187. doi: 10.1016/j.cognition.2019.03.004

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.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership