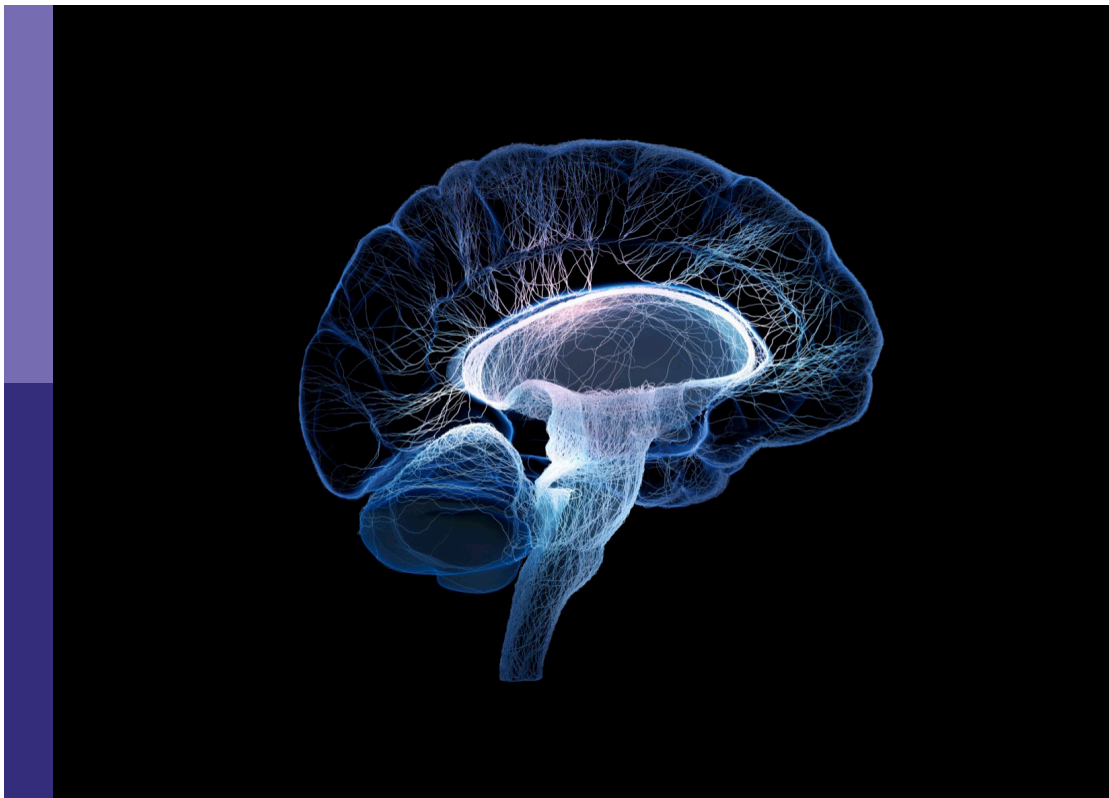
frontiers | Research Topics

Insights in auditory cognitive neuroscience 2021

Edited by
Claude Alain and Marc Schönwiesner

Published in
Frontiers in Neuroscience
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-2241-7
DOI 10.3389/978-2-8325-2241-7

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

Insights in auditory cognitive neuroscience: 2021

Topic editors
Claude Alain — Rotman Research Institute (RRI), Canada
Marc Schönwiesner — Leipzig University, Germany

Citation
Alain, C., Schönwiesner, M., eds. (2023). *Insights in auditory cognitive neuroscience: 2021*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-2241-7

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 11 April 2023
DOI 10.3389/fnins.2023.1192459

OPEN ACCESS

EDITED AND REVIEWED BY
Robert J. Zatorre,
McGill University, Canada

*CORRESPONDENCE
Marc Schönwiesner
✉ marcs@uni-leipzig.de

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

RECEIVED 23 March 2023
ACCEPTED 24 March 2023
PUBLISHED 11 April 2023

CITATION
Schönwiesner M and Alain C (2023) Editorial: Insights in auditory cognitive neuroscience: 2021. *Front. Neurosci.* 17:1192459. doi: 10.3389/fnins.2023.1192459

Editorial: Insights in auditory cognitive neuroscience: 2021

Marc Schönwiesner[1,2]* and Claude Alain[3,4]

[1]Institute of Biology, Faculty of Life Sciences, Leipzig University, Leipzig, Germany, [2]Department of Psychology, Faculté des Arts et des Sciences, Université de Montréal, Montreal, QC, Canada, [3]Department of Psychology, University of Toronto, Toronto, ON, Canada, [4]Rotman Research Institute, Baycrest Hospital, Toronto, ON, Canada

KEYWORDS

auditory, pitch, MMN (mismatch negativity), speech-in-noise, hemispheric asymmetries, voice, what/where system, hearing disorders

Editorial on the Research Topic
Insights in auditory cognitive neuroscience: 2021

Imagine an expensive research and development meeting at a large company. The presenter: "We have our top people working on this. Our top people!" This is how we feel about the many recent breakthroughs in auditory cognitive neuroscience research. Researchers like Tim Griffiths, Robert Zatorre, Andrew Oxenham, and the other contributors to this Frontiers' Research Topic have shaped and advanced the field for years. This collection of ten short perspective papers aims to provide a readable overview of several current (and, in many cases, timeless) topics in auditory cognitive neuroscience through the vantage point of some of the main actors. The papers are best enjoyed as a collection rather than independently because of the many interconnections between the topics they discuss, some of which we will point out here.

We start with topic of processing and representation of critical auditory features. The mechanism of pitch perception is among the oldest such topics in hearing science, going back to Strutt (1907). The brain encoding of time-based pitch cues has seen strong empirical support using delay-and-add noise in brain imaging studies (Griffiths et al., 1998). A classical study by Oxenham et al. (2004) demonstrated that time-based cues are not sufficient and that pitch perception also requires correct cochlear frequency-to-place mapping of the spectral components of the stimulus. After over a 100 years of research, the relationship between these two cues in pitch perception and representation is still under debate. The perspective by Oxenham discusses recent developments and directions in the study of pitch coding and perception.

From pitch extraction is the extraction of voice features: Pascal Belin's discovery of the temporal voice area in 2001 (Belin et al., 2000) opened up new research into the cortical processing of voices and non-speech vocal sounds. This area around the middle of the superior temporal sulcus responds more strongly to voices than other sounds. There is some discussion of whether this area is processing speech rather than voice information, which is reminiscent of the debate around whether the fusiform face area genuinely represents faces or any stimuli that observers have acquired expertise with (Gauthier et al., 1999). Here Trapeau et al. present evidence-based arguments to support the role of the temporal voice area in genuine voice processing.

The mismatch negativity is one of the most popular neural metrics to study preattentive processing, predictive coding mechanisms, auditory memory, and many other phenomena. Its discovery in late 1978 by Finish psychologist Risto Nätäänen created a paradigm shift in auditory neuroscience. Tervaniemi discusses the development of stimulation paradigms from simple sine tones to complex multi-feature sounds and paradigms, including recent efforts to achieve ecological validity in experiments with such tightly controlled and repetitive stimuli. These new developments will ensure that the mismatch negativity remains among the most significant and versatile tools in auditory cognitive neuroscience for years to come.

Our understanding of the function and organization of the human primary (core) auditory cortex needs to catch up to that of the visual cortex. The auditory core is much smaller than V1 and is divided into subfields, nested on the superior temporal gyrus. Several functional and anatomical markers have been discovered and allow some non-invasive access, for example, increased myelination (Sigalovsky et al., 2006), the 40-Hz auditory steady-state response (Gutschalk et al., 1999), or a peak in the slope of the magneto-encephalographic response at about 20 ms (Lütkenhöner et al., 2003). Simon et al. argues that early time-locked high gamma band responses to natural speech can track primary cortical activity, adding a robust and ecologically valid method to study primary auditory cortex function non-invasively.

We now turn to the organization of the auditory system. Zatorre provides a perspective of hemispherical asymmetries in music and speech processing, in which his group has contributed significant theoretical and empirical advances. This is a topic with deep historical roots going back to the recognition of lateralized language areas by Broca and Wernicke in the late 19th century. Zatorre unifies recent results on the processing of musical pitch patterns in auditory networks of the right hemisphere (and complementary lateralization of speech sounds) in the framework of spectrotemporal modulation processing. The paper discusses the importance of low-level differential sensitivity to acoustical features of communication sounds (bottom-up) and high-level modulation of asymmetries by learning, attention, or other top-down factors.

A central concept of sensory processing in the cortex is that of partially segregated streams with different functions. This idea was initially conceived to explain different sensitivities, and latencies in cortical fields along the visual pathway (Schneider, 1969; Mishkin and Ungerleider, 1982; Goodale and Milner, 1992) and later applied to audition by Rauschecker and Tian (2000) with the proposal of "what" and "where" pathways. This idea was reconceptualized several times, and the dual pathways have lost their initial clear functional separation and are now often referred to by location. These ventral and dorsal processing streams originate in the secondary (belt) auditory cortex in rostral and caudal fields, which then connect to different downstream areas in the frontal and parietal cortex. A recurrent functional distinction that has held up since the original studies in non-human primates is that rostral fields tend to be more involved in sound recognition and caudal areas more in sound localization. Scott and Jasmin discuss the origins and recent developments of the dual stream concept and its interaction with speech and voice processing of simultaneous talkers.

The feedback or top-down auditory projections is another principle of brain organization with powerful implications. The cortico-fugal pathway, the thickest efferent projection in the human brain after the pyramidal tract, instructively illustrates this. McAlpine and de Hoz discuss how adaptation in such feedback pathways of the auditory system aids in adaptive en- and decoding of complex sounds by building a representation of their statistical structure at different time scales. Exploring these feedback loops at different granularities, from *in vivo* recording to human neuroimaging, may reveal the fundamental listening processes.

Finally, we turn to topics in more applied auditory neuroscience. Griffiths provides an overview of recent work in the lab on predicting speech-in-noise ability based on performance with non-speech material in basic auditory cognitive tests. Speech in noise perception is the most important human auditory capacity and a consistent problem for persons with hearing disorders. Such tests may reveal the basic auditory factors that determine speech-in-noise understanding and enable more robust, language-independent clinical diagnosis. Rönnberg et al. discusses the ongoing trend of including more cognitive factors in this effort to add to the classical models based on system identification approaches to peripheral hearing mechanisms. He proposes the Ease of Language Understanding model, which models complex interactions of cognitive modules, such as the different memory systems, lexical access, and predictive and postdictive processes. Such models help to understand the perceptual consequences of hearing disorders and mirror the trend to include cognitive factors in hearing aids and rehabilitation.

Author contributions

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

Acknowledgments

We would like to thank all authors and reviewers who participated in the special issue.

Conflict of interest

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

Publisher's note

References

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312 doi: 10.1038/35002078

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., and Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. *Nat. Neurosci.* 2, 568–573 doi: 10.1038/9224

Goodale, M. A., and Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci.* 15, 20–25. doi: 10.1016/0166-2236(92)90344-8

Griffiths, T. D., Büchel, C., Frackowiak, R. S., and Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nat. Neurosci.* 1, 422–427 doi: 10.1038/1637

Gutschalk, A., Mase, R., Roth, R., Ille, N., Rupp, A., Hähnel, S., et al. (1999). Deconvolution of 40 Hz steady-state fields reveals two overlapping source activities of the human auditory cortex. *Clin. Neurophysiol.* 110, 856–868 doi: 10.1016/S1388-2457(99)00019-X

Lütkenhöner, B., Krumbholz, K., Lammertmann, C., Seither-Preisler, A., Steinsträter, O., and Patterson, R. D. (2003). Localization of primary auditory cortex in humans by magnetoencephalography. *Neuroimage* 18, 58–66 doi: 10.1006/nimg.2002.1325

Mishkin, M., and Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behav. Brain Res.* 6, 57–77. doi: 10.1016/0166-4328(82)90081-X

Oxenham, A. J., Bernstein, J. G., and Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proc. Natl. Acad. Sci. U. S. A.* 101, 1421–1425 doi: 10.1073/pnas.0306958101

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U. S. A.* 97, 11800–11806. doi: 10.1073/pnas.97.22.11800

Schneider, G. E. (1969). Two visual systems. *Science* 163, 895–902 doi: 10.1126/science.163.3870.895

Sigalovsky, I. S., Fischl, B., and Melcher, J. R. (2006). Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: a possible marker for primary cortex and hemispheric differences. *Neuroimage* 32, 1524–1537 doi: 10.1016/j.neuroimage.2006.05.023

Strutt, J. W. (1907). On our perception of sound direction. *Philos. Mag.* 13, 214–232.

frontiers | Frontiers in Neuroscience

TYPE Perspective
PUBLISHED 09 January 2023
DOI 10.3389/fnins.2022.1074752

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Chris Plack,
The University of Manchester,
United Kingdom

*CORRESPONDENCE
Andrew J. Oxenham
✉ oxenham@umn.edu

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

RECEIVED 19 October 2022
ACCEPTED 16 December 2022
PUBLISHED 09 January 2023

CITATION
Oxenham AJ (2023) Questions and controversies surrounding the perception and neural coding of pitch.
Front. Neurosci. 16:1074752.
doi: 10.3389/fnins.2022.1074752

Questions and controversies surrounding the perception and neural coding of pitch

Andrew J. Oxenham[1,2]*

[1]Center for Applied and Translational Sensory Science, University of Minnesota Twin Cities, Minneapolis, MN, United States, [2]Department of Psychology, University of Minnesota Twin Cities, Minneapolis, MN, United States

Pitch is a fundamental aspect of auditory perception that plays an important role in our ability to understand speech, appreciate music, and attend to one sound while ignoring others. The questions surrounding how pitch is represented in the auditory system, and how our percept relates to the underlying acoustic waveform, have been a topic of inquiry and debate for well over a century. New findings and technological innovations have led to challenges of some long-standing assumptions and have raised new questions. This article reviews some recent developments in the study of pitch coding and perception and focuses on the topic of how pitch information is extracted from peripheral representations based on frequency-to-place mapping (tonotopy), stimulus-driven auditory-nerve spike timing (phase locking), or a combination of both. Although a definitive resolution has proved elusive, the answers to these questions have potentially important implications for mitigating the effects of hearing loss *via* devices such as cochlear implants.

KEYWORDS

pitch, auditory perception, auditory neuroscience, computational models, cochlear filtering, phase locking

1. Introduction

Pitch—the perceptual correlate of acoustic repetition rate or fundamental frequency (F0)—plays a critical role in both music and speech perception (Plack et al., 2005). Pitch is also thought to be crucial for source segregation—our ability to selectivity hear out and attend to one sound (e.g., a singer or your conversation partner) in the presence of other sounds (e.g., backing instruments or neighboring conversations). Experimental approaches to understanding pitch can be traced back to Seebeck (1841), Ohm (1843), and Helmholtz (1885/1954). Indeed, an early dispute (Turner, 1977) foreshadowed a long-running debate that continues to this day in various forms on what aspects of sound the auditory system extracts in order to derive pitch.

2. A time and a place for pitch

2.1. Historical roots

The classic pitch-evoking stimulus is a harmonic complex tone, which repeats at the fundamental frequency (F0) and consists of pure tones with frequencies at integer multiples of the F0 (F0, 2F0, 3F0, etc.). The components that form the harmonic tone complex are known as harmonics. We perceive a pitch corresponding to the F0 of a harmonic complex tone, even when the component at F0 itself is missing (the so-called pitch of the missing fundamental; Oxenham, 2012). Much of the debate surrounding pitch has focused on whether pitch is extracted *via* the frequency-to-place mapping that occurs along the basilar membrane (place code; e.g., Wightman, 1973; Terhardt, 1974; Cohen et al., 1995), *via* the timing of stimulus-driving spiking activity in the auditory nerve that is phase-locked to the periodicities present in the stimulus (temporal or time code; Licklider, 1951; Cariani and Delgutte, 1996; Meddis and O'Mard, 1997), or *via* some combination of the two (place-time code; Shamma and Klein, 2000; Cedolin and Delgutte, 2010).

Place theories can be likened to a Fourier transform, followed by pattern recognition or template matching to identify the F0 based on the pattern of places along the basilar membrane responding to different harmonics of a complex tone. These theories or models are often referred to as rate-place models, because they are based on the average firing rate and the tonotopic location of auditory-nerve fibers. Time theories have often been implemented *via* an autocorrelation function, again with either a peak-picking or template-matching stage to identify the dominant underlying periodicity. This timing information can be extracted from the temporal fine structure (TFS) of individual spectrally resolved harmonics, as well as from the temporal envelope fluctuations at the F0 produced by the interactions of spectrally unresolved harmonics (Oxenham, 2012). The contrast between the spectral representation and the autocorrelation function goes some way toward explaining why it has been so difficult to distinguish between the two approaches: the power spectral density and the autocorrelation functions are Fourier transforms of each other, meaning that they are mathematically equivalent and any change to one representation will invariably lead to a change in the other.

Aside from being difficult to distinguish between peripheral rate-place and time codes, the question becomes moot by the level of the cortex, because neurons no longer phase-lock to frequencies higher than a few hundred hertz, meaning that any code based on phase-locked information must have been transformed to another code by this stage of processing (Fishman et al., 2013). So why should we be interested in how information is being extracted from the auditory periphery? One strong rationale is that people with sensorineural hearing loss and/or cochlear implants can be severely limited in their perception of pitch. Understanding how pitch is extracted in the normally functioning auditory periphery may provide important insights into how best to improve pitch perception *via* devices such as cochlear implants.

2.2. Rethinking arguments in favor of a time code

A number of arguments exist in favor of a time code for pitch. However, recent work has led to a rethinking of many of these arguments, as listed below.

2.2.1. Pitch is still heard, even in the absence of any place cues

Amplitude-modulated white noise can elicit a pitch (Burns and Viemeister, 1976, 1981), as can a harmonic complex tone that has been highpass filtered to remove any spectrally resolved harmonics (Houtsma and Smurzynski, 1990). The pitch of such sounds is thought to be extracted *via* the periodicity in the temporal envelope of the stimulus, providing *prima facie* evidence that periodic temporal information can be extracted from auditory-nerve activity to encode pitch.

However, *temporal-envelope pitch is fragile*. The resulting pitch is susceptible to interference through noise or reverberation (Qin and Oxenham, 2005), insufficient to convey multiple simultaneous pitches (Carlyon, 1996; Micheyl et al., 2010; Graves and Oxenham, 2019), and produces discrimination thresholds (just-noticeable differences in pitch) that are several times worse than those of complex tones with spectrally resolved harmonics (e.g., Mehta and Oxenham, 2020). This evidence for poor human processing of temporal-envelope pitch suggests that the timing information extracted from the envelope is insufficient to explain the highly salient and accurate perception of pitch we experience with everyday sounds. Indeed, our insensitivity to temporal-envelope pitch poses a problem for timing-based models of pitch, which generally perform too well (relative to human listeners) in cases where only temporal-envelope cues are present (Carlyon, 1998), and require somewhat *ad hoc* assumptions to bring their predictions into line with the perceptual data (Bernstein and Oxenham, 2005; de Cheveigné and Pressnitzer, 2006).

2.2.2. Pitch discrimination is too good to be explained by place cues

We are exquisitely sensitive to small changes in the frequency of pure tones and the F0 of complex tones, to the extent that trained listeners can detect changes of less than 1% (e.g., Micheyl et al., 2006). A place code requires the change in frequency to produce a detectable change in the response level at one or more places along the basilar membrane (leading to a change in average firing rate in one or more auditory-nerve fibers). Standard estimates of human

frequency selectivity (Glasberg and Moore, 1990), combined with estimates of the level change needed to be detectable, lead to predicted thresholds for frequency discrimination and frequency-modulation detection that are considerably higher (worse) than observed in humans (Micheyl et al., 2013). Moreover, computational modeling suggests that the amount of information present in the timing of auditory-nerve fibers can exceed the information present when considering just the spatial distribution of average firing rates by two or more orders of magnitude (Siebert, 1970; Heinz et al., 2001; Guest and Oxenham, 2022).

On the other hand, *place cues may be more accurate than we thought.* Early estimates of peripheral frequency selectivity came from physiological studies in small mammals (e.g., Kiang et al., 1967). More recent work combining otoacoustic emissions with behavioral studies using forward masking has suggested that human cochlear tuning is sharper than that in the most commonly studied smaller mammals by a factor of 2–3 (Shera et al., 2002; Sumner et al., 2018). Sharper tuning implies more accurate place coding of small changes in frequency and pitch. In addition, computational modeling has shown that frequency and intensity discrimination in humans can be explained within the same rate-place framework if the reasonable assumption is made that there exists some non-stimulus-related (noise) correlation between cortical neurons with similar frequency response characteristics (Micheyl et al., 2013; Oxenham, 2018). Finally, the ability to detect small fluctuations in the frequency of pure tones (frequency modulation, or FM) shows a significant correlation with estimates of cochlear tuning in people with a wide range of hearing losses, consistent with expectations based on place-based frequency and pitch coding (Whiteford et al., 2020). Based on these newer results, there may no longer be a need to postulate an additional timing-based code to account for human frequency and pitch sensitivity.

2.2.3. Pitch perception degrades at high frequencies

Our ability to discriminate small changes in the frequency of pure tones degrades at frequencies beyond about 4 kHz (Moore, 1973; Moore and Ernst, 2012), as does our ability to recognize even well-known melodies (Attneave and Olson, 1971). This degradation is at least qualitatively consistent with the loss of phase-locking at frequencies beyond 1–2 kHz observed in other mammalian species, such as cat or guinea pig, and possibly humans (Verschooten et al., 2018). In contrast, the sharpness of cochlear filtering, on which place coding depends, actually improves with increasing frequency (Shera et al., 2002), leading to predictions of better, not worse, pitch discrimination.

However, *changes in pitch at high frequencies may not be due to loss of phase locking.* Several recent strands of evidence suggest that the link between poor high-frequency pitch and degraded phase-locking may not be so clear cut. First, complex pitch perception remains accurate even when spectrally resolved harmonics are all above 8 kHz (and so likely beyond the range of usable phase-locking), so long as the F0 itself remains within the musical pitch range (Oxenham et al., 2011; Lau et al., 2017). This suggests that phase-locked information is not necessary for complex pitch perception. Second, the degradation of frequency and FM sensitivity at high frequencies (and at fast FM rates), which had been ascribed to a loss of usable phase-locked information (Moore and Sek, 1996), is also found for tasks that do not involve TFS but instead involve comparisons of level fluctuations across frequency, as would be needed by a rate-place code for frequency (Whiteford et al., 2020). It may be that sensitivity to frequency changes and pitch at high frequencies is poorer due to cortical, rather than peripheral, limitations because pitch from high frequencies is less common and less relevant to us for everyday communication (Oxenham et al., 2011).

2.2.4. The time code is robust to changes in sound level

Perhaps the most compelling remaining argument is that place cues may be dependent on overall sound level, with cochlear tuning broadening and most auditory-nerve responses saturating at high levels, whereas timing cues are generally less susceptible to non-linearities and saturation (Carney et al., 2015).

However, *human data show level dependencies too.* Behavioral studies show a decrease in the number of spectrally resolved harmonics, and a concomitant decrease in pitch discrimination ability, with increasing sound level, in line with the predicted effects of broader cochlear tuning (Bernstein and Oxenham, 2006a). Also, high-threshold, low-spontaneous-rate auditory-nerve fibers remain unsaturated, even at high sound levels (Liberman, 1978; Winter et al., 1990), leaving open the possibility of rate-place coding over a wide range of sound levels.

In summary, none of the primary arguments in support of phase-locked encoding of TFS cues for pitch remains compelling in light of recent empirical data and computational modeling. Indeed, several aspects of the human data, such as the inability to use timing information when it is presented to the "wrong" place along the cochlea (Oxenham et al., 2004) and the ability to perceive complex pitch with only high-frequency components for which little or no timing information can be extracted (Oxenham et al., 2011; Lau et al., 2017; Mehta and Oxenham, 2022), suggest that timing information may be neither necessary nor sufficient for the perception of pitch.

3. Asking why as well as how: Machine learning approaches

As noted in the previous section, it has been suggested that poorer pitch discrimination for high-frequency pure

tones may be a consequence of less exposure and less ecological relevance of these high-frequency stimuli, rather than a consequence of poorer peripheral encoding (Oxenham et al., 2011). A more comprehensive approach to ecological relevance was taken earlier by Schwartz and Purves (2004), who suggested that many aspects of pitch perception could be explained in terms of the statistics of periodic sounds in our environment, such as voiced speech. This approach can be thought of as asking "why" pitch perception is the way it is, rather than "how" it is represented in the auditory system. A similar approach has been taken more recently by harnessing deep neural networks (DNN) and training them on a large database of over 2 million brief segments of periodic sounds, taken from speech and music recordings embedded in noise (Saddler et al., 2021). Using a well-established computational model of the auditory periphery (cochlea and auditory nerve) as a front end (Bruce et al., 2018), Saddler et al. (2021) found that after training the networks to identify the F0 of these sounds, the networks were able to reproduce a number of "classical" pitch phenomena, supporting the idea of Schwartz and Purves (2004) that many aspects of pitch perception can be explained in terms of the statistics of the sounds we encounter, and extending it by providing quantitative comparisons of the model's predictions and human performance.

Saddler et al.'s approach also extended beyond the "why" and returned to "how" by testing the relative importance of the spectral resolution and phase-locking in their front-end model. Their simulation results suggested that the spectral resolution of their model was not critical to their results, but that phase-locking was. This result, taken at face value, might suggest support for time over place models of pitch. However, the predictions are at odds with empirical data showing that poorer spectral resolution, either *via* hearing loss in humans (Bernstein and Oxenham, 2006b) or *via* broader cochlear filters in other species (Shofner and Chaney, 2013; Walker et al., 2019), does in fact affect pitch perception. This mismatch between model predictions and empirical data may be because the model has complete access to all the timing information in the simulated auditory nerve. In that sense, the conclusion from the DNN model can be treated as a restatement of the earlier findings from optimal-detector or ideal-observer models (Siebert, 1970; Heinz et al., 2001) that timing information from the auditory nerve provides much greater coding accuracy than average firing rate (rate-place code), and so is more likely to influence model performance. Although the DNN approach holds great promise, the implementations so far have not been tested on the most critical pitch conditions (e.g., on spectrally resolved harmonics outside the range of phase locking) and have remained limited to F0s between 100 and 300 Hz. Although this range spans the average F0s of male (~100 Hz) and female (~200 Hz) human voices, it represents less than 2 of the more than 7-octave range of musical pitch, meaning that the majority of our pitch range remains to be explored with this approach.

4. Remaining questions and clinical implications

4.1. Why is timing extracted from the temporal envelope but not TFS?

If the auditory system can extract pitch from the temporal envelope, why not from TFS? A speculative reason is based on the processing that occurs in the brainstem and midbrain. Temporal-envelope modulation produces amplitude fluctuations that are broadly in phase across the entire stimulated length of the basilar membrane. Many types of neurons in the brainstem and beyond are known to integrate information from across auditory nerve fibers with a range of characteristic frequencies (CFs). By receiving input from auditory-nerve fibers that are synchronized with the period of the temporal envelope and are in phase with each other, the responses from such neurons can be more highly synchronized to the waveform (in terms of vector strength) than those in the auditory nerve itself (Joris et al., 2004). In the case of responses to the TFS of a sinusoidal component (a pure tone or a spectrally resolved harmonic), however, the rapid phase transition of the traveling wave around CF (Shamma and Klein, 2000) means that even auditory-nerve fibers with similar CFs are unlikely to be in phase with each other. The outcome could therefore be desynchronized input to brainstem units, and an inability to transmit the phase-locked responses to TFS beyond the auditory nerve. Note that some brainstem units, such as the globular and spherical bushy cells in the cochlear nucleus, do show highly phase-locked responses to low-frequency CF tones (Joris et al., 1994). However, these are only more synchronized than the auditory-nerve fibers below about 1 kHz, and drop off rapidly thereafter, a pattern that reflects behavioral sensitivity to binaural timing differences but not to monaural or diotic pitch. One possibility, therefore, is that sensitivity to temporal-envelope periodicity is based on brainstem and midbrain sensitivity and tuning to amplitude modulation (Joris et al., 2004). Perceptual sensitivity to amplitude modulation deteriorates above about 150 Hz (Kohlrausch et al., 2000), also with an upper limit of around 1 kHz (Viemeister, 1979). In contrast, information regarding the frequency components themselves may be based solely on place or tonotopic information. Therefore, the difference between the strong pitch based on low-number spectrally resolved components and high-numbered unresolved components may reflect a difference between rate-place coding of the former and temporal (phase-locked) coding of the latter.

4.2. Implications for cochlear implants

Cochlear implants are the world's most successful sensorineural prosthetic device, providing hearing to over one million people worldwide (Zeng, 2022). Despite their success, cochlear implants do not provide "normal" hearing to their users, and one major shortcoming involves the transmission of pitch. Pitch has been defined in multiple ways for cochlear implants. "Place pitch" refers to the sensation reported by cochlear-implant users as the place of stimulation is changed by altering which electrode is activated (Nelson et al., 1995); "rate pitch" or "temporal pitch" is the sensation reported by cochlear-implant users when the electrical pulse rate is changed (Pijl and Schwarz, 1995; Zeng, 2002). For pure tones in acoustic hearing, place and rate covary, but for complex tones, they can be dissociated and are typically referred to as pitch (corresponding to the F0) and brightness (an aspect of timbre related to the spectral centroid of the stimulus). The rate pitch experienced by cochlear-implant users is most akin to the temporal-envelope pitch experienced by normal-hearing listeners in the absence of spectrally resolved harmonics (Carlyon et al., 2010; Kreft et al., 2010), whereas cochlear-implant place pitch seems to behave more like brightness in normal-hearing listeners than pitch (Allen and Oxenham, 2014).

The type of pitch that is not available to cochlear-implant users with current devices is the one that normal-hearing listeners rely on: the salient pitch provided by low-numbered, spectrally resolved harmonics. Some efforts have been made to provide this information to cochlear-implant users *via* TFS cues, but while there may be benefits to binaural hearing (Francart et al., 2015), there is no evidence yet to suggest that pitch salience or accuracy comparable to that in normal-hearing listeners can be induced *via* temporal coding (Landsberger, 2008; Kreft et al., 2010; Magnusson, 2011). The failure to induce accurate pitch perception *via* electrical pulse timing is expected, if we accept that pitch is typically conveyed *via* place cues, and that timing cues can only elicit the relatively crude pitch normally produced by temporal-envelope cues. Would it be possible to provide cochlear-implant users with sufficiently accurate place cues to recreate the kind of pitch elicited *via* spectrally resolved harmonics? Recent studies using acoustic vocoder simulations suggest that this will not be possible with current technology (Mehta and Oxenham, 2017; Mehta et al., 2020). These studies suggest that the spectral resolution required to transmit resolved harmonics requires the equivalent of filter slopes that exceed 100 dB/octave. Current cochlear implants have resolution that seems equivalent to slopes somewhere between 6 and 12 dB/octave (Oxenham and Kreft, 2014), perhaps extending to 24 dB/octave when using focused stimulation techniques (DeVries and Arenberg, 2018; Feng and Oxenham, 2018). Thus, the unfortunate conclusion is that the limited spectral resolution of cochlear implants is unlikely to provide the information necessary to elicit a salient pitch. This conclusion provides an additional impetus for the search for new technologies, based perhaps on neurotrophic agents to decrease the distance between electrodes and neurons, a different stimulation site, such as the auditory nerve, or a different stimulation strategy based, for instance, on optogenetic technology (Oxenham, 2018).

Data availability statement

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

Author contributions

AO conceived and carried out the work and approved the submitted version.

Funding

This work was supported by the National Institutes of Health (grant R01 DC005216).

Acknowledgments

Kelly Whiteford and the reviewer provided helpful comments on an earlier version of this manuscript.

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.

Publisher's note

References

Allen, E. J., and Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *J. Acoust. Soc. Am.* 135, 1371–1379. doi: 10.1121/1.4863269

Attneave, F., and Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *Am. J. Psychol.* 84, 147–166. doi: 10.2307/1421351

Bernstein, J. G., and Oxenham, A. J. (2005). An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *J. Acoust. Soc. Am.* 117, 3816–3831. doi: 10.1121/1.1904268

Bernstein, J. G., and Oxenham, A. J. (2006a). The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *J. Acoust. Soc. Am.* 120, 3916–3928. doi: 10.1121/1.2372451

Bernstein, J. G., and Oxenham, A. J. (2006b). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *J. Acoust. Soc. Am.* 120, 3929–3945. doi: 10.1121/1.2372452

Bruce, I. C., Erfani, Y., and Zilany, M. S. A. (2018). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Implications of limited neurotransmitter release sites. *Hear. Res.* 360, 40–54. doi: 10.1016/j.heares.2017.12.016

Burns, E. M., and Viemeister, N. F. (1976). Nonspectral pitch. *J. Acoust. Soc. Am.* 60, 863–869. doi: 10.1121/1.381166

Burns, E. M., and Viemeister, N. F. (1981). Played again SAM: Further observations on the pitch of amplitude-modulated noise. *J. Acoust. Soc. Am.* 70, 1655–1660. doi: 10.1121/1.387220

Cariani, P. A., and Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698–1716. doi: 10.1152/jn.1996.76.3.1698

Carlyon, R. P. (1996). Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *J. Acoust. Soc. Am.* 99, 517–524. doi: 10.1121/1.414510

Carlyon, R. P. (1998). Comments on "A unitary model of pitch perception". *J. Acoust. Soc. Am.* 104, 1118–1121. doi: 10.1121/1.423319

Carlyon, R. P., Deeks, J. M., and Mckay, C. M. (2010). The upper limit of temporal pitch for cochlear-implant listeners: Stimulus duration, conditioner pulses, and the number of electrodes stimulated. *J. Acoust. Soc. Am.* 127, 1469–1478. doi: 10.1121/1.3291981

Carney, L. H., Li, T., and Mcdonough, J. M. (2015). Speech coding in the brain: Representation of vowel formants by midbrain neurons tuned to sound fluctuations. *eNeuro* 2, 1–12. doi: 10.1523/ENEURO.0004-15.2015

Cedolin, L., and Delgutte, B. (2010). Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. *J. Neurosci.* 30, 12712–12724. doi: 10.1523/JNEUROSCI.6365-09.2010

Cohen, M. A., Grossberg, S., and Wyse, L. L. (1995). A spectral network model of pitch perception. *J. Acoust. Soc. Am.* 98, 862–879. doi: 10.1121/1.413512

de Cheveigné, A., and Pressnitzer, D. (2006). The case of the missing delay lines: Synthetic delays obtained by cross-channel phase interaction. *J. Acoust. Soc. Am.* 119, 3908–3918. doi: 10.1121/1.2195291

DeVries, L., and Arenberg, J. G. (2018). Current focusing to reduce channel interaction for distant electrodes in cochlear implant programs. *Trends Hear.* 22:2331216518813811. doi: 10.1177/2331216518813811

Feng, L., and Oxenham, A. J. (2018). Auditory enhancement and the role of spectral resolution in normal-hearing listeners and cochlear-implant users. *J. Acoust. Soc. Am.* 144:552. doi: 10.1121/1.5048414

Fishman, Y. I., Micheyl, C., and Steinschneider, M. (2013). Neural representation of harmonic complex tones in primary auditory cortex of the awake monkey. *J. Neurosci.* 33, 10312–10323. doi: 10.1523/JNEUROSCI.0020-13.2013

Francart, T., Lenssen, A., Buchner, A., Lenarz, T., and Wouters, J. (2015). Effect of channel envelope synchrony on interaural time difference sensitivity in bilateral cochlear implant listeners. *Ear Hear.* 36, e199–e206. doi: 10.1097/AUD.0000000000000152

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

Graves, J. E., and Oxenham, A. J. (2019). Pitch discrimination with mixtures of three concurrent harmonic complexes. *J. Acoust. Soc. Am.* 145:2072. doi: 10.1121/1.5096639

Guest, D. R., and Oxenham, A. J. (2022). Human discrimination and modeling of high-frequency complex tones shed light on the neural codes for pitch. *PLoS Comput. Biol.* 18:e1009889. doi: 10.1371/journal.pcbi.1009889

Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001). Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Comput.* 13, 2273–2316. doi: 10.1162/089976601750541804

Helmholtz, H. L. F. (1885/1954). *On the sensations of tone*. New York, NY: Dover.

Houtsma, A. J. M., and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87, 304–310. doi: 10.1121/1.399297

Joris, P. X., Carney, L. H., Smith, P. H., and Yin, T. C. (1994). Enhancement of neural synchronization in the anteroventral cochlear nucleus. I. Responses to tones at the characteristic frequency. *J. Neurophysiol.* 71, 1022–1036. doi: 10.1152/jn.1994.71.3.1022

Joris, P. X., Schreiner, C. E., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84, 541–577. doi: 10.1152/physrev.00029.2003

Kiang, N. Y., Sachs, M. B., and Peake, W. T. (1967). Shapes of tuning curves for single auditory-nerve fibers. *J. Acoust. Soc. Am.* 42, 1341–1342. doi: 10.1121/1.1910723

Kohlrausch, A., Fassel, R., and Dau, T. (2000). The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J. Acoust. Soc. Am.* 108, 723–734. doi: 10.1121/1.429605

Kreft, H. A., Oxenham, A. J., and Nelson, D. A. (2010). Modulation rate discrimination using half-wave rectified and sinusoidally amplitude modulated stimuli in cochlear-implant users. *J. Acoust. Soc. Am.* 127, 656–659. doi: 10.1121/1.3282947

Landsberger, D. M. (2008). Effects of modulation wave shape on modulation frequency discrimination with electrical hearing. *J. Acoust. Soc. Am.* 124, EL21–EL27. doi: 10.1121/1.2947624

Lau, B. K., Mehta, A. H., and Oxenham, A. J. (2017). Superoptimal perceptual integration suggests a place-based representation of pitch at high frequencies. *J. Neurosci.* 37, 9013–9021. doi: 10.1523/JNEUROSCI.1507-17.2017

Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. *J. Acoust. Soc. Am.* 63, 442–455. doi: 10.1121/1.381736

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia* 7, 128–133. doi: 10.1007/BF02156143

Magnusson, L. (2011). Comparison of the fine structure processing (FSP) strategy and the CIS strategy used in the MED-EL cochlear implant system: Speech intelligibility and music sound quality. *Int. J. Audiol.* 50, 279–287. doi: 10.3109/14992027.2010.537378

Meddis, R., and O'Mard, L. (1997). A unitary model of pitch perception. *J. Acoust. Soc. Am.* 102, 1811–1820. doi: 10.1121/1.420088

Mehta, A. H., and Oxenham, A. J. (2017). Vocoder simulations explain complex pitch perception limitations experienced by cochlear implant users. *J. Assoc. Res. Otolaryngol.* 18, 789–802. doi: 10.1007/s10162-017-0632-x

Mehta, A. H., and Oxenham, A. J. (2020). Effect of lowest harmonic rank on fundamental-frequency difference limens varies with fundamental frequency. *J. Acoust. Soc. Am.* 147:2314. doi: 10.1121/10.0001092

Mehta, A. H., and Oxenham, A. J. (2022). Role of perceptual integration in pitch discrimination at high frequencies. *JASA Express Lett.* 2:084402. doi: 10.1121/10.0013429

Mehta, A. H., Lu, H., and Oxenham, A. J. (2020). The perception of multiple simultaneous pitches as a function of number of spectral channels and spectral spread in a noise-excited envelope vocoder. *J. Assoc. Res. Otolaryngol.* 21, 61–72. doi: 10.1007/s10162-019-00738-y

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219, 36–47. doi: 10.1016/j.heares.2006.05.004

Micheyl, C., Keebler, M. V., and Oxenham, A. J. (2010). Pitch perception for mixtures of spectrally overlapping harmonic complex tones. *J. Acoust. Soc. Am.* 128, 257–269. doi: 10.1121/1.3372751

Micheyl, C., Schrater, P. R., and Oxenham, A. J. (2013). Auditory frequency and intensity discrimination explained using a cortical population rate code. *PLoS Comput. Biol.* 9:e1003336. doi: 10.1371/journal.pcbi.1003336

Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *J. Acoust. Soc. Am.* 54, 610–619. doi: 10.1121/1.1913640

Moore, B. C. J., and Ernst, S. M. (2012). Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code. *J. Acoust. Soc. Am.* 132, 1542–1547. doi: 10.1121/1.4739444

Moore, B. C. J., and Sek, A. (1996). Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *J. Acoust. Soc. Am.* 100, 2320–2331. doi: 10.1121/1.417941

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (1995). Electrode ranking of "place pitch" and speech recognition in electrical hearing. *J. Acoust. Soc. Am.* 98, 1987–1999. doi: 10.1121/1.413317

Ohm, G. S. (1843). Über die definition des tones, nebst daran geknüpfter theorie der sirene und ähnlicher tonbildender vorrichtungen [On the definition of tones, including a theory of sirens and similar tone-producing apparatuses]. *Ann. Phys. Chem.* 59, 513–565. doi: 10.1002/andp.18431350802

Oxenham, A. J. (2012). Pitch perception. *J. Neurosci.* 32, 13335–13338. doi: 10.1523/JNEUROSCI.3815-12.2012

Oxenham, A. J. (2018). How we hear: The perception and neural coding of sound. *Annu. Rev. Psychol.* 69, 27–50. doi: 10.1146/annurev-psych-122216-011635

Oxenham, A. J., and Kreft, H. A. (2014). Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends Hear.* 18:2331216514553783. doi: 10.1177/2331216514553783

Oxenham, A. J., Bernstein, J. G. W., and Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1421–1425. doi: 10.1073/pnas.0306958101

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., and Santurette, S. (2011). Pitch perception beyond the traditional existence region of pitch. *Proc. Natl. Acad. Sci. U.S.A.* 108, 7629–7634. doi: 10.1073/pnas.1015291108

Pijl, S., and Schwarz, D. W. (1995). Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. *J. Acoust. Soc. Am.* 98, 886–895. doi: 10.1121/1.413514

Plack, C. J., Oxenham, A. J., Fay, R., and Popper, A. N. (eds) (2005). *Pitch: Neural coding and perception*. New York, NY: Springer Verlag. doi: 10.1007/0-387-28958-5

Qin, M. K., and Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear.* 26, 451–460. doi: 10.1097/01.aud.0000179689.79868.06

Saddler, M. R., Gonzalez, R., and Mcdermott, J. H. (2021). Deep neural network models reveal interplay of peripheral coding and stimulus statistics in pitch perception. *Nat. Commun.* 12:7278. doi: 10.1038/s41467-021-27366-6

Schwartz, D. A., and Purves, D. (2004). Pitch is determined by naturally occurring periodic sounds. *Hear. Res.* 194, 31–46. doi: 10.1016/j.heares.2004.01.019

Seebeck, A. (1841). Beobachtungen über einige bedingungen der entstehung von tönen [Observations on some conditions for the formation of tones]. *Ann. Phys. Chem.* 53, 417–436. doi: 10.1002/andp.18411290702

Shamma, S., and Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *J. Acoust. Soc. Am.* 107, 2631–2644. doi: 10.1121/1.428649

Shera, C. A., Guinan, J. J., and Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc. Natl. Acad. Sci. U.S.A.* 99, 3318–3323. doi: 10.1073/pnas.032675099

Shofner, W. P., and Chaney, M. (2013). Processing pitch in a nonhuman mammal (*Chinchilla laniger*). *J. Comp. Psychol.* 127, 142–153. doi: 10.1037/a0029734

Siebert, W. M. (1970). Frequency discrimination in the auditory system: Place or periodicity mechanisms. *Proc. IEEE* 58, 723–730. doi: 10.1109/PROC.1970.7727

Sumner, C. J., Wells, T. T., Bergevin, C., Sollini, J., Kreft, H. A., Palmer, A. R., et al. (2018). Mammalian behavior and physiology converge to confirm sharper cochlear tuning in humans. *Proc. Natl. Acad. Sci. U.S.A.* 115, 11322–11326. doi: 10.1073/pnas.1810766115

Terhardt, E. (1974). Pitch, consonance, and harmony. *J. Acoust. Soc. Am.* 55, 1061–1069. doi: 10.1121/1.1914648

Turner, R. S. (1977). The ohm-seebeck dispute, Hermann von Helmholtz, and the origins of physiological acoustics. *Br. J. Hist. Sci.* 10, 1–24. doi: 10.1017/S0007087400015089

Verschooten, E., Desloovere, C., and Joris, P. X. (2018). High-resolution frequency tuning but not temporal coding in the human cochlea. *PLoS Biol.* 16:e2005164. doi: 10.1371/journal.pbio.2005164

Viemeister, N. F. (1979). Temporal modulation transfer functions based on modulation thresholds. *J. Acoust. Soc. Am.* 66, 1364–1380. doi: 10.1121/1.383531

Walker, K. M., Gonzalez, R., Kang, J. Z., Mcdermott, J. H., and King, A. J. (2019). Across-species differences in pitch perception are consistent with differences in cochlear filtering. *Elife* 8:e41626. doi: 10.7554/eLife.41626

Whiteford, K. L., Kreft, H. A., and Oxenham, A. J. (2020). The role of cochlear place coding in the perception of frequency modulation. *Elife* 9:e58468. doi: 10.7554/eLife.58468

Wightman, F. L. (1973). The pattern-transformation model of pitch. *J. Acoust. Soc. Am.* 54, 407–416. doi: 10.1121/1.1913592

Winter, I. M., Robertson, D., and Yates, G. K. (1990). Diversity of characteristic frequency rate-intensity functions in guinea pig auditory nerve fibres. *Hear. Res.* 45, 203–220. doi: 10.1016/0378-5955(90)90120-E

Zeng, F. G. (2002). Temporal pitch in electric hearing. *Hear. Res.* 174, 101–106. doi: 10.1016/S0378-5955(02)00644-5

Zeng, F. G. (2022). Celebrating the one millionth cochlear implant. *JASA Express Lett.* 2:077201. doi: 10.1121/10.0012825

frontiers | Frontiers in Neuroscience

TYPE Perspective
PUBLISHED 04 January 2023
DOI 10.3389/fnins.2022.1075288

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Deborah Levy,
University of California, San Francisco, United States
Christopher I. Petkov,
Newcastle University, United Kingdom

*CORRESPONDENCE
Pascal Belin
✉ pascal.belin@univ-amu.fr

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

RECEIVED 20 October 2022
ACCEPTED 06 December 2022
PUBLISHED 04 January 2023

CITATION
Trapeau R, Thoret E and Belin P (2023) The Temporal Voice Areas are not "just" Speech Areas.
Front. Neurosci. 16:1075288.
doi: 10.3389/fnins.2022.1075288

The Temporal Voice Areas are not "just" Speech Areas

Régis Trapeau[1], Etienne Thoret[2,3] and Pascal Belin[1,4]*

[1]La Timone Neuroscience Institute, CNRS and Aix-Marseille University, UMR 7289, Marseille, France, [2]Aix-Marseille University, CNRS, UMR7061 PRISM, UMR7020 LIS, Marseille, France, [3]Institute of Language, Communication and the Brain (ILCB), Marseille, France, [4]Department of Psychology, Montreal University, Montreal, QC, Canada

The Temporal Voice Areas (TVAs) respond more strongly to speech sounds than to non-speech vocal sounds, but does this make them Temporal "Speech" Areas? We provide a perspective on this issue by combining univariate, multivariate, and representational similarity analyses of fMRI activations to a balanced set of speech and non-speech vocal sounds. We find that while speech sounds activate the TVAs more than non-speech vocal sounds, which is likely related to their larger temporal modulations in syllabic rate, they do not appear to activate additional areas nor are they segregated from the non-speech vocal sounds when their higher activation is controlled. It seems safe, then, to continue calling these regions the Temporal Voice Areas.

KEYWORDS

voice, speech, Temporal Voice Areas, functional MRI, humans, decoding, representational similarity analysis

1. Introduction

It is a well-replicated finding that the Temporal Voice Areas (TVAs) of secondary auditory cortex are significantly more active in response to human voices compared to non-vocal environmental sounds (Belin et al., 2000; Kriegstein and Giraud, 2004; Andics et al., 2010; Frhholz and Grandjean, 2013; Pernet et al., 2015).

Neuroimaging voice localizers typically include speech in the human voice category of stimuli, as well as vocalizations with minimal linguistic content (here after, non-speech vocal sounds) such as coughs, laughs, or simple sustained vowels. TVA responses to non-speech vocal sounds are typically smaller than speech sounds (Belin et al., 2002; Fecteau et al., 2004; Bodin et al.'s, 2021), and in some cases not significantly stronger than control sounds (Belin et al., 2002). This has led some researchers to doubt that the TVAs are sensitive to vocal sounds, in general, and suggest that they are in fact Speech Areas, that is, responsive to the phonemic and/or semantic content of the input signal [e.g., component 5 in Norman-Haignere et al. (2015) study].

Yet, other results indicate that even non-speech vocal sounds induce greater TVA activity than control sounds (Bodin et al.'s, 2021) or lead to above chance classification into vocal/non-vocal categories (Rupp et al., 2022), suggesting a selectivity to this category of sounds in the TVAs.

Here, we provide a perspective on this issue by performing additional analyses of a published dataset (Bodin et al.'s, 2021), in which the same number ($n = 12$) of individual speech and non-speech vocal sounds were used along with 24 non-vocal sounds.

Visualization using symmetrical colormaps (−max < *t*-value < max; allowing easy visual comparison of activation location differences between contrasts irrespective of significance threshold) of whole brain fixed-effects group t-maps of speech sounds vs. non-vocal sounds contrast (Figure 1A) and non-speech vocal sounds vs. non-vocal sounds contrast (Figure 1B) reveals topographically similar patterns of activation in both contrasts, suggesting that TVA activity is not limited to speech sounds. T-maps of both contrasts closely resemble those obtained by contrasting human voices

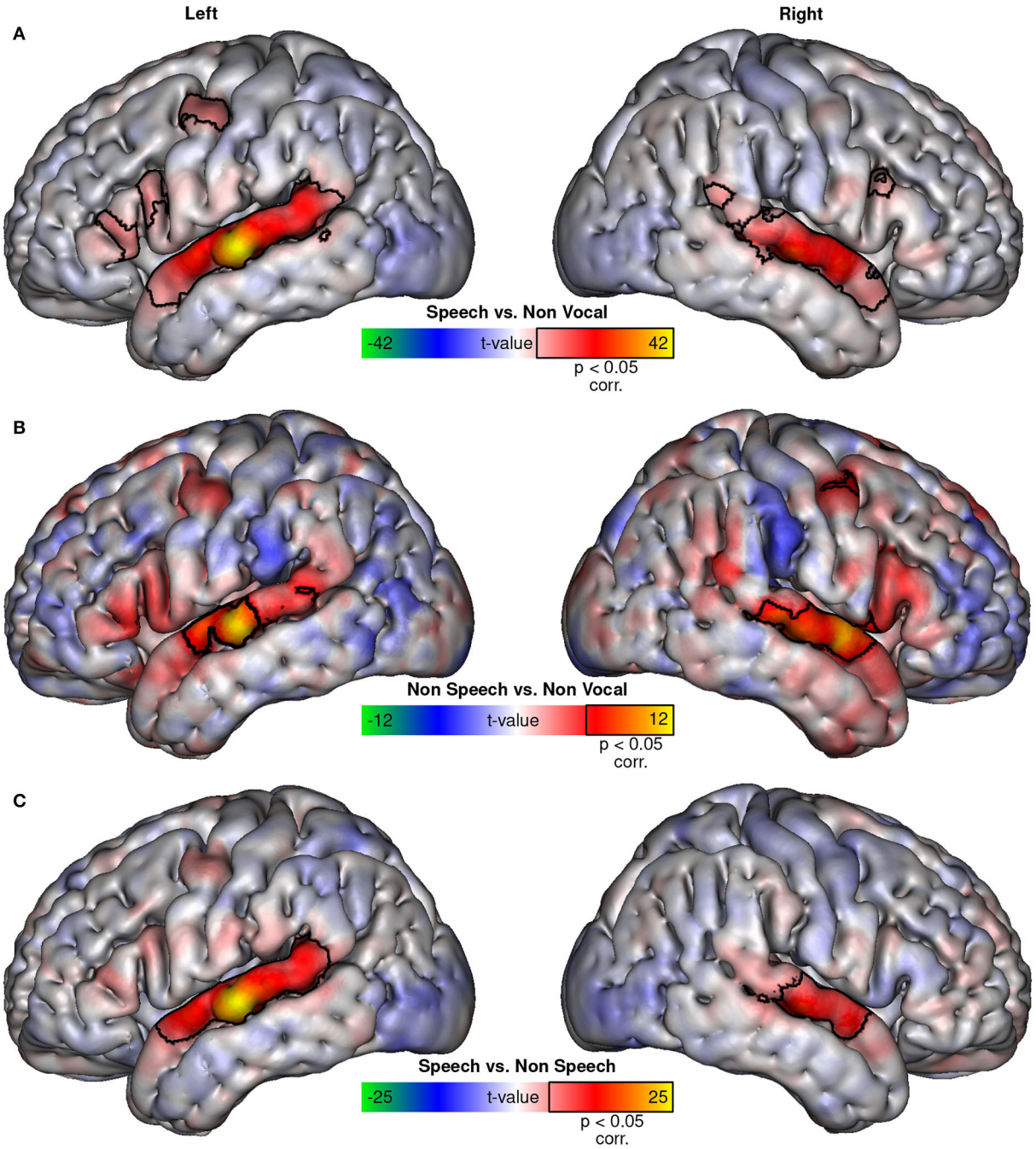


FIGURE 1
BOLD activations. Fixed-effects t-maps projected on the MNI152 surface of the speech vs. non-vocal contrast **(A)**, non-speech vs. non-vocal contrast **(B)**, and the speech vs. non-speech contrast **(C)**. Colormaps were adjusted to be symmetrical, with limits corresponding to the maximal *t*-value in each contrast. Areas with significant ($p < 0.05$, corrected) activation to each contrast are outlined in black.

vs. other types of sounds [compared with figure 1G from Bodin et al.'s (2021) study]. There is no clear visual evidence for supplementary regions recruited by speech stimuli, and both contrasts share the same maximum of activation in the left superior temporal gyrus. The main difference between the two contrasts is the higher general level of activation when using speech instead of non-speech vocal sounds. The speech vs. non-speech vocal stimuli contrast (Figure 1C) confirms this observation, as well as the apparent absence of additional regions activated by speech.

The larger general activation elicited by speech compared to non-speech vocal sounds might imply that speech sounds have a special status in the TVAs. To further investigate the role of speech and non-speech vocal sounds in the TVAs, we examined how a voice/non-voice decoder based on TVA activation performs for speech and non-speech vocal sounds, even when controlling for activation level differences between speech and non-speech. We also examined whether the representational geometry in the TVAs groups together speech and non-speech relative to non-vocal sounds.

2. Materials and methods

This analysis was performed on data collected in a previous study, which was designed for comparative neuroimaging between humans and nun-human primates (explaining the small sample size), but allowed distinct analyses of the activity evoked by speech and non-speech vocal sounds (Bodin et al.'s, 2021). Please refer to that study for a detailed description of materials and methods. The following sections present methodology that is specific to the present analysis.

2.1. Participants

Five native French human speakers were scanned [one man (author RT) and four women; 23–38 years of age]. Participants gave written informed consent and were paid for their participation.

2.2. Auditory stimuli

The analysis was performed on fMRI events corresponding to a subset of the stimulus set used in Bodin et al.'s (2021) study. Two main categories of sounds were used: human voices and non-vocal sounds, each containing 24 stimuli, for a total of 48 sound stimuli. Each main category was divided into two subcategories of 12 stimuli, forming four subcategories in total (cf. Supplementary Table 1). Human voices contained both speech [sentence segments from the set of stimuli used in Moerel et al.'s (2012) study, $n = 12$] and non-speech vocal sounds [vocal affect bursts selected from the Montreal Affective Voices dataset (Belin et al., 2008), $n = 12$].

Non-vocal sounds included both natural and artificial sounds from previous studies from our group (Belin et al., 2000; Capilla et al., 2013) or kindly provided by Petkov et al. (2008) and Moerel et al.'s (2012). Supplementary Figure 1 shows spectrograms and waveforms of the speech and non-speech vocal stimuli.

2.3. fMRI protocol

Detailed description of the fMRI protocol can be found in Bodin et al.'s (2021) study. In brief, functional scanning was done using an event-related paradigm with clustered-sparse acquisitions on a 3-Tesla MRI scanner (Prisma, Siemens Healthcare), equipped with a 64-channel matrix head-coil. To avoid interference between sound stimulation and scanner noise, the scanner stopped acquisitions such that three repetitions of a 500-ms stimulus (inter-stimulus interval of 250 ms) were played on a silent background. Then, seven whole-head functional volumes were acquired (TR = 0.945 s). Two functional runs, each containing one repetition of each stimulus, were acquired for each participant. Participants were instructed to stay still in the scanner while passively listening to the stimuli.

2.4. fMRI general linear modeling

General linear model estimates of responses to speech stimuli vs. non-vocal sounds, to non-speech vocal stimuli vs. non-vocal sounds, and to speech stimuli vs. non-speech vocal sounds were computed using fMRISTAT (Worsley et al., 2002).

2.5. Decoding

We tested whether support vector classification with a linear kernel [SVC: Chang and Lin (2011)] was able to predict, from beta values in primary auditory cortex (A1) and TVAs, whether fMRI events corresponded to the presentation of vocal or non-vocal sounds. We first tried this decoding using only speech vocal sounds and then using only non-speech vocal sounds. To have a balanced frequency in each category tested ($n = 12$), only half of the non-vocal sounds were used during classification. As the dataset consisted of sessions containing two functional runs during which a repetition of each stimulus was presented, we used a two-fold cross-validation, with each run serving successively as train and test sets. For each participant, the classifier was first trained on data from one functional run and tested on the other, and the other way around in a second fold. The reported classification accuracy is the average of the scores obtained in two-fold cross-validation. Above significance

threshold in classification accuracy was determined by building a bootstrapped distribution of classification scores obtained on 100,000 iterations of two-fold dummy classification tests with random labels. Comparisons between different classification results were tested using Wilcoxon signed-rank tests.

2.6. Representational similarity analysis

Representations of dissimilarities within the stimulus set in A1 and TVAs were assessed using the representational similarity analysis (RSA) framework (Kriegeskorte et al., 2008; Nili et al., 2014). Representational dissimilarity matrices (RDMs) capturing the pattern of dissimilarities in fMRI responses, and generated by computing the Euclidean distance between stimuli in multi-voxel activity space, were compared with three binary categorical models: (1) a "human" model in which human voices are categorized separately from non-vocal sounds, with an equal contribution of speech and non-speech vocal stimuli; (2) a "speech" model categorizing speech apart from all other sounds (i.e., non-vocal and non-speech vocal stimuli); and (3) a "non-speech" model categorizing non-speech human voices apart from other sounds (i.e., non-vocal and speech stimuli).

We also compared brain RDMs with an acoustical RDM reflecting the pattern of differences between the modulation power spectra [Thoret et al. (2016); MPS: quantifies amplitude and frequency modulations present in a sound] of the 48 stimuli (see Supplementary Figure 2).

Planned comparisons were performed using two-sample bootstrapped *t*-tests (100,000 iterations, one-tailed) that compared the within vs. between portions of the brain and acoustical RDMs, as shown in Supplementary Figure 3.

2.7. Regions of interest

RSA and SVC were performed in two regions of interest (ROI): primary auditory cortex (A1) and Temporal Voice Areas (TVAs) in each hemisphere.

In each participant and hemisphere, the center of the A1 ROI was defined as the maximum value of the probabilistic map (non-linearly registered to each participant functional space) of Heschl's gyri provided with the MNI152 template (Penhune et al., 1996). The 57 voxels in the functional space that were the closest to this point and above 50% in the probabilistic maps constituted the A1 ROI.

In each participant and hemisphere, the TVAs' ROI was the conjunction of three TVAs (posterior, middle, and anterior). TVA locations vary from one individual to another and were therefore located functionally. The center of each TVA region corresponded to the local maximum of the *human voice > all other sounds* t-map [computed using both speech and non-speech events, see Bodin et al.'s (2021)], whose coordinates were the closest to the corresponding TVA reported in the study of Aglieri et al. (2018). The 19 voxels in the functional space that were the closest to this point and above significance threshold in *human voice > all other sounds* t-map constituted a TVA ROI. The TVAs' ROI for one hemisphere was the conjunction of the three TVA ROIs of 19 voxels, forming a ROI of 57 voxels.

2.8. Standardization

To assess the contribution of either categorical or topographical differences in stimulus activation, activity patterns of each ROI (*RSA*: 48 stimuli × 57 voxels; *SVC*: 96 events × 57 voxels) were standardized using two methods before running RSA and SVC: a standardization *along stimuli*, where *z*-scores were computed for each voxel along the stimulus (RSA) or event (SVC) dimension [which is the default standardization in machine learning packages; Pedregosa et al. (2011)], and a standardization *along voxels*, where *z*-scores were computed for each stimulus (or event) along the voxel dimension (see Supplementary Figure 4). For RSA, standardization was performed on activity patterns before computing RDMs. For SVC, standardization was performed on all events (both runs) before splitting data in train-test sets.

3. Results

3.1. Decoding stimulus categories

Decoding results are shown in Supplementary Figure 5. For both standardization methods, when attempting to classify fMRI events in speech or non-vocal categories, the SVC performed poorly in A1 and well above significance level in TVAs (mean scores for standardization method along stimuli and along voxels, respectively. A1: $\overline{x} = 0.58$ and 0.57; TVAs: $\overline{x} = 0.89$ and 0.84). When using non-speech events instead of speech events, performance in A1 remained poor and performance in TVAs dropped to values close to significance level (A1: $\overline{x} = 0.56$ and 0.61; TVAs: $\overline{x} = 0.65$ and 0.64). The differences in SVC performance when using speech or non-speech vocal stimuli were not significant for both A1 and TVAs. However, in the TVAs, classification accuracy was higher for speech than for non-speech vocal sounds for all the participants, suggesting that this difference may become significant with a larger sample of participants. The differences in SVC performance between standardization methods were not significant for both A1 and TVAs.

3.2. Representational similarity analysis

The visual representation of the pattern of Spearman correlations among brain RDMs

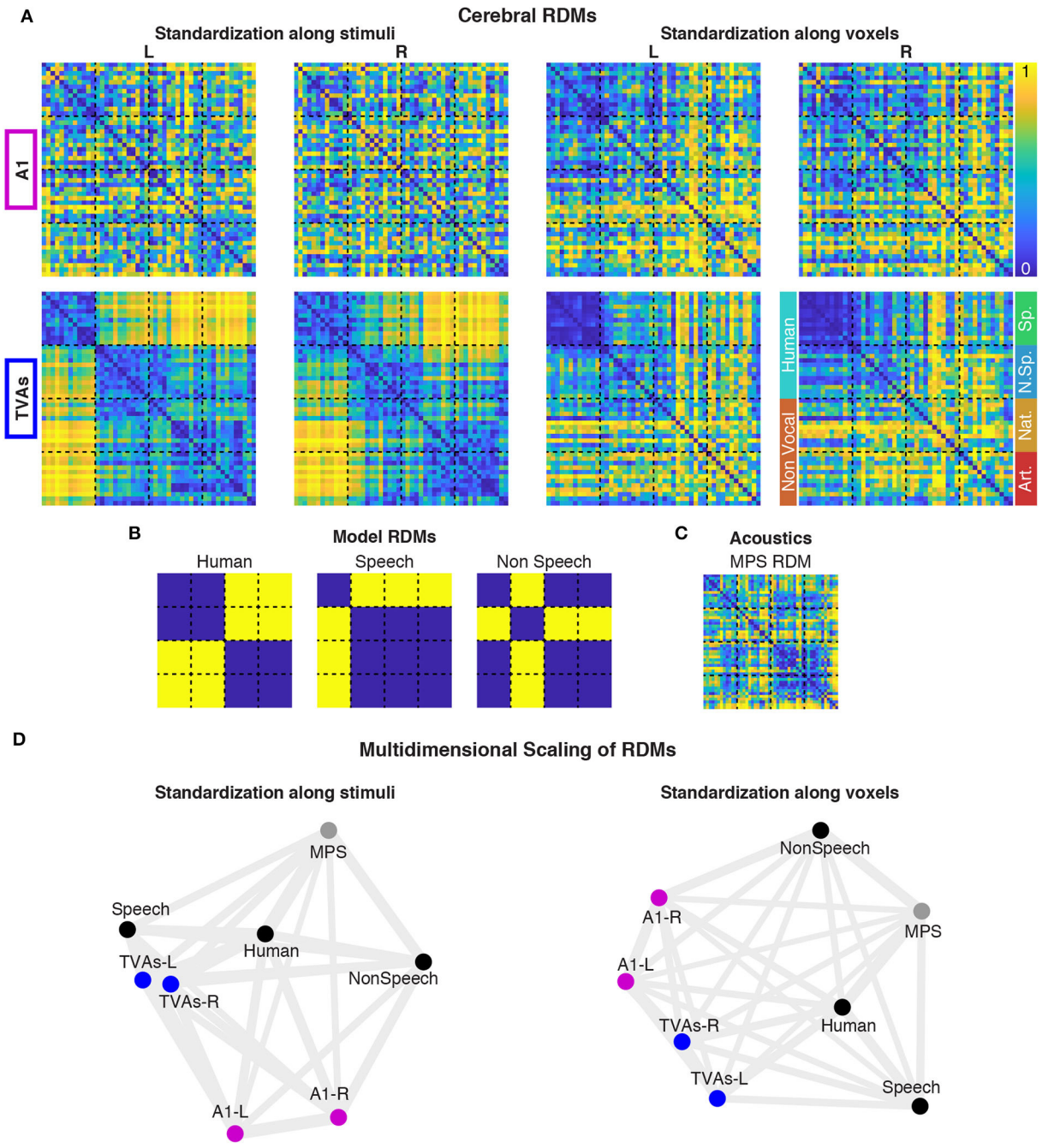


FIGURE 2
Representational similarity analysis (RSA) in A1 and the TVAs. Cerebral RDMs showing percentile dissimilarities in pairwise fMRI response to the 48 stimuli, for both ROIs and standardization methods **(A)**. Portions of the RDMs corresponding to the main and sub-categories of stimuli are indicated next to the bottom right RDM. Cerebral RDMs were compared (Spearman correlations) with three categorical model RDMs **(B)** and one acoustical RDM **(C)**, for each standardization method. These comparisons are represented *via* multidimensional scaling **(D)**.

(Figure 2A), categorical models (Figure 2B), and acoustical RDM (Figure 2C) was performed *via* multidimensional scaling (MDS, Figure 2D) for both standardization methods.

Using standardization along stimuli, cerebral RDMs computed in the left and right TVAs cluster together close to the "speech" (especially for TVAs-L) and "human" categorical models, and separated from the "non-speech" categorical model,

the acoustical model, or the A1 brain RDMs. All three planned comparisons (see Supplementary Figure 3) were significant in the TVA RDMs (all *p*-values are below 0.01 after Bonferroni correction for 24 comparisons), while nothing was significant in A1.

Using standardization along voxels, TVA RDMs are less separated from A1 RDMs and closer to the "human" than the "speech" model. Only speech vs. non-speech test was significant in the TVAs, while nothing was significant in A1.

4. Perspective

The univariate analysis suggests that speech sounds activate the same set of regions as non-speech vocal sounds, simply more strongly. There is no clear evidence of additional areas activated specifically by speech sounds, as shown in Figure 1C, in which the contrasts of speech vs. non-speech vocal sounds show the same distribution of regions as the classical speech vs. non-vocal sounds contrast. This voice network appears to be recruited by both speech and non-speech vocal sounds, but more strongly by speech sounds.

The classification analysis confirms this notion: while classification accuracy for vocal vs. non-vocal sounds was larger on average for speech than for non-speech vocal sounds, the difference was not significant (likely due, though, to our small number of participants), and both were above chance level. Controlling for differences in activation level between stimuli with the standardization along voxels did not change this pattern (Supplementary Figure 5).

The Representational Similarity Analysis helped refine this picture. While A1 RDMs did not show any similarity with any of the categorical model RDMs (Figure 2B), the TVA RDMs were strongly associated, in both hemispheres, with the "speech" model, categorizing speech apart from all other sounds including non-speech voice. However, when controlling for stimulus activation levels *via* the voxelwize standardization (Supplementary Figure 4), the picture changed and the "human" model, grouping speech and non-speech vocal sounds together and apart from the non-vocal sounds, was the most closely associated to both left and right TVAs.

Overall, our analyses indicate that speech does not have a special status compared to non-speech vocal sounds in the TVAs, apart from the fact that they drive them to a higher activation level. This particular result needs to be further investigated in future studies, but is likely related to the more complex spectro-temporal structure of speech compared to non-speech vocal sounds (Supplementary Figure 1), with more pronounced temporal modulations around 4 Hz, close to the syllabic rate in English, (Supplementary Figure 2). Spectro-temporal complexity is indeed known to increase the strength of activation in non-primary auditory fields (Samson et al., 2011). It seems safe, then, to continue calling these regions the Temporal Voice Areas. Furthermore, using the more encompassing term of "voice" instead of "speech" to name these areas, opens up more questions and hypotheses for future studies using dedicated experimental designs with larger sample size, that will help to understand how spectro-temporal complexity, linguistic content, or attention to distinct voice features (von Kriegstein et al., 2003) modulate the cortical processing of voice.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: Zenodo: https://doi.org/10.5281/zenodo.5071389.

Ethics statement

The studies involving human participants were reviewed and approved by Ethical board of Institut de Neurosciences de la Timone. The participants provided their written informed consent to participate in this study.

Author contributions

PB and RT contributed to the conception and design of the study. RT and ET performed the statistical analysis. RT wrote the first draft of the manuscript. PB and ET wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was funded by the Fondation pour la Recherche Médicale (AJE201214 to PB), the Agence Nationale de la Recherche grants ANR-16-CE37-0011-01 (PRIMAVOICE), ANR-16-CONV-0002 (Institute for Language, Communication and the Brain) and ANR-11-LABX-0036 (Brain and Language Research Institute), the Excellence Initiative of Aix-Marseille University (A∗MIDEX), and the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation program (grant agreement no. 788240).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships

that could be construed as a potential conflict of interest.

Publisher's note

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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.1075288/full#supplementary-material

References

Aglieri, V., Chaminade, T., Takerkart, S., and Belin, P. (2018). Functional connectivity within the voice perception network and its behavioural relevance. *Neuroimage* 183, 356–365. doi: 10.1016/j.neuroimage.2018.08.011

Andics, A., McQueen, J. M., Petersson, K. M., Gál, V., Rudas, G., Vidnynszky, Z., et al. (2010). Neural mechanisms for voice recognition. *Neuroimage* 52, 1528–1540. doi: 10.1016/j.neuroimage.2010.05.048

Belin, P., Fillion-Bilodeau, S., and Gosselin, F. (2008). The Montreal affective voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav. Res. Methods* 40, 531–539. doi: 10.3758/BRM.40.2.531

Belin, P., Zatorre, R. J., and Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cogn. Brain Res.* 13, 17–26. doi: 10.1016/S0926-6410(01)00084-2

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., and Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature* 403, 309–312. doi: 10.1038/35002078

Bodin, C., Trapeau, R., Nazarian, B., Sein, J., Degiovanni, X., Baurberg, J., et al. (2021). Functionally homologous representation of vocalizations in the auditory cortex of humans and macaques. *Curr. Biol.* 31, 4839–4844. doi: 10.1016/j.cub.2021.08.043

Capilla, A., Belin, P., and Gross, J. (2013). The early spatio-temporal correlates and task independence of cerebral voice processing studied with MEG. *Cereb. Cortex* 23, 1388–1395. doi: 10.1093/cercor/bhs119

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* 2, 1–27. doi: 10.1145/1961189.1961199

Fecteau, S., Armony, J. L., Joanette, Y., and Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage* 23, 840–848. doi: 10.1016/j.neuroimage.2004.09.019

Frhholz, S., and Grandjean, D. (2013). Multiple subregions in superior temporal cortex are differentially sensitive to vocal expressions: a quantitative meta-analysis. *Neurosci. Biobehav. Rev.* 37, 24–35. doi: 10.1016/j.neubiorev.2012.11.002

Kriegeskorte, N., Mur, M., and Bandettini, P. A. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Front. Syst. Neurosci.* 2, 4. doi: 10.3389/neuro.06.004.2008

Kriegstein, K. V., and Giraud, A.-L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage* 22, 948–955. doi: 10.1016/j.neuroimage.2004.02.020

Moerel, M., De Martino, F., and Formisano, E. (2012). Processing of natural sounds in human auditory cortex: tonotopy, spectral tuning, and relation to voice sensitivity. *J. Neurosci.* 32, 14205–14216. doi: 10.1523/JNEUROSCI.1388-12.2012

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen-Wilson, W., Kriegeskorte, N., et al. (2014). A toolbox for representational similarity analysis. *PLoS Comput. Biol.* 10, e1003553. doi: 10.1371/journal.pcbi.1003553

Norman-Haignere, S., Kanwisher, N. G., and McDermott, J. H. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. doi: 10.1016/j.neuron.2015.11.035

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830. Available online at: http://jmlr.org/papers/v12/pedregosa11a.html

Penhune, V. B., Zatorre, R. J., MacDonald, J. D., and Evans, A. C. (1996). Interhemispheric anatomical differences in human primary auditory cortex: probabilistic mapping and volume measurement from magnetic resonance scans. *Cereb. Cortex* 6, 661–672. doi: 10.1093/cercor/6.5.661

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E., et al. (2015). The human voice areas: spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage* 119, 164–174. doi: 10.1016/j.neuroimage.2015.06.050

Petkov, C. I., Kayser, C., Steudel, T., Whittingstall, K., Augath, M., Logothetis, N. K., et al. (2008). A voice region in the monkey brain. *Nat. Neurosci.* 11, 367–374. doi: 10.1038/nn2043

Rupp, K., Hect, J. L., Remick, M., Ghuman, A., Chandrasekaran, B., Holt, L. L., et al. (2022). Neural responses in human superior temporal cortex support coding of voice representations. *PLoS Biol.* 20, e3001675. doi: 10.1371/journal.pbio.3001675

Samson, F., Zeffiro, T. A., Toussaint, A., and Belin, P. (2011). Stimulus complexity and categorical effects in human auditory cortex: an activation likelihood estimation meta-analysis. *Front. Psychol.* 1, 241. doi: 10.3389/fpsyg.2010.00241

Thoret, E., Depalle, P., and McAdams, S. (2016). Perceptually salient spectrotemporal modulations for recognition of sustained musical instruments. *J. Acoust. Soc. Am.* 140, EL478–EL483. doi: 10.1121/1.4971204

von Kriegstein, K., Eger, E., Kleinschmidt, A., and Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cogn. Brain Res.* 17, 48–55. doi: 10.1016/S0926-6410(03)00079-X

Worsley, K. J., Liao, C. H., Aston, J., Petre, V., Duncan, G. H., Morales, F., et al. (2002). A general statistical analysis for fMRI data. *Neuroimage* 15, 1–15. doi: 10.1006/nimg.2001.0933

frontiers | Frontiers in Neuroscience

TYPE Perspective
PUBLISHED 17 November 2022
DOI 10.3389/fnins.2022.1025763

OPEN ACCESS

EDITED BY
Claude Alain,
Rotman Research Institute, Canada

REVIEWED BY
Robert J. Zatorre,
McGill University, Canada
Megumi Takasago,
The University of Tokyo, Japan

*CORRESPONDENCE
Mari Tervaniemi
mari.tervaniemi@helsinki.fi

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

RECEIVED 23 August 2022
ACCEPTED 24 October 2022
PUBLISHED 17 November 2022

CITATION
Tervaniemi M (2022) Mismatch negativity–stimulation paradigms in past and in future. *Front. Neurosci.* 16:1025763. doi: 10.3389/fnins.2022.1025763

Mismatch negativity–stimulation paradigms in past and in future

Mari Tervaniemi[1,2]*

[1]Center of Excellence in Music, Mind, Body, and Brain, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland, [2]Cognitive Brain Research Unit, Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

Mismatch negativity (MMN) studies were initiated as part of a well-controlled experimental research tradition with the aim to identify some key principles of auditory processing and memory. During the past two decades, empirical paradigms have moved toward more ecologically valid ones while retaining rigid experimental control. In this paper, I will introduce this development of MMN stimulation paradigms starting from the paradigms used in basic science and then moving to paradigms that have been particularly relevant for studies on music learning and musical expertise. *Via* these historical and thematic perspectives, I wish to stimulate paradigm development further to meet the demands of naturalistic ecologically valid studies also when using MMN in the context of event-related potential technique that necessarily requires averaging across several stimulus presentations.

KEYWORDS

music, audition, musical learning, cognition, EEG, fMRI

Introduction

Thanks to versatile development in theoretical and methodological domains, auditory cognitive neuroscience has witnessed immense progress in past decades. When considering the development of methodology in the field, the main emphasis of scientific discussion is commonly given on methods in data acquisition and analyses. However, when considering the key questions of the field (specifically brain basis underlying neuroplasticity particularly in the domains of auditory learning, development, and aging), it is evident that validity of the stimulation paradigms is also of utmost importance. If these paradigms (that is, their sounds and the auditory soundscapes created by them) fail to address the neurocognitive processes of interest, the results are of minimal use in scientific or applied perspectives.

Notably, while a transition from well-controlled laboratory-based studies toward ecologically valid stimulation and recording paradigms has occurred in several related research traditions such as social and emotion neuroscience, it is questionable whether this is a feasible framework for studies in auditory cognitive neuroscience, particularly

when event-related potential (ERP) technique and the mismatch negativity (MMN) are of interest. This perspectives paper aims to offer a framework for observing the development of stimulation paradigms of the MMN field since the 1970s and to propose some future novel advancements. The discussion will be divided into two main sections, the first on basic MMN studies and the second on MMN studies in music-related contexts. After them, the brain generators of the MMN will be briefly illuminated. In the end of the paper, future directions of the MMN will be discussed.

Historical overview on mismatch negativity studies in oddball and multi-feature paradigms

When pioneering studies that launched mismatch negativity (MMN) were conducted (Näätänen et al., 1978), the fundamental question of the highest theoretical relevance was actually quite simple: is it possible to isolate a difference signal from the human brain? In other words, is there a neural signal that can differentiate acoustically different frequent standard and rare deviant sounds from each other? At that time, EEG recording and sound stimulation technologies were rather limited, and studies were conducted using sinusoidal sounds in an oddball paradigm. Once MMN had been established as a general index of the difference monitoring and sensory memory, empirical studies were conducted to indicate those sound parameters that are encoded in the sensory memory (e.g., Paavilainen et al., 1993 for duration, and Näätänen et al., 1987 for intensity). Further, parametric studies were conducted to indicate the accuracy of the sensory memory in this encoding (Sams et al., 1985 for frequency) and the correspondence between the MMN parameters and perceptual accuracy (Tiitinen et al., 1994 for frequency; Amenedo and Escera, 2000, for duration).

The next generation of studies aimed to avoid the co-occurrence of acoustical deviance and rareness of the deviant stimulus. This may sound simple, but it is less so since perceptual deviance is most often coupled by acoustical features. The solutions were diverse. First, Yabe et al. (1997) and Tervaniemi et al. (1994a) used *sound omission* as the deviant stimulus in isochronous sequences and in tone pairs, respectively. They both showed that MMN can be generated by a sound omission but only within a definite window enabling integration of incoming auditory information for some hundreds of milliseconds only. Second, Winkler et al. (1995) used a phenomenon called *missing fundamental* that denotes an "illusion" of the sound's fundamental frequency being identified even if this specific frequency is not present in the sound at all; it is computed in the brain based on the spectrum of the harmonic overtones. They showed that the MMN indeed reflects perceived fundamental frequency that can be created by several combinations of overtones, while a subset of the same overtones in a different constellation causes a perception of a different fundamental frequency and, subsequently, the MMN. Third, Tervaniemi et al. (1994b) utilized another auditory illusion created by Shepard tones. They can be presented in an ascending or descending manner in a loop to give an impression of *an endlessly ascending or descending pitch* (**Figure 1**). In the MMN experiment, these Shepard tones were looped to create an illusion of continuous pitch decrement that was eventually interrupted by a pitch repetition or by an ascending pitch. It was found that both pitch repetition and ascending pitch evoked the MMN when using Shepard tones. This was taken as evidence of the MMN being an index of violated prediction of the pitch of the sound-to-come rather than an index of sensory memory representation only.

Despite the theoretical relevance of the paradigms mentioned above, they had less to offer for applications of the MMN in clinical studies or studies with child participants. In traditional oddball paradigms, one sequence had one or maximally three deviants, making the studies rather long and repetitive, particularly if the signal-to-noise ratio was to be optimized by maximizing the number of sound presentations. As a solution, Näätänen introduced the idea of having several deviants in one sequence with one standard. Here, the basic assumption is that a standard sound is encoded as a sum of its acoustic features. Thus, one deviant can differ from this standard "template" independently by one or several features, as shown by so-called additivity studies by Schröger (1995) in which the MMN parameters sensitively reflected the number of violated sound features. When MMN recorded in a traditional oddball paradigm was compared with an MMN recorded in this multi-feature paradigm, there was no significant difference in the MMN parameters (Pakarinen et al., 2010). However, the recording time was remarkably shorter and thus the MMN

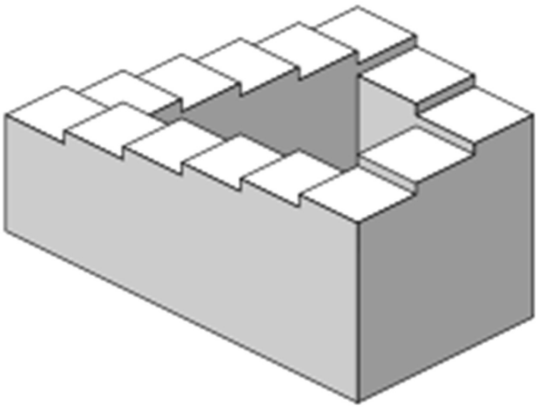


FIGURE 1
Visual analogue (endless Penrose stairs) of ever ascending/descending sound sequence created by Shepard (reproduced from Wikipedia).

recordings became more feasible with many clinical populations and in children.

Mismatch negativity paradigms in music-related studies

Based on the paradigm development described above, about 15 years ago interests emerged to develop "musical" MMN stimulation paradigms to probe the neural basis of musical skills. The first of these paradigms was based on the idea about multiple acoustical features being encoded in parallel and thus being behind the generation of the MMN. In the group of prof. Vuust, the starting point was an Alberti bass—a looped sound sequence often used in the classical era as an arpeggio. There, the sounds of a given triad chord were presented in the order "lowest, highest, middle, highest" in a looped manner (Vuust et al., 2012; **Figure 2A**). In this paradigm, the recording time was less than 15 min for a total of six different deviants, thus data collection is considerably faster than in traditional paradigms. In the melodic multi-feature paradigm developed by prof. Huotilainen, a looped 2-s melody was used as the starting point (Putkinen et al., 2014; **Figure 2B**). This melody also included a total of six deviants, three of which modulated the structure of the melody for its successive presentations. The data collection here also took less than 15 min.

By employing these musical multi-feature paradigms, it was shown that the MMN reflects the musical expertise and their participant background in a genre-specific manner; the sound parameters that are most important in a performance of a given musician evoke the largest MMN, P3a response, or both [for a review, see Putkinen and Tervaniemi (2018)]. The MMN was also shown to emerge in a gradual feature-specific manner during music training in children learning to play an instrument during their school years from 9 to 13 years of age (Putkinen et al., 2014). Furthermore, implicit vs. explicit forms of expertise were shown to have different neural trajectories as reflected by the MMN; while enthusiastic jazz listeners had a diminished MMN to a slide deviant, professional jazz performers showed an enlarged MMN to this deviant and to timbre and pitch deviants (Kliuchko et al., 2019). Thus, these paradigms highlighted the complexity of music learning and have also been helpful in differentiating implicit and explicit profiles in music listeners vs. performers.

In addition to looped melodic and chordal paradigms, various MMN studies have also been conducted using randomized chord sequences consisting of two or more triad chords (e.g., major chords as standards and minor chords as deviants). These studies have been conducted using several paradigms and there is no paradigm we could nominate as the prevalent paradigm (unlike in looped musical paradigms). Here, the first paradigms only used two chords and thus had the co-occurrence of acoustic and musical deviance; major and minor chords were different from each other in both manners [Tervaniemi et al., 1999; Brattico et al., 2009; and Tervaniemi et al., 2011 with magnetoencephalography (MEG) and Tervaniemi et al., 2000 with positron emission tomograpy (PET)]. More recently, Virtala et al. (2011) with EEG; **Figure 2C** created a paradigm in which the contribution of acoustical deviance could be excluded. This was accomplished by creating the stimulus chords from various tones at several frequency levels. By this design it was possible to control and balance how often each tone was presented either as part of a major chord or as part of a minor chord. Thus, any difference in the MMN evoked by the chords was a result of its category (major/minor) and not its acoustical composition. Using this chord-MMN paradigm, it was observed that already newborn infants can differentiate major and minor chords from each other (Virtala et al., 2013) and that music training enhances this differentiation in adolescents and in adults (Virtala et al., 2012, 2014).

In addition to major/minor mode, another dimension of any musical interval or chord is its consonance or dissonance. This attribute is often reduced as the pleasantness and unpleasantness of the intervals or chords, respectively, even if this nomenclature is not accurate since some individuals prefer dissonant "unpleasant" intervals, chords, and music excerpts over consonant "pleasant" intervals (see next paragraph). To investigate the effects of musical expertise on consonance/dissonance discrimination, Linnavalli et al. (2020) created two types of dissonant chords and introduced them in the context of consonant chords. They included groups of professional musicians and non-musicians as their participants. It was found that both groups of participants discriminated dissonant chords from consonant ones both neurally and behaviorally. In the behavioral task, the musicians were more accurate than the non-musicians without a group difference in the MMN elicitation. As the dissonant chords elicited MMN responses for both groups, sensory dissonance seems to be discriminated in an early sensory level, irrespective of musical expertise, and the facilitating effects of musical expertise for this discrimination seems to be activated only in later stages of auditory processing, as reflected by performance in the behavioral auditory task.

As the last example of the use of MMN in music-related studies, a recent paradigm developed by Sarasso et al. (2022) will be introduced. Sarasso and colleagues used intervals of two kinds: consonant (perfect fifth) and dissonant (tritones) at low and high frequency levels. The novel aspect in their study is that the data were analyzed based on the participants' preference for these intervals; half of them preferred consonant intervals, half of them dissonant intervals. It was found that irrespective of the acoustical and musical characteristics of the intervals, it was the most preferred

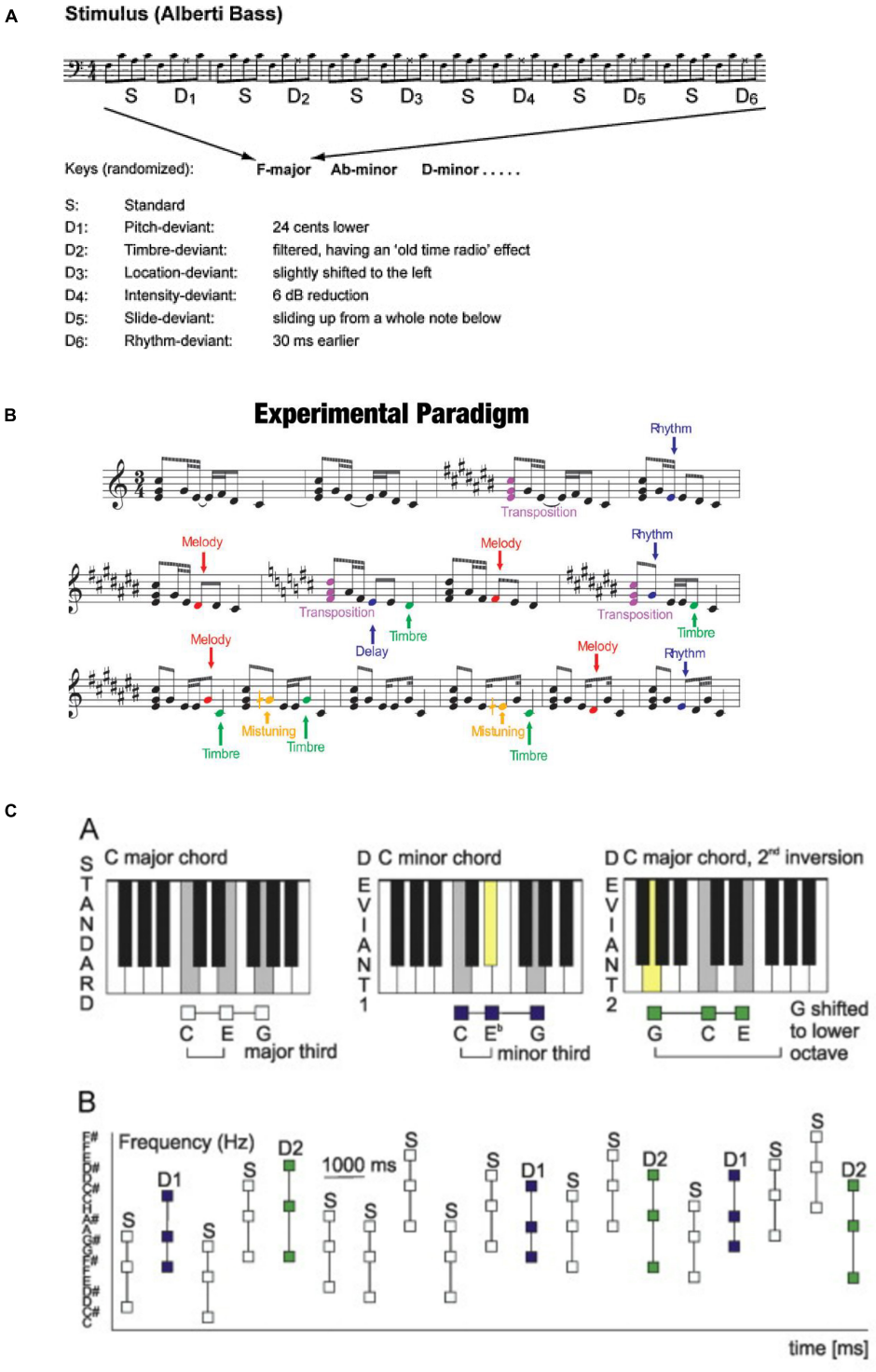


FIGURE 2
(A) Musical multifeature paradigm that includes sound patterns with six different deviant tones as indicated below the score. The sequence is presented in one key for six bars and then transposed to a new key, in other words, it was presented at various pitch levels. Reprinted from Vuust et al. (2012) with permission from Elsevier. **(B)** Melodic multifeatured paradigm that includes short melodies with three different acoustic deviances and three different cognitive deviances as indicated on the right. Cognitive deviants change the content of the melody while acoustic deviants do not. One of the cognitive deviants is transposition, meaning that the melody is presented at various pitch levels. Reproduced by permission from Tervaniemi et al. (2014). **(C)** Chord paradigm with standard and two different deviant chords as indicated in the upper row. During the experiment, these three chords are presented at randomly varying pitch levels (reproduced from Virtala et al. (2014) under CC-BY license).

and "attractive" interval that evoked larger MMN when compared with the other, less attractive interval. Moreover, computational Bayesian surprise index was associated with both MMN and behavioral indices, suggesting that (early) auditory learning is related to higher-order aesthetic processing of music sounds.

Mismatch negativity generators

Main contribution to the scalp recorded auditory MMN originates from the auditory cortices with an additional generator in the right frontal lobe [for a review, see Näätänen et al. (2010); see below]. Important in the current context is to note that the MMN generator source within the auditory areas may also vary as a function of the stimulus complexity: when an identical pitch change was embedded in an oddball sequence of sinusoidal tones versus musical chords, the MEG recordings indicated the MMN generator to be more medially located when more complex (musical) stimuli were used (Alho et al., 1996). Furthermore, in non-musicians, the left vs. right auditory cortices may adopt different roles as a function of the stimulus type: in PET and MEG experiments, the left auditory areas responded more strongly to changes in phonemes (Tervaniemi et al., 2000) and rhythm (Vuust et al., 2005) while the right auditory areas respond more strongly to changes in chords (Tervaniemi et al., 1999, 2000). However, this asymmetry may also be modulated by musical expertise: musicians were found to have predominantly left-hemispheric (MEG counterpart of) MMN to chord changes (Tervaniemi et al., 2011).

In addition to the auditory areas, also frontal areas, particularly the right inferior frontal gyrus, can be activated by the deviants when presented in an oddball paradigm, at least when the stimulation has acoustically small deviances (Opitz et al., 2002). Recently, using the melodic multifeature paradigm, it was shown that while the sensory deviants (e.g., timbre) were primarily processed in the auditory areas, the cognitively more demanding deviants (e.g., transposition) were primarily processed in the frontal areas (Bonetti et al., 2022). Together, these findings point to the multifaceted characteristics of the MMN generation along the sensory-cognitive-axis of our auditory neurocognition and, respectively, in the brain.

Finally, the deviance detection as indexed by the MMN may be initiated already below cortical areas. This was shown by fMRI findings using naturalistic stimuli (pseudoword/ba:ba/and its close acoustical musical counterpart produced by saxophone) in a semi-attend paradigm (Tervaniemi et al., 2006). There, non-musicians were instructed to indicate by a button press whether each sound was speech or music sound but not to pay attention to slight deviances. It was found that in addition to BOLD activations in the temporal and frontal areas, deviances in sound pitch and duration activated also thalamic structures. This finding is in line with increasing body of the literature highlighting the roles of ascending auditory pathways in deviance detection (Escera and Malmierca, 2014).

Future directions

This current perspective paper sought to highlight the developments in past decades in paradigms that have been developed in MMN studies for basic science and music-related research projects (due to the space limitations of this paper, clinical studies had to be ignored despite their high relevance). Even if the MMN was originally considered as a tool for investigating learning and neurocognition of simple sounds in simple contexts, there are now several paradigms that enable investigating higher-order phenomena, such as musical development, musical expertise, and appreciation. Thus, the progress of the paradigm development(s) enables theoretical advancement that is needed in the larger field of auditory cognitive neuroscience.

In the future, it is likely that also in MMN studies the stimulus material will include elements of real music instead of only isolated sounds or repetitive computer-generated sound sequences. Even if this sounds implausible, there are possibilities already available that enable such studies. One means of meeting this challenge is offered by music information retrieval (MIR) technology. Using a MIR toolbox (Lartillot and Toiviainen, 2007) it is possible to identify acoustical and musical events (sounds or sound sequences) and code with trigger pulses any sound of interest, be it repetitive or surprising in its context. This can be done before or after an experiment to recorded music or after the experiment to a music recording based on live performance during a study. MIR-based ERP analyses were already conducted by Poikonen et al. (2016) for sounds that had the largest computational values related to timbre, harmony, and dynamics. After averaging the ERPs following each of these sound categories, N100 and P200 responses were computed and compared between three different compositions. More recently, Haumann et al. (2021) elaborated and further tested the feasibility of such analyses with different musical excerpts, again with focus on P1-N1-P2 responses.

Naturally, it should be considered that to utilize MIR-based analysis in MMN studies, it is necessary to include some repetitive sound features in the music excerpts. However, this repetitiveness can also be interpreted in abstract terms, such that sounds to be used as one category in the analyses differ from each other in their exact acoustical features but also simultaneously form a distinct category of the other sounds of a given musical excerpt up to a sufficient degree (e.g., instrumental sounds that form a category "novel instrument" even if they differ from each other acoustically). By careful behavioral screening of the participants' cognitive, emotional,

and aesthetic ratings of the sounds as by Sarasso et al. (2022), we can additionally categorize the sounds and subsequent ERP/MMN responses not only based on their acoustical or musical features but also by their perceptual loadings. By these procedures, we can continue developing the MMN study paradigms on sounds as part of music and not merely on sounds as such.

Data availability statement

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

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

Writing has been supported by Business Finland (the governmental organization for innovation funding; project CREDU 9214/31/2019) and by the Center of Excellence in Music, Mind, Body, and Brain (Academy of Finland).

Acknowledgments

The author was thankful for the comments of Dr. Linnavalli (University of Helsinki) on a draft version of this manuscript.

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.

Publisher's note

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

References

Alho, K., Tervaniemi, M., Huotilainen, M., Lavikainen, J., Tiitinen, H., Ilmoniemi, R. J., et al. (1996). Processing of complex sounds in the human auditory cortex as revealed by magnetic brain responses. *Psychophysiology* 33, 369–375.

Amenedo, E., and Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *Eur. J. Neurosci.* 12, 2570–2574. doi: 10.1046/j.1460-9568.2000.00114.x

Bonetti, L., Carlomagno, F., Kliuchko, M., Gold, B., Palva, S., Haumann, N., et al. (2022). Whole-brain computation of cognitive versus acoustic errors in music. *Neuroimage Rep.*

Brattico, E., Pallesen, K. J., Varyagina, O., Bailey, C., Anourova, I., Järvenpää, M., et al. (2009). Neural discrimination of nonprototypical chords in music experts and laymen: An MEG study. *J. Cogn. Neurosci.* 21, 2230–2244. doi: 10.1162/jocn.2008.21144

Escera, C., and Malmierca, M. S. (2014). The auditory novelty system: An attempt to integrate human and animal research. *Psychophysiology* 51, 111–123. doi: 10.1111/psyp.12156

Haumann, N. T., Lumaca, M., Kliuchko, M., Santacruz, J. L., Vuust, P., and Brattico, E. (2021). Extracting human cortical responses to sound onsets and acoustic feature changes in real music, and their relation to event rate. *Brain Res.* 1754:147248. doi: 10.1016/j.brainres.2020.147248

Kliuchko, M., Brattico, E., Gold, B. P., Tervaniemi, M., Bogert, B., Toiviainen, P., et al. (2019). Fractionating auditory priors: A neural dissociation between active and passive experience of musical sounds. *PLoS One* 14:e0216499. doi: 10.1371/journal.pone.0216499

Lartillot, O., and Toiviainen, P. (2007). "MIR in Matlab (II): A toolbox for musical feature extraction from audio," in *Proceedings of the international conference on music information retrieval* (Vienna: Osterreichische Computer Gesellschaft).

Linnavalli, T., Haveri, L., Ojala, J., Putkinen, V., Kostilainen, K., Seppänen, S., et al. (2020). Musical expertise facilitates dissonance detection on behavioural, not on early sensory level. *Music Percept.* 38, 78–98. doi: 10.1525/mp.2020.38.1.78

Näätänen, R., Astikainen, P., Ruusuvirta, T., and Huotilainen, M. (2010). Automatic auditory intelligence: An expression of the sensory-cognitive core of cognitive processes. *Brain Res. Rev.* 64, 123–136. doi: 10.1016/j.brainresrev.2010.03.001

Näätänen, R., Gaillard, A. W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., and Sams, M. (1987). The mismatch negativity to intensity changes in an auditory stimulus sequence. *Electroencephalogr. Clin. Neurophysiol. Suppl.* 40, 125–131.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., and Schröger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *Neuroimage* 15, 167–174. doi: 10.1006/nimg.2001.0970

Paavilainen, P., Jiang, D., Lavikainen, J., and Näätänen, R. (1993). Stimulus duration and the sensory memory trace: An event-related potential study. *Biol. Psychol.* 35, 139–152. doi: 10.1016/0301-0511(93)90010-6

Pakarinen, S., Huotilainen, M., and Näätänen, R. (2010). The mismatch negativity (MMN) with no standard stimulus. *Clin. Neurophysiol.* 121, 1043–1050. doi: 10.1016/j.clinph.2010.02.009

Poikonen, H., Alluri, V., Brattico, E., Lartillot, O., Tervaniemi, M., and Huotilainen, M. (2016). Event-related brain responses while listening to entire pieces of music. *Neuroscience* 312, 58–73. doi: 10.1016/j.neuroscience.2015.10.061

Putkinen, V., and Tervaniemi, M. (2018). "Neuroplasticity in music learning," in *Oxford handbook of music and brain research: The neural basis of music perception, performance, learning, and music in therapy and medicine*, eds M.

Thaut and D. Hodges (Oxford: Oxford University Press). doi: 10.1093/oxfordhb/9780198804123.013.22

Putkinen, V., Tervaniemi, M., Saarikivi, K., de Vent, N., and Huotilainen, M. (2014). Investigating the effects of musical training on functional brain development with a novel melodic MMN paradigm. *Neurobiol. Learn. Mem.* 110, 8–15. doi: 10.1016/j.nlm.2014.01.007

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol.* 62, 437–448. doi: 10.1016/0168-5597(85)90054-1

Sarasso, P., Neppi-Modona, M., Rosaia, N., Perna, P., Barbieri, P., Del Fante, E., et al. (2022). Nice and easy: Mismatch negativity responses reveal a significant correlation between aesthetic appreciation and perceptual learning. *J. Exp. Psychol. Gen.* 151, 1433–1445. doi: 10.1037/xge0001149

Schröger, E. (1995). Processing of auditory deviants with changes in one versus two stimulus dimensions. *Psychophysiology* 32, 55–65. doi: 10.1111/j.1469-8986.1995.tb03406.x

Tervaniemi, M., Huotilainen, M., and Brattico, E. (2014). Melodic multi-feature paradigm reveals auditory profiles in music-sound encoding. *Front. Hum. Neurosci.* 8:496. doi: 10.3389/fnhum.2014.00496

Tervaniemi, M., Kujala, A., Alho, K., Virtanen, J., Ilmoniemi, R. J., and Näätänen, R. (1999). Functional specialization of the human auditory cortex in processing phonetic and musical sounds: A magnetoencephalographic (MEG) study. *Neuroimage* 9, 330–336.

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., and Näätänen, R. (1994a). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biol. Psychol.* 38, 157–167.

Tervaniemi, M., Maury, S., and Näätänen, R. (1994b). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport* 5, 844–846.

Tervaniemi, M., Medvedev, S. V., Alho, K., Pakhomov, S. V., Roudas, M. S., van Zuijen, T. L., et al. (2000). Lateralized automatic auditory processing of phonetic versus musical information: A PET study. *Hum. Brain Mapp.* 10, 74–79.

Tervaniemi, M., Sannemann, C., Salonen, J., Nöyränen, M., and Pihko, E. (2011). Importance of the left auditory areas in chord discrimination in music experts as evidenced by MEG. *Eur. J. Neurosci.* 34, 517–523. doi: 10.1111/j.1460-9568.2011.07765.x

Tervaniemi, M., Szameitat, A. J., Kruck, S., Schröger, E., Alter, K., De Baene, W., et al. (2006). From air oscillations to music and speech: Functional magnetic resonance imaging evidence for fine-tuned neural networks in audition. *J. Neurosci.* 26, 8647–8652. doi: 10.1523/JNEUROSCI.0995-06.2006

Tiitinen, H., May, P., Reinikainen, K., and Näätänen, R. (1994). Attentive novelty detection in humans is governed by pre-attentive sensory memory. *Nature* 372, 90–92. doi: 10.1038/372090a0

Virtala, P., Berg, V., Kivioja, M., Purhonen, J., Salmenkivi, M., Paavilainen, P., et al. (2011). The preattentive processing of major vs. minor chords in the human brain. An event-related potential study. *Neurosci. Lett.* 487, 406–410. doi: 10.1016/j.neulet.2010.10.066

Virtala, P., Huotilainen, M., Partanen, E., and Tervaniemi, M. (2014). Musicianship facilitates the processing of Western music chords – an ERP and behavioural study. *Neuropsychologia* 61, 247–258. doi: 10.1016/j.neuropsychologia.2014.06.028

Virtala, P., Huotilainen, M., Partanen, E., Fellman, V., and Tervaniemi, M. (2013). Newborn infants' auditory system is sensitive to Western music chord categories. *Front. Psychol.* 4:492. doi: 10.3389/fpsyg.2013.00492

Virtala, P., Putkinen, V., Huotilainen, M., Makkonen, T., and Tervaniemi, M. (2012). Musical training facilitates the neural discrimination of major vs. minor chords in 13-year-old children. *Psychophysiology* 49, 1125–1132. doi: 10.1111/j.1469-8986.2012.01386.x

Vuust, P., Brattico, E., Seppänen, M., Näätänen, R., and Tervaniemi, M. (2012). The sound of music: Differentiating musicians using a fast, musical multi-feature mismatch negativity paradigm. *Neuropsychologia* 50, 1432–1443.

Vuust, P., Pallesen, K. J., Bailey, C., van Zuijen, T. L., Gjedde, A., Roepstorff, A., et al. (2005). To musicians, the message is in the meter: Pre-attentive neuronal responses to incongruent rhythm are left-lateralized in musicians. *Neuroimage* 24, 560–564. doi: 10.1016/j.neuroimage.2004.08.039

Winkler, I., Tervaniemi, M., Huotilainen, M., Ilmoniemi, R. J., Ahonen, A., Salonen, O., et al. (1995). From objective to subjective: Pitch representation in the human auditory cortex. *Neuroreport* 6, 2317–2320.

Yabe, H., Tervaniemi, M., Reinikainen, K., and Näätänen, R. (1997). The temporal window of integration in auditory system as revealed by omission MMN. *Neuroreport* 8, 1971–1974. doi: 10.1097/00001756-199705260-00035

frontiers | Frontiers in Neuroscience

TYPE Perspective
PUBLISHED 08 December 2022
DOI 10.3389/fnins.2022.1075369

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Kirill Vadimovich Nourski,
The University of Iowa,
United States

*CORRESPONDENCE
Jonathan Z. Simon
jzsimon@umd.edu

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

RECEIVED 20 October 2022
ACCEPTED 24 November 2022
PUBLISHED 08 December 2022

CITATION
Simon JZ, Commuri V and Kulasingham JP (2022) Time-locked auditory cortical responses in the high-gamma band: A window into primary auditory cortex. *Front. Neurosci.* 16:1075369. doi: 10.3389/fnins.2022.1075369

Time-locked auditory cortical responses in the high-gamma band: A window into primary auditory cortex

Jonathan Z. Simon[1,2,3]*, Vrishab Commuri[1] and Joshua P. Kulasingham[4]

[1]Department of Electrical and Computer Engineering, University of Maryland, College Park, College Park, MD, United States, [2]Department of Biology, University of Maryland, College Park, College Park, MD, United States, [3]Institute for Systems Research, University of Maryland, College Park, College Park, MD, United States, [4]Department of Electrical Engineering, Linköping University, Linköping, Sweden

Primary auditory cortex is a critical stage in the human auditory pathway, a gateway between subcortical and higher-level cortical areas. Receiving the output of all subcortical processing, it sends its output on to higher-level cortex. Non-invasive physiological recordings of primary auditory cortex using electroencephalography (EEG) and magnetoencephalography (MEG), however, may not have sufficient specificity to separate responses generated in primary auditory cortex from those generated in underlying subcortical areas or neighboring cortical areas. This limitation is important for investigations of effects of top-down processing (e.g., selective-attention-based) on primary auditory cortex: higher-level areas are known to be strongly influenced by top-down processes, but subcortical areas are often assumed to perform strictly bottom-up processing. Fortunately, recent advances have made it easier to isolate the neural activity of primary auditory cortex from other areas. In this perspective, we focus on time-locked responses to stimulus features in the high gamma band (70–150 Hz) and with early cortical latency (~40 ms), intermediate between subcortical and higher-level areas. We review recent findings from physiological studies employing either repeated simple sounds or continuous speech, obtaining either a frequency following response (FFR) or temporal response function (TRF). The potential roles of top-down processing are underscored, and comparisons with invasive intracranial EEG (iEEG) and animal model recordings are made. We argue that MEG studies employing continuous speech stimuli may offer particular benefits, in that only a few minutes of speech generates robust high gamma responses from bilateral primary auditory cortex, and without measurable interference from subcortical or higher-level areas.

KEYWORDS

phase locked response, medial geniculate body, high frequency, envelope following response, cortical FFR

Introduction

Primary auditory cortex plays a key role in the human brain's processing of sounds, being a major gateway between auditory subcortical areas, including the inferior colliculus (midbrain) and thalamus, and higher order auditory cortical areas, including secondary auditory areas, associative auditory areas, and language areas. While the neurophysiology of primary auditory cortex has been studied for decades in animal models, there are still many unanswered questions. One of the hallmarks of primary auditory cortex in animal models is its sluggishness compared to subcortical areas, since its typical neurons time-lock[1] to acoustic modulations only up to a few tens of Hz (Lu et al., 2001; Joris et al., 2004), though at the same time it does respond very reliably (temporally) to brief acoustic features, with a spiking precision of milliseconds both for punctate features (Phillips and Hall, 1990; Heil and Irvine, 1997) and ongoing spectrotemporally dynamic features (Elhilali et al., 2004).

Less is known about temporal processing in *human* primary auditory cortex, where neurophysiological recording techniques for healthy subjects are restricted to non-invasive methods, primarily electroencephalography (EEG) and magnetoencephalography (MEG). Neither EEG nor MEG has very fine spatial resolution (typically a few centimeters) and so may not be able to distinguish different neural sources based purely on their anatomical origin. Both, however, have sufficient temporal resolution to distinguish typical response latencies of primary auditory cortex (~40 ms) from subcortical (shorter latency) and non-primary (longer latency) auditory areas.

Beyond these commonalities, EEG and MEG have distinctive strengths and weaknesses. EEG is sensitive to neural sources throughout the brain at both low frequencies (tens of Hz) and high frequencies (hundreds of Hz) (Kraus et al., 2017; White-Schwoch et al., 2019). It is therefore relatively straightforward to record time-locked activity from any auditory area of the brain, but it may be difficult to distinguish contributions from multiple areas, at least without additional information (e.g., response latency, which can be used to distinguish between the sources giving rise to the auditory P1 and N1 components). In contrast, MEG is insensitive to subcortical neural sources (Hämäläinen et al., 1993), though not entirely unresponsive, as seen below. Perhaps counterintuitively, this insensitivity gives MEG an advantage over EEG, by allowing recordings from auditory cortical sources without substantial subcortical interference (Ross et al., 2020). Nevertheless, MEG responses from different auditory cortical areas can still interfere with each other.

Another consideration is that EEG's sensitivity to most auditory sources holds for both low and high frequencies, but because of MEG's cortical bias and because cortical responses are usually sluggish, MEG typically only captures cortical sources at low frequencies. An important counterexample, however, is the case of fast (~100 Hz) auditory time-locked cortical responses (Hertrich et al., 2012; Coffey et al., 2016). At these frequencies there are few, if any, cortical sources aside from primary auditory cortex. In this sense, MEG recordings of fast time-locked auditory cortical responses act as an exquisite window into primary auditory cortex, without interference from subcortical or other cortical areas. Therefore, it may be especially suited for questions regarding how primary auditory cortical responses are affected by cognitive processes, whether modulated by top-down neural activity (e.g., selective attention or task-specific processing) or supplemented by super-auditory aspects of the stimulus (e.g., processing of speech sounds using language-based information).

One newly established method to analyze neural responses to continuous speech (Hamilton and Huth, 2018) is temporal response function (TRF) analysis (Lalor et al., 2009; Ding and Simon, 2012). TRFs are an effective tool to disambiguate neural sources based on their characteristic latencies, as will be discussed below.

Results

Fast (~100 Hz) cortical time-locked auditory responses are typically investigated using one of two different stimulus paradigms. The more time-honored paradigm is the frequency following response (FFR) (Kraus et al., 2017), for which a typical stimulus is either acoustically simple, such as click trains or amplitude modulated tones (e.g., Gorina-Careta et al., 2021), or consists of many repetitions of a short but more complex stimulus, such as a single syllable (e.g., Coffey et al., 2016).

The well-established FFR paradigm (or really, family of paradigms, including the envelope following response; EFR) has been used to great effect with EEG to investigate midbrain responses to acoustic stimuli. Near 100 Hz, midbrain sources dominate the EEG FFR over cortical sources, and well above 100 Hz there is little to no cortical EEG FFR contribution at all (Coffey et al., 2019). Until the MEG FFR investigations of Coffey et al. (2016), however, it was not widely appreciated how substantial the cortical FFR contributions might be near 100 Hz. In this seminal paper, the investigators presented the 120-ms syllable/da/, synthesized with a 98 Hz fundamental frequency in the vowel portion, for 14,000 repetitions (sufficient to also obtain responses from subcortical sources despite the cortical bias of MEG). The cortical responses, whose sources were consistent with primary auditory cortex, were prominent and showed a significant lateralization to the right hemisphere, with a longer latency profile compared to subcortical components. This work firmly established the measurability of distinct cortical

1 We employ the term "time-locked" neural responses rather than "phase-locked" since phase is only defined when the coupled stimulus/response is analyzed in a narrow frequency band. The term "time-locking", sometimes called "neural tracking" when applied to low frequency responses to speech, applies equally well to narrowband and broadband cases.

contributions to the FFR near 100 Hz. In comparison, Gorina-Careta et al. (2021) demonstrated that the MEG FFR at the much higher frequency of 333 Hz (15,200 tone-burst repetitions) originated solely from subcortical sources (**Figure 1**). Note that both these studies demonstrate that, while MEG is not incapable of measuring high frequency FFR from subcortical sources, the number of repetitions required is considerable, with an associated experimental design cost (e.g., limited to a small number of stimulus types).

One of the limitations of the FFR paradigm is that accessing the different latencies of distinct sources may not be straightforward, since the FFR is ultimately just the evoked response to a sustained stimulus: a linear sum of overlapping responses from multiple sources with different latencies (Teichert et al., 2022). A more recently developed paradigm uses neural responses to continuous speech, such as individual sentences (e.g., Hertrich et al., 2012) or longer narrated story passages (e.g., Kulasingham et al., 2020). The use of the continuous speech stimulus paradigm, combined with TRF analysis, sidesteps this temporal overlap issue by deconvolving the sustained response from the stimulus, which often allows direct comparison of neural source peak latencies. Though typical uses of TRF analysis employ the slow (<10 Hz) acoustic envelope as the stimulus feature with which to deconvolve (Di Liberto et al., 2015; Cervantes Constantino and Simon, 2018), the TRF methodology generalizes well to other stimulus features (Brodbeck and Simon, 2020). This includes responses from high frequency stimulus features processed in subcortical areas (Maddox and Lee, 2018; Polonenko and Maddox, 2021).

High frequency (70–200 Hz) MEG TRFs were first investigated by Kulasingham et al. (2020) using only 6 mins of continuous speech as the stimulus. Responses source-localized to bilateral primary auditory cortex, with a small but significant lateralization to the right hemisphere (**Figure 2A**). The peak latency of the cortical response, 40 ms, is consistent

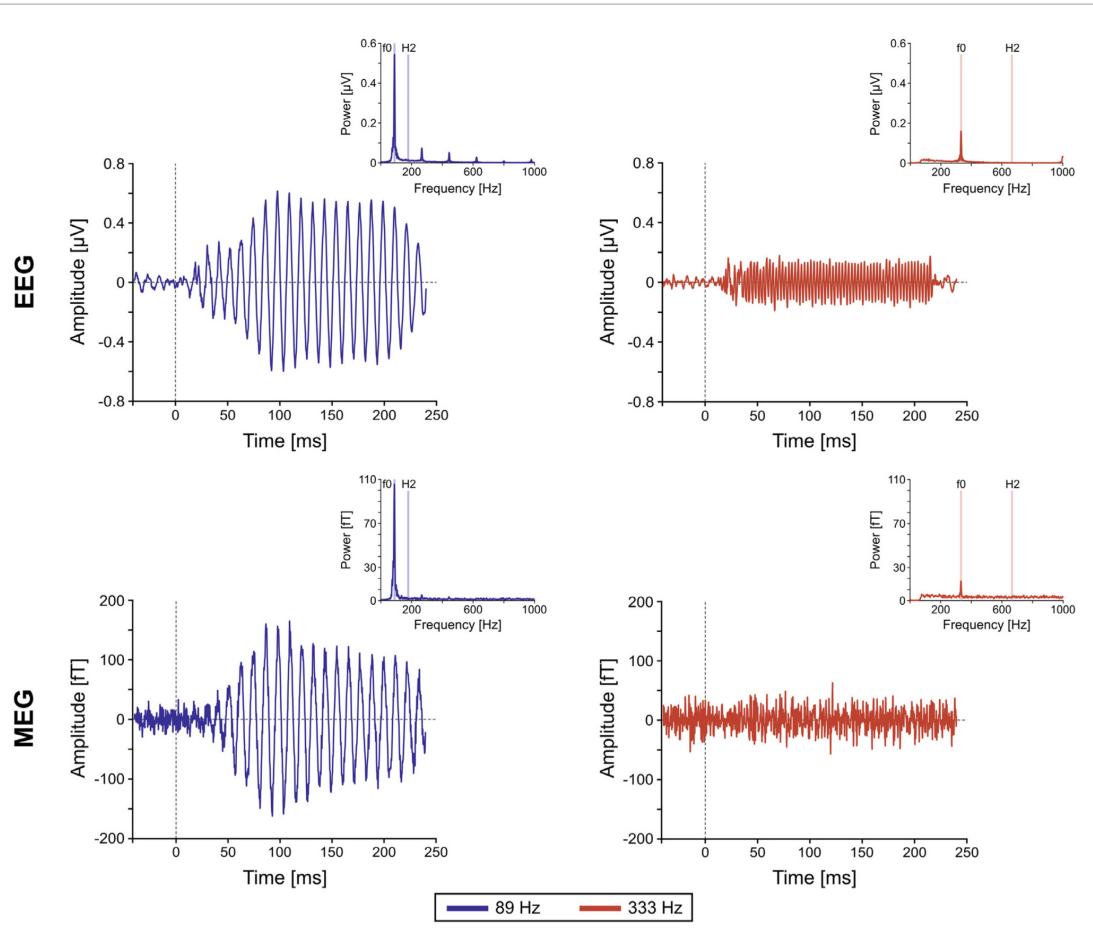


FIGURE 1
Example frequency following responses (FFRs). Grand-averaged FFR time course and spectral representations (insets) of single-channel EEG and magnetoencephalography (MEG) elicited in the high gamma frequency range (89 Hz; blue) and the very high gamma range (333 Hz; red). It can be shown that the very high gamma frequency (333 Hz; red) FFR is almost entirely subcortical for both EEG and MEG. In contrast, the high gamma frequency (89 Hz; blue) FFR is almost entirely cortical for MEG and a mix of cortical and subcortical for EEG [from Gorina-Careta et al. (2021), Figure 1].

with a primary auditory cortical origin. Analysis additionally revealed that frequencies contributing to time-locking fell off substantially above 100 Hz. This demonstration that such a short recording can reveal responses localized to primary auditory cortex serves several purposes. It allows future experiments to include multiple stimulus conditions (e.g., presenting stimuli under different task conditions or at different SNRs), and at the same time ensures that the responses do not contain measurable subcortical interference.

High frequency (70–200 Hz) EEG TRFs with cortical contributions have also been recently investigated by Kegler et al. (2022). These TRFs show a pair of peaks with distinguishable latencies allowing inference of separate sources, each with a separate anatomical origin and auditory processing role (analogous to traditional P1 and N1 peaks arising from separate cortical sources). In this case, the earlier peak at 18 ms is consistent with a subcortical origin, and the later peak at 45 ms is consistent with a dominantly cortical origin (**Figure 2B**).

It should not be surprising that invasive iEEG recordings had already demonstrated similar high gamma time-locked cortical responses almost a decade earlier (Brugge et al., 2009; Steinschneider et al., 2013), using click trains and isolated speech sounds. What is surprising is that such responses could be seen even non-invasively. The most robust time-locked high gamma iEEG responses are seen in primary auditory cortex, specifically posteromedial Heschl's gyrus (Nourski, 2017), but smaller time-locked high gamma responses are also seen in other auditory cortical areas. As such, iEEG remains a premiere electrophysiological method for obtaining responses known to originate in primary auditory cortex, but only for a fraction of subjects relative to those eligible for MEG or EEG recordings.

Discussion

As indicated above, a physiological window into human primary auditory cortex allows the investigation of the extent to which primary auditory cortex is influenced by higher order cortical areas. How, and under which circumstances, are primary auditory cortical responses modulated by top-down neural activity, or affected by language-specific non-auditory features of the stimulus? A related question is to what extent subcortical auditory areas might be influenced by cortical processing. Neither can be answered without first identifying the specific sources of neural activity (e.g., midbrain vs. thalamus vs. primary auditory cortex) being modulated by distant cortical activity.

Using MEG, Hartmann and Weisz (2019) demonstrated that the FFR near 100 Hz from right hemisphere primary auditory cortex is modulated by intermodal (auditory vs. visual) attention. Most FFR investigations use EEG, which is well-suited to separate responses from primary auditory cortex from those originating in other cortical areas, but, as indicated above, has difficulty in separating auditory subcortical and primary auditory cortical contributions. Intriguing results include: modulation of the EEG FFR by selective attention for frequencies near 100 Hz but not above 200 Hz (Holmes et al., 2018); modulation by overall level of attention near 150 Hz (Price and Bidelman, 2021); and, at 100 Hz, modulation by whether a continuous-speech masker is in a known vs. unknown (but acoustically similar) language (Presacco et al., 2016; Zan et al., 2019). There has also been a report of selective attentional modulation of subcortical auditory responses to continuous speech (Forte et al., 2017); the result has not yet been replicated, however, and due to the specialty of the analysis method it

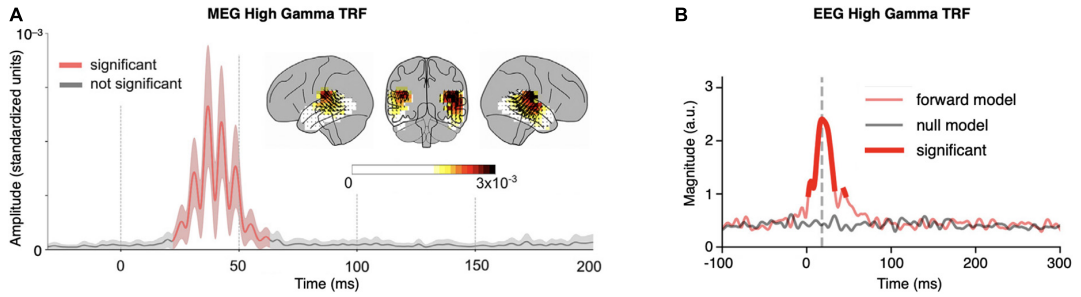


FIGURE 2
Example high gamma temporal response functions (TRFs). **(A)** High frequency (70–200 Hz) magnetoencephalography (MEG) TRF from 6 mins of continuous speech. The grand-averaged amplitude of TRF source localized current-dipole vectors, averaged across voxels in the cortical ROI, is shown (±standard error across subjects; red indicates amplitude significantly greater than noise). The TRF has a peak latency of ~40 ms and oscillates with a frequency of ~80 Hz (note that since only the TRF amplitude is shown, and not signed current values, signal troughs and peaks both appear as peaks). Inset: the distribution of TRF current-dipole vectors in the brain at each voxel at the moment of the maximum response; color represents response amplitude (standardized units) and arrows represent TRF current-dipole orientations [modified from Kulasingham et al. (2020), Figure 3]. **(B)** High frequency (70–200 Hz) EEG TRF from 40 mins of continuous speech. The grand-averaged magnitude of the Hilbert transform of the TRF, averaged across channels, is shown; bright red indicates magnitude significantly greater than the null model. The TRF magnitude significantly exceeds that of the null model in two latency ranges: between 2 and 33 ms with a peak at 18 ms (dominantly subcortical; grey dashed line), and between 44 and 46 ms with a peak at 45 ms (dominantly cortical) [modified from Kegler et al. (2022), Figure 3].

is as yet difficult to rule out entirely whether the result might be due to cortical response leakage.

More recently, using EEG with a continuous speech stimulus, Kegler et al. (2022) demonstrated that the high gamma EEG TRF arising from a combination of subcortical and primary auditory cortical sources (illustrated in **Figure 2B**) is modulated by word-boundary effects. This is strong evidence that a linguistic (super-acoustic) feature can modulate either primary auditory cortical or auditory subcortical processing (or both). Kulasingham et al. (2022) have also recently demonstrated that the high gamma MEG TRF, originating solely from bilateral primary auditory cortex, is indeed modulated by selective attention, using re-analysis of previously published data (Kulasingham et al., 2021).

There is additional evidence that human primary auditory cortical responses exhibit modulation arising from other cortical areas, but the effects are subtle. Using iEEG and employing selective attention to one of two competing talkers, O'Sullivan et al. (2019) did not observe modulation of cortical responses in Heschl's gyrus (the anatomical location of primary auditory cortex), while, in contrast, they did find modulation in non-primary areas, as expected. Using a similar paradigm to investigate the role of selective attention on MEG low frequency cortical TRFs, Brodbeck et al. (2018), did see evidence of significant TRF modulation at short latencies consistent with a primary auditory cortex origin (in addition to the expected strong modulation at longer latencies), but only under limited conditions.

In animal studies, top-down (task-dependent) modulation of neural activity in primary auditory cortex has been seen as far back as two decades ago (Fritz et al., 2003). Despite the robustness and reproducibility of these results, however, the effect size is nevertheless small, and it has not been clear until recently whether such modulations would ever be observable non-invasively.

What is the physiological origin of the high gamma time-locked responses from primary auditory cortex? Two theories have been put forward. The first concerns the physics underlying the generators of EEG and MEG signals, which are dominantly driven by dendritic currents produced by synaptic inputs (Hämäläinen et al., 1993; Buzsaki et al., 2012), i.e., the same mechanisms that also give rise to the local field potential (LFP). For primary auditory cortex, the most significant neural input is the spiking output of the medial geniculate body (MGB) of the thalamus, whose spiking rates can reach up to 100 Hz (Miller et al., 2002), and whose thalamocortical fibers show ensemble-wide time-locking up to 300 Hz (Steinschneider et al., 1998), in animal models. A second theory, strongly tied to the first, is that the spikes of primary auditory cortex, which can only fire at rates well below 100 Hz, can nevertheless fire with temporal precision of the order of milliseconds (Elhilali et al., 2004). It has been recently shown by Downer et al. (2021) that these precise but infrequent spikes are actually highly synchronous across the local population, even to the point of acting as a time-locked *population* model for fast acoustic features (almost up to 200 Hz). Indeed, Gnanateja et al. (2021) recently demonstrated a connection between both these explanations, using intracortical FFR (90–140 Hz) recordings from multiple species, to show both an LFP FFR and a multi-unit (spiking) FFR, in the thalamorecipient layers of primary auditory cortex.

In conclusion, recent advances in auditory neuroscience have opened up new non-invasive windows into the neurophysiology of primary auditory cortex. Using EEG FFR techniques, responses are dominantly subcortical but also contain strong contributions from primary auditory cortex at frequencies near 100 Hz. Using MEG FFR techniques, responses are dominantly from primary auditory cortex for frequencies near 100 Hz (though at higher frequencies subcortical responses can also be detected given sufficient recording time). EEG TRF studies have the potential to show both auditory subcortical and primary auditory cortical contributions to the time-locked high gamma responses to continuous speech, but, unlike FFR, segregated in time/latency. Finally, MEG time-locked high gamma TRF studies may hold great promise in isolating primary auditory cortical responses from other areas, due to its insensitivity to subcortical sources and its ability to differentiate competing cortical sources in both time and anatomical location.

Data availability statement

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

Author contributions

JZS wrote the initial draft of the manuscript. All authors contributed to the interpretations of results and discussions, were involved in manuscript revision, and approved the final version.

Funding

This work was supported by grants from the National Institute of Deafness and Other Communication Disorders (R01-DC019394), the National Institute on Aging (P01-AG055365), the National Science Foundation (SMA-1734892), and the William Demant Foundation (20-0480).

Conflict of interest

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

Publisher's note

References

Brodbeck, C., and Simon, J. (2020). Continuous speech processing. *Curr. Opin. Physiol.* 18, 25–31. doi: 10.1016/j.cophys.2020.07.014

Brodbeck, C., Hong, L., and Simon, J. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Curr. Biol.* 28, 3976–3983.e5. doi: 10.1016/j.cub.2018.10.042

Brugge, J., Nourski, K., Oya, H., Reale, R., Kawasaki, H., Steinschneider, M., et al. (2009). Coding of repetitive transients by auditory cortex on Heschl's gyrus. *J. Neurophysiol.* 102, 2358–2374. doi: 10.1152/jn.91346.2008

Buzsaki, G., Anastassiou, C., and Koch, C. (2012). The origin of extracellular fields and currents–EEG, ECoG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241

Cervantes Constantino, F., and Simon, J. (2018). Restoration and efficiency of the neural processing of continuous speech are promoted by prior knowledge. *Front. Syst. Neurosci.* 12:56. doi: 10.3389/fnsys.2018.00056

Coffey, E., Herholz, S., Chepesiuk, A., Baillet, S., and Zatorre, R. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7:11070. doi: 10.1038/ncomms11070

Coffey, E., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., et al. (2019). Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* 10:5036. doi: 10.1038/s41467-019-13003-w

Di Liberto, G., O'Sullivan, J., and Lalor, E. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Curr. Biol.* 25, 2457–2465. doi: 10.1016/j.cub.2015.08.030

Ding, N., and Simon, J. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11854–11859. doi: 10.1073/pnas.1205381109

Downer, J., Bigelow, J., Runfeldt, M., and Malone, B. (2021). Temporally precise population coding of dynamic sounds by auditory cortex. *J. Neurophysiol.* 126, 148–169. doi: 10.1152/jn.00709.2020

Elhilali, M., Fritz, J., Klein, D., Simon, J., and Shamma, S. (2004). Dynamics of precise spike timing in primary auditory cortex. *J. Neurosci.* 24, 1159–1172. doi: 10.1523/JNEUROSCI.3825-03.2004

Forte, A., Etard, O., and Reichenbach, T. (2017). The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *Elife* 6:e27203. doi: 10.7554/eLife.27203

Fritz, J., Shamma, S., Elhilali, M., and Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nat. Neurosci.* 6, 1216–1223. doi: 10.1038/nn1141

Gnanateja, G., Rupp, K., Llanos, F., Remick, M., Pernia, M., Sadagopan, S., et al. (2021). Frequency-following responses to speech sounds are highly conserved across species and contain cortical contributions. *eNeuro* 8. doi: 10.1523/ENEURO.0451-21.2021

Gorina-Careta, N., Kurkela, J., Hamalainen, J., Astikainen, P., and Escera, C. (2021). Neural generators of the frequency-following response elicited to stimuli of low and high frequency: A magnetoencephalographic (MEG) study. *Neuroimage* 231:117866. doi: 10.1016/j.neuroimage.2021.117866

Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., and Lounasmaa, O. (1993). Magnetoencephalography–theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys.* 65, 413–497. doi: 10.1103/RevModPhys.65.413

Hamilton, L., and Huth, A. (2018). The revolution will not be controlled: Natural stimuli in speech neuroscience. *Lang. Cogn. Neurosci.* 35, 573–582. doi: 10.1080/23273798.2018.1499946

Hartmann, T., and Weisz, N. (2019). Auditory cortical generators of the frequency following response are modulated by intermodal attention. *Neuroimage* 203:116185. doi: 10.1016/j.neuroimage.2019.116185

Heil, P., and Irvine, D. (1997). First-spike timing of auditory-nerve fibers and comparison with auditory cortex. *J. Neurophysiol.* 78, 2438–2454. doi: 10.1152/jn.1997.78.5.2438

Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., and Ackermann, H. (2012). Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. *Psychophysiology* 49, 322–334. doi: 10.1111/j.1469-8986.2011.01314.x

Holmes, E., Purcell, D., Carlyon, R., Gockel, H., and Johnsrude, I. (2018). Attentional modulation of envelope-following responses at lower (93-109 Hz) but not higher (217-233 Hz) modulation rates. *J. Assoc. Res. Otolaryngol.* 19, 83–97. doi: 10.1007/s10162-017-0641-9

Joris, P., Schreiner, C., and Rees, A. (2004). Neural processing of amplitude-modulated sounds. *Physiol. Rev.* 84, 541–577. doi: 10.1152/physrev.00029.2003

Kegler, M., Weissbart, H., and Reichenbach, T. (2022). The neural response at the fundamental frequency of speech is modulated by word-level acoustic and linguistic information. *Front. Neurosci.* 16:915744. doi: 10.3389/fnins.2022.915744

Kraus, N., Anderson, S., and White-Schwoch, T. (2017). *The frequency-following response: A window into human communication, the frequency-following response.* New York, NY: Springer, 1–15.

Kulasingham, J., Brodbeck, C., Presacco, A., Kuchinsky, S., Anderson, S., and Simon, J. (2020). High gamma cortical processing of continuous speech in younger and older listeners. *Neuroimage* 222:117291. doi: 10.1016/j.neuroimage.2020.117291

Kulasingham, J., Commuri, V., and Simon, J. (2022). "High gamma time-locked cortical responses to continuous speech," in *Proceedings of the 6th international conference on cognitive hearing science for communication (CHSCOM)*, Linköping.

Kulasingham, J., Joshi, N., Rezaeizadeh, M., and Simon, J. (2021). Cortical processing of arithmetic and simple sentences in an auditory attention task. *J. Neurosci.* 41, 8023–8039. doi: 10.1523/JNEUROSCI.0269-21.2021

Lalor, E., Power, A., Reilly, R., and Foxe, J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *J. Neurophysiol.* 102, 349–359. doi: 10.1152/jn.90896.2008

Lu, T., Liang, L., and Wang, X. (2001). Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. *Nat. Neurosci.* 4, 1131–1138. doi: 10.1038/nn737

Maddox, R., and Lee, A. (2018). Auditory brainstem responses to continuous natural speech in human listeners. *eNeuro* 5. doi: 10.1523/ENEURO.0441-17.2018

Miller, L., Escabi, M., Read, H., and Schreiner, C. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *J. Neurophysiol.* 87, 516–527. doi: 10.1152/jn.00395.2001

Nourski, K. (2017). Auditory processing in the human cortex: An intracranial electrophysiology perspective. *Laryngoscope Investig. Otolaryngol.* 2, 147–156. doi: 10.1002/lio2.73

O'Sullivan, J., Herrero, J., Smith, E., Schevon, C., McKhann, G., Sheth, S., et al. (2019). Hierarchical encoding of attended auditory objects in multi-talker speech perception. *Neuron* 104, 1195-1209.e3. doi: 10.1016/j.neuron.2019.09.007

Phillips, D., and Hall, S. (1990). Response timing constraints on the cortical representation of sound time structure. *J. Acoust. Soc. Am.* 88, 1403–1411. doi: 10.1121/1.399718

Polonenko, M., and Maddox, R. (2021). Exposing distinct subcortical components of the auditory brainstem response evoked by continuous naturalistic speech. *Elife* 10:e62329. doi: 10.7554/eLife.62329

Presacco, A., Simon, J., and Anderson, S. (2016). Effect of informational content of noise on speech representation in the aging midbrain and cortex. *J. Neurophysiol.* 116, 2356–2367. doi: 10.1152/jn.00373.2016

Price, C., and Bidelman, G. (2021). Attention reinforces human corticofugal system to aid speech perception in noise. *Neuroimage* 235:118014. doi: 10.1016/j.neuroimage.2021.118014

Ross, B., Tremblay, K., and Alain, C. (2020). Simultaneous EEG and MEG recordings reveal vocal pitch elicited cortical gamma oscillations in young and older adults. *Neuroimage* 204:116253. doi: 10.1016/j.neuroimage.2019.116253

Steinschneider, M., Nourski, K., and Fishman, Y. (2013). Representation of speech in human auditory cortex: Is it special? *Hear Res.* 305, 57–73. doi: 10.1016/j.heares.2013.05.013

Steinschneider, M., Reser, D., Fishman, Y., Schroeder, C., and Arezzo, J. (1998). Click train encoding in primary auditory cortex of the awake monkey: Evidence for two mechanisms subserving pitch perception. *J. Acoust. Soc. Am.* 104, 2935–2955. doi: 10.1121/1.423877

Teichert, T., Gnanateja, G., Sadagopan, S., and Chandrasekaran, B. (2022). A linear superposition model of envelope and frequency following responses may help identify generators based on latency. *Neurobiol. Lang.* 3, 441–468.

White-Schwoch, T., Anderson, S., Krizman, J., Nicol, T., and Kraus, N. (2019). Case studies in neuroscience: Subcortical origins of the frequency-following response. *J. Neurophysiol.* 122, 844–848. doi: 10.1152/jn.00112.2019

Zan, P., Presacco, A., Anderson, S., and Simon, J. (2019). Mutual information analysis of neural representations of speech in noise in the aging midbrain. *J. Neurophysiol.* 122, 2372–2387. doi: 10.1152/jn.00270.2019

frontiers | Frontiers in Neuroscience

TYPE Mini Review
PUBLISHED 20 December 2022
DOI 10.3389/fnins.2022.1075511

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Jonathan Z. Simon,
University of Maryland, College Park,
United States

*CORRESPONDENCE
Robert J. Zatorre
robert.zatorre@mcgill.ca

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

RECEIVED 20 October 2022
ACCEPTED 30 November 2022
PUBLISHED 20 December 2022

CITATION
Zatorre RJ (2022) Hemispheric
asymmetries for music and speech:
Spectrotemporal modulations
and top-down influences.
Front. Neurosci. 16:1075511.
doi: 10.3389/fnins.2022.1075511

Hemispheric asymmetries for music and speech: Spectrotemporal modulations and top-down influences

Robert J. Zatorre*

International Laboratory for Brain, Music, and Sound Research, Montreal Neurological Institute, McGill University, Montreal, QC, Canada

Hemispheric asymmetries in auditory cognition have been recognized for a long time, but their neural basis is still debated. Here I focus on specialization for processing of speech and music, the two most important auditory communication systems that humans possess. A great deal of evidence from lesion studies and functional imaging suggests that aspects of music linked to the processing of pitch patterns depend more on right than left auditory networks. A complementary specialization for temporal resolution has been suggested for left auditory networks. These diverse findings can be integrated within the context of the spectrotemporal modulation framework, which has been developed as a way to characterize efficient neuronal encoding of complex sounds. Recent studies show that degradation of spectral modulation impairs melody perception but not speech content, whereas degradation of temporal modulation has the opposite effect. Neural responses in the right and left auditory cortex in those studies are linked to processing of spectral and temporal modulations, respectively. These findings provide a unifying model to understand asymmetries in terms of sensitivity to acoustical features of communication sounds in humans. However, this explanation does not account for evidence that asymmetries can shift as a function of learning, attention, or other top-down factors. Therefore, it seems likely that asymmetries arise both from bottom-up specialization for acoustical modulations and top-down influences coming from hierarchically higher components of the system. Such interactions can be understood in terms of predictive coding mechanisms for perception.

KEYWORDS

lateralization, music, speech, neuroimaging, spectrotemporal modulation

Introduction

We have known since observations in the mid-19th century about aphasia that the two cerebral hemispheres of the human brain do not have identical functions (Manning and Thomas-Antérion, 2011). Yet, debate continues to this day on the underlying principles that govern these differences. Asymmetries have been described in many domains, including visuospatial, motor, and affective functions. But here I will focus on asymmetries related to auditory processes. A great deal of work has been carried out on the linguistic functions of the left hemisphere, in part because those earliest observations showed such salient effects of left-hemisphere lesions on language in general and speech in particular. But it is instructive to compare speech to that other auditory-motor communication system that we humans possess: music.

Comparisons between music and language are extremely valuable for many reasons (Patel, 2010), and can be carried out at many different levels of analysis. In this mini-review I will focus on certain acoustical features that I argue are critical for important aspects of musical processing, and contrast them with those most relevant for speech, to show that auditory networks within each hemisphere are specialized in terms of sensitivity to those features. However, one of the main points I wish to make is that those input-driven specializations interact with top-down mechanisms to yield a complex interplay between the two hemispheres.

Specialization for spectral features

A great deal of evidence supports the idea that certain aspects of musical perceptual functions depend to a greater extent on auditory networks in the right hemisphere than the left. This conclusion is supported by a recent meta-analysis of the effects of vascular lesions on musical perceptual skills (Sihvonen et al., 2019), as well as by early experimental studies of the consequences of temporal-lobe excisions (Milner, 1962; Samson and Zatorre, 1988; Zatorre, 1988; Liégeois-Chauvel et al., 1998). Apart from these effects of acquired lesions, deficits in congenital amusia (also termed tone-deafness) also seem to be linked to a disruption in the organization of connections between right auditory cortex and right inferior frontal regions. Evidence for this conclusion comes from studies of functional activation (Albouy et al., 2013) and functional connectivity (Hyde et al., 2011; Albouy et al., 2015), as well as anatomical measures of cortical thickness (Hyde et al., 2007) and of white-matter fiber connections (Loui et al., 2009; Albouy et al., 2013).

These findings are compelling, but what particular aspects of perception are most relevant in eliciting these asymmetries? A hint comes from the amusia literature, where several authors have found that the ability to process fine pitch differences seems to be particularly impaired (Hyde and Peretz, 2004; Tillmann et al., 2016). Those results are echoed in surgical lesion studies showing that damage to an area adjacent to right primary auditory cortex specifically leads to elevated pitch-direction discrimination thresholds compared to equivalent lesions on the left side (Johnsrude et al., 2000). Fine pitch resolution is important for processing musical features such as melody and harmony (Zatorre and Baum, 2012), which is why if that function is impaired, amusia typically follows (Peretz et al., 2002; Hyde and Peretz, 2004).

Many neuroimaging studies also align well with the idea that the right auditory cortical system is specialized for fine pitch processing. Several experiments have found that functional MRI responses in right auditory cortex scale more strongly than those on the left as pitch distance is manipulated from smaller to larger in a tone pattern; that is, the right side is more sensitive to variation of this parameter (Zatorre and Belin, 2001; Jamison et al., 2006; Hyde et al., 2008; Zatorre et al., 2012). Supportive findings also come from an MEG experiment examining spectral and temporal deviant detection (Okamoto and Kakigi, 2015). Importantly, the asymmetry of response seems to be linked to individual differences in pitch perception skill, thus showing a direct brain-behavior link. For example, functional MRI activity in the right (but not the left) auditory cortex of a group of musicians was correlated with their individual pitch discrimination thresholds (Bianchi et al., 2017). A correlation between individual pitch discrimination thresholds and the amplitude of the frequency-following response measured from the right (but not the left) auditory cortex was also observed using MEG (Coffey et al., 2016).

If spectral resolution on the right is better than on the left, what could be the physiological mechanism behind it? One possible answer was provided by an analysis of local functional connectivity patterns in relation to frequency tuning (Cha et al., 2016). This study found that the interconnectivity between voxels in auditory cortex is greater for those whose frequency tuning is more similar than for voxels which are tuned to more distant frequencies. But of greater relevance is that this pattern was more marked within right than left core auditory regions. In other words, frequency selectivity played a greater role on the right than the left, which would then lead to sharper tuning on the right, since there would be summation of activity from neurons with similar response properties. This conclusion is in line with electrophysiological recordings indicating that sharp tuning of neurons to frequency in early auditory cortex depends on excitatory intracortical inputs, rather than thalamic inputs (Liu et al., 2007).

Specialization for temporal features

The evidence favoring a relative enhancement of frequency resolution in the right auditory networks is paralleled by

evidence favoring a relative enhancement of temporal resolution in the left hemisphere. Several functional neuroimaging studies have shown that parametric variation of temporal features of stimuli is better tracked by responses coming from the left auditory cortex and adjacent regions compared to the right (Zatorre and Belin, 2001; Schönwiesner et al., 2005; Jamison et al., 2006; Obleser et al., 2008). Causal evidence in favor of this concept was also provided by a brain stimulation experiment showing increased thresholds for gap detection, after left, but not right auditory cortex disruption (Heimrath et al., 2014).

Spectrotemporal modulations

A theoretically powerful way to integrate these findings is by considering how these patterns fit with models of spectrotemporal modulation. Many neurophysiological studies exist showing that the response properties of auditory cortical neurons across species are well described in terms of joint sensitivity to spectral and temporal modulations found in the stimulus (Shamma, 2001). This mechanism is thought to enable efficient encoding of complex real-world sounds (Singh and Theunissen, 2003), especially those that are an important part of the animal's communicative repertoire (Gehr et al., 2000; Woolley et al., 2005). Sensitivity to spectrotemporal modulations in auditory cortex has also been described using both neuroimaging (Schonwiesner and Zatorre, 2009; Santoro et al., 2014; Venezia et al., 2019) and intracortical recordings in humans (Mesgarani et al., 2014; Hullett et al., 2016).

Two recent studies have brought together the research questions surrounding hemispheric differences with the spectrotemporal modulation hypothesis to yield evidence that functional asymmetries map well onto this theoretical framework. One study (Flinker et al., 2019) used MEG to measure brain activity associated either with the verbal content, or the timbre (male vs. female voice, which is largely based on spectral cues) of spoken sentences. Behaviorally, they reported that when temporal modulations were filtered out, speech comprehension was affected but vocal timbre was not, and vice-versa for filtering of spectral modulations. The imaging data showed greater left auditory cortex response for the temporal cues in speech, and a right, albeit weaker lateralization effect for the spectral cues. The second study (Albouy et al., 2020) used sung sentences whose speech and melodic content had been fully orthogonalized ensuring independence of the two types of cues (**Figure 1**). Behavioral data showed a double dissociation such that degradation of temporal cues affected comprehension of the words to the song but not the melody, whereas degradation of spectral cues affected discrimination of the melodies but had no effect on the speech component. The functional imaging data reflected the behavioral data in that speech content could only be decoded from left auditory cortex, but was abolished by temporal degradation, whereas melodic content could only be decoded from right auditory cortex, but was abolished by spectral degradation.

These converging findings from experiments using different techniques strongly support the idea that left and right auditory cortices are linked to heightened resolution in temporal and spectral modulation, respectively. This explanation fits with a broader idea that the nervous system optimizes its representations according to the properties of the physical environment that are most relevant, as has been proposed for vision (Simoncelli and Olshausen, 2001), and for speech (Gervain and Geffen, 2019). I suggest expanding this concept to encompass hemispheric asymmetries on the grounds that humans have two main auditory communication systems, speech and music (Zatorre et al., 2002; Mehr et al., 2021), and that they each exploit, to some extent at least, opposite ends of the temporal-spectral continuum; so the best way to accommodate the competing requirements of the two types of signals is by segregating the necessary specializations within each hemisphere. Thus, rather than think in terms of specializations at the cognitive domain level (speech vs. music), we can reconceptualize it in terms of specialization at the acoustical feature level.

This interpretation predicts that the two domains are lateralized only to the extent that they make greater use of one or another of those cues. We need to keep in mind that both music and speech utilize both temporal and spectral modulations. In the case of speech, spectral modulations are important in carrying prosodic information, and, in tonal languages, lexical information. Interestingly, a good amount of evidence suggests that prosodic processing depends more on right-hemisphere structures, in accord with our model (Sammler et al., 2015). These kinds of spectral cues are important for some aspects of communication, but they do not seem to be quite as important for speech comprehension as those driven by temporal modulations, based on the fact that degradation of temporal but not spectral cues abolishes speech comprehension in the two studies mentioned earlier (Flinker et al., 2019; Albouy et al., 2020). That conclusion was already known from an early influential study (Shannon et al., 1995) that demonstrated that comprehension was well-preserved when normal speech was replaced by amplitude-modulated noise passed through as few as three or four filter banks centered at different frequencies. This procedure degraded the spectral content but preserved most of the temporal modulations. Indeed, this property is what enables cochlear implants to transmit comprehensible speech despite poor representation of spectral modulations due to the limited number of channels available.

Music also contains both spectral and temporal modulations. The latter are obviously critical for transmitting information that can be used to perceive rhythm and metrical organization, and hence the importance of temporal modulations may vary depending on the nature of the music and of the instruments used to generate it (e.g., percussion vs.

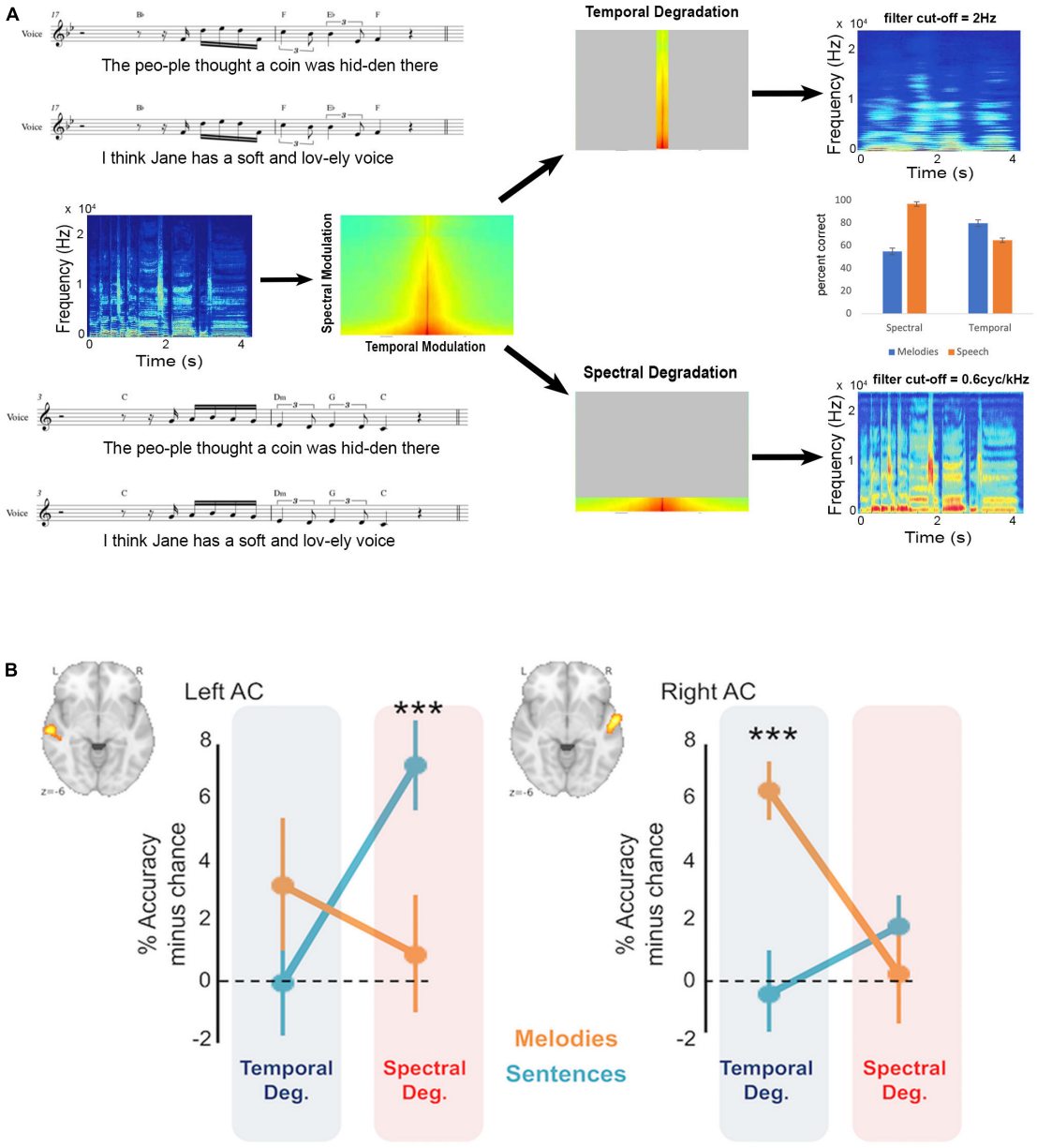


FIGURE 1
Behavioral and neural effects of spectrotemporal degradation in music and speech. **(A)** Sung stimuli (music notation) consisted of the same tunes sung to different phrases or vice-versa, yielding an orthogonal set of songs with matched melodic and speech content. These songs (spectrogram and spectrotemporal plots in middle panel) were then degraded either in the temporal domain, leaving spectral modulation intact (top), or vice-versa (bottom). The effect of this manipulation can be seen in the resulting spectrograms (right side) where the temporal degradation smears the temporal information but leaves spectral information intact, while spectral degradation smears the spectral information but leaves temporal information intact. The behavioral result (middle panel bar graph) shows that behavioral performance for melodic content is severely reduced after spectral compared to temporal degradation (blue bars) while performance for speech is reduced after temporal compared to spectral degradation (orange bars). **(B)** In the left auditory cortex, functional MRI classification performance for decoding speech content is reduced to chance only after temporal degradation; while in the right auditory cortex functional MRI classification performance for decoding melodic content is reduced to chance only after spectral degradation, paralleling the behavioral effects shown in panel **(A)**. Adapted with permission from Albouy et al. (2020). ***Refers to significantly above chance performance.

song). Moreover, pitch information and temporal information interact in interesting, complex ways in music cognition (Jones, 2014). So it is simplistic to think of the two dimensions are entirely independent of one another. However, the fact remains that, as mentioned above, the poor spectral resolution that can be observed with congenital amusia seems to lead to a more global inability to learn the relevant rules of music, and results in a fairly global deficit. So this observation would argue that

even if both types of cues are present and important for music, spectral cues seem to play a more prominent role.

Top-down effects

One might conclude from all the foregoing that hemispheric differences are driven *exclusively* by low-level acoustical features. But that does not seem to be the whole story. There are in fact numerous experiments showing that even when acoustics are held constant, hemispheric responses can be modulated. A good example is provided by studies showing that sine-wave speech analogs elicit left auditory cortex responses only after training that led to them being perceived as speech, and not in the naive state when they were perceived as just weird sounds (Liebenthal et al., 2003; Dehaene-Lambertz et al., 2005; Möttönen et al., 2006). A complementary phenomenon can be seen with speech sounds that when looped repeatedly begin to sound like music (Deutsch et al., 2011; Falk et al., 2014). Once the stimulus was perceived as music, more brain activity was seen in some right-hemisphere regions that were not detected before the perceptual transformation (Tierney et al., 2013). Tracking of pitch contours in speech can also shift from right to bilateral auditory regions as a function of selective attention (Brodbeck and Simon, 2022).

These kinds of results have sometimes been interpreted as evidence in favor of domain-specific models, on the grounds that bottom-up mechanisms cannot explain the results since the inputs are held constant in those studies. However, given the strength of the findings reviewed above that spectrotemporal tuning is asymmetric, another way to interpret these effects is that they represent interactions between feedforward and feedback systems that interconnect auditory areas with higher-order processing regions, especially in the frontal cortex. Although this idea remains to be worked out in any detail, it would be compatible with known control functions of the frontal cortex, which is reciprocally connected with auditory cortical processing streams.

The idea that interactions occur between ascending, stimulus-driven responses, and descending, more cognitive influences can also be thought of in the context of predictive coding models (Friston, 2010). A great deal of work has recently been devoted to this framework, which essentially proposes that perception is enabled by the interface between predictions generated at higher levels of the hierarchy that influence stimulus-driven encoding processes at lower levels of the hierarchy. When the latter signals do not match the prediction, an error signal is generated, which can be used for updating of the internal model (that is, learning). These models have gained prominence because they can explain many phenomena not easily accounted for by more traditional bottom-up driven models of perception, even if they also raise questions that are not yet fully answered (Heilbron and Chait, 2018).

As applied to the question at hand, the idea would be that as a complex stimulus like speech or music is being processed, continuous predictions and confirmations/errors would be generated at different levels of the system. Depending on the spectrotemporal content of the signal, neuronal networks in the left or right auditory cortex would predominate in the initial processing; but as top-down predictions are generated that are based on higher-order features, then the activity could shift from one side to another. So, in the case of sine-wave speech for instance, initial, naïve processing would presumably involve right auditory cortex since the stimulus contains a great deal of spectral modulation. But once the listener is able to apply top-down control to disambiguate how those sounds could fit into a linguistic pattern, then more language-relevant predictions would be generated that could inhibit spectral-based processing in favor of temporal-based processing. By the same token, hemispheric differences could be amplified by these interactions even if initial processing differences in early parts of the auditory system are only slightly asymmetric. This scenario remains largely speculative at the moment, but at least sets up some testable hypotheses for future research.

Author contributions

RZ wrote the manuscript and approved the submitted version.

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.

Publisher's note

References

Albouy, P., Benjamin, L., Morillon, B., and Zatorre, R. J. (2020). Distinct sensitivity to spectrotemporal modulation supports brain asymmetry for speech and melody. *Science* 367:1043. doi: 10.1126/science.aaz3468

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.-E., et al. (2013). Impaired pitch perception and memory in congenital amusia: The deficit starts in the auditory cortex. *Brain* 136, 1639–1661. doi: 10.1093/brain/awt082

Albouy, P., Mattout, J., Sanchez, G., Tillmann, B., and Caclin, A. (2015). Altered retrieval of melodic information in congenital amusia: Insights from dynamic causal modeling of MEG data. *Front. Hum. Neurosci.* 9:20. doi: 10.3389/fnhum.2015.00020

Bianchi, F., Hjortkjaer, J., Santurette, S., Zatorre, R. J., Siebner, H. R., and Dau, T. (2017). Subcortical and cortical correlates of pitch discrimination: Evidence for two levels of neuroplasticity in musicians. *Neuroimage* 163, 398–412. doi: 10.1016/j.neuroimage.2017.07.057

Brodbeck, C., and Simon, J. Z. (2022). Cortical tracking of voice pitch in the presence of multiple speakers depends on selective attention. *Front. Neurosci.* 16:828546. doi: 10.3389/fnins.2022.828546

Cha, K., Zatorre, R. J., and Schonwiesner, M. (2016). Frequency selectivity of voxel-by-voxel functional connectivity in human auditory cortex. *Cereb. Cortex* 26, 211–224. doi: 10.1093/cercor/bhu193

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., and Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7:11070. doi: 10.1038/ncomms11070

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., and Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *Neuroimage* 24, 21–33. doi: 10.1016/j.neuroimage.2004.09.039

Deutsch, D., Henthorn, T., and Lapidis, R. (2011). Illusory transformation from speech to song. *J. Acoust. Soc. Am.* 129, 2245–2252. doi: 10.1121/1.3562174

Falk, S., Rathcke, T., and Dalla Bella, S. (2014). When speech sounds like music. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 1491–1506. doi: 10.1037/a0036858

Flinker, A., Doyle, W. K., Mehta, A. D., Devinsky, O., and Poeppel, D. (2019). Spectrotemporal modulation provides a unifying framework for auditory cortical asymmetries. *Nat. Hum. Behav.* 3, 393–405. doi: 10.1038/s41562-019-0548-z

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Gehr, D. D., Komiya, H., and Eggermont, J. J. (2000). Neuronal responses in cat primary auditory cortex to natural and altered species-specific calls. *Hear. Res.* 150, 27–42. doi: 10.1016/s0378-5955(00)00170-2

Gervain, J., and Geffen, M. N. (2019). Efficient neural coding in auditory and speech perception. *Trends Neurosci.* 42, 56–65. doi: 10.1016/j.tins.2018.09.004

Heilbron, M., and Chait, M. (2018). Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience* 389, 54–73. doi: 10.1016/j.neuroscience.2017.07.061

Heimrath, K., Kuehne, M., Heinze, H. J., and Zaehle, T. (2014). Transcranial direct current stimulation (tDCS) traces the predominance of the left auditory cortex for processing of rapidly changing acoustic information. *Neuroscience* 261, 68–73. doi: 10.1016/j.neuroscience.2013.12.031

Hullett, P. W., Hamilton, L. S., Mesgarani, N., Schreiner, C. E., and Chang, E. F. (2016). Human superior temporal gyrus organization of spectrotemporal modulation tuning derived from speech stimuli. *J. Neurosci.* 36, 2014–2026. doi: 10.1523/JNEUROSCI.1779-15.2016

Hyde, K. L., Lerch, J. P., Zatorre, R. J., Griffiths, T. D., Evans, A. C., and Peretz, I. (2007). Cortical thickness in congenital amusia: When less is better than more. *J. Neurosci.* 27, 13028–13032. doi: 10.1523/JNEUROSCI.3039-07.2007

Hyde, K. L., and Peretz, I. (2004). Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360. doi: 10.1111/j.0956-7976.2004.00683.x

Hyde, K. L., Peretz, I., and Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* 46, 632–639. doi: 10.1016/j.neuropsychologia.2007.09.004

Hyde, K. L., Zatorre, R. J., and Peretz, I. (2011). Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299. doi: 10.1093/cercor/bhq094

Jamison, H. L., Watkins, K. E., Bishop, D. V., and Matthews, P. M. (2006). Hemispheric specialization for processing auditory nonspeech stimuli. *Cereb. Cortex* 16, 1266–1275. doi: 10.1093/cercor/bhj068

Johnsrude, I. S., Penhune, V. B., and Zatorre, R. J. (2000). Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* 123(Pt 1), 155–163. doi: 10.1093/brain/123.1.155

Jones, M. R. (2014). "Dynamics of musical patterns: How do melody and rhythm fit together?," in *Psychology and music*, eds T. J. Tighe and W. J. Dowling (London: Psychology Press), 67–92.

Liebenthal, E., Binder, J. R., Piorkowski, R. L., and Remez, R. E. (2003). Short-term reorganization of auditory analysis induced by phonetic experience. *J. Cogn. Neurosci.* 15, 549–558. doi: 10.1162/089892903321662930

Liégeois-Chauvel, C., Peretz, I., Babaï, M., Laguitton, V., and Chauvel, P. (1998). Contribution of different cortical areas in the temporal lobes to music processing. *Brain* 121, 1853–1867. doi: 10.1093/brain/121.10.1853

Liu, B.-H., Wu, G. K., Arbuckle, R., Tao, H. W., and Zhang, L. I. (2007). Defining cortical frequency tuning with recurrent excitatory circuitry. *Nat. Neurosci.* 10, 1594–1600. doi: 10.1038/nn2012

Loui, P., Alsop, D., and Schlaug, G. (2009). Tone deafness: A new disconnection syndrome? *J. Neurosci.* 29, 10215–10220. doi: 10.1523/JNEUROSCI.1701-09.2009

Manning, L., and Thomas-Antérion, C. (2011). Marc Dax and the discovery of the lateralisation of language in the left cerebral hemisphere. *Rev. Neurol.* 167, 868–872. doi: 10.1016/j.neurol.2010.10.017

Mehr, S. A., Krasnow, M. M., Bryant, G. A., and Hagen, E. H. (2021). Origins of music in credible signaling. *Behav. Brain Sci.* 44:e60.

Mesgarani, N., Cheung, C., Johnson, K., and Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. *Science* 343, 1006–1010. doi: 10.1126/science.1245994

Milner, B. (1962). "Laterality effects in audition," in *Interhemispheric relations and cerebral dominance*, ed. V. B. Mountcastle (Baltimore, MD: The Johns Hopkins Press), 177–195.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., Tuomainen, J., et al. (2006). Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. *Neuroimage* 30, 563–569. doi: 10.1016/j.neuroimage.2005.10.002

Obleser, J., Eisner, F., and Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *J. Neurosci.* 28, 8116–8123. doi: 10.1523/JNEUROSCI.1290-08.2008

Okamoto, H., and Kakigi, R. (2015). Hemispheric asymmetry of auditory mismatch negativity elicited by spectral and temporal deviants: A magnetoencephalographic study. *Brain Topogr.* 28, 471–478.

Patel, A. D. (2010). *Music, language, and the brain*. New York, NY: Oxford university press.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191. doi: 10.1016/S0896-6273(01)00580-3

Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G., and Belin, P. (2015). Dorsal and ventral pathways for prosody. *Curr. Biol.* 25, 3079–3085. doi: 10.1016/j.cub.2015.10.009

Samson, S., and Zatorre, R. J. (1988). Melodic and harmonic discrimination following unilateral cerebral excision. *Brain Cogn.* 7, 348–360. doi: 10.1016/0278-2626(88)90008-5

Santoro, R., Moerel, M., De Martino, F., Goebel, R., Ugurbil, K., Yacoub, E., et al. (2014). Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput. Biol.* 10:e1003412. doi: 10.1371/journal.pcbi.1003412

Schönwiesner, M., Rübsamen, R., and Von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* 22, 1521–1528. doi: 10.1111/j.1460-9568.2005.04315.x

Schonwiesner, M., and Zatorre, R. J. (2009). Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14611–14616. doi: 10.1073/pnas.0907682106

Shamma, S. (2001). On the role of space and time in auditory processing. *Trends Cogn. Sci.* 5, 340–348. doi: 10.1016/S1364-6613(00)01704-6

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303

Sihvonen, A. J., Särkämö, T., Rodríguez-Fornells, A., Ripollés, P., Münte, T. F., and Soinila, S. (2019). Neural architectures of music–insights from acquired amusia. *Neurosci. Biobehav. Rev.* 107, 104–114.

Simoncelli, E. P., and Olshausen, B. A. (2001). Natural image statistics and neural representation. *Annu. Rev. Neurosci.* 24, 1193–1216.

Singh, N. C., and Theunissen, F. E. (2003). Modulation spectra of natural sounds and ethological theories of auditory processing. *J. Acoust. Soc. Am.* 114, 3394–3411. doi: 10.1121/1.1624067

Tierney, A., Dick, F., Deutsch, D., and Sereno, M. (2013). Speech versus song: Multiple pitch-sensitive areas revealed by a naturally occurring musical illusion. *Cereb. Cortex* 23, 249–254. doi: 10.1093/cercor/bhs003

Tillmann, B., Lévêque, Y., Fornoni, L., Albouy, P., and Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Res.* 1640, 251–263. doi: 10.1016/j.brainres.2015.10.035

Venezia, J. H., Thurman, S. M., Richards, V. M., and Hickok, G. (2019). Hierarchy of speech-driven spectrotemporal receptive fields in human auditory cortex. *Neuroimage* 186, 647–666. doi: 10.1016/j.neuroimage.2018.11.049

Woolley, S. M. N., Fremouw, T. E., Hsu, A., and Theunissen, F. E. (2005). Tuning for spectro-temporal modulations as a mechanism for auditory discrimination of natural sounds. *Nat. Neurosci.* 8, 1371–1379. doi: 10.1038/nn1536

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *J. Acoust. Soc. Am.* 84, 566–572. doi: 10.1121/1.396834

Zatorre, R. J., and Baum, S. R. (2012). Musical melody and speech intonation: Singing a different tune. *PLoS Biol.* 10:e1001372.

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946

Zatorre, R. J., Belin, P., and Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends Cogn. Sci.* 6, 37–46.

Zatorre, R. J., Delhommeau, K., and Zarate, J. M. (2012). Modulation of auditory cortex response to pitch variation following training with microtonal melodies. *Front. Psychol.* 3:544. doi: 10.3389/fpsyg.2012.00544

frontiers | Frontiers in Neuroscience

TYPE Mini Review
PUBLISHED 15 December 2022
DOI 10.3389/fnins.2022.1076374

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Sonja A. Kotz,
Maastricht University, Netherlands
Eliane Schochat,
University of São Paulo, Brazil
Erik Edwards,
Zeit Medical, Inc., United States

*CORRESPONDENCE
Sophie K. Scott
sophie.scott@ucl.ac.uk
Kyle Jasmin
kyle.jasmin@rhul.ac.uk

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

RECEIVED 21 October 2022
ACCEPTED 28 November 2022
PUBLISHED 15 December 2022

CITATION
Scott SK and Jasmin K (2022) Rostro-caudal networks for sound processing in the primate brain. *Front. Neurosci.* 16:1076374. doi: 10.3389/fnins.2022.1076374

Rostro-caudal networks for sound processing in the primate brain

Sophie K. Scott[1]* and Kyle Jasmin[2]*

[1]Institute of Cognitive Neuroscience, University College London, London, United Kingdom, [2]Department of Psychology, Royal Holloway, University of London, Egham, United Kingdom

Sound is processed in primate brains along anatomically and functionally distinct streams: this pattern can be seen in both human and non-human primates. We have previously proposed a general auditory processing framework in which these different perceptual profiles are associated with different computational characteristics. In this paper we consider how recent work supports our framework.

KEYWORDS

speech perception, auditory cortex, neuroanatomy, auditory recognition, sensorimotor processing

> "Hearing is a form of touch. You feel it through your body, and sometimes it almost hits your face".
>
> — Evelyn Glennie
>
> "Intermittently she caught the gist of his sentences and supplied the rest from her subconscious, as one picks up the striking of a clock in the middle with only the rhythm of the first uncounted strokes lingering in the mind".
>
> — F. Scott Fitzgerald, Tender is the Night

Auditory processing in primates is neuroanatomically and functionally bifurcated. There are several models of speech and auditory processing in the human brain built around this principle (Alain et al., 2001; Hickok and Poeppel, 2004; Rauschecker and Scott, 2009; Jasmin et al., 2019), which originated in work on non-human primates (NHP). The NHP literature showed that rostral and caudal auditory cortical fields have distinctly different patterns of anatomical connectivity and different functional properties. For example, cells in rostral superior temporal sulcus were shown to be sensitive to the different kinds of non-human primate vocalizations (recognizing "monkey calls") while those in the caudal fields were sensitive to the spatial location of the vocalizations (Rauschecker and Tian, 2000). These different functions have been described as "what" and "where/how" pathways within the rostral and caudal fields, respectively. Thus it was discovered that, in the visual system, auditory perception entails more than one kind of processing, with more than one functional goal.

This discovery was transformational for functional imaging studies of human speech processing, not just in terms of the neuroanatomical findings, but because it indicated

that different speech perception tasks might recruit different elements of the auditory perception network depending on the task. Tasks that required speech recognition networks, such as single word and sentence perception, consistently show recruitment of rostral temporal lobe fields (Mummery et al., 1999; Wise et al., 1999; Scott et al., 2000). By contrast, tasks that required motor engagement—e.g., speaking aloud, reading aloud in synchrony with other people (Jasmin et al., 2016), or when one's own voice is acoustically altered during speech production (Meekings and Scott, 2021), caudal auditory fields in humans are recruited. There is also a clear role for caudal auditory fields in representing the spatial location of voices (Hunter et al., 2002): all of these findings are consistent with a role for posterior auditory fields in guiding action.

Auditory neuroscience has made strides to move beyond mere description to computational mechanisms. Indeed, there have been significant advances in our understanding of the potential computational properties that underlie the functional differences seen in rostral/caudal auditory fields. In terms of anatomical connectivity, work by Scott et al. (2017) has shown convincingly that rostral and caudal auditory core, belt and parabelt areas receive different inputs from thalamic nuclei, which follows a caudal-rostral distinction: caudal auditory areas receive input mainly from the auditory thalamus and from the somatosensory thalamus (Hackett et al., 2007): moving rostrally, the medial geniculate body (the auditory thalamic input) drops, proportionally, and rostral auditory fields receive proportionally more input from the medial pulvinar, which receives input from the ascending visual pathway. Moving from caudal to rostral fields, the proportion of responses from subnuclei of the medial geniculate body also changes—from a ventral medial geniculate body dominance in caudal and mid-core auditory cortex, to a rough equivalence of inputs from the ventral medial geniculate body and the posterior dorsal medial geniculate body. Given the sheer complexity of the mammalian ascending auditory pathway, an important step in exploring the computational basis of different patterns of auditory processing is going to entail engaging with the nature of the representations of sound in these cortico-thalamic interactions. Some work on the stimulation of brain stem nuclei has suggested that there may even be processing pathways as early as the cochlear nucleus that have critical importance for speech perception (Moore and Shannon, 2009).

Scott et al. (2011) also showed that the caudal core field (A1) shows more detailed temporal response characteristics than the rostral temporal core area (RT): Neurons in caudal A1 respond faster to the onsets of sounds than rostral RT, and they are also accurate at tracking both fast and slow amplitude modulations. This stands in contrast to rostral RT, which responds more slowly to sound onsets and can only track slower amplitude modulations. Recent electrocorticography (ECoG) in humans are consistent with this macaque findings. Across human auditory cortex, regardless of the nature of the auditory stimuli, the neural responses in caudal auditory fields are fast, transient, and linked to the onsets of sounds, while the neural responses in rostral auditory fields are slow and sustained (Hamilton et al., 2018). We argued in 2019 that these findings suggested a critical role for neuronal temporal responses in different kinds of computational processes on incoming sounds. In caudal fields, the responses to sound onsets are fast and temporally accurate, but not sustained, as responses that are critical to the control of action would need to be. By contrast, in rostral fields, the responses to sound onsets are slow and sustained, which potentially reflects hierarchical patterns of perceptual processing that interact with higher order linguistic and predictive processes.

This work has been recently replicated and extended in humans using fMRI. Zulfiqar et al. (2021) modeled fMRI BOLD responses for different temporal and spectral characteristics of the responses to stimuli. They found that caudal belt regions of the auditory cortex showed responses to natural sound stimuli that were fast but not frequency specific, responding to a broad spectral range. In contrast, rostral belt regions showed more specific spectral responses, and slower onset responses. Further support for this comes from another ECoG paper from Hamilton et al. (2021), which reported the shortest onset responses (generally less than 100 ms) in caudal Heschl's Gyrus (the location of primary auditory cortex in humans) and posterior superior temporal gyrus fields, and longer onset responses (up to 500 ms) in anterior superior temporal gyrus fields and the planum polare.

These findings strongly suggest that, as we hypothesized in 2019, the caudal/posterior "what/how" auditory pathway is underpinned by distinct computational processes from those of the anterior/rostral "what" pathway. Caudal fields (core and non-core) have responses that are generally fast, transient, and not necessarily specifically associated with particular stimulus characteristics: The responses in rostral fields (core and non-core) are generally slow and sustained and can be much more driven by stimulus specific properties. These distinctions are generalities—as can be seen in the Hamilton et al. (2021) paper, there is some overlap of these responses, but the general pattern is clear: Fast transient caudal responses reflect feed forward networks which are critical to the fast sensory guidance of action; slow, sustained responses in rostral fields likely reflect recognition processes which are slower as they require feedback processes from higher order language areas, which can have a profound effect on speech intelligibility (Obleser et al., 2007). This pattern reflects the overall cortical thickness gradient in the temporal lobes, such that primary auditory cortex is thin, with fewer feedback connections that cross cortical layers, whereas moving rostrally the cortex is thicker and has a higher ratio of feedback connections (Wagstyl et al., 2015).

Several studies have now shown that the rostral recognition "what" pathway, seen for intelligibility in speech, is not only seen for speech: music and other identifiable environmental sounds

also recruit the anterior temporal lobes in humans. There is compelling evidence that sound recognition is processed by parallel and distinct streams within these anterior fields (Norman-Haignere et al., 2015, 2022; Boebinger et al., 2021). This strongly suggests that while speech may often appear to dominate in these regions, that may be a function of the predominance of studies that focus on speech, and of the well-established speech processing problems that arise due to damage in left middle temporal artery territory. Using non-speech stimuli can show how speech fits within a wider range of auditory stimuli—in a recent ECoG study, song showed greater responses than speech or instrumental music within these fields (Norman-Haignere et al., 2022). However, a computational framework based on the temporal response properties we have described could be applied to a wide range of auditory stimuli—not necessarily specific to speech, as we have discussed (Jasmin et al., 2019). A challenge for further studies will be to determine the degree to which speech, song, instrumental music and other sound sources recruit distinct pathways, and what the computational properties are that may underlie these. This is all the more critical since there is good evidence that when we hear sounds in normal environments, they are rarely in silence, and rostral auditory areas seem to be key for simultaneously representing different sound sources (Evans et al., 2016).

These different auditory perceptual networks also interact with distributed systems throughout the human brain, including both other perceptual networks (including visual, somatosensory systems), and non-perceptual (including linguistic, emotional, musical networks): In many everyday auditory environments one would imagine that both auditory pathways are continually recruited. For example, during conversational speech, we have suggested that the rostral pathway is recruited to process the voice of the other speaker, feeding into language networks that are also engaged in generating a response, while the caudal pathway is recruited to track the features of the other speaker's voice (e.g., the rate and the rhythm), such that the planned response is aligned with the talkers voice and a smooth turn taking can managed (Scott et al., 2009). Auditory perception requires multiple kinds of perceptual processes, because the brain needs both to track the meaning of our auditory environments and to guide our production of sound into those environments.

Author contributions

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

Conflict of interest

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

Publisher's note

References

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., and Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proc. Natl. Acad. Sci. U.S.A.* 98, 12301–12306. doi: 10.1073/pnas.211209098

Boebinger, D., Norman-Haignere, S., McDermott, J., and Kanwisher, N. (2021). Music-selective neural populations arise without musical training. *J. Neurophysiol.* 125, 2237–2263. doi: 10.1152/jn.00588.2020

Evans, S., McGettigan, C., Agnew, Z., Rosen, S., and Scott, S. (2016). Getting the cocktail party started: Masking effects in speech perception. *J. Cogn. Neurosci.* 28, 483–500. doi: 10.1162/jocn_a_00913

Hackett, T. A., De La Mothe, L. A., Ulbert, I., Karmos, G., Smiley, J., and Schroeder, C. E. (2007). Multisensory convergence in auditory cortex, II. Thalamocortical connections of the caudal superior temporal plane. *J. Comp. Neurol.* 502, 924–952. doi: 10.1002/cne.21326

Hamilton, L. S., Edwards, E., and Chang, E. F. (2018). A spatial map of onset and sustained responses to speech in the human superior temporal gyrus. *Curr. Biol.* 28, 1860–1871. doi: 10.1016/j.cub.2018.04.033

Hamilton, L. S., Oganian, Y., Hall, J., and Chang, E. F. (2021). Parallel and distributed encoding of speech across human auditory cortex. *Cell* 184, 4626–4639. doi: 10.1016/j.cell.2021.07.019

Hickok, G., and Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99. doi: 10.1016/j.cognition.2003.10.011

Hunter, M. D., Griffiths, T. D., Farrow, T. F., Zheng, Y., Wilkinson, I. D., Hegde, N., et al. (2002). A neural basis for the perception of voices in external auditory space. *Brain* 126, 161–169. doi: 10.1093/brain/awg015

Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., et al. (2016). Cohesion and joint speech: Right hemisphere contributions to synchronized vocal production. *J. Neurosci.* 36, 4669–4680. doi: 10.1523/JNEUROSCI.4075-15.2016

Jasmin, K., Lima, C. F., and Scott, S. K. (2019). Understanding rostral–caudal auditory cortex contributions to auditory perception. *Nat. Rev. Neurosci.* 20, 425–434. doi: 10.1038/s41583-019-0160-2

Meekings, S., and Scott, S. K. (2021). Error in the superior temporal gyrus? A systematic review and activation likelihood estimation meta-analysis of speech production studies. *J. Cogn. Neurosci.* 33, 422–444. doi: 10.1162/jocn_a_01661

Moore, D. R., and Shannon, R. V. (2009). Beyond cochlear implants: Awakening the deafened brain. *Nat. Neurosci.* 12, 686–691. doi: 10.1038/nn.2326

Mummery, C. J., Ashburner, J., Scott, S. K., and Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* 106, 449–457. doi: 10.1121/1.427068

Norman-Haignere, S., Feather, J., Boebinger, D., Brunner, P., Ritaccio, A., McDermott, J., et al. (2022). A neural population selective for song in human auditory cortex. *Curr. Biol.* 32, 1470–1484. doi: 10.1016/j.cub.2022.01.069

Norman-Haignere, S., Kanwisher, N., and McDermott, J. (2015). Distinct cortical pathways for music and speech revealed by hypothesis-free voxel decomposition. *Neuron* 88, 1281–1296. doi: 10.1016/j.neuron.2015.11.035

Obleser, J., Wise, R. J., Dresner, M. A., and Scott, S. K. (2007). Functional integration across brain regions improves speech perception under adverse listening conditions. *J. Neurosci.* 27, 2283–2289. doi: 10.1523/JNEUROSCI.4663-06.2007

Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724. doi: 10.1038/nn.2331

Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing of "what" and "where" in auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806. doi: 10.1073/pnas.97.22.11800

Scott, B. H., Malone, B. J., and Semple, M. N. (2011). Transformation of temporal processing across auditory cortex of awake macaques. *J. Neurophysiol.* 105, 712–730. doi: 10.1152/jn.01120.2009

Scott, B. H., Saleem, K. S., Kikuchi, Y., Fukushima, M., Mishkin, M., and Saunders, R. C. (2017). Thalamic connections of the core auditory cortex and rostral supratemporal plane in the macaque monkey. *J. Comp. Neurol.* 525, 3488–3513. doi: 10.1002/cne.24283

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400

Scott, S., McGettigan, C., and Eisner, F. (2009). A little more conversation, a little less action–candidate roles for the motor cortex in speech perception. *Nat. Rev. Neurosci.* 10, 295–302. doi: 10.1038/nrn2603

Wagstyl, K., Ronan, L., Goodyer, I. M., and Fletcher, P. C. (2015). Cortical thickness gradients in structural hierarchies. *Neuroimage* 111, 241–250. doi: 10.1016/j.neuroimage.2015.02.036

Wise, R. J., Greene, J., Büchel, C., and Scott, S. K. (1999). Brain regions involved in articulation. *Lancet* 353, 1057–1061. doi: 10.1016/s0140-6736(98)07491-1

Zulfiqar, I., Havlicek, M., Moerel, M., and Formisano, E. (2021). Predicting neuronal response properties from hemodynamic responses in the auditory cortex. *Neuroimage* 244:118575. doi: 10.1016/j.neuroimage.2021.118575

frontiers | Frontiers in Neuroscience

TYPE Perspective
PUBLISHED 16 March 2023
DOI 10.3389/fnins.2023.1081295

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Anahita Mehta,
University of Michigan, United States

*CORRESPONDENCE
David McAlpine
✉ david.mcalpine@mq.edu.au

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

RECEIVED 27 October 2022
ACCEPTED 17 February 2023
PUBLISHED 16 March 2023

CITATION
McAlpine D and de Hoz L (2023) Listening loops and the adapting auditory brain. *Front. Neurosci.* 17:1081295. doi: 10.3389/fnins.2023.1081295

Listening loops and the adapting auditory brain

David McAlpine[1]* and Livia de Hoz[2,3]

[1]Department of Linguistics, Macquarie University, Sydney, NSW, Australia, [2]Neuroscience Research Center, Charité – Universitätsmedizin Berlin, Berlin, Germany, [3]Bernstein Center for Computational Neuroscience, Berlin, Germany

Analysing complex auditory scenes depends in part on learning the long-term statistical structure of sounds comprising those scenes. One way in which the listening brain achieves this is by analysing the statistical structure of acoustic environments over multiple time courses and separating background from foreground sounds. A critical component of this statistical learning in the auditory brain is the interplay between feedforward and feedback pathways—"listening loops"—connecting the inner ear to higher cortical regions and back. These loops are likely important in setting and adjusting the different cadences over which learned listening occurs through adaptive processes that tailor neural responses to sound environments that unfold over seconds, days, development, and the life-course. Here, we posit that exploring listening loops at different scales of investigation—from *in vivo* recording to human assessment—their role in detecting different timescales of regularity, and the consequences this has for background detection, will reveal the fundamental processes that transform hearing into the essential task of listening.

KEYWORDS

auditory, listen, loops, feedback, adaptation, prediction

The act of listening

Our brain is continuously interpreting the soundscape, it is listening even when we are not. Listening is essential to understanding. Without listening, sound is meaningless to us—a wash of noise, reflections, and competing sources vying for our attention. Many of our listening environments are challenging—from restaurants to railway stations, we listen in complex, multi-sensory and multi-dimensional spaces. Compared to even the most advanced listening technologies, however, we navigate these spaces with relative ease, and it is not obvious how we do so. We evolved to deal with listening in an embodied manner but our experimental approaches, and often our listening technologies, pay little regard to the immersive and embodied qualities of listening. A reductionist approach to our exploration of the listening brain will limit the development of algorithms, devices, and therapies that seek to establish or re-establish listening—in humans and machines—as an immersive experience.

Here, we posit that advancing our understanding of the listening brain requires a reframing of our investigative neuroscience to include both the multi-layered soundscape with its noisy background as well as its complex foreground. In doing so, we will have to contend with the complexities of an extensive neural circuit and the specific features of the auditory pathway—evident from cochlea to cortex and back—the "listening loops" responsible for setting the cadences of our listening lives (Winer, 2005; Asilador and Llano, 2020). Sensitivity to salient foreground acoustic cues is important for processing speech information, for example, but background features such as multi-talker babble or the flurry

of late-arriving reflections from walls and other surfaces in a room also need to be integrated into our listening experience. Exploring how the listening brain parses background features of the soundscape is critical to survival—fight or flight—since this sensitivity to the statistical structure of background sounds may also enhance our capacity to attend to foreground sounds. Here we posit that studying the mechanisms underlying the detection and coding of the background is essential to understand listening (McWalter and McDermott, 2018). What are the statistics of the background that facilitate its detection? How is it coded? What is the role of feedback? And on which time scale? How and when does its coding depend on contextual information (spatial context, movement, visual stimuli)? Learning the longer-term statistical structure of acoustic environments involves an interplay between feedforward and feedback pathways—the listening loops—including to the level of the inner ear, which takes us directly to the issue of how to explore listening through, and in the context of the complex neural circuits that constitutes the auditory brain. Though afferent, or feedforward, pathways in the auditory brain are rightly considered vital at the juncture between hearing and cognition, feedback (efferent) fibres outnumber feedforward in the auditory brain to influence every station in the pathway, including mechanical and neural structures within the middle and inner ear (Saldaña et al., 1996; Terreros and Delano, 2015). The functional understanding of these cortico-subcortical loops lags well behind our knowledge of their anatomy. Overall, it seems reasonable to assume that the act of listening arises from activity generated in a rich subcortical network replete with bilateral and feedback connectivity, and that this activity operates over progressively wider time windows along the ascending pathway (Ding et al., 2016; Kell and McDermott, 2019; Asokan et al., 2021; Henin et al., 2021), with feedback from relatively higher centres in the auditory pathway modulating neural activity at lower centres over potentially progressively longer epochs (Robinson et al., 2016; **Figure 1**). Understanding the functional role of cortico-subcortical listening loops in the human brain could support the many autonomous listening devices—from hearing aids and cochlear implants to Amazon's "Alexa"—that currently provide little of the capacity of human listening abilities. Striving for signal fidelity on millisecond, and even sub-millisecond, timescales, they often struggle to perform in even moderately noisy environments, and fail to operate over the multiple, and much slower, cadences of listening that make effective communication possible. The dominance of rapid signal-processing techniques in the development of hearing technologies and therapies, also surfaces in machine-learning and artificial intelligence approaches to listening. Performance remains distinctly *subpar* but progress on this front will be critical if autonomous listening devices.

Tools for exploring listening loops

If we are to take advantage of listening loops to explore the timescales over which sensory information is integrated in the auditory brain (Ding et al., 2016) we may need to implement some new tools to do so. One difficulty in studying cortico-subcortical loops has been the sampling and targeting of, not only deep-sitting neurons, but also those specifically involved in the loop. Thanks to the development of genetic tools, combined with the creation of manipulation tools, we can now opto- and chemo-genetically target deep and superficial cells specifically involved in cortico-subcortical interactions in awake rodents (Clayton et al., 2021; Souffi et al., 2021). The use of the newly developed large-scale high-density recording probes (Jun et al., 2017) allows to record activity from deep and superficial neurons simultaneously across structures (Kleinfeld et al., 2019). The advent of brain-imaging techniques, with the potential to sample from wide populations of neurons within and across brain structures (Bathellier et al., 2012; Silva, 2017), has put hearing on a more equal footing to other sensory systems, particularly vision, for which an understanding of cortical structure and function was well advanced through *in vivo* experimentation (Hübener and Bonhoeffer, 2005). Imaging of the auditory brain has rapidly advanced from employing simple sounds that build on our understanding of sensory reception and the importance of spectral analysis—tonotopy is widely accepted as the primary representation of the cochlea (Marin et al., 2022)—to more-naturalistic listening assessments permitted by advances in audio technologies (Filipchuk et al., 2022). The downside of current brain-imaging techniques, however, is that they still favour a cortico-centric perspective, with some exceptions (Barnstedt et al., 2015), at a time when subcortical structures and efferent pathways are increasingly understood to be critical to the act of listening (Cruces-Solís et al., 2018). In *in vivo* experimental settings, two-photon imaging is generally confined to the exploration of cortical structures—though this is changing with the implementation of mesoscale imaging techniques. However, access to subcortical structures—some deep within the brainstem—as well investigations of the efferent pathways, remain limited, especially in humans. Further, many practical limitations of imaging arise beyond the inability to access subcortical structures. The (dangerously loud) sounds generated by magnetic resonance imaging (MRI) scanners pose a specific challenge to structural and functional investigations of the listening brain *per se*, but MRI as well as magnetoencephalography (MEG) are contra-indicated for the use of the very listening devices that might provide powerful insights to hearing and listening in health and disease.

Listening loops and the adapting brain

Exploring the auditory brain in terms of listening loops conditioned for effective sensing and communication with the outside world is, in fact, how the research field is starting to align (Bajo et al., 2010; Robinson et al., 2016; Weible et al., 2020; Yudintsev et al., 2021; Wang et al., 2022), powered by a combination of new technologies applied generally across sensory neuroscience (e.g. Zingg et al., 2017; Williamson and Polley, 2019), and a specific re-imagining of the structure and function of subcortical auditory structures (Xiong et al., 2015; Bidelman et al., 2018; Lohse et al., 2021). Freed from a cortico-centric approach, the concept of listening loops provides the time-dimensional perspective to understanding, or at least exploring, the different cadences of

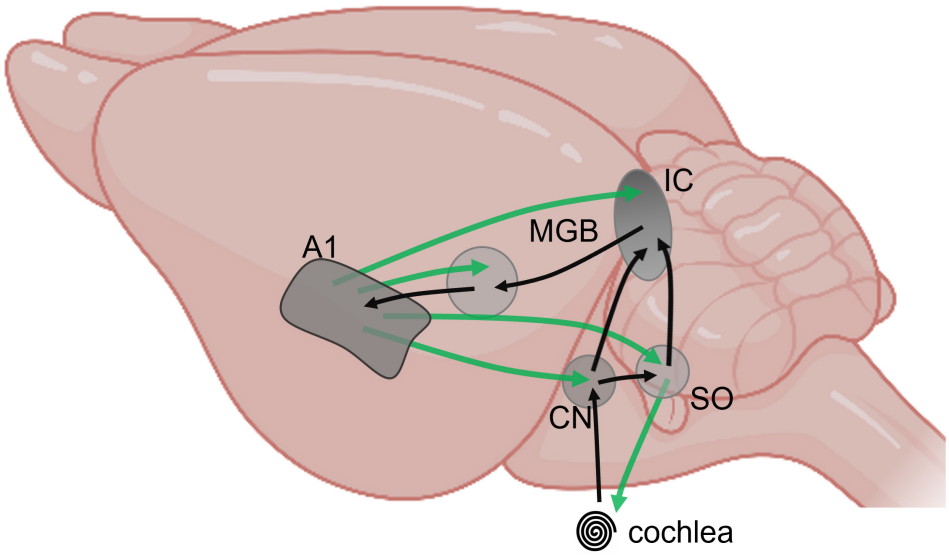


FIGURE 1
Schematic representation of major feedforward (black) and feedback (green) pathways between subcortical and cortical auditory structures, from cochlea to primary cortex, projected on a mouse brain (BioRender).

listening (Antunes and Malmierca, 2021; Homma and Bajo, 2021), and connects with the well-developed concept of the predictive brain. Indeed, despite the technical challenges of accessing sub-cortical structures, the concept of listening loops that operate over distinct feedforward and feedback pathways provides an excellent framework in which to investigate fundamental principles of brain processing such as predictive coding that might be applied to other sensory systems, not generally a role the auditory system has performed.

One means by which the temporal dynamics of the listening brain might be investigated, including its capacity for prediction, is by assessing how it adapts over time to enhance the flow of information (Latimer et al., 2019). We can define adaptation to mean changes (usually a reduction) in neural firing in response to sustained stimulation, though definitions of the term are plentiful. From a functional perspective, firing-rate adaptation seems important in the listening brain's ability to adjust dynamically to the listening environments in response to changes in that environment, or in response to internal changes that alter its overall sensitivity or dynamics. Adaptive coding is a common phenomenon throughout the brain, and a recent review article provides an excellent primer for understanding the different cadences over which adaptation in the auditory brain unfolds, from the range of milliseconds to over the life-course, as well as potential mechanisms by which these cadences are set or arise (Willmore and King, 2022).

Continuous adaptation within listening loops likely sets and adjusts the cadences over which learned listening occurs, tailoring neural responses to sound environments that unfold over seconds, days, development, and the life-course. Exploring these loops in the context of the adapting auditory brain—from single neurons in animal models to human behavioural assessments—will help us understand the immersive quality of listening, as well as advance the many technologies currently available or under development that purport to listen to us.

Data availability statement

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

Author contributions

Both authors conceptualised, wrote, and edited the manuscript.

Funding

Einstein Foundation (Grant number: EVF-2021-618) funded the research for which LH was the local applicant and DM was the visiting fellow.

Conflict of interest

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

Publisher's note

References

Antunes, F. M., and Malmierca, M. S. (2021). Corticothalamic pathways in auditory processing: Recent advances and insights from other sensory systems. *Front. Neural Circuits* 15:721186. doi: 10.3389/fncir.2021.721186

Asilador, A., and Llano, D. A. (2020). Top-down inference in the auditory system: Potential roles for corticofugal projections. *Front. Neural Circuits* 14:615259. doi: 10.3389/fncir.2020.615259

Asokan, M. M., Williamson, R. S., Hancock, K. E., and Polley, D. B. (2021). Inverted central auditory hierarchies for encoding local intervals and global temporal patterns. *Curr. Biol.* 31, 1762–1770.e4. doi: 10.1016/j.cub.2021.01.076

Bajo, V. M., Nodal, F. R., Moore, D. R., and King, A. J. (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat. Neurosci.* 13, 253–260. doi: 10.1038/nn.2466

Barnstedt, O., Keating, P., Weissenberger, Y., King, A. J., and Dahmen, J. C. (2015). Functional microarchitecture of the mouse dorsal inferior colliculus revealed through in vivo two-photon calcium imaging. *J. Neurosci.* 35, 10927–10939. doi: 10.1523/JNEUROSCI.0103-15.2015

Bathellier, B., Ushakova, L., and Rumpel, S. (2012). Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* 76, 435–449. doi: 10.1016/j.neuron.2012.07.008

Bidelman, G. M., Davis, M. K., and Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hear. Res.* 367, 149–160. doi: 10.1016/j.heares.2018.05.018

Clayton, K. K., Williamson, R., Hancock, K., Tasaka, G., Mizrahi, A., Hackett, T., et al. (2021). Auditory corticothalamic neurons are recruited by motor preparatory inputs. *Curr. Biol.* 31, 310–321.e5. doi: 10.1016/j.cub.2020.10.027

Cruces-Solís, H., Jing, Z., Babaev, O., Rubin, J., Gür, B., Krueger-Burg, D., et al. (2018). Auditory midbrain coding of statistical learning that results from discontinuous sensory stimulation. *PLoS Biol.* 16:e2005114. doi: 10.1371/journal.pbio.2005114

Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nat. Neurosci.* 19, 158–164.

Filipchuk, A., Schwenkgrub, J., Destexhe, A., and Bathellier, B. (2022). Awake perception is associated with dedicated neuronal assemblies in the cerebral cortex. *Nat. Neurosci.* 25, 1327–1338. doi: 10.1038/s41593-022-01168-5

Henin, S., Turk-Browne, N. B., Friedman, D., Liu, A., Dugan, P., Flinker, A., et al. (2021). Learning hierarchical sequence representations across human cortex and hippocampus. *Sci. Adv.* 7:eabc4530. doi: 10.1126/sciadv.abc4530

Homma, N. Y., and Bajo, V. M. (2021). Lemniscal corticothalamic feedback in auditory scene analysis. *Front. Neurosci.* 15:723893. doi: 10.3389/fnins.2021.723893

Hübener, M., and Bonhoeffer, T. (2005). Visual cortex: Two-photon excitement. *Curr. Biol.* 15, R205–R208.

Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236.

Kell, A. J. E., and McDermott, J. H. (2019). Invariance to background noise as a signature of non-primary auditory cortex. *Nat. Commun.* 10:3958. doi: 10.1038/s41467-019-11710-y

Kleinfeld, D., Luan, L., Mitra, P. P., Robinson, J. T., Sarpeshkar, R., Shepard, K., et al. (2019). Can one concurrently record electrical spikes from every neuron in a mammalian brain? *Neuron* 103, 1005–1015. doi: 10.1016/j.neuron.2019.08.011

Latimer, K. W., Barbera, D., Sokoletsky, M., Awwad, B., Katz, Y., Nelken, I., et al. (2019). Multiple timescales account for adaptive responses across sensory cortices. *J. Neurosci.* 39, 10019–10033. doi: 10.1523/JNEUROSCI.1642-19.2019

Lohse, M., Dahmen, J. C., Bajo, V. M., and King, A. J. (2021). Subcortical circuits mediate communication between primary sensory cortical areas in mice. *Nat. Commun.* 12:3916. doi: 10.1038/s41467-021-24200-x

Marin, N., Lobo Cerna, F., and Barral, J. (2022). Signatures of cochlear processing in neuronal coding of auditory information. *Mol. Cell. Neurosci.* 120:103732. doi: 10.1016/j.mcn.2022.103732

McWalter, R., and McDermott, J. H. (2018). Adaptive and selective time averaging of auditory scenes. *Curr. Biol.* 28, 1405–1418.e10. doi: 10.1016/j.cub.2018.03.049

Robinson, B. L., Harper, N. S., and McAlpine, D. (2016). Meta-adaptation in the auditory midbrain under cortical influence. *Nat. Commun.* 7:13442. doi: 10.1038/ncomms13442

Saldaña, E., Feliciano, M., and Mugnaini, E. (1996). Distribution of descending projections from primary auditory neocortex to inferior colliculus mimics the topography of intracollicular projections. *J. Comp. Neurol.* 371, 15–40. doi: 10.1002/(SICI)1096-9861(19960715)371:1<15::AID-CNE2>3.0.CO;2-O

Silva, A. C. (2017). Anatomical and functional neuroimaging in awake, behaving marmosets. *Dev. Neurobiol.* 77, 373–389.

Souffi, S., Nodal, F. R., Bajo, V. M., and Edeline, J.-M. (2021). When and how does the auditory cortex influence subcortical auditory structures? New insights about the roles of descending cortical projections. *Front. Neurosci.* 15:690223. doi: 10.3389/fnins.2021.690223

Terreros, G., and Delano, P. H. (2015). Corticofugal modulation of peripheral auditory responses. *Front. Syst. Neurosci.* 9:134. doi: 10.3389/fnsys.2015.00134

Wang, X., Zhang, Y., Zhu, L., Bai, S., Li, R., Sun, H., et al. (2022). Selective corticofugal modulation on sound processing in auditory thalamus of awake marmosets. *Cereb. Cortex*. bhac278. doi: 10.1093/cercor/bhac278

Weible, A. P., Yavorska, I., and Wehr, M. (2020). A cortico-collicular amplification mechanism for gap detection. *Cereb. Cortex* 30, 3590–3607. doi: 10.1093/cercor/bhz328

Williamson, R. S., and Polley, D. B. (2019). Parallel pathways for sound processing and functional connectivity among layer 5 and 6 auditory corticofugal neurons. *Elife* 8:e42974. doi: 10.7554/eLife.42974

Willmore, B. D. B., and King, A. J. (2022). Adaptation in auditory processing. *Physiol. Rev.* 103, 1025–1058. doi: 10.1152/physrev.00011.2022

Winer, J. A. (2005). Decoding the auditory corticofugal systems. *Hear. Res.* 207, 1–9.

Xiong, X. R., Liang, F., Zingg, B., Ji, X., Ibrahim, L. A., Tao, H. W., et al. (2015). Auditory cortex controls sound-driven innate defense behaviour through corticofugal projections to inferior colliculus. *Nat. Commun.* 6:7224. doi: 10.1038/ncomms8224

Yudintsev, G., Asilador, A. R., Sons, S., Sekaran, N. V., Coppinger, M., Nair, K., et al. (2021). Evidence for layer-specific connectional heterogeneity in the mouse auditory corticocollicular system. *J. Neurosci.* 41, 9906–9918. doi: 10.1523/JNEUROSCI.2624-20.2021

Zingg, B., Chou, X., Zhang, Z., Mesik, L., Liang, F., Tao, H. W., et al. (2017). AAV-mediated anterograde transsynaptic tagging: Mapping corticocollicular input-defined neural pathways for defense behaviors. *Neuron* 93, 33–47. doi: 10.1016/j.neuron.2016.11.045

frontiers | Frontiers in Neuroscience

TYPE Perspective
PUBLISHED 07 February 2023
DOI 10.3389/fnins.2023.1077344

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
David R. Moore,
Cincinnati Children's Hospital Medical Center, United States

*CORRESPONDENCE
Timothy D. Griffiths
✉ t.d.griffiths@ncl.ac.uk

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

RECEIVED 22 October 2022
ACCEPTED 23 January 2023
PUBLISHED 07 February 2023

CITATION
Griffiths TD (2023) Predicting speech-in-noise ability in normal and impaired hearing based on auditory cognitive measures.
Front. Neurosci. 17:1077344.
doi: 10.3389/fnins.2023.1077344

Predicting speech-in-noise ability in normal and impaired hearing based on auditory cognitive measures

Timothy D. Griffiths*

Biosciences Institute, Newcastle University Medical School, Newcastle upon Tyne, United Kingdom

Problems with speech-in-noise (SiN) perception are extremely common in hearing loss. Clinical tests have generally been based on measurement of SiN. My group has developed an approach to SiN based on the auditory cognitive mechanisms that subserve this, that might be relevant to speakers of any language. I describe how well these predict SiN, the brain systems for them, and tests of auditory cognition based on them that might be used to characterise SiN deficits in the clinic.

KEYWORDS

auditory cognition, speech in noise, behaviour, cortex, brain

Introduction

The ability to hear speech in noisy listening situations is the most important aspect of natural listening carried out by humans. Problems with speech-in-noise (SiN) are ubiquitous in peripheral hearing loss due to cochlear damage, and also in common brain disorders including stroke and dementia. SiN ability is dependent on cochlear function and can be predicted to an extent by the audiogram, but also depends on cortical analysis: even for aspects of auditory pattern analysis that are independent of language.

From first principles, SiN might depend on mechanisms that allow separation of foreground from background elements, the grouping together of foreground elements over time, selective attention to these, and linguistic analysis. I focus here on auditory cognitive mechanism that are responsible for the first two processes. This represents an effort to characterise mechanisms beyond the cochlea for the detection of sound that might account for the large variance in SiN ability that is not due to the audiogram. I do not dismiss the importance and relevance of linguistic factors: the aim of the exercise is to define generic brain mechanisms relevant to speakers of any language of any ability. The data suggest a large amount of the variance can be defined in this way. I will describe behavioural measures of auditory cognition that predict SiN ability and the brain basis for these.

Clinically, behavioural and brain measures of this level of auditory cognition provide a potential means to characterise cortical mechanisms for auditory cognition that explain variation in the SiN listening that is not accounted for by the audiogram. Such measures have potential use in the prediction of hearing outcome after restoration by hearing aids and cochlear implantation. They will not replace SiN tests clinically but suggest a means to partition the causes of SiN impairment that might guide intervention and rehabilitation.

Auditory cognitive mechanisms for speech-in-noise analysis

Listening to speech in noise is complicated even at the level of auditory analysis before linguistic processing. Speech is a complex broadband signal that contains features in frequency-time space that change over time. This must be separated from background noise that overlaps in frequency and in time. **Figure 1** shows a stimulus developed by my group to define the ability of subjects to carry out the figure-ground separation that is required at the initial stage of SiN analysis. The ground part of the stimulus is based on tonal elements that are distributed randomly in frequency-time space. At a certain point in time, we constrain a certain number of elements to remain constant from one time frame to the next. When there are enough elements that are on for long enough subjects hear a figure that emerges from the ground stimulus.

In the original version of the stimulus (Teki et al., 2013) it is impossible to say whether a figure is present or not based on the distribution of tonal elements over frequency at a single time point. The perceptual mechanism therefore requires a basis that operates over time. One possibility is a local mechanism based on adaptation within the frequency bands that remain constant. An argument against this is the fact that detection of the figure increases as a similar function of the number of elements, irrespective of whether the time window is 25 or 50 ms: an adaptation mechanism would be expected to depend on the absolute duration of the figure. Further evidence against the adaptation model is provided in our original study based on manipulations of the stimulus including placing broadband noise in alternate time frames and using "ramped" stimuli containing a systematic change over time in the frequencies comprising the figure. The detection mechanism is robust to these manipulations. We have developed a model for the process based on a mechanism that "looks" at the activity in auditory cortical neurons tuned to different frequencies to seek coherence between their activity (Teki et al., 2013). We can derive a single metric corresponding to the coherence between all frequency bands that predicts psychophysical performance well. I will argue in the next section that this process first occurs in high-level auditory cortex.

Work on listeners without a history of hearing symptoms has demonstrated correlation between the detection of SiN in the form of sentences in noise and both audiometry and figure-ground analysis (Holmes and Griffiths, 2019). The subjects were "normal listeners" (defined in terms of the average hearing levels over frequency) but showed variable threshold increases in the high-frequency (4–8 kHz) range of the audiogram. We demonstrated a significant correlation between the high-frequency audiogram and SiN ability. A version of the original figure-detection task showed a weak correlation with SiN ability with marginal significance, and a version of the task that required discrimination of figures based on a feature (a temporal gap in the figure) showed a medium correlation that was significant ($r = 0.32$, $p \leq 0.01$, $n = 97$). Essentially, subjects had to discriminate two intervals containing the same figure with and without a gap in the middle. Hierarchal regression demonstrated that the audiogram and figure discrimination tasks together accounted for approximately half of the explainable variance in SiN and that the audiogram and figure-ground tasks accounted for independent variance.

The figure-ground task can also be used to assess cross-frequency grouping mechanisms in subjects with electrical hearing. **Figure 2** shows the relationship between figure-ground detection and hearing sentences in noise in 47 subjects with cochlear implants. The implants were a mixture of conventional long electrodes that stimulate most of the cochlear partition and short electrodes that preserve low frequency acoustic hearing and stimulate the high-frequency basal region. We tested using a figure with components in the range above 1 kHz that was always in the electrical range even for users with the short devices. We see greater effect size ($r = 0.45$, $p < 0.01$, $n = 47$) for the relationship between figure-ground analysis and SiN compared to normal listeners, which is remarkable given that the figures are in a restricted range that does not include the whole speech range. Multiple linear regression demonstrated a significant effect of figure detection (standardised beta 0.29, $p < 0.05$) even after accounting for spectral modulation discrimination and temporal modulation detection as measures of cochlear function. A model containing all three of these non-linguistic factors accounted for 46% of the variance in SiN ability.

The tests of figure-ground analysis in both normal hearing and hearing-impaired listeners demonstrate a mechanism that explains variance in SiN ability independently of peripheral encoding of the stimulus. The idea is that a central grouping mechanism allows "pop out" or the formation of an auditory gestalt as a central process operating after peripheral analysis. The brain basis and clinical application of this work is considered below. This process is plausibly related to the perception of individual words in noise (although the experiments all measured correlations at the sentence-in-noise level). Mechanisms that contribute to the grouping together of words in sentences might also correlate with sentence-in-noise ability but not words in noise. The importance of grouping at this level was first suggested by a link between phonological working memory (WM) and SiN (Akeroyd, 2008; Dryden et al., 2017; Kim et al., 2020). From first principles, sentence comprehension has to require phonological WM at some point to allow the elements of the sentence to form the whole sentence. For sentences in noise, mechanisms might be based on separation of each word from noise followed by linking together of the elements by phonological WM, or linking together of the elements to form a "sentence gestalt" that is separate from the background noise. Correlation between phonological WM ability and SiN is better explained by the second mechanism. Debate about this correlation has centred on whether it holds in all listeners or whether age and hearing status moderate the relationship (Füllgrabe and Rosen, 2016). Moreover, there is ongoing discussion about the degree to which traditional phonological WM tasks depend on language skills (Schwering and MacDonald, 2020). We have been interested to develop non-linguistic paradigms to assess mechanisms that assess the grouping of acoustic elements over the timescale of sentences (seconds) that might contribute to SiN skill.

Recent work by my group examined the relationship between WM for non-verbal sounds and sentences in noise in a group of young listeners (Lad et al., 2020). The studies estimate WM capacity based on the precision with which non-speech sounds are held in memory. We use a delayed adjustment paradigm in which participants hear a sound, and then after a delay of several seconds adjust a second sound to match the sound heard in memory. The reciprocal of the standard deviation of the adjusted sound measures the precision of memory. In the context of distributed resource models of WM that have been applied to the visual and auditory domains (Bays and Husain, 2008; Kumar et al., 2013), this yields a measure of the resource available for WM: the greater the resource, the greater the precision. The study of Lad et al. (2020) showed a significant correlation between ability for sentences in

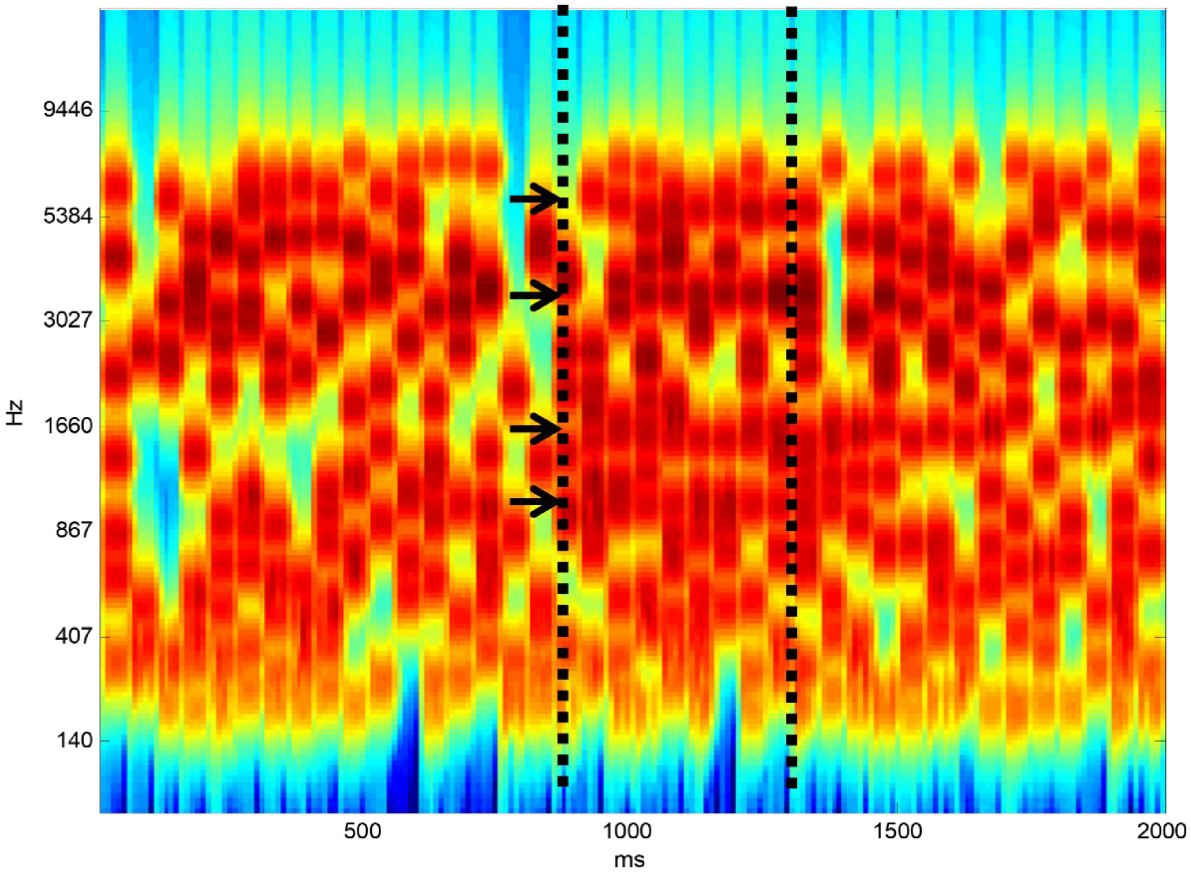


FIGURE 1
Figure-ground stimulus. The "ground" stimulus contains random elements distributed in frequency-time space. At a certain point in time, shown by the first vertical dashed line, a certain number of elements (four here–shown by arrows) are constrained to remain constant from one time frame to the next. If there are enough elements and they are on for long enough a "figure" emerges perceptually from the background.

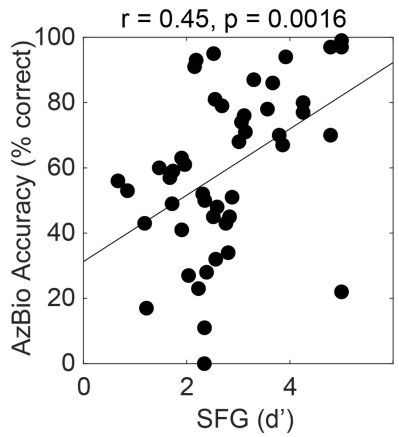


FIGURE 2
Correlation between performance for the figure ground task (SFG, stochastic figure ground) and a widely used US measure of sentence-in-noise perception, AzBio. The data are for 47 cochlear-implant users (see text).

noise (subjects had to match heard sentences to a matrix of written possibilities) and frequency precision ($r = 0.36$, $p < 0.05$, $n = 44$) but not amplitude-modulation-rate precision (where precision is defined for both measures based on the distribution of responses as above). No correlation was seen between the audiogram and the sentence-in-noise task. A possible interpretation of the dissociation between frequency and modulation precision is in terms of a critical importance of WM for the grouping of sources (like voices) that need to be yoked together in the foreground and are defined by stable frequency properties, as opposed to shorter events (like words). But further data on over 100 listeners has shown a significant correlation with WM for both frequency and amplitude-modulation rate with a moderate effect size (Meher Lad–unpublished observation). This suggests an alternative possibility–that there might be a common WM resource for storing acoustic features that is a determinant of SiN ability.

As an aside, the study of Lad et al. (2022) also examined links between musicality and acoustic WM and found a significant correlation. There is a longstanding debate about whether musicians have greater perceptual abilities relevant to music such as frequency discrimination: see (Moore et al., 2019) for a recent investigation and discussion. In a further study (Lad et al., 2022) the correlation between musicianship (based on the Goldsmith's Musical Sophistication Index) and perceptual discrimination and WM for frequency was examined. The study showed a correlation with frequency WM but not perception of frequency. The data highlight an interesting specific link between musicality and WM for frequency for which a causal relationship could in principle be in either direction.

Brain mechanisms for auditory cognition relevant to speech-in-noise perception

The behavioural experiments above suggest auditory cognitive bases for speech in noise at a pre-linguistic level relevant to words in noise (figure-ground analysis) and sentences in noise (WM for sound). These explain variance in SiN separate to that associated with cochlear measures. Linguistic factors represent another cause of variance in SiN. There are strong priors that implicate cortex in these mechanisms. In the case of figure-ground analysis a mechanism is required "looks" between widely separated frequencies that are, in general, represented in separate neurons in the ascending auditory pathway and primary auditory cortex. This suggests a mechanism in cortex beyond primary auditory cortex. In the case of WM for non-speech sounds a basis in auditory cortex or frontal cortex might be considered, based on early studies using musical stimuli (Zatorre et al., 1994).

Studies of the brain basis for figure-ground analysis have been based on a primate model that we have developed and human studies using functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG) and invasive neurophysiology. The studies support a system based on high-level auditory cortex and parietal cortex.

Unlike SiN, studies of the underlying auditory cognitive bases for SiN that do not require linguistic processing can be studied in the macaque. Macaques have a similar frequency range for hearing to humans (Heffner and Heffner, 1986), and a similar lower limit of pitch (Joly et al., 2014). The auditory cortex is situated in the superior temporal plane at the top of the temporal lobe as in humans. The model allows systematic neurophysiology in a way that would never be possible in humans. Recordings of multiunit activity demonstrate tonic responses to figure onset that are present in all three auditory core areas in the superior temporal plane (Schneider et al., 2021). We can also carry out fMRI in the macaque to measure BOLD activity in neuronal ensembles: this allows a direct comparison with human studies. The macaque studies show activity associated with figure perception in high-level cortex over the lateral part of the superior temporal gyrus in the superior temporal lobe in the region of parabelt cortex, at a higher level in the auditory hierarchy than the core areas (Schneider et al., 2018).

Human fMRI also demonstrates BOLD activity over the lateral part of the superior temporal lobe corresponding to the presence of figures (Teki et al., 2011), in a human homolog of auditory parabelt. The human work also shows activity in the intraparietal sulcus. MEG has also demonstrated tonic response to figure onset arising from auditory cortex and intraparietal sulcus (Teki et al., 2016). Local-field-potential recordings from twelve neurosurgical candidates have demonstrated local-field potentials to figure onset that arise from early auditory cortex (human core homologs in the superior temporal plane) and high frequency oscillatory activity in the gamma band that arise from the lateral part of the superior temporal lobe (Gander et al., 2017).

Studies of the brain basis for acoustic WM analysis have been based on human studies using fMRI, and invasive neurophysiology. The studies support a system based on auditory cortex, inferior frontal cortex and the hippocampus. The studies have been based on a paradigm in which subjects hear two tones and are required to remember one after a retro-cue. After a delay of seconds, they are required to recall the tone.

fMRI has shown activity in auditory cortex during the memory maintenance period of this paradigm (Kumar et al., 2016). The activity was present in human core and belt homologs in superior temporal plane. This is not surprising but is not a given based on studies of visual WM using a similar retro-cue in which decoding of memory content from delay activity in visual cortex was possible but where activity levels did not increase (Harrison and Tong, 2009). I would also point out "activity silent" models for WM maintenance in which WM maintenance is based on synaptic strength rather than ongoing activity *per se* (e.g., Wolff et al., 2017). fMRI also demonstrated involvement of the inferior frontal cortex in WM maintenance (Kumar et al., 2016), which is also not surprising given the musical studies referred to above.

What was less anticipated in the fMRI WM study was the involvement of the hippocampus in WM maintenance (Kumar et al., 2016). The hippocampus is conventionally regarded as a part of the system for episodic rather than WM. We needed to use a long delay period in the fMRI study because of the sluggish BOLD response and one idea is that episodic measures might have been engaged during the BOLD experiment. But we have now carried out six sets of intracranial recordings on neurosurgical candidates with a much shorter delay and demonstrated consistent low-frequency oscillatory activity during WM maintenance in medial temporal lobe structures: hippocampus and parahippocampal gyrus (Kumar et al., 2021). Readers interested in a general account of how the computational machinery of the hippocampus might be used for auditory analysis are referred to Billig et al. (2022).

In summary, although fundamental auditory cognition relevant to SiN analysis need not explicitly engage the language system, the auditory-pattern analysis required engages cortical mechanism well beyond what is conventionally regarded as auditory cortex. I suggest that a complete account of the brain bases for speech in noise needs to consider auditory cognition in addition to higher-level linguistic analysis.

Final comments: Possible clinical implications

I have developed an argument that problems with auditory cognitive mechanisms explain difficulties with speech in noise that cannot be accounted for by the audiogram. Defined in this way, auditory cognitive deficits might be considered a type of "hidden hearing loss." The area is controversial. Hidden hearing loss is sometimes used as a synonym for cochlear synaptopathy (Kujawa and Liberman, 2009): the loss of synapses between inner hair cells and the afferent auditory nerve caused by noise exposure. Valderrama et al. (2022) consider other possible bases including auditory nerve demyelination and elevated central gain and mal-adaptation in brainstem auditory centres. Despite the controversy, the debate about bases for hidden hearing loss has consistently focussed on mechanisms in the ascending pathway. The cortical mechanisms I have described here add another level of complexity. Further work is required to examine the contribution of cortical figure-ground analysis and acoustic WM to SiN when both conventional

measures of cochlear function and measures of hidden hearing loss due to brainstem factors are taken into account.

A major driver of this work is to develop new behavioural and brain tools that might allow better prediction of the potential success of hearing restoration using hearing aids or cochlear implants. The stimulus in **Figure 1** might be thought of as an audiogram for acoustic scene analysis that might realistically be used alongside conventional pure tone and speech audiograms in the audiology clinic. We have developed simple brain measures of figure-ground analysis based on EEG that could also be used in any clinical centre (Guo et al., 2022).

Finally, understanding of auditory cognition relevant to speech in noise can potentially shed light on how hearing loss in middle life explains 9% of dementia cases (Livingston et al., 2017, 2020). We consider possible models in Griffiths et al. (2020), including the idea that this might be due to interaction between high-level mechanisms for auditory cognition beyond the auditory cortex, that are stressed by natural listening in subjects with hearing loss, and the pathological processes responsible for dementia. The idea that follows is that speech in noise and its auditory cognitive determinants, rather than simple hearing loss, is the critical determinant of dementia.

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.

Ethics statement

The studies involving human participants were reviewed and approved by the IRB, University of Iowa Hospitals and Clinics. The patients/participants provided their written informed consent to participate in this study. The animal study was reviewed and approved by Home Office Project Licence to A. Thiele.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

This study was supported by the Wellcome Trust, United Kingdom (WT106964MA and 210567/Z/18/Z), the Medical Research Council, United Kingdom (MR/T032553/1), and the National Institutes of Health, United States (2R01DC004290 and 5P50DC000242-33).

Acknowledgments

These ideas evolved from work and discussion with my excellent colleagues in Newcastle, UCL and Iowa: F. Balezeau, E. Benzaquen, J. Berger, A. Billig, I. Choi, P. Gander, B. Gantz, X. Guo, M. Hansen, E. Holmes, M. Howard, Y. Kikuchi, S. Kumar, M. Lad, E. Maguire, B. McMurray, C. Petkov, F. Schneider, W. Sedley, and A. Thiele.

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.

Publisher's note

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int J Audiol* 47 Suppl 2, S53–71. doi: 10.1080/14992020802301142

Bays, P. M., and Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science* 321, 851–854. doi: 10.1126/science.1158023

Billig, A. J., Lad, M., Sedley, W., and Griffiths, T. D. (2022). The hearing hippocampus. *Prog. Neurobiol.* 218:102326. doi: 10.1016/j.pneurobio.2022.102326

Dryden, A., Allen, H. A., Henshaw, H., and Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends Hear.* 21:2331216517744675. doi: 10.1177/2331216517744675

Füllgrabe, C., and Rosen, S. (2016). On the (Un)importance of working memory in speech-in-noise processing for listeners with normal hearing thresholds. *Front. Psychol.* 7:1268. doi: 10.3389/fpsyg.2016.01268

Gander, P., Kumar, S., Nourski, K., Oya, H., Kawasaki, H., Howard, M., et al. (2017). *Direct electrical recordings of neural activity related to auditory figure-ground segregation in the human auditory cortex. Nashville*. Nashville, TN: Association for Research in Otolaryngology.

Griffiths, T. D., Lad, M., Kumar, S., Holmes, E., Mcmurray, B., Maguire, E. A., et al. (2020). How can hearing loss cause dementia? *Neuron* 108, 401–412. doi: 10.1016/j.neuron.2020.08.003

Guo, X., Dheerendra, P., Benzaquén, E., Sedley, W., and Griffiths, T. D. (2022). EEG responses to auditory figure-ground perception. *Hear Res.* 422:108524. doi: 10.1016/j.heares.2022.108524

Harrison, S. A., and Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458, 632–635. doi: 10.1038/nature07832

Heffner, H. E., and Heffner, R. S. (1986). Hearing loss in Japanese macaques following bilateral auditory cortex lesions. *J. Neurophysiol.* 55, 256–271. doi: 10.1152/jn.1986.55.2.256

Holmes, E., and Griffiths, T. D. (2019). 'Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. *Sci. Rep.* 9:16771. doi: 10.1038/s41598-019-53353-5

Joly, O., Baumann, S., Poirier, C., Patterson, R. D., Thiele, A., and Griffiths, T. D. (2014). A perceptual pitch boundary in a non-human primate. *Front. Psychol.* 5:998. doi: 10.3389/fpsyg.2014.00998

Kim, S., Choi, I., Schwalje, A. T., Kim, K., and Lee, J. H. (2020). Auditory working memory explains variance in speech recognition in older listeners under adverse listening conditions. *Clin. Interv. Aging* 15, 395–406. doi: 10.2147/CIA.S241976

Kujawa, S. G., and Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J. Neurosci.* 29, 14077–14085. doi: 10.1523/JNEUROSCI.2845-09.2009

Kumar, S., Gander, P. E., Berger, J. I., Billig, A. J., Nourski, K. V., Oya, H., et al. (2021). Oscillatory correlates of auditory working memory examined with human electrocorticography. *Neuropsychologia* 150:107691. doi: 10.1016/j.neuropsychologia.2020.107691

Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., and Griffiths, T. D. (2016). A brain system for auditory working memory. *J. Neurosci.* 36, 4492–4505. doi: 10.1523/JNEUROSCI.4341-14.2016

Kumar, S., Joseph, S., Pearson, B., Teki, S., Fox, Z. V., Griffiths, T. D., et al. (2013). Resource allocation and prioritization in auditory working memory. *Cogn. Neurosci.* 4, 12–20. doi: 10.1080/17588928.2012.716416

Lad, M., Billig, A. J., Kumar, S., and Griffiths, T. D. (2022). A specific relationship between musical sophistication and auditory working memory. *Sci. Rep.* 12:3517. doi: 10.1038/s41598-022-07568-8

Lad, M., Holmes, E., Chu, A., and Griffiths, T. D. (2020). Speech-in-noise detection is related to auditory working memory precision for frequency. *Sci. Rep.* 10:13997. doi: 10.1038/s41598-020-70952-9

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., et al. (2020). Dementia prevention, intervention, and care: 2020 report of the lancet commission. *Lancet* 396, 413–446. doi: 10.1016/S0140-6736(20)30367-6

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734. doi: 10.1016/S0140-6736(17)31363-6

Moore, B. C. J., Wan, J., Varathanathan, A., Naddell, S., and Baer, T. (2019). No effect of musical training on frequency selectivity estimated using three methods. *Trends Hear.* 23:2331216519841980. doi: 10.1177/2331216519841980

Schneider, F., Balezeau, F., Distler, C., Kikuchi, Y., Van Kempen, J., Gieselmann, A., et al. (2021). Neuronal figure-ground responses in primate primary auditory cortex. *Cell Rep.* 35:109242. doi: 10.1016/j.celrep.2021.109242

Schneider, F., Dheerendra, P., Balezeau, F., Ortiz-Rios, M., Kikuchi, Y., Petkov, C. I., et al. (2018). Auditory figure-ground analysis in rostral belt and parabelt of the macaque monkey. *Sci. Rep.* 8:17948. doi: 10.1038/s41598-018-36903-1

Schwering, S. C., and MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Front. Hum. Neurosci.* 14:68.

Teki, S., Barascud, N., Picard, S., Payne, C., Griffiths, T. D., and Chait, M. (2016). Neural correlates of auditory figure-ground segregation based on temporal coherence. *Cereb. Cortex* 26, 3669–3680. doi: 10.1093/cercor/bhw173

Teki, S., Chait, M., Kumar, S., Shamma, S., and Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife* 2:e00699.

Teki, S., Chait, M., Kumar, S., Von Kriegstein, K., and Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *J. Neurosci.* 31, 164–171. doi: 10.1523/JNEUROSCI.3788-10.2011

Valderrama, J. T., De La Torre, A., and Mcalpine, D. (2022). The hunt for hidden hearing loss in humans: From preclinical studies to effective interventions. *Front. Neurosci.* 16:1000304. doi: 10.3389/fnins.2022.1000304

Wolff, M. J., Jochim, J., Akyurek, E. G., and Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nat. Neurosci.* 20, 864–871. doi: 10.1038/nn.4546

Zatorre, R. J., Evans, A. C., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908–1919. doi: 10.1523/JNEUROSCI.14-04-01908.1994

frontiers | Frontiers in Psychology

TYPE Review
PUBLISHED 01 September 2022
DOI 10.3389/fpsyg.2022.967260

OPEN ACCESS

EDITED BY
Marc Schönwiesner,
Leipzig University, Germany

REVIEWED BY
Rebecca E. Millman,
The University of Manchester,
United Kingdom
Arthur Wingfield,
Brandeis University, United States

*CORRESPONDENCE
Jerker Rönnberg
jerker.ronnberg@liu.se

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

RECEIVED 12 June 2022
ACCEPTED 08 August 2022
PUBLISHED 01 September 2022

CITATION
Rönnberg J, Signoret C, Andin J and Holmer E (2022) The cognitive hearing science perspective on perceiving, understanding, and remembering language: The ELU model. *Front. Psychol.* 13:967260. doi: 10.3389/fpsyg.2022.967260

The cognitive hearing science perspective on perceiving, understanding, and remembering language: The ELU model

Jerker Rönnberg*, Carine Signoret, Josefine Andin and Emil Holmer

Linnaeus Centre HEAD, Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

The review gives an introductory description of the successive development of data patterns based on comparisons between hearing-impaired and normal hearing participants' speech understanding skills, later prompting the formulation of the Ease of Language Understanding (ELU) model. The model builds on the interaction between an input buffer (RAMBPHO, Rapid Automatic Multimodal Binding of PHOnology) and three memory systems: working memory (WM), semantic long-term memory (SLTM), and episodic long-term memory (ELTM). RAMBPHO input may either match or mismatch multimodal SLTM representations. Given a match, lexical access is accomplished rapidly and implicitly within approximately 100–400 ms. Given a mismatch, the prediction is that WM is engaged explicitly to repair the meaning of the input – in interaction with SLTM and ELTM – taking seconds rather than milliseconds. The multimodal and multilevel nature of representations held in WM and LTM are at the center of the review, being integral parts of the prediction and postdiction components of language understanding. Finally, some hypotheses based on a selective use-disuse of memory systems mechanism are described in relation to mild cognitive impairment and dementia. Alternative speech perception and WM models are evaluated, and recent developments and generalisations, ELU model tests, and boundaries are discussed.

KEYWORDS

the ELU model, working memory, semantic long-term memory, episodic long-term memory, adverse listening conditions, age-related hearing loss, dementia

Background

Cognitive hearing science builds on the principle that individual cognitive functions play an important role from very early subcortical auditory processing (Stenfelt and Rönnberg, 2009; Sörqvist et al., 2012) to interactions among memory systems at cortical levels of listening, language understanding, and dialogue (Rudner et al., 2008, 2009; Rönnberg et al., 2021, 2022). Anatomically, several precise downstream corticofugal pyramidal cell axons from neocortical layers 5 and 6 (Usrey and Sherman, 2019) allow for early cognitive impact at subcortical levels, even down to the cochlea (cf. the early filter model, Marsh and Campbell, 2016). This neural organization sets the stage for deep cognitive penetration of the very early sensory and perceptual windows of our experience—a possibility that had not been systematically scrutinized in the audiological and hearing research field before the advent of the Ease of Language Understanding model (ELU, Rönnberg, 2003). Generally, the ELU model (Rönnberg et al., 2008, 2013, 2019) is about rapid abstraction of the meaning of multimodal linguistic input, mediated by working memory (WM) in adverse listening conditions (Mattys et al., 2012).

Our early studies of speech perception and speech understanding focused on speech-reading, or lip-reading (a narrower term not including gesture and body language), adult individuals with normal hearing, impaired hearing, and deafness. The question that we posed was whether persons with hearing impairment—through their increased reliance on visual speech–produced superior, compensatory, visual speech perception/understanding skills. We also investigated the presentation modality, type of materials, type of task (Hygge et al., 1992; even including more ecological tasks Rönnberg et al., 1983), and whether hearing-impairment related variables like the duration of impairment and/or the degree of hearing loss played a role (e.g., Rönnberg et al., 1982, 1983; Lyxell and Rönnberg, 1987; Rönnberg, 1990). The answer was surprising since no compensatory signs were empirically observed.

The emergence of cognitive hearing science

With further data collections, the data pattern took a radically different turn: First, we tried to examine why some people were such excellent speech-readers (e.g., Rönnberg, 1993; Rönnberg et al., 1999). In a set of case studies of extreme speech-reading skills, we demonstrated that instead of a compensatory effect due to the hearing impairment, it was about cognitive skill in processing and storage of perceived information, measured by the reading span test (RST, Daneman and Carpenter, 1980; Daneman and Merikle, 1996). High RST performance described the cases who in their daily life relied on poorly conveyed auditory speech information, but who still were very competent communicatively: it could be lip-reading only (the case of SJ: Lyxell, 1994), tactile-visually conveyed speech information (the case of GS: Rönnberg, 1993), or a hearing-impaired person with a speech-sign bilingual background (the case of MM: Rönnberg et al., 1999). They all used very different communication strategies, but effectively so. The common denominator of the different case studies reported was that each person was well-equipped cognitively, and that cognitive functions seemed to operate over and above the variables we had studied up to that point. More specifically, it was demonstrated and replicated from the case studies that not only did high WM capacity (WMC) play a significant role in holding information alive, thus presumably mitigating the prediction of upcoming events, but it also represented a cognitive workbench for reconstructing misperceived linguistic units (i.e., postdiction, Rönnberg et al., 2019, 2021, 2022). In the same vein, we found that other related kinds of cognitive functions also contributed to the picture.

It was observed that cognitive functions like lexical access speed (Rönnberg, 1990), executive functions (Andersson and Lidestam, 2005), and inference-making capacity (Lyxell and Rönnberg, 1987) were associated with speech perception and understanding (reviewed in Rönnberg et al., 1998, 2021; Rönnberg, 2003). The data pattern withstood many experimental variations, especially in difficult speech-in-noise conditions (reviewed by Lyxell et al., 1996; Gatehouse et al., 2003, 2006; Akeroyd, 2008; Arlinger et al., 2009; Lunner et al., 2009; Besser et al., 2013), where WMC played the dominating role. Thus, it was (and still is, see e.g., Mishra et al., 2021) hard to escape the general conclusion that poor hearing and/or poorly specified or fragmented speech stimuli depend on individual cognitive processing skills to fill in the gaps of incomplete input to the perceptual and cognitive systems. These findings make up the foundation of Cognitive Hearing Science. For a more complete and historical account of the emergence of the field, see Arlinger et al. (2009).

Early studies of cross-modal language plasticity

Early neurophysiological evidence spoke to the issue of plasticity of brain tissue and prerequisites for commonalities in central perceptual and cognitive functions. Many studies testified to cross-modal language activations, suggesting for example that visual areas are recruited in pre-lingually deaf cochlear implant users (Giraud et al., 2000, 2001; Zatorre, 2001; Kral and Sharma, 2012). In addition, tactile stimuli in the congenitally deaf tactile aid user activate secondary auditory areas (Levänen, 1998). Also, the duration of deprivation plays a key role in the reorganization of the sensory cortices, such as early sensory deprivation will result in better neural plasticity adaptation (Tillberg et al., 1996; Bernstein et al., 1998;

MacSweeney et al., 2001). In normal-hearing listeners, it was suggested that silent lip-reading activates the auditory cortex (Calvert et al., 1997), especially for skilled speech-readers (Ludman et al., 2000). Furthermore, signed and auditory to-be recalled story materials perceived by sign-speech bilinguals (i.e., sign-language interpreters) have been shown to activate temporal areas to a similar extent if a visual component was involved for both modalities (Söderfeldt et al., 1994). However, compared to auditory-only, specific bilateral temporal areas activated by sign-language were involved, specifically the addition of the left area V5, later replicated across imaging techniques (e.g., Söderfeldt et al., 1997; Rudner et al., 2007, but see further under *boundaries*). However, when it came to WM for sign and speech, we still found that there were similarities for left inferior frontal and inferior temporal gyri, which subserve phonological and semantic processing areas (Rönnberg et al., 2004)—areas that also were similarly activated in the early Söderfeldt et al. (1994, 1997) studies. Finally, sign language phonological awareness and word reading ability have also been demonstrated to be associated (Holmer et al., 2016). Again, these data suggest that the brain rapidly transcends the "raw" sensory codes and rapidly abstracts input into modality compatible representations.

In all, there were many early studies that suggested commonalities and plasticity in brain activation independent of language modality and presentation modality. In addition, individual cognitive factors like WMC determined performance on language perception, and the WM system seemed to have modality-independent properties. These neurophysiological data patterns—in combination with the behavioral data—prompted the formulation of a modality-independent ELU model based on individual differences in specific perceptual and cognitive components (Rönnberg, 2003; Rönnberg et al., 2008).

The ELU model takes shape

In the formulation of the original ELU model (Rönnberg, 2003; Rönnberg et al., 2008), we were quite bold in the sense that the assumption of an occurring mismatch between perceived language input and long-term memory representations of linguistic units was supposed to hold across sensory modalities (auditory, visual, and tactile) as well as language modality (spoken and signed). The "language processor" in the brain was assumed to have a multimodal combinatorial capacity, typically occurring at the "syllabic" or sublexical level across language and presentation modes (Rönnberg, 2003; Stenfelt and Rönnberg, 2009). All cases of mismatch were supposed to trigger an increased dependence on WMC for reconstructive, postdictive purposes.

In some more detail, the ELU system assumes that the perceptual input is conceptualized as an input buffer which Rapidly, Automatically, Multimodally Binds PHOnological information together (RAMBPHO, cf. Baddeley, 2000, 2012; Stenfelt and Rönnberg, 2009). This binding, or integration process, presupposes a rapid "default mode" of abstraction into a multimodal input, where the main task of the system is to implicitly and directly unlock multi-attribute phonological representations in Semantic Long-Term Memory (SLTM), leading to access of lexical meaning (Bernstein et al., 1998; Bavelier and Neville, 2002; Stenfelt and Rönnberg, 2009; Rönnberg et al., 2013, 2019). This process typically occurs during a short time window from 100 to 400 ms, depending on the paradigm, if the chain of events runs smoothly, implicitly, and without effort. In general, this RAMBPHO process is reminiscent of Gibson's (1966) direct perception approach in that the senses should be considered as interacting perceptual systems, without short-lived intermediary representations.

However, for hearing-impaired participants, or when listening conditions are adverse (e.g., when competing noises or foreign accents are present, or the signal processing in the hearing aid is suboptimal), RAMBPHO-delivered attributes may be fuzzy and too few in numbers to surpass a hypothetical threshold to unlock lexical representations in SLTM (see Rönnberg et al., 2013, for details). The consequence of such a mismatch is that more deliberate, explicit and WM-based storage and processing functions are assumed to be triggered. These WM functions purportedly aim to piece together and infer what was communicated (Lunner et al., 2009; Rönnberg et al., 2013, 2019, 2021). These explicit functions may depend on several inference-making operations within WM but also include several interactions with SLTM and Episodic Long-Term Memory (ELTM), hence taking a relatively longer time than effortless implicit processing. The implicit processes typically operate on a millisecond scale, and the explicit processes may take seconds (Stenfelt and Rönnberg, 2009), and recent evidence suggests that different brain oscillations can dissociate the two (e.g., Gray et al., 2022). There will always be a ratio between the two, which is assumed to vary dynamically from moment to moment due to turn-taking and interlocutor responses in a conversation (Rönnberg et al., 2019).

In general terms, prediction (and postdiction) processes affect the probability that RAMBPHO will match or mismatch with SLTM representations. On a general time scale, RAMBPHO-delivered information always precedes and then affects WM storage and processing operations. If a mismatch occurs, slower postdiction processes in WM feed back to RAMBPHO until comprehension is reached (see more under theoretical implications) or not, for example in cases where the listener is not sufficiently motivated to allocate resources required for further speech processing (Pichora-Fuller et al., 2016). Thus, it is important to acknowledge that RAMBPHO is an obligatory part of an ELU/WM system, feeding linguistic information to the match/mismatch mechanism—which is at the heart of the system. The ELU model describes a

communication system which relies on interacting memory systems and mechanisms.

Experimental evidence

The first experimental manipulations of habitual vs. non-habitual signal processing in hearing aids successfully tested the cognitive consequence of the mismatch notion. We developed several different kinds of methods to trigger a mismatch. One example, and probably the most evident demonstration, was that of the studies by Rudner et al. (2008, 2009). For example, experimental acclimatization to a non-habitual kind of aggressive hearing-aid signal processing for 9 weeks (i.e., FAST or SLOW wide dynamic range compression) and then subsequent testing in a previously non-acclimatized mode of signal processing (i.e., SLOW-FAST or FAST-SLOW), produced strong reliance on WM in those two mismatching conditions, compared to the matching FAST-FAST and SLOW-SLOW conditions. As a matter of fact, the effect of just shifting from the regular hearing-aid settings to new ones produced higher reliance on WMC (for replications and relevant supporting studies/reviews, see Souza and Sirow, 2014; Souza and Arehart, 2015; Rönnberg et al., 2019; Souza et al., 2019). These findings also expose the interplay between WM and SLTM in the development of representations, as introduced to the ELU framework in recent years (Holmer et al., 2016; Holmer and Rudner, 2020; Rönnberg et al., 2022). In response to a novel input signal, such as that produced by new settings in a hearing aid or learning a new word, the language system is likely to treat the input as something unfamiliar, i.e., a mismatch condition, which WM resources are used to solve. However, each time a mismatch condition is resolved, this has the potential of producing an adjustment to the exemplars associated with the representational space in SLTM.

A second example is the investigation of competing effects of different kinds of maskers. Many studies agree with our view that so-called energetic maskers (Brungart, 2001) produce distraction but not to the same extent as informational maskers engaging SLTM (e.g., Rönnberg et al., 2010; Mattys et al., 2012; Sörqvist and Rönnberg, 2012; Kilman et al., 2014), and that distraction is more pronounced if the masker was in the participants' native language (Kilman et al., 2014; Ng et al., 2015). It should be noted that the original data of WM dependence (using "speech-like" maskers) had already been observed and discussed (Lunner, 2003; Lunner and Sundewall-Thorén, 2007; see a review by Rönnberg et al., 2010). Again, in retrospect, the effects of informational or speech-like maskers were related to partial activation of SLTM, e.g., phonologically similar neighbors (Luce and Pisoni, 1998) or possibly, with ELTM repetitions of SLTM contents.

Thirdly, as noted above, WMC is also an important predictor of performance in ELTM in such circumstances of initial speech-in-speech maskers (e.g., four-talker babble, 4T, Sörqvist and Rönnberg, 2012; Ng and Rönnberg, 2020). Notably, the robustness of the high dependence on WM was not influenced by the duration of hearing-aid use, at least up to 10 years of hearing-aid use for four-talker (4T) maskers (Ng and Rönnberg, 2020). This raises the question of whether some kinds of conversational environments are too dynamic to allow for a lessened dependence on WM. That is, highly dynamic input—or input with poorly defined phonological information—might stress the boundaries of how well the system can adjust its representational space (Han et al., 2019). The 4T-masker results hold irrespective of signal processing in the hearing aid. However, contextual support (Rönnberg et al., 2016) or plausibility/predictability of sentences may override the need for WM resources (Moradi et al., 2013; Amichetti et al., 2016), which in turn makes the signal-context interactions determine the potential need for postdictive processing. The overall idea is that the brain should not invoke WM resources unnecessarily, e.g., when context drives a more rapid and implicit route to comprehension. In that sense, the brain is "lazy" and economical in spending effort and processing energy, using a principle of least effort (cf. Ayasse et al., 2021; Silvestrini et al., 2022).

A final study points to constraints when bimodally combining CI-listening in one ear with listening with a hearing aid in the other (Hua et al., 2017). These two types of signals reaching the brain will not necessarily be RAMBPHO compatible. From an ELU perspective, an electric and a physical-neurostimulation might be harder to convert into some more abstract representation than naturally occurring multimodal sensory stimuli. Analogous to habitual vs. non-habitual signal processing in the hearing aid, this combination of inherently different signals to the brain does not seem to combine easily. One indication is that RST was the most sensitive predictor variable for bimodal sentence materials compared to unimodal listening conditions, where the trail-making test (primarily measuring cognitive speed) was more critical to unimodal conditions and single word identification (Hua et al., 2017). But, it should also be noted that the bimodal condition facilitated speech-in-noise performance, although the cost in terms of WM engagement and effort may create a balancing act that the individual and clinician must decide from the individual WMC data.

Related to these difficult (or mismatching) speech processing tasks is a couple of recent studies corroborating the ELU hypothesis about the engagement of WM in challenging listening conditions. Mishra et al. (2021) found that WMC accounted for large portions of variance (up to 80%) of word recognition in speech noise of spondees and phonetically balanced word lists presented at dB SNR 0, −10, and −20, whereas in quiet the pure tone average explained 78% of the word recognition scores and not the RST scores (see also Kurthen et al., 2020). However, there is also evidence for

specificity in correlations between the demands of the WM task and the complexity of the criterion task (Heinrich et al., 2015; see Rönnberg et al., 2021 for a discussion). Thus, the actual principle of mismatching is replicated in Mishra et al. (2021) with the concomitant demand on WM resources, but further task analyses of both tests of WMC and the type of outcome remain to be carried out (Heinrich et al., 2015). Nevertheless, see our first attempts involving interactions among memory systems (Rönnberg et al., 2011, 2014, 2021).

Linguistic abstraction and WM capacity

In gating tasks (Grosjean, 1980), the participants are required to identify a consonant/vowel or a final word in a sentence, based on the presentation of successive bits of initial phonetic information of the speech token (Moradi et al., 2013, 2014a,b). These studies have demonstrated that early and successful linguistic identification requires WMC. WM is involved in the rapid identification of e.g., a consonant when the semantic context is lacking. Around 90 msec is necessary to identify speech tokens in the auditory gating paradigm (Moradi et al., 2014b). Audiovisual presentation reduces this identification time to 40–50 msec, as does identification of final words in highly predictable sentences (Moradi et al., 2013). This time reduction presumably implies very rapid neuronal communication between different brain regions that, in turn, activate different memory systems. However, hearing loss (even when compensated with hearing aids), age, and noisy signals, all slow down the identification process (Moradi et al., 2013, 2014a).

A further example of rapid abstraction is a kind of priming paradigm which has demonstrated the so-called "pop-out" effect (see a general discussion in Davis et al., 2005), where presenting the written version of a sentence on a computer screen creates an enhanced perceived clarity and understanding of spoken, noise-vocoded, sentences that are otherwise incomprehensible due to the vocoding. In the studies by Signoret et al. (2018), Signoret and Rudner (2019), the written version of the sentence was presented word by word (200 msec before the vocoded version) until a sentence was complete, and required the participant to rate the perceptual clarity of the vocoded sentence. Perceptual clarity is enhanced for semantically coherent and phonologically primed sentences, but when compatibility is low, WM processes entered into play. For example, WM was invoked in conditions with non-matching primes, substantial vocoding, and low semantic coherence (Signoret and Rudner, 2019).

In a recent study using magnetoencephalography (MEG, Signoret et al., 2020), participants were to decide whether the final word in a sentence was expected or not. The participants had studied the sentences before the actual experiment, so expectations on the final word were strong. Different kinds of deviants were used as final words (in background noise): either the final word was semantically different but rhymed with the expected word, or was semantically related but did not rhyme, or was different in both aspects of similarity. Notably, WMC negatively correlated with the number of false alarms to meaning deviants that rhymed, such that participants with high WMC were less lured into accepting a deviant via the RAMBPHO to the SLTM matching process. Further, participants with higher WMC had processed meaning deviants more easily (smaller N400 effect) compared to participants with lower WM capacity. Participants with high WMC, also processed the semantic mismatching more easily (smaller N400 effects) and showed better performance at the behavioral level. WM therefore seems part and parcel of the prediction mechanism.

In sum, the hypothesis of RAMBPHO-like multimodal representations that together with high WMC activate SLTM, receive support from the above examples.

Theoretical implications

Theoretically and neurophysiologically, the RAMBPHO-SLTM interaction necessarily utilizes very rapid subcortical and cortical connections that allow for the implicit initial matching process. At this stage, the *prediction* aspect of the ELU model allow for matching processes most likely affecting attention to certain aspects of the input signal, which naturally varies in form, content, and modality (Samuelsson and Rönnberg, 1993; Sörqvist et al., 2012; Sörqvist and Rönnberg, 2012). Because of the correlations with WMC cited above, we can assume that RAMBPHO rapidly "constructs" a multimodal channel to WM that matches multi-attribute or multimodal representations in SLTM. Thus, a high WMC may facilitate early attention fine-tuning of auditory processing but may also reflect a highly synchronized brain network (Fell and Axmacher, 2011). The conclusion about some kind of early fine-tuning is reinforced by the finding that WM processes are positively interconnected with the effects of practice on auditory music skills (Kraus and Chandrasekaran, 2010) and their corresponding neural brain stem signatures (Kraus et al., 2012).

Thus, we argue that the brain always has, as its primary aim, to abstract meaning, i.e., a cognitive hearing, sense-making, organism. This aim holds regardless of whether the stimuli are represented by sublexical or lexical items, or grammatical constraints in a sentence (Ayasse and Wingfield, 2020). And, if there is a mismatch, there is the advantage of already existing multimodal representations maintained in WM—given a sufficiently capacious system—which can be deconstructed and/or reconstructed among brain networks that belong to SLTM and ELTM (i.e., postdiction).

Furthermore, since we initially have a presumed representation which is rapidly constructed but which also can be deconstructed by cues, and then reconstructed again,

perhaps several times, it necessarily takes longer to execute explicit interactions among brain networks of the brain that belong to the WM, SLTM, and ELTM systems. The explicit postdiction process may take seconds, while the prediction process is on the msec scale. The neural mechanism diverges between them: while the postdiction process is proposed to relate to theta-activity, the prediction process is proposed to relate to alpha-activity (Gray et al., 2022). Typically, the postdiction process becomes more explicit and slower the neural process (i.e., probably related to enhanced theta activity), whereas the prediction processes are more implicit and probably related to faster neural activity, such as decreased alpha activity (Gray et al., 2022) or beta activity (Signoret et al., 2013). Although predictions are often primed implicitly in everyday conversation in their effect on RAMBPHO, contextual cues can be explicitly held in WM prior to the experimental materials (Zekveld et al., 2013), in a similar vein to when postdiction feeds back into RAMBPHO. This priming mechanism likely aims to pre-activate specific knowledge at different levels of processing depending on the environmental context and on individual abilities and skills. Generally speaking, the ELU model assumes an overarching prediction-postdiction system, both dependent on WMC, albeit in different ways (Rönnberg et al., 2019, 2022).

Output from the ELU system

The output of the ELU system is lexical access, the grasp of what was communicated, and what consequently may be recalled from ELTM. Further, the output might involve a change to SLTM, either during development (Holmer et al., 2016) or when adjusting to novel listening conditions (e.g., Ng and Rönnberg, 2020). We have also shown that WMC contributes to ELTM performance in terms of sentence recognition (e.g., Sörqvist and Rönnberg, 2012; Zekveld et al., 2013). In Sörqvist and Rönnberg (2012), it was demonstrated that performance on the Size Comparison (*SIC*) span WM test predicted higher immediate and delayed recall of fictitious stories masked by a person reading from another story at encoding—*SIC* span made significantly better predictions in hierarchical regression analyses the RST. This presumably comes from the fact that apart from processing and storage, as in the RST, *SIC* span involves an additional inhibition component. For example, the task could be to compare four-footed animals in a list of comparisons and an additional to-be-remembered word for each comparison: Is a zebra larger than a mouse? (the Comparison) + the word (lion). Each list of comparisons and words belong to the same semantic category, which could cause confusion between comparison words and to-be-remembered words at the recall of the list. Obviously, an inhibition factor comes into play, and mediates better recall.

Although not a prime purpose of that study (Sörqvist and Rönnberg, 2012), the clinical implications are important as well. As it happened in this study, WMC was important for both immediate performance and for ELTM. The ELTM aspect is crucial from a listening perspective. If the actual signal processing in the hearing aid or the adversity of the listening situation drains too much processing capacity in the "here and now" situation, then less is left over for storage. This implies that conditions for change might also be circumscribed in noisy conditions, perhaps because little WM resources are left for successful encoding into change and development of SLTM. To compare with a long discussion about task demands on storage and processing, see Lunner et al. (2009) and Rönnberg et al. (2021).

In other words, in the dialogue between a hearing-impaired person and an interlocutor, the hearing-impaired person must allocate explicit attention to mismatches when they occur to be able to extract meaning, and perhaps learn something new, from the conversation. This is not required of the normal hearing person to the same extent. This is exactly the reason why a measure of what is left in ELTM after smaller amounts of storage resources remain in WM (e.g., tested a day or two after the conversation) would clinically be a very important ecological aspect of what it means to approach ease of listening and understanding for a hearing-impaired person. Without going into detail, a test that tapped into storage, semantic processing functions, and inhibition at the same time, would seem to be a suitable candidate based on the data we have collected (e.g., Sörqvist and Rönnberg, 2012; Stenbäck et al., 2015). Furthermore, what you remember from a conversation has obvious personal and social consequences, and such consequences might sometimes lead to developing depression (Keidser et al., 2015).

In further support of the inhibition component, recent studies by Stenbäck et al. (2015, 2021) verify that especially WMC (measured with the RST) but also the Swedish Hayling test (Stenbäck et al., 2015), which measures inhibition, were significant predictors of performance in speech-in-noise (SPIN, Hällgren et al., 2006) and Hagerman matrix sentences (Hagerman, 1982). The Hayling test builds on the ability to inhibit sentence completion of the last semantically correct word instead of providing a semantically incorrect but grammatically correct word. Thus, WMC and executive functions are part and parcel of listening, understanding, and recalling in adverse conditions.

In general, inhibition and turn-taking in real dyads or conversation/discussion groups put large demands on the timing of turns. If you are, e.g., interrupting too many times, it might just be the case that you do not have sufficient WMC to follow the line of thought in the conversation. To do that, you need to keep in mind what was just said and process it, at the same time as you are planning to latch on with your own turn, and what you are going to respond to. Therefore, taking another person's perspective in a dialogue demands WMC functions of maintenance, timing and

dual storage and phonological/semantic processing. However, taking the perspective of someone else also involves the cognitive function of theory-of-mind (ToM), which is about decoding and understanding other people's intentions and feelings, and not necessarily just grasping what was actually said in the conversation (Hagoort and Levinson, 2014). That is, the intended meaning might sometimes not be coded in the meaning of the exact wording of a sentence, and therefore, ease of language understanding is about multilevel understanding in dialogue.

Recent findings

Füllgrabe et al. (2014) and Füllgrabe and Rosen (2016) claimed that WMC only accounted for significant amounts of variance in elderly hearing-impaired participants' SPIN performance, whereas for younger normal-hearing participants, only a small percent of the variance was accounted for by WMC. Nevertheless, Vermeire et al. (2019) clearly showed that for elderly normal-hearing participants, RST was a significant predictor of SPIN performance as well. Indeed, Gordon-Salant and Cole (2016) showed the same results to hold across age groups, with RST as part of the most prominent predictors. In the same vein, with large samples, Marsja et al. (2022) used the n200 database (Rönnberg et al., 2016) to study potential differences in cognitive involvement due to hearing loss. Marsja et al. (2022) used a multi-group structural equation model (SEM) approach where the purpose was to assess whether the contribution of a "Cognition" latent variable (based on RST, a visuospatial WM test and a semantic WM word-pair test, and Raven's matrices) was equally related to a SPIN criterion (Hagerman matrix sentences, Hagerman, 1982) for hearing-impaired hearing-aid users compared to normal-hearing participants. The results, based on 200 participants per group, show that the Cognition variable accounted for identical beta weights (-0.32) in both groups of equal average age (60 years), when the groups were compared on an outcome latent construct based on Hagerman matrix sentences. Thus, the cognitive contribution to SPIN perception is not specific to elderly hearing-impaired participants. The statistical models were partialed out for age and hearing loss, and significant on all relevant model fit parameters. This is generally supportive of the initial claims of the ELU model, viz. that there is a communality in cognitive abstraction and cognitive prediction across adult groups with different hearing status (Rönnberg, 2003).

The devil is in the details

It is important to note that the interaction between the fine details of task demands may make a large difference in terms of predictability of outcome in a SPIN task. For example, in the original RST (Daneman and Carpenter, 1980), participants always recalled the final words in each sentence set—which obviously invites a strategic component compared to the version we use (Rönnberg et al., 2016), where participants are post-cued to recalling *either* the first *or* the last word of the sentences to be verified. In the latter case where the strategic component is reduced, the "raw" WMC is more likely to be revealed. Our research builds on this latter task version and that "detail" may be a clue as to why some researchers get higher involvement of WM than in some other studies (e.g., Ng et al., 2013; Souza et al., 2019; Ng and Rönnberg, 2020). Other aspects relate to contextual support either at the prediction stage or in the sentence materials themselves; high contextual support renders lower correlations with WMC, and vice versa (Moradi et al., 2013; Rönnberg et al., 2016). Dependence also varies with age, hearing status, and a host of other factors related to hearing aid signal processing and habitual processing demands, and not least the interplay amongst the speed, phonology, and WM factors depends on the level of adversity of the listening situation (Homman et al., submitted).

Structural equation modeling

In a recent study by Homman et al. (submitted) on the hearing-impaired participants in the n200 study (Rönnberg et al., 2016), we used Structural Equation Modeling (SEM) based on the original cognitive parameters in the ELU model: speed, phonology, and WM (Rönnberg, 2003). Thus, one latent speed parameter (i.e., physical matching and lexical access speed), one latent phonological parameter (auditory and audiovisual gating conditions and rhyme tests, i.e., measures of RAMBPHO), and one latent WMC parameter (i.e., based on the RST, visuospatial WM, and semantic word-pairs), were included, while age and hearing loss were partialed out. The results show that phonology always contributed to the performance in the different Hagerman conditions (irrespective of noise type, performance level, and type of signal processing). Speed did not directly predict the Hagerman outcome, but speed always predicted WM, and the WM to Hagerman path was significant only in the more difficult listening conditions involving 4T maskers. Thus, the new and interesting result of this mediation analysis was that speed contributed *via WMC* to Hagerman in the difficult conditions, i.e., where higher degrees of mismatch can be assumed. The general interpretation is that when being exposed to adverse listening conditions, it is important that WM is capacious because it takes more time to reconstruct what was perceived, i.e., when a more laborious explicit mode of processing is needed. An alternative interpretation would be that the adversity of the listening situation primarily strikes at RAMBPHO. Nevertheless, optimizing speed in WM operations becomes critical in both cases. This result agrees with previously reported results in Rönnberg et al. (2016), where WM was more

strongly related to Hagerman matrix sentences than to HINT sentences, which are contextually driven everyday sentences. By virtue of the semantic coherence in HINT sentences, the prediction mechanism is improved, hence lessening the demand on WM resources for postdiction (cf. also Moradi et al., 2014b).

In a similar SEM approach, Janse and Andringa (2021) modeled word recognition performance in degraded low-pass filtered conditions. They used cognitive speed, vocabulary, hearing acuity, and WM as latent constructs. In their model, WM was the strongest latent construct relating to word recognition in noise, replicating our research. The RST was the test that loaded the highest on the WM factor, compared to digit span and non-word recall (cf. Rönnberg et al., 2016). In our current model (Rönnberg et al., 2021, 2022) we only used WM tests that emphasize storage and processing in dual task formats. We noted that speed of access from SLTM was our mediating factor. However, their mediation model of vocabulary via WM to word recognition (Janse and Andringa, 2021) is not directly comparable to ours, as we did not use vocabulary, but interesting indeed. The communality is that WM predictive capacity is only predictive of SPIN performance, via some back-up parameter such as SLTM speed or SLTM vocabulary. It is obvious that these mechanisms support WM when more complex interactions between WM, SLTM, and ELTM are required for postdiction purposes.

Comparison with other models

Perception and understanding: multimodal and multilevel aspects

Speech perception models are less comprehensive than the ELU model, but in some cases more specific. For example, in the Neighborhood activation model (NAM model, Luce and Pisoni, 1998), lexical access is clearly dependent on how input stimuli matches/mismatches with the lexicon due to phonological similarity and semantic parameters like word frequency. In the initial word cohort model (e.g., Marslen-Wilson, 1987), the initial information that enters the ear is assumed to activate a set of competitors in a cohort of possible candidates, a functional parallelism in the activation of the lexicon. As information successively enters the auditory system, activation and selection of candidates proceeds until only one lexical candidate remains. There are many manipulations of the selection process e.g., by priming, word length, or word endings that in different ways manipulate the word recognition point, which often occurs before the whole word has been perceived.

Furthermore, the probability of lexical access is not an all or none process (as described in Rönnberg et al., 2013); it depends on the RAMBPHO input and the kinds of representations it meets in LTM. And, at some hypothetical threshold of matching attributes lexical access is triggered. The types of error responses we obtain may well be captured by the NAM (Luce and Pisoni, 1998), but in addition we have also made clear that the prediction- postdiction cycle may prime or direct the individual ELU system to other aspects of representation in LTM that then helps the system to surpass the threshold and retrieve the correct lexical candidate. A good example of such priming by sentence context can be found in a MEG study by Signoret et al. (2020), where phonological and semantic error responses were in focus.

Also related to RAMBPHO, we focused on a special form of priming. We dubbed the hypothesis "perceptual doping" (Lidestam et al., 2014; Moradi et al., 2019). In brief, the priming effects of exposure to two initial conditions (auditory only, or audio-visually presented materials) on later auditory perception of consonants, vowels, and sentence materials generally demonstrated a multimodal facilitation ("doping") effect. The interpretation is that there is a recalibration/remapping of the initial audiovisual presentation mode affecting the SLTM representation of phonological and lexical attributes. With the advantage of hindsight, a discussion of the data based on RAMBPHO may also have been possible.

Related to our mismatch concept, earlier basic auditory perception studies outlined the basic properties of the mismatch negativity (MMN) effect measured with EEG (Näätänen, 1995; Näätänen and Escera, 2000). In our research, we have emphasized the *consequence* of mismatch in terms WM involvement (Rönnberg, 2003). Relevant to the current paper is that the mismatch notion has also been applied to grammatical levels of language processing (Federmeier, 2007), as well as phonological/semantic processing (Signoret et al., 2020). The ELU model is here proposed to be about levels of linguistic mismatch, from RAMBPHO and lexical access, via grammar to semantic coherence. Therefore, the fact that the functional role of the frontal cortex in pre-attentive auditory change detection has been shown for grammatical deviations is of high importance (Hanna et al., 2014). The mismatch negativity function automatically detects grammatical anomalies around 200 msec, after the grammatical violation point (subject–verb agreement violations or word category violations, Hasting et al., 2007; cf. Signoret et al., 2020 for different violation types). Tse et al. (2013) and Hanna et al. (2014) have both demonstrated and discussed the very early, pre-attentive and automatic Broca/inferior frontal signals of mismatch negativity for grammatical violations (around 200 msec), which could be an inspiration to our new, more elaborated ELU proposal, of co-occurring mismatch signals possible at different linguistic levels (see below).

In our current view, the multimodal phonological level is crucial to SPIN performance, but if implicit processing occurs at higher levels of language like syntax, keeping auditory characteristics under control (Hasting et al., 2007), it makes our claim about the necessity of rapid WM interactions with SLTM and ELTM even more important. Otherwise, these extra steps would probably prolong the extra time for reconstruction and

postdiction. This specification of the ELU model is that not only is WM involved in rapid RAMBPHO-delivered multimodal abstraction, but it is also involved in multilevel language interactions, given that there is some central mismatch time widow for several levels of language, and which can be processed in parallel (cf. Marslen-Wilson, 1987). *We submit that the prediction-RAMBPHO-SLTM-postdiction interaction demands a "moving time window" within the confines of WMC, the contents of which are rapidly abstracted at multimodal and multilevel aspects of input. In a more generalized form, it may be stated that: for any given aspect of RAMBPHO-delivered linguistic information, the cognitive consequence of a mismatch with SLTM representations, is bound to initiate WM-based postdiction.*

Furthermore, the cognitive consequence of a central multimodal/multilevel mismatch mechanism has not been fully realized in the ELU model, but comparisons with Central Auditory Processing Disorder (CAPD) research could inspire (Gates et al., 1996; Gates, 2012). However, several such multimodal and multilevel interactions will need to exploit all the storage and processing capacities of WM. Therefore, the postdiction processes will necessarily take more time than implicit predictions. But if there are rapid multilevel mismatch functional capacities of the brain, it will allow for an advanced analysis-by-synthesis kind of model, demanding such on-line revisions of what is misperceived (cf. Hickok and Poeppel, 2007). It also demands parallel processing not just at the word level. For example, in the Moradi et al. (2014b) study, high predictability sentences were completed with only minimal initial phonemic information of the final word (40 msec).

Thus, we still assume that relatively context-free perception is dependent on RAMBPHO-based lexical retrieval. But, context-bound, grammatically incorrect sentences can also induce mismatch at some violation point in the sentence. This implies that the functional parallelism (cf. Marslen-Wilson, 1987) is not only realized through multimodal streams of information, as in the ELU, but also in parallel streams at different levels of language that act in concert to optimize implicit understanding of the discourse. This may at later stages demand cognitive functions to keep track and focus on the "winning stream" of information processing (cf. Moradi et al., 2013). Again, presumably the brain is optimizing speed of mental operations in WM even in mismatch situations. However, to our knowledge, the mismatch studies have not emphasized the communicative feedback, which the ELU model denotes as postdiction, which in turn is assumed to feed back into predictive RAMBPHO processing. This postdiction feedback may not only alter predictions but also induce SLTM changes.

Thus, even when well-organized and linguistically interactive and smooth processing is taking place, mismatch at some linguistic levels will demand reconstruction and postdiction, typically taking more time. Parallel levels of processing (without any mismatch) may on the other hand synergistically integrate input and reduce processing time (Moradi et al., 2014a,b; Signoret et al., 2020).

Working memory

In comparison with other working memory models, the ELU model is very much inspired by two working memory traditions, the Baddeley and Hitch (1974; Baddeley, 2012) tradition and its many developments (e.g., the episodic buffer Baddeley, 2000, cf. RAMBPHO), as well as the tradition following a more general resource model tradition, with less structural assumptions on dedicated loops and modular functions (cf. our use of the RST, Daneman and Carpenter, 1980; Just and Carpenter, 1992; Daneman and Merikle, 1996; Barrouillet and Camos, 2020).

More specific capacity models sometimes have taken the form of activation of LTM relevant information, not seldom related to expertise (Ericsson and Kintsch, 1995; Cowan, 2005; Jones et al., 2007). The activation capacity of several representations in LTM thus becomes a measure of expertise, or WMC. In the ELU model, there are two roads to LTM in principle, one implicit and one explicit. This conceptualization and difference to the above models is dependent on the assumption that the ELU model is primarily conceived for communication purposes, where mismatch disturbs the flow of rapid phonologically mediated lexical access, but where WM must engage SLTM and ELTM to optimize explicit postdictions, as well as predictions. Seen from this horizon, the ELU model captures what we believe to be a human propensity, viz. the system is "lazy" or economical (cf. Richter, 2013); it does not spend explicit resources unless sub-threshold levels of language input cause mismatch (especially the phonologically mediated lexical access function).

There have been several recent attempts at refining the component concepts of the ELU-model. The ELU-model has generated several important scientific hypotheses and ways of investigating and testing them.

Model refinements

Edwards (2016) suggests that just before RAMBPHO processing occurs, a process is needed that accomplishes early perceptual segregation of the auditory object from the background, so called Auditory Scene Analysis (Dolležal et al., 2014). His discussion is based on the Rönnberg et al. (2008) version of the ELU model, where RAMBPHO processing focuses on how different streams of sensory information are integrated and bound into a phonological representation (see also Stenfelt and Rönnberg, 2009). Nevertheless, in Rönnberg et al. (2019, 2022), it is made more explicit that the system may feedback via postdiction processes, which may prime the prediction process (Sörqvist and Rönnberg, 2012), including

fine-tuning of attention (Holmer and Rudner, 2020; Andin et al., 2021) and selection processes to specific features of the input (Rönnberg et al., 2013). This seems to be rather close to stream segregation, but the theoretical languages differ. By inference, postdiction may then calibrate the selection of the auditory object, comparable to "perceptual doping" (Moradi et al., 2019).

The second aspect is that RAMBPHO is assumed to be primarily dedicated to phonologically relevant information, embedded in lexical and semantic representations in SLTM. Lexical access and semantic meaning of sentences are tightly tied to the mismatch mechanism—and by default—finding of a linguistic object. Thus, that aspect of Edwards (2016) proposal does not necessarily demand model change (Rönnberg et al., 2019).

In the D-ELU model (Holmer et al., 2016), the development of language representations in SLTM is in focus. The original ELU model focused on the system's input side and the WM-LTM interactions, but the development of appropriate SLTM representations has hitherto received less interest. Nevertheless, it has been demonstrated that vocabulary is very important to speech perception in noise (Kennedy-Higgins et al., 2020), either via WMC (cf. Janse and Andringa, 2021), for hearing-impaired listeners (Signoret and Rudner, 2019), or in how language is represented in bilinguals (Kilman et al., 2014; Bsharat-Maalouf and Karawani, 2022). According to the D-ELU model, existing lexical representations in SLTM shapes further lexical growth, i.e., novel representations build upon existing representations (cf. Jones et al., 2021). Novel words that are rich in lexical attributes are more likely to be successfully encoded into SLTM, and thus learning rates are predicted to be steeper. Further, learning for persons with hearing loss is predicted to be worse than for controls when the perceptual platform at the learning stage is too dynamic (Ng and Rönnberg, 2020).

In the study by Kilman et al. (2014), and of relevance for how representations develop, we found in Swedish native speakers who also knew English, that the most interfering speech in noise condition was when the speech masker was in the same (Swedish) native language as the target. The Swedish babble was interfering more than the English babble in stationary noise, and in fluctuating noise. The interference from language maskers replicates previous work (Van Engen and Bradlow, 2007; Calandruccio et al., 2010).

A recent study by Bsharat-Maalouf and Karawani (2022) examined the speech perception of 60 Arabic-Hebrew bilinguals and a control group of native Hebrew speakers during degraded (speech in noise, vocoded speech) and quiet listening conditions. There was a clear interaction in the data such that performance in the bilinguals was on a par with the native Hebrew speakers in quiet conditions, whereas performance in the babble noise conditions (same language of the noise and targets) was substantially lower. Explaining these and other effects, in terms of proficiency (Kilman et al., 2014) of second language, age of acquisition, propensity to learn in vocoding conditions (Bsharat-Maalouf and Karawani, 2022), and what kinds of SLTM representations mediate these findings is of importance for the bilingual and developmental aspects of the ELU model. Future publications will tell.

Aging, cognitive impairment, and dementia

The ELU emphasis on a meaning-related focus of the brain's perceptual-cognitive system is assumed to prioritize multi-attribute representation and multilevel mismatch processing. In other words, both children and adults are primarily tuned in to understanding language and intended communication but can of course be instructed to learn or memorize what has been communicated. In terms of a use/disuse principle (Rönnberg et al., 2011, 2021), WM is on top, always dealing with both pre- and postdiction processes on-line; the next memory system is SLTM due to the natural semantic bias in interpretation of conversation and discourse, and ELTM will be relatively less used for two reasons: (1) a non-prioritized bias in communication, and (2) denied encoding and retrieval due to hearing loss or other adverse conditions. Thus, disuse can be a key to why WMC is relatively spared when it comes to cognitive decline studies (Rönnberg et al., 2011, 2014), whereas semantic and especially ELTM decline becomes a marker of mild cognitive impairment, which might develop into dementia.

The disuse notion of memory systems is mainly supported by two major studies by our team: (1) In Rönnberg et al. (2011), based on the Betula prospective cohort study (Nilsson et al., 1997), we found that hearing loss did not selectively affect different ELTM encoding tasks in different sensory modalities (i.e., motorically, by text and simultaneous auditory presentation, and auditory only compensated with hearing aids). If anything, the hearing loss–ELTM encoding task correlations were higher with the motorically encoded task. This may seem counterintuitive, unless one assumes multimodal representations and that the multimodal memory system level is negatively affected, not memory via a specific encoding modality. (2) In addition, long-term memory, especially ELTM was affected by hearing loss, but not by visual impairment. The fact that the cognitive aging and dementia-related literature suggests that ELTM is the most sensitive predictor variable among memory systems to mild cognitive impairment and dementia (Bäckman et al., 2001; Fortunato et al., 2016; Younan et al., 2020) makes our case strong. Combining (1) and (2), we may infer that hearing loss is an important risk factor for accelerated dementia progression (Livingston et al., 2017).

Common cause accounts (e.g., Baltes and Lindenberger, 1997; Humes, 2013; Powell et al., 2021) may predict that hearing loss affects several encoding modalities, but they do not predict selectivity of memory systems. In Rönnberg et al. (2014)—building on 138098 participants from the UK Biobank

resource—we observed that ELTM was more affected by hearing loss than WM, thus replicating the data from Rönnberg et al. (2011). In terms of encoding modality, the Rönnberg et al. (2014) study employed visuospatial tests only. Still, we obtain a negative effect of hearing loss on ELTM and not on WM. This further supports and replicates a memory systems account of relative use/disuse as a potentially viable ELU explanation of the data pattern.

In addition, the potential risk of cognitive decline due to hearing loss cannot be explained by the information degradation hypotheses (Pichora-Fuller, 2003; McCoy et al., 2005), nor by attention costs for the hearing-impaired person (Sarampalis et al., 2009; Tun et al., 2009; Heinrich and Schneider, 2011). This could have been the case had the auditory encoding condition been negatively affected by hearing loss, even if the participants wore hearing-aids at testing (Rönnberg et al., 2011). A further important aspect of the 2011 data is that testing the same models, replacing hearing loss with estimated visual impairment (legibility of font size, on a scale from 6 to 24, Rönnberg et al., 2011; wearing eye glasses or not, or having a diagnosis, Rönnberg et al., 2014), did not replicate the memory system selectivity of the hearing loss results. As a matter of fact, the models tested were not acceptable by the structural equation model criteria used. What is also true of the above two data sets is that the hearing losses were only of the mild to moderate kinds (assessed by the pure tone audiogram in Rönnberg et al., 2011, and by the digit triplets test in Rönnberg et al., 2014), suggesting that early prevention with hearing aids should be employed (Arlinger, 2003), although the data for treatment by hearing aids is relatively meager when it comes to dementia.

At any rate, hearing loss, not visual impairment, is a very sensitive predictor variable of especially ELTM impairment, hearing loss being the largest modifiable factor of the development of dementia (Livingston et al., 2017). However, that is the overall picture and the more specific underlying mechanism as to why hearing loss is a risk factor for dementia is still argued to be unclear (Wayne and Johnsrude, 2015; Hewitt, 2017). Other independent analyses from the UK Biobank resource suggest that subclinical small variations in hearing acuity may still be associated with loss of gray matter volumes in the brain, especially in areas related to cognition and hearing (Rudner et al., 2019; however, see further about brain atrophy and cognitive reserve Uchida et al., 2021).

Generalizations

The previously mentioned study by Marsja et al. (2022) suggests impressively similar (if not identical) cognitive predictions from one hearing-impaired group compared to a normal-hearing group on a matrix sentence latent construct. This finding suggests a powerful generalization of the case that Cognitive Hearing Science—and the ELU model—applies to anyone, regardless of hearing status.

Moreover, recent studies of different speech distortions (Kennedy-Higgins et al., 2020) show that WM and vocabulary (i.e., SLTM) come out as the main predictors, irrespective of the type of distortion (time-compressed and noise-vocoded signals, and speech in noise). This informs us that the cognitive machinery underlying speech perception and speech understanding is rather invariant in its reliance on certain cognitive building blocks irrespective of how underspecified or distorted target stimuli are. As already argued, it presupposes that rapid abstraction into formats suitable for WM is a prerequisite for the system to work.

An interesting extension of the generalization aspect is the study by Blomberg et al. (2019) of adults with Attention Deficit Hyperactivity Disorder (ADHD). She also used different kinds of speech distortions (normal vs. noise-vocoded), orthogonally combined with type of background noise (clear speech, white noise, and speech babble). Materials were taken from the Swedish HINT sentence corpus, which consists of everyday sentences (Hällgren et al., 2006). Results showed that compared to an age-matched control group there was no interaction between group and type of masker or stimulus distortion (but main effects were observed), generalizing the Kennedy-Higgins et al. (2020) findings to another group of participants, with similar kinds of distortion manipulations. This pattern may depend on the possibility that the cognitive analysis and representations are multimodal and information-based rather than modality-specific. Importantly, different assessments of WM were used to construct a cognitive factor that heavily influenced performance across the distortion/noise conditions, supporting the ELU model.

As long as information collated or bound by RAMBPHO is incomplete in some of the many ways that will cause mismatch, dependence of WM tests indicate that a certain level of generalization is possible to make. But, we would not argue that RAMBPHO processing of e.g., vocoded speech is exactly the same as e.g., RAMBPHO processing of rapid wide dynamic range compression of speech. The general point is that at a cognitive postdiction level you must (for different reasons) infer, manipulate, and "mentally fill in" some pieces of information that demand WM processing, as well as retrieval of LTM information, to reconstruct poorly specified stimulus materials. Any other model that emphasizes the cognitive work needed in degraded, distorted, or perceptually demanding conditions would also be supported by such findings across groups and stimulus conditions (e.g., the FUEL framework, Pichora-Fuller et al., 2016).

Finally, another experiment seems to indicate that load on WM is reflected in larger pupil dilation responses (assumed to reflect cognitive load) than the physical characteristics (SNR) of the task (Zekveld et al., 2018), implying that high level cognitive processing in WM is accomplished, but pushes the system to its

limits for participants with low WMC. This study was followed up by a study on an auditory Stroop task (measuring executive control), where pupil size was higher in conflict conditions (e.g., saying "left" in the right ear). This connects well with our early observations that not only did WMC play an important role in speech understanding, but also executive functions or cognitive control seemed important (Badre, 2021). In ELU terms, the postdictive phase of inferring what was uttered, may use both the processing capacities of WM but also of related or overlapping executive and cognitive control functions.

Testing the boundaries of the ELU model

Neuroimaging studies show that sensory deprivation, e.g., deafness, during development cause reorganization of superior temporal regions (auditory cortex; Bavelier and Neville, 2002; Merabet and Pascual-Leone, 2010; Andin and Holmer, 2022). Theories behind such cross-modal reorganization differs, with some suggesting pure neural processes and some suggesting behaviorally driven processes. In the case of early deafness, the latter has gained most attention, with two main lines of explanations (see extensive review in Cardin et al., 2020). The first explanation proposes functional preservation, where the type of processing in a sensory deprived region, i.e., auditory cortex, is preserved but applied to a different modality (e.g., visual instead of auditory). This notion finds support in results suggesting that superior temporal regions, which respond to speech in hearing individuals, are activated in response to sign language in deaf but not hearing signers (MacSweeney et al., 2001; Cardin et al., 2013). Such reorganization supports an extension of the ELU model to the manual-visual language modality.

The second proposal is that reorganization reflects a functional shift. This idea is supported by studies reporting activation in superior temporal regions during cognitive tasks (Twomey et al., 2017), and suggests modality-dependent differences in cognitive processes. This perspective speaks against one of the original claims of the ELU model, i.e., that there is a modality-independent "language processor" in the brain. This is of course given that superior temporal regions are exclusively engaged in language processing. However, the empirical evidence to date lends support for both explanations and it has also been suggested that they can coexist (Cardin et al., 2020).

In WM studies using sign-language material (e.g., Rönnberg et al., 2004; Bola et al., 2017; Cardin et al., 2018; Andin et al., 2021), the superior temporal regions (auditory cortex) and occipito-parietal regions (in speech-sign bilinguals, Rönnberg et al., 2004) are activated to a greater extent for deaf compared to hearing individuals. However, in a recent neuroimaging study from our lab, we found that the activation of auditory cortex did not increase with increasing WM load in a sign language-based task, suggesting a general sensory-perceptual processing role in response to visual linguistic material (Andin et al., 2021) in line with functional preservation. Further, we found support of a modality-specific pattern in relation to the degradation of the sign-language signal. In previous studies on auditory signal degradation in individuals with normal-hearing and impaired hearing, similar changes in neural activation have been identified for both increased WM load (amount of information needed to be kept in memory) and acoustic degradation. These findings have been taken as evidence for resource models of WM in general and the ELU model for language processing in particular (Obleser et al., 2012; Petersen et al., 2015; Peelle, 2018; Rönnberg et al., 2019). Although visual degradation of the language signal resulted in similar effects at the behavioral level, the neural overlap was absent for sign language in deaf early signers (Andin et al., 2021). Hence, while increasing WM load was reflected in increased engagement of the frontoparietal working memory network, as predicted, the degradation of the visual signal instead caused activation of bilateral inferior occipital and temporal cortices. The lack of neural overlap, might challenge the validity of the ELU model, potentially reflecting modality-specificity. However, it should be noted that the same effect was found for hearing non-signers. Hence, the effect might be related to presentation modality rather than the language modality. Further studies investigating the auditory and visual domain within the same paradigm are needed to further evaluate the modality-generality of the ELU model.

Conclusion

1. Cognitive and communicative data patterns preceding the formulation of the ELU model (Rönnberg, 2003) were described. Individual cognitive ability was (and is) important for communicative competence.
2. Rapid multimodal and multilevel abstraction by means of RAMBPHO is supported by recent and previous experiments. WM stores these types of information in an on-line "moving window."
3. Parallel levels of mismatch negativity make the system extremely effective and rapid in deconstruction and reconstruction, prediction and postdiction.
4. A use-disuse principle was introduced and combined with a multimodal memory systems account to suggest why hearing loss strikes at ELTM, SLTM, and WM in that order of decreasing negative impact.
5. Recent preliminary modelling gives strong and more nuanced support of a mediation model of the original ELU parameters, which takes into account that processing speed is important for WM operations only in adverse SPIN conditions Phonology (i.e., RAMBPHO) is a

basic predictor variable of SPIN performance under all circumstances.
6. New models, such as the D-ELU were discussed. SLTM adaptations show acclimatization to certain non-habitual signal processing strategies, as well as to "perceptual doping."
7. Language proficiency and bilingualism are further factors discussed in the D-ELU context.
8. Generalization studies have shown that hearing-impaired and normal hearing persons equally on a cognition factor as predictor of SPIN performance. Moreover, the reliance on WM across different signal distortion conditions is equal when comparing persons with ADHD and normal hearing persons.
9. Boundary conditions are discussed in a sign language context in terms of preserved brain functions which are applied to another language modality; or in terms of a functional shift, where deaf participants' sign language use is assumed to change brain organization.

Author contributions

JR conceived and wrote a primary draft of ELU-related research and model development. JA was especially responsible for the sign language research. CS was responsible for the ERP and MEG studies. EH was responsible for overall coherence and the D-ELU model. All authors contributed to the article and approved the submitted version.

Funding

Writing of this manuscript by JR was supported by the Swedish Research Council (Grant No. 2017-06092; held by Anders Fridberger).

Conflict of interest

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

Publisher's note

References

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *Int. J. Audiol.* 47, S53–S71. doi: 10.1080/14992020802301142

Amichetti, N. M., White, A. G., and Wingfield, A. (2016). Multiple solutions to the same problem: Utilization of plausibility and syntax in sentence comprehension by older adults with impaired hearing. *Fron. Psychol.* 7:789. doi: 10.3389/fpsyg.2016.00789

Andersson, U., and Lidestam, B. (2005). Bottom-up driven speechreading in a speechreading expert: The case of AA (JK023). *Ear Hear.* 26, 214–224. doi: 10.1097/00003446-200504000-00008

Andin, J., and Holmer, E. (2022). Reorganization of large-scale brain networks in deaf signing adults: The role of auditory cortex in functional reorganization following deafness. *Neuropsychologia* 166:108139. doi: 10.1016/j.neuropsychologia.2021.108139

Andin, J., Holmer, E., Schönström, K., and Rudner, M. (2021). Working memory for signs with poor visual resolution: FMRI evidence of reorganization of auditory cortex in deaf signers. *Cerebral Cortex* 31, 3165–3176. doi: 10.1093/cercor/bhaa400

Arlinger, S. (2003). Negative consequences of uncorrected hearing loss—A review. *Int. J. Audiol.* 42, S17–S20. doi: 10.3109/14992020309074639

Arlinger, S., Lunner, T., Lyxell, B., and Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scand. J. Psychol.* 50, 371–384. doi: 10.1111/j.1467-9450.2009.00753.x

Ayasse, N. D., Hodosn, A., and Wingfield, A. (2021). The principle of least effort and comprehension of spoken sentences by younger and older adults. *Front. Psychol. Lang. Sci.* 12:629464. doi: 10.3389/fpsyg.2021.629464

Ayasse, N. D., and Wingfield, A. (2020). The two sides of linguistic context: Eye-tracking as a measure of semantic competition in spoken word recognition among younger and older adults. *Front. Hum. Neurosci.* 14:132. doi: 10.3389/fnhum.2020.00132

Bäckman, L., Small, B. J., and Fratiglioni, L. (2001). Stability of the preclinical episodic memory deficit in Alzheimer's disease. *Brain A J. Neurol.* 124, 96–102. doi: 10.1093/brain/124.1.96

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends Cogn. Sci.* 4, 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422

Baddeley, A., and Hitch, G. (1974). "Working memory," in *Psychology of learning and motivation*, 8, ed. G. H. Bower (Cambridge, MA: Academic Press), 47–89.

Badre, D. (2021). "Brain networks for cognitive control: Four unresolved questions," in *Intrusive thinking: From molecules to free will Strüngmann forum reports*, eds P. W. Kalivas and M. P. Paulus (Cambridge, MA: The MIT Press), 203–228. doi: 10.7551/mitpress/13875.001.0001

Baltes, P. B., and Lindenberger, U. (1997). Emergence of a powerful connection between sensory and cognitive functions across the adult life span: A new window to the study of cognitive aging? *Psychol. Aging* 12, 12–21. doi: 10.1037//0882-7974.12.1.12

Barrouillet, P., and Camos, V. (2020). "The time-based resource-sharing model of working memory," in *Working memory*, eds P. Barrouillet and V. Camos (Oxford: Oxford University Press), doi: 10.1093/oso/9780198842286.003.0004

Bavelier, D., and Neville, H. J. (2002). Cross-modal plasticity: Where and how? *Nat. Rev. Neurosci.* 3, 443–452. doi: 10.1038/nrn848

Bernstein, L. E., Tucker, P. E., and Auer, E. T. (1998). Potential perceptual bases for successful use of a vibrotactile speech perception aid. *Scand. J. Psychol.* 39, 181–186. doi: 10.1111/1467-9450.393076

Besser, J., Koelewijn, T., Zekveld, A. A., Kramer, S. E., and Festen, J. M. (2013). How linguistic closure and verbal working memory relate to speech recognition in noise—A review. *Trends Amplif.* 17, 75–93. doi: 10.1177/1084713813495459

Blomberg, R., Danielsson, H., Rudner, M., Söderlund, G. B. W., and Rönnberg, J. (2019). Speech processing difficulties in attention deficit hyperactivity disorder. *Front. Psychol.* 10:1536. doi: 10.3389/fpsyg.2019.01536

Bola, Ł, Zimmermann, M., Mostowski, P., Jednoróg, K., Marchewka, A., Rutkowski, P., et al. (2017). Task-specific reorganization of the auditory cortex in deaf humans. *Proc. Natl. Acad. Sci. U.S.A.* 114, E600–E609. doi: 10.1073/pnas.1609000114

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

Bsharat-Maalouf, D., and Karawani, H. (2022). Learning and bilingualism in challenging listening conditions: How challenging can it be? *Cognition* 222:105018. doi: 10.1016/j.cognition.2022.105018

Calandruccio, L., Dhar, S., and Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *J. Acoust. Soc. Am.* 128, 860–869. doi: 10.1121/1.3458857

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., et al. (1997). Activation of auditory cortex during silent lipreading. *Science (New York, N.Y.)* 276, 593–596. doi: 10.1126/science.276.5312.593

Cardin, V., Grin, K., Vinogradova, V., and Manini, B. (2020). Crossmodal reorganisation in deafness: Mechanisms for functional preservation and functional change. *Neurosci. Biobehav. Rev.* 113, 227–237. doi: 10.1016/j.neubiorev.2020.03.019

Cardin, V., Orfanidou, E., Rönnberg, J., Capek, C. M., Rudner, M., and Woll, B. (2013). Dissociating cognitive and sensory-neural plasticity in human superior temporal cortex. *Nat. Commun.* 4:1473. doi: 10.1038/ncomms2463

Cardin, V., Rudner, M., De Oliveira, R. F., Andin, J., Su, M. T., Beese, L., et al. (2018). The organization of working memory networks is shaped by early sensory experience. *Cerebral Cortex.* 28, 3540–3554. doi: 10.1093/cercor/bhx222

Cowan, N. (2005). *Working memory capacity*. London: Psychology Press. doi: 10.4324/9780203342398

Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learn.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6

Daneman, M., and Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonom. Bull. Rev.* 3, 422–433. doi: 10.3758/BF03214546

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A. G., Taylor, K., and McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *J. Exp. Psychol.* 134, 222–241. doi: 10.1037/0096-3445.134.2.222

Dolležal, L.-V., Brechmann, A., Klump, G. M., and Deike, S. (2014). Evaluating auditory stream segregation of SAM tone sequences by subjective and objective psychoacoustical tasks, and brain activity. *Front. Neurosci.* 8:119. doi: 10.3389/fnins.2014.00119

Edwards, B. (2016). A model of auditory-cognitive processing and relevance to clinical applicability. *Ear Hear.* 37, 85S–91S. doi: 10.1097/AUD.0000000000000308

Ericsson, K. A., and Kintsch, W. (1995). Long-term working memory. *Psychol. Rev.* 102, 211–245. doi: 10.1037/0033-295X.102.2.211

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x

Fell, J., and Axmacher, N. (2011). The role of phase synchronization in memory processes. *Nat. Rev. Neurosci.* 12, 105–118. doi: 10.1038/nrn2979

Fortunato, S., Forli, F., Guglielmi, V., De Corso, E., Paludetti, G., Berrettini, S., et al. (2016). A review of new insights on the association between hearing loss and cognitive decline in ageing. *Acta Otorhinolaryngol.* 36, 155–166. doi: 10.14639/0392-100X-993

Füllgrabe, C., Moore, B. C. J., and Stone, M. A. (2014). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Front. Aging Neurosci.* 6:347. doi: 10.3389/fnagi.2014.00347

Füllgrabe, C., and Rosen, S. (2016). On the (un)importance of working memory in speech-in-noise processingfor listeners with normal hearing thresholds. *Front. Psychol.* 7:1268. doi: 10.3389/fpsyg.2016.01268

Gatehouse, S., Naylor, G., and Elberling, C. (2003). Benefits from hearing aids in relation to the interaction between the user and the environment. *Int. J. Audiol.* 42, S77–S85. doi: 10.3109/14992020309074627

Gatehouse, S., Naylor, G., and Elberling, C. (2006). Linear and nonlinear hearing aid fittings—1: Patterns of benefit. *Int. J. Audiol.* 45, 130–152. doi: 10.1080/14992020500429518

Gates, G. A. (2012). Central presbycusis: An emerging view. *Otolaryngol. Head Neck Surg.* 147, 1-2. doi: 10.1177/0194599812446282

Gates, G. A., Cobb, J. L., Linn, R. T., Rees, T., Wolf, P. A., and D'Agostino, R. B. (1996). Central auditory dysfunction, cognitive dysfunction, and dementia in older people. *Arch. Otolaryngol. Head Neck Surg.* 122, 161–167. doi: 10.1001/archotol.1996.01890140047010

Gibson, J. J. (1966). *The senses considered as perceptual systems*. London: GEORGE ALLEN & UNWIN LTD.

Giraud, A. L., Truy, E., and Frackowiak, R. (2001). Imaging plasticity in cochlear implant patients. *Audiol. Neuro Otol.* 6, 381–393. doi: 10.1159/000046847

Giraud, A. L., Truy, E., Frackowiak, R. S., Grégoire, M. C., Pujol, J. F., and Collet, L. (2000). Differential recruitment of the speech processing system in healthy subjects and rehabilitated cochlear implant patients. *Brain* 123, 1391–1402. doi: 10.1093/brain/123.7.1391

Gordon-Salant, S., and Cole, S. S. (2016). Effects of age and working memory capacity on speech recognition performance in noise among listeners with normal hearing. *Ear Hear.* 37, 593–602. doi: 10.1097/AUD.0000000000000316

Gray, R., Sarampalis, A., Başkent, D., and Harding, E. E. (2022). Working-memory, alpha-theta oscillations and musical training in older age: Research perspectives for speech-on-speech perception. *Front. Aging Neurosci.* 14:806439. doi: 10.3389/fnagi.2022.806439

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Percept. Psychophys.* 28, 267–283. doi: 10.3758/BF03204386

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scand. Audiol.* 11, 79-87. doi: 10.3109/01050398209076203

Hagoort, P., and Levinson, S. C. (2014). "Neuropragmatics," in *The cognitive neurosciences*, 5th Edn, eds M. S. Gazzaniga and G. R. Mangun (Cambridge, MA: MIT Press), 667–674.

Hällgren, M., Larsby, B., and Arlinger, S. (2006). A Swedish version of the hearing in noise test (HINT) for measurement of speech recognition. *Int. J. Audiol.* 45, 227–237. doi: 10.1080/14992020500429583

Han, M. K., Storkel, H., and Bontempo, D. E. (2019). The effect of neighborhood density on children's word learning in noise. *J. Child Lang.* 46, 153–169. doi: 10.1017/S0305000918000284

Hanna, J., Mejias, S., Schelstraete, M.-A., Pulvermüller, F., Shtyrov, Y., and Van der Lely, H. K. J. (2014). Early activation of Broca's area in grammar processing as revealed by the syntactic mismatch negativity and distributed source analysis. *Cogn. Neurosci.* 5, 66–76. doi: 10.1080/17588928.2013.860087

Hasting, A. S., Kotz, S. A., and Friederici, A. D. (2007). Setting the stage for automatic syntax processing: The mismatch negativity as an indicator of syntactic priming. *J. Cogn. Neurosci.* 19, 386–400. doi: 10.1162/jocn.2007.19.3.386

Heinrich, A., Henshaw, H., and Ferguson, M. A. (2015). The relationship of speech intelligibility with hearing sensitivity, cognition, and perceived hearing difficulties varies for different speech perception tests. *Front. Psychol.* 6:782. doi: 10.3389/fpsyg.2015.00782

Heinrich, A., and Schneider, B. A. (2011). Elucidating the effects of ageing on remembering perceptually distorted word pairs. *Q. J. Exp. Psychol.* 64, 186–205. doi: 10.1080/17470218.2010.492621

Hewitt, D. (2017). Age-related hearing loss and cognitive decline: You haven't heard the half of it. *Front. Aging Neurosci.* 9:112. doi: 10.3389/fnagi.2017.00112

Hickok, G., and Poeppel, D. (2007). The cortical organization of speech processing. *Nat. Rev.* 8, 393–402. doi: 10.1038/nrn2113

Holmer, E., Heimann, M., and Rudner, M. (2016). Imitation, sign language skill and the developmental ease of language understanding (D-ELU) model. *Front Psychol.* 7:107. doi: 10.3389/fpsyg.2016.00107

Holmer, E., and Rudner, M. (2020). *Developmental ease of language understanding model and literacy acquisition: Evidence from deaf and hard-of-hearing signing children*. Washington, DC: Gallaudet University Press, 153–173.

Homman, L., Danielsson, H., and Rönnberg, J. (submitted). A structural equation mediation model captures the predictions amongst the parameters of the Ease of Language Understanding model. *Front. Psychol. Auditory Cogn. Neurosci.*

Hua, H., Johansson, B., Magnusson, L., Lyxell, B., and Ellis, R. J. (2017). Speech recognition and cognitive skills in bimodal cochlear implant users. *J. Speech Lang. Hear. Res.* 60, 2752–2763. doi: 10.1044/2017_JSLHR-H-16-0276

Humes, L. E. (2013). Understanding the speech-understanding problems of older adults. *Am. J. Audiol.* 22, 303–305. doi: 10.1044/1059-0889(2013/12-0066

Hygge, S., Rönnberg, J., Larsby, B., and Arlinger, S. (1992). Normal-hearing and hearing-impaired subjects' ability to just follow conversation in competing speech, reversed speech, and noise backgrounds. *J. Speech Hear. Res.* 35, 208–215. doi: 10.1044/jshr.3501.208

Janse, E., and Andringa, S. (2021). The roles of cognitive abilities and hearing acuity in older adults' recognition of words taken from fast and spectrally reduced speech. *Appl. Psycholinguist.* 42, 1–28. doi: 10.1017/S0142716421000047

Jones, G., Cabiddu, F., Andrews, M., and Rowland, C. (2021). Chunks of phonological knowledge play a significant role in children's word learning and explain effects of neighborhood size, phonotactic probability, word frequency and word length. *J. Memory Lang.* 119:104232. doi: 10.1016/j.jml.2021.104232

Jones, G., Gobet, F., and Pine, J. M. (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Dev. Sci.* 10, 853–873. doi: 10.1111/j.1467-7687.2007.00638.x

Just, M. A., and Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychol. Rev.* 99, 122–149. doi: 10.1037/0033-295x.99.1.122

Keidser, G., Seeto, M., Rudner, M., Hygge, S., and Rönnberg, J. (2015). On the relationship between functional hearing and depression. *Int. J. Audiol.* 54, 653–664. doi: 10.3109/14992027.2015.1046503

Kennedy-Higgins, D., Devlin, J. T., and Adank, P. (2020). Cognitive mechanisms underpinning successful perception of different speech distortions. *J. Acoust. Soc. Am.* 147, 2728–2740. doi: 10.1121/10.0001160

Kilman, L., Zekveld, A., Hällgren, M., and Rönnberg, J. (2014). The influence of non-native language proficiency on speech perception performance. *Front. Psychol.* 5:651. doi: 10.3389/fpsyg.2014.00651

Kral, A., and Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends Neurosci.* 35, 111–122. doi: 10.1016/j.tins.2011.09.004

Kraus, N., and Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nat. Rev.* 11, 599–605. doi: 10.1038/nrn2882

Kraus, N., Strait, D. L., and Parbery-Clark, A. (2012). Cognitive factors shape brain networks for auditory skills: Spotlight on auditory working memory. *Ann. N Y Acad. Sci.* 1252, 100–107. doi: 10.1111/j.1749-6632.2012.06463.x

Kurthen, I., Meyer, M., Schlesewsky, M., and Bornkessel-Schlesewsky, I. (2020). Individual differences in peripheral hearing and cognition reveal sentence processing differences in healthy older adults. *Front. Neurosci.* 14:573513. doi: 10.3389/fnins.2020.573513

Levänen, S. (1998). Neuromagnetic studies of human auditory cortex function and reorganization. *Scand. Audiol. Suppl.* 49, 1–6. doi: 10.1080/010503998420595

Lidestam, B., Moradi, S., Pettersson, R., and Ricklefs, T. (2014). Audiovisual training is better than auditory-only training for auditory-only speech-in-noise identification. *J. Acoust. Soc. Am.* 136, EL142–EL147. doi: 10.1121/1.4890200

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., et al. (2017). Dementia prevention, intervention, and care. *Lancet* 390, 2673–2734. doi: 10.1016/S0140-6736(17)31363-6

Luce, P. A., and Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear.* 19, 1–36. doi: 10.1097/00003446-199802000-00001

Ludman, C. N., Summerfield, A. Q., Hall, D., Elliott, M., Foster, J., Hykin, J. L., et al. (2000). Lip-reading ability and patterns of cortical activation studied using fMRI. *Br. J. Audiol.* 34, 225–230. doi: 10.3109/03005364000000132

Lunner, T. (2003). Cognitive function in relation to hearing aid use. *Int. J. Audiol.* 42, 49–58. doi: 10.3109/14992020309074624

Lunner, T., Rudner, M., and Rönnberg, J. (2009). Cognition and hearing aids. *Scand. J. Psychol.* 50, 395–403. doi: 10.1111/j.1467-9450.2009.00742.x

Lunner, T., and Sundewall-Thorén, E. (2007). Interactions between cognition, compression, and listening conditions: Effects on speech-in-noise performance in a two-channel hearing aid. *J. Am. Acad. Audiol.* 18, 604–617. doi: 10.3766/jaaa.18.7.7

Lyxell, B. (1994). Skilled speechreading: A single-case study. *Scand. J. Psychol.* 35, 212–219. doi: 10.1111/j.1467-9450.1994.tb00945.x

Lyxell, B., Arlinger, S., Andersson, J., Harder, H., Näsström, E., Svensson, H., et al. (1996). Information-processing capabilities and cochlear implants. Pre-operative predictors for speech understanding. *J. Deaf Stud. Deaf Educ.* 11, 190–201. doi: 10.1093/oxfordjournals.deafed.a014294

Lyxell, B., and Rönnberg, J. (1987). Guessing and speechreading. *Br. J. Audiol.* 21, 13–20. doi: 10.3109/03005368709077769

MacSweeney, M., Campbell, R., Calvert, G. A., McGuire, P. K., David, A. S., Suckling, J., et al. (2001). Dispersed activation in the left temporal cortex for speech-reading in congenitally deaf people. *Proc. Biol. Sci.* 268, 451–457. doi: 10.1098/rspb.2000.0393

Marsh, J. E., and Campbell, T. A. (2016). Processing complex sounds passing through the rostral brainstem: The new early filter model. *Front. Neurosci.* 10:136. doi: 10.3389/fnins.2016.00136

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition* 25, 71–102. doi: 10.1016/0010-0277(87)90005-9

Marsja, E., Stenbäck, V., Moradi, S., Danielsson, H., and Rönnberg, J. (2022). Is having hearing loss fundamentally different? Multigroup structural equation modeling of the effect of cognitive functioning on speech identification. *Ear Hear.* doi: 10.1097/AUD.0000000000001196 [Epub ahead of print].

Mattys, S. L., Davis, M. H., Bradlow, A. R., and Scott, S. K. (2012). Speech recognition in adverse conditions: A review. *Lang. Cogn. Process.* 27, 953–978. doi: 10.1080/01690965.2012.705006

McCoy, S. L., Tun, P. A., Cox, L. C., Colangelo, M., Stewart, R. A., and Wingfield, A. (2005). Hearing loss and perceptual effort: Downstream effects on older adults' memory for speech. *Q. J. Exp. Psychol. A Hum. Exp. Psychol.* 58, 22–33. doi: 10.1080/02724980443000151

Merabet, L. B., and Pascual-Leone, A. (2010). Neural reorganization following sensory loss: The opportunity of change. *Nat. Rev. Neurosci.* 11, 44–52. doi: 10.1038/nrn2758

Mishra, S., Shubhadarshan, A., Behera, D., and Sahoo, R. (2021). Can working memory capacity predict speech perception in presence of noise in older adults? *Int. J. Educ. Res.* 10, 18–23.

Moradi, S., Lidestam, B., Saremi, A., and Rönnberg, J. (2014b). Gated auditory speech perception: Effects of listening conditions and cognitive capacity. *Front. Psychol.* 5:531. doi: 10.3389/fpsyg.2014.00531

Moradi, S., Lidestam, B., Hällgren, M., and Rönnberg, J. (2014a). Gated auditory speech perception in elderly hearing aid users and elderly normal-hearing individuals: Effects of hearing impairment and cognitive capacity. *Trends Hear.* 18:2331216514545406. doi: 10.1177/2331216514545406

Moradi, S., Lidestam, B., Ng, E. H. N., Danielsson, H., and Rönnberg, J. (2019). Perceptual doping: An audiovisual facilitation effect on auditory speech processing, from phonetic feature extraction to sentence identification in noise. *Ear Hear.* 40, 312–327. doi: 10.1097/AUD.0000000000000616

Moradi, S., Lidestam, B., and Rönnberg, J. (2013). Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. *Front. Psychol.* 4:359. doi: 10.3389/fpsyg.2013.00359

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear Hear.* 16, 6–18.

Näätänen, R., and Escera, C. (2000). Mismatch negativity: Clinical and other applications. *Audiol. Neuro Otol.* 5, 105–110. doi: 10.1159/000013874

Ng, E. H. N., and Rönnberg, J. (2020). Hearing aid experience and background noise affect the robust relationship between working memory and speech recognition in noise. *Int. J. Audiol.* 59, 208–218. doi: 10.1080/14992027.2019.1677951

Ng, E. H., Rudner, M., Lunner, T., Pedersen, M. S., and Rönnberg, J. (2013). Effects of noise and working memory capacity on memory processing of speech for hearing-aid users. *Int. J. Audiol.* 52, 433–441. doi: 10.3109/14992027.2013.776181

Ng, E. H. N., Rudner, M., Lunner, T., and Rönnberg, J. (2015). Noise reduction improves memory for target language speech in competing native but not foreign language speech. *Ear Hear.* 36, 82–91. doi: 10.1097/AUD.0000000000000080

Nilsson, L.-G., Bäckman, L., Erngrund, K., Nyberg, L., Adolfsson, R., Bucht, G., et al. (1997). The Betula prospective cohort study: Memory, health, and aging. *Aging Neuropsychol. Cogn.* 4, 1–32. doi: 10.1080/13825589708256633

Obleser, J., Wostmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* 32, 12376–12383. doi: 10.1523/JNEUROSCI.4908-11.2012

Peelle, J. E. (2018). Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior. *Ear Hear.* 39, 204–214. doi: 10.1097/AUD.0000000000000494

Petersen, E. B., Wöstmann, M., Obleser, J., Stenfelt, S., and Lunner, T. (2015). Hearing loss impacts neural alpha oscillations under adverse listening conditions. *Front. Psychol.* 6:177. doi: 10.3389/fpsyg.2015.00177

Pichora-Fuller, M. K. (2003). Processing speed and timing in aging adults: Psychoacoustics, speech perception, and comprehension. *Int. J. Audiol.* 42, 59–67. doi: 10.3109/14992020309074625

Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., et al. (2016). Hearing impairment and cognitive energy: The framework for understanding effortful listening (FUEL). *Ear Hear.* 37, 5S–27S. doi: 10.1097/AUD.0000000000000312

Powell, D. S., Oh, E. S., Reed, N. S., Lin, F. R., and Deal, J. A. (2021). Hearing loss and cognition: What we know and where we need to go. *Front. Aging Neurosci.* 13:769405. doi: 10.3389/fnagi.2021.769405

Richter, M. (2013). A closer look into the multi-layer structure of motivational intensity theory. *Soc. Pers. Psychol. Compass* 7, 1–12. doi: 10.1111/spc3.12007

Rönnberg, J. (1990). Cognitive and communicative function: The effects of chronological age and "handicap age". *Eur. J. Cogn. Psychol.* 2, 253–273. doi: 10.1080/09541449008406207

Rönnberg, J. (1993). Cognitive characteristics of skilled tactiling: The case of GS. *Eur. J. Cogn. Psychol.* 5, 19–33. doi: 10.1080/09541449308406512

Rönnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *Int. J. Audiol.* 42, S68–S76. doi: 10.3109/14992020309074626

Rönnberg, J., Andersson, J., Andersson, U., Johansson, K., Lyxell, B., and Samuelsson, S. (1998). Cognition as a bridge between signal and dialogue: Communication in the hearing impaired and deaf. *Scand. Audiol.* 27, 101–108. doi: 10.1080/010503998420720

Rönnberg, J., Andersson, J., Samuelsson, S., Söderfeldt, B., Lyxell, B., and Risberg, J. (1999). A speechreading expert: The case of MM. *J. Speech Lang. Hear. Res.* 42, 5–20. doi: 10.1044/jslhr.4201.05

Rönnberg, J., Danielsson, H., Rudner, M., Arlinger, S., Sternäng, O., Wahlin, A., et al. (2011). Hearing loss is negatively related to episodic and semantic long-term memory but not to short-term memory. *J. Speech Lang. Hear. Res.* 54, 705–726. doi: 10.1044/1092-4388(2010/09-0088

Rönnberg, J., Holmer, E., and Rudner, M. (2019). Cognitive hearing science and ease of language understanding. *Int. J. Audiol.* 58, 247–261. doi: 10.1080/14992027.2018.1551631

Rönnberg, J., Holmer, E., and Rudner, M. (2021). Cognitive hearing science: Three memory systems, two approaches, and the ease of language understanding model. *J. Speech Lang. Hear. Res.* 64, 359–370. doi: 10.1044/2020_JSLHR-20-00007

Rönnberg, J., Holmer, E., and Rudner, M. (2022). "Working memory and the ease of language understanding model," in *The Cambridge handbook of working memory*, eds W. Schwieter and Z. E. Wen (Cambridge: Cambridge University Press), 197–218. doi: 10.1017/9781108955638.013

Rönnberg, J., Hygge, S., Keidser, G., and Rudner, M. (2014). The effect of functional hearing loss and age on long- and short-term visuospatial memory: Evidence from the UK biobank resource. *Front. Aging Neurosci.* 6:326. doi: 10.3389/fnagi.2014.00326

Rönnberg, J., Lunner, T., Ng, E. H. N., Lidestam, B., Zekveld, A. A., Sörqvist, P., et al. (2016). Hearing impairment, cognition and speech understanding: Exploratory factor analyses of a comprehensive test battery for a group of hearing aid users, the n200 study. *Int. J. Audiol.* 55, 623–642. doi: 10.1080/14992027.2016.1219775

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., et al. (2013). The Ease of Language Understanding (ELU) model: Theoretical, empirical, and clinical advances. *Front. Syst. Neurosci.* 7:31. doi: 10.3389/fnsys.2013.00031

Rönnberg, J., Öhngren, G., and Nilsson, L. G. (1982). Hearing deficiency, speechreading and memory functions. *Scand. Audiol.* 11, 261–268. doi: 10.3109/01050398209087477

Rönnberg, J., Öhngren, G., and Nilsson, L. G. (1983). Speechreading performance evaluated by means of TV and real-life presentation. A comparison between a normally hearing, moderately and profoundly hearing-impaired group. *Scand. Audiol.* 12, 71–77. doi: 10.3109/01050398309076227

Rönnberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). Cognition counts: A working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 47, S99–S105. doi: 10.1080/14992020802301167

Rönnberg, J., Rudner, M., and Ingvar, M. (2004). Neural correlates of working memory for sign language. *Brain Res. Cogn. Brain Res.* 20, 165–182. doi: 10.1016/j.cogbrainres.2004.03.002

Rönnberg, J., Rudner, M., Lunner, T., and Zekveld, A. (2010). When cognition kicks in: Working memory and speech understanding in noise. *Noise Health* 12:263. doi: 10.4103/1463-1741.70505

Rudner, M., Foo, C., Rönnberg, J., and Lunner, T. (2009). Cognition and aided speech recognition in noise: Specific role for cognitive factors following nine-week experience with adjusted compression settings in hearing aids. *Scand. J. Psychol.* 50, 405–418. doi: 10.1111/j.1467-9450.2009.00745.x

Rudner, M., Foo, C., Sundewall-Thorén, E., Lunner, T., and Rönnberg, J. (2008). Phonological mismatch and explicit cognitive processing in a sample of 102 hearing-aid users. *Int. J. Audiol.* 47, S91–S98. doi: 10.1080/14992020802304393

Rudner, M., Fransson, P., Ingvar, M., Nyberg, L., and Rönnberg, J. (2007). Neural representation of binding lexical signs and words in the episodic buffer of working memory. *Neuropsychologia* 45, 2258–2276. doi: 10.1016/j.neuropsychologia.2007.02.017

Rudner, M., Seeto, M., Keidser, G., Johnson, B., and Rönnberg, J. (2019). Poorer speech reception threshold in noise is associated with lower brain volume in auditory and cognitive processing regions. *J. Speech Lang. Hear. Res.* 62, 1117–1130. doi: 10.1044/2018_JSLHR-H-ASCC7-18-0142

Samuelsson, S., and Rönnberg, J. (1993). Implicit and explicit use of scripted constraints in lip-reading. *Eur. J. Cogn. Psychol.* 5, 201–233. doi: 10.1080/09541449308520116

Sarampalis, A., Kalluri, S., Edwards, B., and Hafter, E. (2009). Objective measures of listening effort: Effects of background noise and noise reduction. *J. Speech Lang. Hear. Res.* 52, 1230–1240. doi: 10.1044/1092-4388(2009/08-0111

Signoret, C., Andersen, L. M., Dahlström, Ö, Blomberg, R., Lundqvist, D., Rudner, M., et al. (2020). The influence of form- and meaning-based predictions on cortical speech processing under challenging listening conditions: A MEG study. *Front. Neurosci.* 14:573254. doi: 10.3389/fnins.2020.573254

Signoret, C., Gaudrain, E., and Perrin, F. (2013). Similarities in the neural signature for the processing of behaviorally categorized and uncategorized speech sounds. *Eur. J. Neurosci.* 37, 777–785. doi: 10.1111/ejn.12097

Signoret, C., Johnsrude, I., Classon, E., and Rudner, M. (2018). Combined effects of form- and meaning-based predictability on perceived clarity of speech. *Hum. Percept. Perform.* 44, 277–285. doi: 10.1037/xhp0000442

Signoret, C., and Rudner, M. (2019). Hearing impairment and perceived clarity of predictable speech. *Ear Hear.* 40, 1140–1148. doi: 10.1097/AUD.0000000000000689

Silvestrini, N., Musslick, S., Berry, A. S., and Vassena, E. (2022). An integrative effort: Bridging motivational intensity theory and recent neurocomputational and neuronal models of effort and control allocation. *Psychol. Rev.* doi: 10.1037/rev0000372

Söderfeldt, B., Ingvar, M., Rönnberg, J., Eriksson, L., Serrander, M., and Stone-Elander, S. (1997). Signed and spoken language perception studied by positron emission tomography. *Neurology* 49, 82–87. doi: 10.1212/wnl.49.1.82

Söderfeldt, B., Rönnberg, J., and Risberg, J. (1994). Regional cerebral blood flow in sign language users. *Brain* 46, 59-68. doi: 10.1006/brln.1994.1004

Sörqvist, P., and Rönnberg, J. (2012). Episodic long-term memory of spoken discourse masked by speech: What is the role for working memory capacity? *J. Speech Lang. Hear. Res.* 55, 210–218. doi: 10.1044/1092-4388(2011/10-0353

Sörqvist, P., Stenfelt, S., and Rönnberg, J. (2012). Working memory capacity and visual-verbal cognitive load modulate auditory-sensory gating in the brainstem: Toward a unified view of attention. *J. Cogn. Neurosci.* 24, 2147–2154. doi: 10.1162/jocn_a_00275

Souza, P., and Arehart, K. (2015). Robust relationship between reading span and speech recognition in noise. *Int. J. Audiol.* 54, 705–713. doi: 10.3109/14992027.2015.1043062

Souza, P., Arehart, K., Schoof, T., Anderson, M., Strori, D., and Balmert, L. (2019). Understanding variability in individual response to hearing aid signal processing in wearable hearing aids. *Ear Hear.* 40, 1280–1292. doi: 10.1097/AUD.0000000000000717

Souza, P., and Sirow, L. (2014). Relating working memory to compression parameters in clinically fit hearing AIDS. *Am. J. Audiol.* 23, 394–401. doi: 10.1044/2014_AJA-14-0006

Stenbäck, V., Hällgren, M., Lyxell, B., and Larsby, B. (2015). The Swedish hayling task, and its relation to working memory, verbal ability, and speech-recognition-in-noise. *Scand. J. Psychol.* 56, 264–272. doi: 10.1111/sjop.12206

Stenbäck, V., Marsja, E., Hällgren, M., Lyxell, B., and Larsby, B. (2021). The contribution of age, working memory capacity, and inhibitory control on speech recognition in noise in young and older adult listeners (world). *J Speech Lang Hear Res.* 64, 4513–4523. doi: 10.1044/2021_JSLHR-20-00251

Stenfelt, S., and Rönnberg, J. (2009). The signal-cognition interface: Interactions between degraded auditory signals and cognitive processes. *Scand. J. Psychol.* 50, 385–393. doi: 10.1111/j.1467-9450.2009.00748.X

Tillberg, I., Rönnberg, J., Svärd, I., and Ahlner, B. (1996). Audio-visual speechreading in a group of hearing aid users. The effects of onset age, handicap age, and degree of hearing loss. *Scand. Audiol.* 25, 267–272. doi: 10.3109/01050399609074966

Tse, C.-Y., Rinne, T., Ng, K. K., and Penney, T. B. (2013). The functional role of the frontal cortex in pre-attentive auditory change detection. *Neuroimage* 83, 870–879. doi: 10.1016/j.neuroimage.2013.07.037

Tun, P. A., McCoy, S., and Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychol. Aging* 24, 761–766. doi: 10.1037/a0014802

Twomey, T., Waters, D., Price, C. J., Evans, S., and MacSweeney, M. (2017). How auditory experience differentially influences the function of left and right superior temporal cortices. *J. Neurosci.* 37, 9564–9573. doi: 10.1523/JNEUROSCI.0846-17.2017

Uchida, Y., Nishita, Y., Otsuka, R., Sugiura, S., Sone, M., Yamasoba, T., et al. (2021). Aging brain and hearing: A mini-review. *Front. Aging Neurosci.* 13:791604. doi: 10.3389/fnagi.2021.791604

Usrey, W. M., and Sherman, S. M. (2019). Corticofugal circuits: Communication lines from the cortex to the rest of the brain. *J. Comp. Neurol.* 527, 640–650. doi: 10.1002/cne.24423

Van Engen, K. J., and Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *J. Acoust. Soc. Am.* 121, 519–526. doi: 10.1121/1.2400666

Vermeire, K., Knoop, A., De Sloovere, M., Bosch, P., and van den Noort, M. (2019). Relationship between working memory and speech-in-noise recognition in young and older adult listeners with age-appropriate hearing. *J. Speech Lang. Hear. Res.* 62, 3545–3553. doi: 10.1044/2019_JSLHR-H-18-0307

Wayne, R. V., and Johnsrude, I. S. (2015). A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Ageing Res. Rev.* 23, 154–166. doi: 10.1016/j.arr.2015.06.002

Younan, D., Petkus, A. J., Widaman, K. F., Wang, X., Casanova, R., Espeland, M. A., et al. (2020). Particulate matter and episodic memory decline mediated by early neuroanatomic biomarkers of Alzheimer's disease. *Brain* 143, 289–302. doi: 10.1093/brain/awz348

Zatorre, R. J. (2001). Do you see what I'm saying? Interactions between auditory and visual cortices in cochlear implant users. *Neuron* 31, 13–14. doi: 10.1016/s0896-6273(01)00347-6

Zekveld, A. A., Pronk, M., Danielsson, H., and Rönnberg, J. (2018). Reading behind the lines: The factors affecting the text reception threshold in hearing aid users. *J. Speech Lang. Hear. Res.* 61, 762–775. doi: 10.1044/2017_JSLHR-H-17-0196

Zekveld, A. A., Rudner, M., Johnsrude, I. S., and Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *J. Acoust. Soc. Am.* 134, 2225–2234. doi: 10.1121/1.4817926

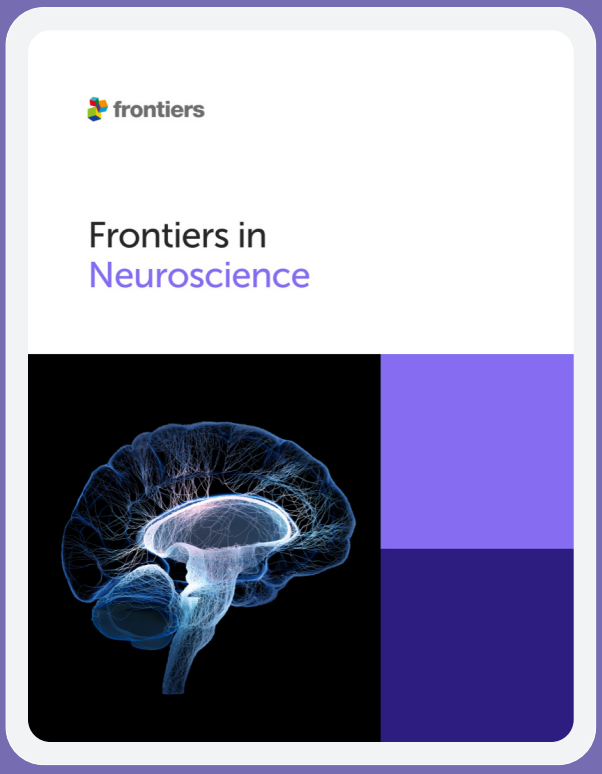
Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact