Speech-DPOAEs for probing speech processing in the inner ear

Marina Saiz-Alia,[1, a)] Peter Miller,[1, b)] and Tobias Reichenbach[2, 1, c)]

[1)]*Department of Bioengineering, Imperial College London, South Kensington Campus, SW7 2BX, U.K.*
[2)]*Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Konrad-Zuse-Strasse 3/5, 91052 Erlangen, Germany*

[a)]*Electronic mail: marina.saiz-alia15@imperial.ac.uk*
[b)]*Electronic mail: peter.miller15@imperial.ac.uk*
[c)]*Corresponding author: tobias.j.reichenbach@fau.de*

Abstract. Speech is a complex real-world stimulus that consists of time-varying contributions in many frequency bands. The inner ear, or cochlea, not only transduces the acoustic vibrations into electrical signals, but also spatially segregates the acoustic waveform into different frequency components, as well as mechanically amplifies the vibrations in frequency bands where the amplitude is low. This processing of the speech signal can potentially be controlled through efferent nerve fibers that extend from higher processing centers in the brain to the cochlea. However, we are currently lacking methodological tools to non-invasively investigate the putative role of top-down feedback on speech processing in humans. Here we develop such a method that builds on distortion-product otoacoustic emissions (DPOAEs) related to the temporal fine structure of the voiced parts of speech (speech-DPOAEs).

INTRODUCTION

Understanding speech is crucial for human interaction and communication [1]. Humans are indeed experts at understanding speech even in adverse listening conditions such as loud background noise in a noisy restaurant, coffee shop or in traffic.

Our ability to understand speech in noise is, however, vulnerable. Hearing impairments affect more than 16% of the adult population in the E.U., and more than 5% in children [1, 2, 3, 4, 5]. Moreover, hearing impairments progress with age and are hence a particular problem in the E.U.'s aging society: 40% of people above age 50 and 70% of those above age 70 have some form of hearing disorder. Such hearing impairment impacts first and foremost the ability to understand speech in noisy backgrounds, and current hearing aids are unfortunately of little help [6]. The resulting difficulties in communicating in many social settings have major effects on an afflicted person's personal, social and economic development [4, 7].

A better diagnosis of hearing impairment as well as the development of corresponding treatments require a better understanding of the neural pathways of hearing, and in particular of the neural mechanisms that allow a healthy person to understand speech in noise. This issue has traditionally been tackled by investigating the underlying processes in the ear and brain through simple acoustic signals such as pure tones or clicks. As an example, pure-tone audiometry uses pure tones of different frequencies to assess the hearing threshold of a subject at those frequencies. As another example, click-evoked auditory brainstem responses record the electrical activity of the brainstem at different latencies when a subject hears a series of short clicks. These methods are highly useful for assessing the basic functioning of different parts of the auditory system, such as the inner ear and the auditory brainstem. However, they do not fully reveal the neural mechanisms that allow us to process more complex signals such as speech in background noise.

Recent research has therefore sought to measure neural responses to natural speech, and therefrom learn more about the neural mechanisms of speech processing. An important example concerns the neural response to the temporal fine structure of speech. Many parts of speech are voiced, originating from vibrations of the vocal fold. The vibration occurs at the so-called fundamental frequency, typically between 100 - 300 Hz. This frequency as well as its many higher harmonics constitute the so-called temporal fine structure and carry most of the acoustic energy of a voiced speech segment. When measuring the neural response to running speech through electroencephalography (EEG) or magnetoencephalography (MEG), a strong signal can be found at the fundamental frequency, as well as, to a much lower degree, at the higher harmonics. This response presumably arises predominantly from the auditory brainstem, although recent work has also identified cortical contributions as well [8, 9, 10, 11].

The neural response at the fundamental frequency of speech has traditionally been studied using many repeated presentations of the same short speech token, such as a single syllable [12]. We as well as others, however, recently developed the statistical methodology to measure this neural response in running speech [13, 14]. To deal with the time-varying nature of the fundamental frequency in natural speech, we first extracted a waveform that, at each time point, oscillates at the fundamental frequency. In a second step, we then related that waveform to the EEG recordings

Nonlinearity and Hearing: Advances in Theory and Experiment
AIP Conf. Proc. 3062, 030010-1–030010-6; https://doi.org/10.1063/5.0189618
Published by AIP Publishing. 978-0-7354-4844-5/$30.00

obtained when a person listened to the corresponding speech. The relation allowed to infer both magnitude and latency of the neural response [13, 15].

Importantly, our work on the neural response at the fundamental frequency of speech showed that this neural signal is consistently modulated by selective attention [16]. Because the neural response to the fundamental frequency of speech arises predominantly in subcortical structures such as the inferior colliculus, the modulation of this response through selective attention likely involves efferent feedback from the auditory cortex to these subcortical areas [17, 18].

An important part of the auditory system for which responses to natural speech have not yet been measured is the inner ear. The inner ear houses the mechanosensitive hair cells that transduce the mechanical sound vibration into electrical signals. Strikingly, through a so-called active process, these hair cells also mechanically amplify weak sound stimulation [19, 20]. Otoacoustic emissions (OAEs) are generated as a byproduct of the amplification [21]. As an example, when stimulated with two frequencies, so-called distortion-product otoacoustic emissions (DPOAEs) emerge at other frequencies that correspond to linear combinations of the primary ones. These sounds can be measured with a sensitive microphone in the ear canal.

DPOAEs have so far almost exclusively been elicited by pure tones, such that the emissions are pure tones as well. These measurements have established DPOAEs as a highly useful diagnostic test of the inner ear's functioning that is widely applied to assess hearing in newborns [22].

Importantly, DPOAEs as well as other types of OAEs have also been employed to show that the inner ear's activity can be influenced by efferent feedback from higher neural processing centers. In particular, the inner ear's outer hair cells that provide mechanical feedback and generate the DPOAEs are innervated by the medial olivocochlear (MOC) bundle. Activation of this bundle, for instance through ipsilateral or contralateral noise, leads to a reduction of mechanical activity of the outer hair cells and thereby to a reduction of DPOAEs (MOC reflex) [23, 24].

Computational studies have shown that the regulation of cochlear activity through the efferent feedback can aid speech processing [25]. Such efferent feedback can, for instance, help to filter out unwanted background noise by reducing the amplification at cochlear locations that lie inbetween locations at which resolved harmonics of the target speech are detected. The efferent feedback to the cochlea could therefore contribute significantly to speech-in-noise comprehension, and could explain a portion of the attentional modulation of the neural response to the temporal fine structure of speech that we have measured earlier [13, 15]. On the other hand, impairment with this pathway could constitute one form of speech-in-noise difficulty.

However, due to the complexity of speech, the contribution of the active feedback in the inner ear to speech processing has not yet been investigated.

Methods and Results

Our approach to measure speech-DPOAEs, that is, DPOAEs related to the temporal fine structure of speech, was as follows (Figure 1A). We first extracted two waveforms w_{P1} and w_{P2} that corresponded to two nearby harmonics of the fundamental frequency $f(t)$ of a speech signal, say to $a(f(t)$ and $bf(t)$ with a and b two close integers. The two waveforms were obtained by bandpass-filtering the speech signal in a narrow region of the frequency range of the corresponding harmonics.

If the speech signal was monotone, that is, if the fundamental frequency $f(t)$ was constant at a value f, the waveforms w_{P1} and w_{P2} would simply be sinusoidal oscillations at the constant frequencies af and bf of the corresponding harmonics. In natural speech, however, the fundamental frequency varies over time, such that the frequencies of the harmonics vary as well. The waveforms w_{P1} and w_{P2} are therefore not sinusoidal, although, on a short temporal scale, they appear similar to a sinusoidal oscillation at the waveform's instantaneous frequency $af(t)$ or $bf(t)$.

Because the waveforms w_{P1} and w_{P2} are narrow-band signals, the cochlear activity elicited by them is likely similar to that in response to a pure tone (Figure 1B). In particular, the two waveforms will elicit traveling waves on the basilar membrane that peak at near the characteristic place of the instantaneous frequencies $af(t)$ and $bf(t)$ of the two waveforms. In the overlap regions of the peak, the cochlear nonlinearity will then produce distortion, such as at the cubic distortion frequencies $2af(t)-bf(t)=(2a-b)f(t)$ and $2bf(t)-af(t)=(2b-a)f(t)$. These cubic distortion frequencies are therefore again harmonics of the speech signal's fundamental frequency $f(t)$.

In the case of monotone speech, the cubic distortion signals would be pure tones, at the constant frequencies $(2a-b)f$ and $(2b-a)f$. Both the stimulation signals w_{P1} and w_{P2} and the speech-DPOAEs could then be measured through the Fourier transform of a microphone recording in the ear canal. In particular, the stimulation signals would

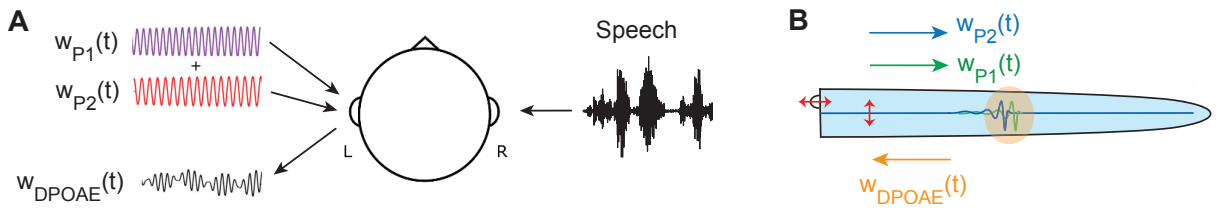

FIGURE 1. Principle of speech-DPOAEs. A, One ear is stimulated with speech. DPOAEs related to the harmonic structure of that speech signal can then be evoked and recorded from the contralateral ear. To this end, two waveforms w_{P1} and w_{P2} that correspond to two nearby harmonics of the fundamental frequency of the speech signal are delivered. The nonlinear activity in the cochlea leads to the generation of distortion products, and hence to the emission of a wave w_{DPOAE}. This emission can be measured by correlation the acoustic signal in the ear canal with the waveform w_{DPOAE}. **B**, The distortion product is generated by the cochlear nonlinearity, presumably in the overlap region of the peaks of the traveling waves elicited by w_{P1} and w_{P2}.

be evident through peaks at the frequencies af and bf, and the cubic DPOAEs through peaks at the frequencies $(2a-b)f$ and $(2b-a)f$.

Because the fundamental frequency of natural speech varies, neither the stimulation waveforms nor the resulting speech-DPOAEs will be apparent in the Fourier transform of the microphone recording. To assess these emissions, we instead employed a technique that we recently used to measure the neural response at the fundamental frequency [13]. There, we computed the cross-correlation of the expected waveform w_{DPOAE} with the microphone recording. If the microphone recording contains an emission at the time-varying distortion frequency, the cross-correlation will exhibit a peak at the delay of the emission. The peak will be the broader the more narrow-band the waveform w_{DPOAE} is. In particular, for monotone speech where the emission consists of a pure tone, the peak will be infinitely broad. For natural speech, however, where the emission varies across a certain frequency range, the peak in the cross-correlation will have a finite width, allowing to assess the presence of the emission as well as to obtain an estimate of its latency.

Of course, this technique can not only be employed to measure the presence of the emission waveform w_{DPOAE} in the microphone recording, but also those of the stimulation waveforms w_{P1} and w_{P2}.

To put these ideas to the test, we constructed waveforms that oscillated at three nearby harmonics of the fundamental frequency of a voiced speech signal, the 7th, the 9th, and the 11th (Figure 2). We stimulated one ear (contralateral to the ear to which the speech stimuli were presented) with two of these three harmonics, the 9th and the 11th. We measured the signal in the ear canal with a very sensitive microphone (ER10X, Etymotics).)We verified the presence of the stimulation waveforms w_9 and w_{11} in the ear canal through computing the cross-correlation of the waveforms with the microphone recording. Both cross-correlation showed pronounced peaks at the delay of 0 ms, evidencing that both waveforms were present instantaneously in the ear canal where they were generated by small speakers.

The inner ear's nonlinear distortion then created a signal, amongst others, at the 7th harmonic. We measured this signal by cross-correlating the microphone recording with the waveform w_7 that corresponded to the 7th harmonic. The cross-correlation showed a peak at a delay of about 2.2 ms, in line with estimates of the temporal delay that is required for the stimulation signals to propagate into the inner ear and for the distortion product to propagate back from its generation site to the middle ear and the ear canal. This showed that the distortion signal was indeed present in the ear canal.

We repeated the measurement, but this time positioning the DPOAE probe with the speakers and microphone outside the ear canal (Figure 2B, D, F). The stimulation waveforms w_9 and w_{11} were still present in the microphone recording, as evidence by pronounced peaks in the corresponding cross-correlation. The waveform w_7 that corresponded to a distortion product, however, did not yield a peak in the cross-correlation. The speech-DPOAE could therefore not be detected in the microphone recording. This verified that the speech-DPOAE observed above was indeed produced by the inner ear and not by the DPOAE probe itself.

Discussion

We have proposed a method to measure DPOAEs related to the temporal fine structure of natural speech. The method extracts waveforms of the harmonics of the time-varying fundamental frequency of a speech signal. It then employs two waveforms at nearby frequencies to stimulate the cochlea. The resulting speech-DPOAE emerges at another

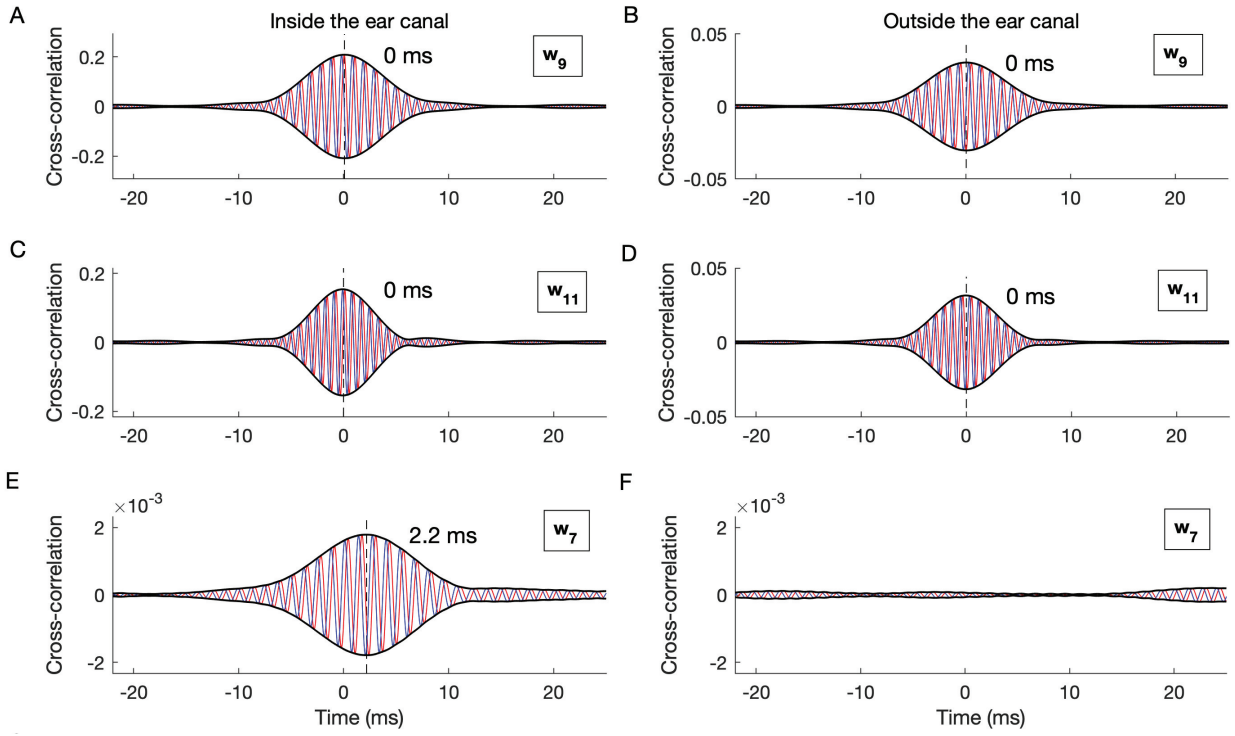

FIGURE 2. An example of recorded speech-DPOAEs. **A,C,E**, We presented a subject with the waveforms w_9 and w_{11}, corresponding to the 9th and 11th harmonics of a speech signal, in the contralateral ear, and recorded the sound with a microphone in the same ear. **A,C**, We show the correlation of the microphone recording with the waveforms w_9 and w_{11} as red lines. We also show the correlation with these waveforms shifted by 90 degrees (blue line) as well as the envelope of the cross-correlation (black line). The cross-correlations of the microphone exhibits a peak at 0 ms, confirming that the microphone is picking up the sound. **E**, The cross-correlation of the microphone recording with the waveform w_7 of the corresponding DPOAE shows a peak at a latency of 2.2 ms, evidencing an otoacoustic emission at this delay. **B, D, F**, When the OAE probe is placed outside the ear canal, the microphone in the probe still detects the signals that are played by the probe (**B,D**), but not the otoacoustic emission (**F**), confirming that the speech-DPOAE w_7 is generated by the inner ear and not by the measurement device.

waveform and can be measured through cross-correlation of the waveform with a microphone recording from the ear canal.

In our work on verifying this methodology, we have stimulated one ear with the speech signal, and measured the speech-DPOAEs from the contralateral ear. Future work may extend this method to measure speech-DPOAEs from the ipsilateral ear as well.

Importantly, the speech-DPOAEs allow to assess cochlear responses to natural speech in humans in a non-invasive manner. Natural speech has recently received increased attention. In particular, we have shown that, through efferent feedback, the early neural response at the fundamental waveform is already modulated by selective attention to one of two competing talkers [13]. The attentional modulation can allow to decode the attentional focus of a listener from short segments of EEG recordings, suggesting potential applications in mind-controlled auditory prosthesis that decode a listener's attention from neural recordings and then use it to steer the prosthesis towards the target of attention [26]. In addition, and pointing towards clinical applications, we have shown that the strength of the attentional modulation relates to a subject's ability to understand speech in noise [27].

Because the efferent feedback extends to the inner ear, the latter might be involved in selective attention to speech as well. We have already started to investigate this issue through the use of our speech-DPOAEs [28]. Our results provide tentative evidence that the speech-DPOAEs are larger when the corresponding speech signal is attended then-https://www.overleaf.com/project/620b719042daa8a0a93cd1ec when it is ignored. In the future, it will be important to further investigate such attentional effects, for instance regarding the issue of resolved versus unresolved harmonics of speech. Moreover, it will be interesting to see how such attentional effects link to deficits with speech processing,

such as speech-in-noise impairments.

For potential medical applications of speech-DPOAEs, it is worth considering the clinical implications of a recent breakthrough in our understanding of neural speech processing, namely the tracking of speech rhythms in the auditory cortex. In this regard, the neural activity in the theta frequency band (4 - 8 Hz) has been shown by my research group and others to reflect largely the processing of lower-level acoustic aspects of speech, while the slower delta frequency band (1 - 4 Hz) contains information on linguistic aspects of speech processing, including on speech comprehension [29, 30, 31, 32]. The neural tracking of speech rhythms is further modulated by selective attention: when listening to one of two competing talkers, for instance, the cortical activity tracks the speech rhythms of the attended speaker more than those of the unattended talker [33, 34, 35].

These novel insights are now being employed towards a better understanding of neurological impairments. Both autism and dyslexia, for instance, have already been linked to impairments with the cortical tracking of speech rhythms [36, 37, 38]. Moreover, disorders of consciousness have been shown to be better assessible when including the neural tracking of speech in the diagnostic toolkit [39]. While such applications are still in their infancy, they demonstrate the potential of using running speech, rather than simpler speech tokens or yet simpler acoustic signals, for better understanding our hearing system as well as for diagnosing neurological impairments. Whether efferent feedback to the cochlea is involved as well, and if so, in which manner, is an exciting research topic that can now be studied.

ACKNOWLEDGMENTS

This research was supported by the Royal British Legion Centre for Blast Injury Studies, from the 'la Caixa' Foundation (LCF/BQ/EU15/10350044) as well as by EPSRC grants EP/M026728/1 and EP/R032602/1.

REFERENCES

1. C. Cherry, "On human communication," (1966).
2. A. C. DAvIs, "The prevalence of hearing impairment and reported hearing disability among adults in great britain," International journal of epidemiology **18**, 911–917 (1989).
3. A. Davis and S. Wood, "The epidemiology of childhood hearing impairment: factors relevant to planning of services," British Journal of Audiology **26**, 77–90 (1992).
4. D. R. Moore, S. Rosen, D.-E. Bamiou, N. G. Campbell, and T. Sirimanna, "Evolving concepts of developmental auditory processing disorder (apd): a british society of audiology apd special interest group 'white paper'," International Journal of Audiology **52**, 3–13 (2013).
5. A. C. Davis and H. J. Hoffman, "Hearing loss: rising prevalence and impact," Bulletin of the World Health Organization **97**, 646 (2019).
6. R. Plomp, "A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired," Journal of Speech, Language, and Hearing Research **29**, 146–154 (1986).
7. D. R. Moore, "Auditory processing disorder (apd): Definition, diagnosis, neural basis, and intervention," Audiological Medicine **4**, 4–11 (2006).
8. E. Skoe and N. Kraus, "Auditory brainstem response to complex sounds: a tutorial," Ear and Hearing **31**, 302 (2010).
9. E. B. Coffey, S. C. Herholz, A. M. Chepesiuk, S. Baillet, and R. J. Zatorre, "Cortical contributions to the auditory frequency-following response revealed by meg," Nature Communications **7**, 1–11 (2016).
10. E. B. Coffey, T. Nicol, T. White-Schwoch, B. Chandrasekaran, J. Krizman, E. Skoe, R. J. Zatorre, and N. Kraus, "Evolving perspectives on the sources of the frequency-following response," Nature Communications **10**, 1–10 (2019).
11. G. M. Bidelman, "Subcortical sources dominate the neuroelectric auditory frequency-following response to speech," Neuroimage **175**, 56–69 (2018).
12. N. Kraus and T. Nicol, "Brainstem encoding of speech and music sounds in humans," Oxford Handbook of the Auditory Brainstem. Oxford, UK: Oxford University Press. Advance online publication (2018).
13. A. E. Forte, O. Etard, and T. Reichenbach, "The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention," elife **6**, e27203 (2017).
14. R. K. Maddox and A. K. Lee, "Auditory brainstem responses to continuous natural speech in human listeners," Eneuro **5** (2018).
15. O. Etard, M. Kegler, C. Braiman, A. E. Forte, and T. Reichenbach, "Decoding of selective attention to continuous speech from the human auditory brainstem response," Neuroimage **200**, 1–11 (2019).
16. M. Kegler, O. Etard, A. Forte, and T. Reichenbach, "Pyhton code for the computation of complex trfs (ctrfs). github," (2018).
17. R. F. Huffman and O. Henson Jr, "The descending auditory pathway and acousticomotor systems: connections with the inferior colliculus," Brain Research Reviews **15**, 295–323 (1990).
18. M. Saiz-Alía and T. Reichenbach, "Computational modeling of the auditory brainstem response to continuous speech," Journal of Neural Engineering **17**, 036035 (2020).
19. T. Reichenbach and A. Hudspeth, "The physics of hearing: fluid mechanics and the active process of the inner ear," Reports on Progress in Physics **77**, 076601 (2014).
20. P. Dallos, "The active cochlea," Journal of Neuroscience **12**, 4575–4585 (1992).

21. R. Probst, B. L. Lonsbury-Martin, and G. K. Martin, "A review of otoacoustic emissions," The Journal of the Acoustical Society of America **89**, 2027–2067 (1991).
22. J. Pickles, *An introduction to the physiology of hearing* (Brill, 1998).
23. J. J. Guinan Jr, "Olivocochlear efferents: Their action, effects, measurement and uses, and the impact of the new conception of cochlear mechanical responses," Hearing Research **362**, 38–47 (2018).
24. J. J. Guinan Jr, "Olivocochlear efferents: anatomy, physiology, function, and the measurement of efferent effects in humans," Ear and Hearing **27**, 589–607 (2006).
25. N. R. Clark, G. J. Brown, T. Jürgens, and R. Meddis, "A frequency-selective feedback model of auditory efferent suppression and its implications for the recognition of speech in noise," The Journal of the Acoustical Society of America **132**, 1535–1541 (2012).
26. O. Etard, M. Kegler, C. Braiman, A. E. Forte, and T. Reichenbach, "Real-time decoding of selective attention from the human auditory brainstem response to continuous speech," bioRxiv , 259853 (2018).
27. M. Saiz-Alía, A. E. Forte, and T. Reichenbach, "Individual differences in the attentional modulation of the human auditory brainstem response to speech inform on speech-in-noise deficits," Scientific Reports **9**, 1–10 (2019).
28. M. Saiz-Alía, P. Miller, and T. Reichenbach, "Otoacoustic emissions evoked by the time-varying harmonic structure of speech," Eneuro **8** (2021).
29. A. Kösem and V. Van Wassenhove, "Distinct contributions of low-and high-frequency neural oscillations to speech comprehension," Language, Cognition and Neuroscience **32**, 536–544 (2017).
30. N. Ding, L. Melloni, H. Zhang, X. Tian, and D. Poeppel, "Cortical tracking of hierarchical linguistic structures in connected speech," Nature Neuroscience **19**, 158–164 (2016).
31. M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse, and E. C. Lalor, "Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech," Current Biology **28**, 803–809 (2018).
32. O. Etard and T. Reichenbach, "Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise," Journal of Neuroscience **39**, 5750–5759 (2019).
33. N. Ding and J. Z. Simon, "Neural coding of continuous speech in auditory cortex during monaural and dichotic listening," Journal of Neurophysiology **107**, 78–89 (2012).
34. S. Srinivasan, A. Keil, K. Stratis, A. F. Osborne, C. Cerwonka, J. Wong, B. L. Rieger, V. Polcz, and D. W. Smith, "Interaural attention modulates outer hair cell function," European Journal of Neuroscience **40**, 3785–3792 (2014).
35. J. A. O'sullivan, A. J. Power, N. Mesgarani, S. Rajaram, J. J. Foxe, B. G. Shinn-Cunningham, M. Slaney, S. A. Shamma, and E. C. Lalor, "Attentional selection in a cocktail party environment can be decoded from single-trial eeg," Cerebral Cortex **25**, 1697–1706 (2015).
36. D. Jochaut, K. Lehongre, A. Saitovitch, A.-D. Devauchelle, I. Olasagasti, N. Chabane, M. Zilbovicius, and A.-L. Giraud, "Atypical coordination of cortical oscillations in response to speech in autism," Frontiers in Human Neuroscience **9**, 171 (2015).
37. M. Lizarazu, M. Lallier, M. Bourguignon, M. Carreiras, and N. Molinaro, "Impaired neural response to speech edges in dyslexia," Cortex **135**, 207–218 (2021).
38. G. M. Di Liberto, V. Peter, M. Kalashnikova, U. Goswami, D. Burnham, and E. C. Lalor, "Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia," Neuroimage **175**, 70–79 (2018).
39. C. Braiman, E. A. Fridman, M. M. Conte, H. U. Voss, C. S. Reichenbach, T. Reichenbach, and N. D. Schiff, "Cortical response to the natural speech envelope correlates with neuroimaging evidence of cognition in severe brain injury," Current Biology **28**, 3833–3839 (2018).
40. A.-L. Giraud and D. Poeppel, "Cortical oscillations and speech processing: emerging computational principles and operations," Nature Neuroscience **15**, 511–517 (2012).
41. N. Ding and J. Z. Simon, "Cortical entrainment to continuous speech: functional roles and interpretations," Frontiers in Human Neuroscience **8**, 311 (2014).
42. J. E. Peelle, J. Gross, and M. H. Davis, "Phase-locked responses to speech in human auditory cortex are enhanced during comprehension," Cerebral Cortex **23**, 1378–1387 (2013).
43. G. M. Di Liberto, J. A. O'Sullivan, and E. C. Lalor, "Low-frequency cortical entrainment to speech reflects phoneme-level processing," Current Biology **25**, 2457–2465 (2015).
44. C. Micheyl and A. J. Oxenham, "Across-frequency pitch discrimination interference between complex tones containing resolved harmonics," The Journal of the Acoustical Society of America **121**, 1621–1631 (2007).
45. M. J. Polonenko and R. K. Maddox, "Exposing distinct subcortical components of the auditory brainstem response evoked by continuous naturalistic speech," Elife **10**, e62329 (2021).
46. T. Duke and F. Jülicher, "Active traveling wave in the cochlea," Physical Review Letters **90**, 158101 (2003).
47. F. Jülicher, D. Andor, and T. Duke, "Physical basis of two-tone interference in hearing," Proceedings of the National Academy of Sciences U.S.A. **98**, 9080–9085 (2001).
48. N. Ciganović, A. Wolde-Kidan, and T. Reichenbach, "Hair bundles of cochlear outer hair cells are shaped to minimize their fluid-dynamic resistance," Scientific Reports **7**, 1–9 (2017).
49. L. Sumner, J. Mestel, and T. Reichenbach, "Steady streaming as a method for drug delivery to the inner ear," Scientific Reports **11**, 1–12 (2021).
50. D. McFadden, "Sex differences in the auditory system," Developmental Neuropsychology **14**, 261–298 (1998).
51. D. McFadden, G. K. Martin, B. B. Stagner, and M. M. Maloney, "Sex differences in distortion-product and transient-evoked otoacoustic emissions compared," The Journal of the Acoustical Society of America **125**, 239–246 (2009).