



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

COPYRIGHT
© 2023 Oxenham. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Questions and controversies surrounding the perception and neural coding of pitch

Andrew J. Oxenham^{1,2*}

¹Center for Applied and Translational Sensory Science, University of Minnesota Twin Cities, Minneapolis, MN, United States, ²Department 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 selectively 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 (Kohlerusch 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

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

References

- Allen, 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.000000000000152
- 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
- Lieberman, 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