ANNUAL REVIEWS

Annual Review of Psychology

How We Hear: The Perception and Neural Coding of Sound

Andrew J. Oxenham

Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455;
email: oxenham@umn.edu

ANNUAL REVIEWS **Further**

Click here to view this article's online features:
• Download figures as PPT slides
• Navigate linked references
• Download citations
• Explore related articles
• Search keywords

Keywords

auditory perception, frequency selectivity, pitch, auditory scene analysis, hearing loss

Abstract

Auditory perception is our main gateway to communication with others via speech and music, and it also plays an important role in alerting and orienting us to new events. This review provides an overview of selected topics pertaining to the perception and neural coding of sound, starting with the first stage of filtering in the cochlea and its profound impact on perception. The next topic, pitch, has been debated for millennia, but recent technical and theoretical developments continue to provide us with new insights. Cochlear filtering and pitch both play key roles in our ability to parse the auditory scene, enabling us to attend to one auditory object or stream while ignoring others. An improved understanding of the basic mechanisms of auditory perception will aid us in the quest to tackle the increasingly important problem of hearing loss in our aging population.

Contents

INTRODUCTION

Hearing provides us with access to the acoustic world, including the fall of raindrops on the roof, the chirping of crickets on a summer evening, and the cry of a newborn baby. It is the primary mode of human connection and communication via speech and music. Our ability to detect, localize, and identify sounds is astounding given the seemingly limited sensory input: Our eardrums move to and fro with tiny and rapid changes in air pressure, providing us only with a continuous measure of change in sound pressure at two locations in space, about 20 cm apart, on either side of the head. From this simple motion arises our rich perception of the acoustic environment around us. The feat is even more impressive when one considers that sounds are rarely presented in isolation: The sound wave that reaches each ear is often a complex mixture of many sound sources, such as the conversations at surrounding tables of a restaurant, mixed with background music and the clatter of plates. All that reaches each eardrum is a single sound wave, and yet, in most cases, we are able to extract from that single waveform sufficient information to identify the different sound sources and direct our attention to the ones that currently interest us.

Deconstructing a waveform into its original sources is no simple matter; in fact, the problem is mathematically ill posed, meaning that there is no unique solution. Similar to solutions in the visual domain (e.g., Kersten et al. 2004), our auditory system is thought to use a combination of information learned during development and more hardwired solutions developed over evolutionary time to solve this problem. Decades of psychological, physiological, and computational research have gone into unraveling the processes underlying auditory perception. Understanding basic auditory processing, auditory scene analysis (Bregman 1990), and the ways in which humans solve the "cocktail party problem" (Cherry 1953) has implications not only for furthering

fundamental scientific progress but also for audio technology applications. Such applications include low-bit-rate audio coding (e.g., MP3) for music storage, broadcast and cell phone technology, automatic speech recognition, and the mitigation of the effects of hearing loss through hearing aids and cochlear implants.

This review focuses on recent trends and developments in the area of auditory perception, as well as on relevant computational and neuroscientific studies that shed light on the processes involved. The areas of focus include the peripheral mechanisms that enable the rich analysis of the auditory scene, the perception and coding of pitch, and the interactions between attention and auditory scene analysis. The review concludes with a discussion of hearing loss and the efforts underway to understand and alleviate its potentially devastating effects.

EARLY STAGES: PERCEPTION AND THE COCHLEA

Frequency Range of Hearing

Just as the visual system is sensitive to oscillations in the electromagnetic spectrum, the auditory system is sensitive to oscillations in the acoustic spectrum. There are, however, interesting quantitative differences in the ranges of sensitivity. For instance, the human visual system is sensitive to light wavelengths between approximately 380 and 750 nm (or frequencies between 400 and 790 THz), spanning just under 1 octave (a doubling in frequency), which we perceive as the spectrum of colors. In contrast, the human auditory system is sensitive to sound frequencies between 20 Hz and 20,000 Hz, or approximately 10 octaves, which we perceive along the dimension of pitch. Just as important as the ability to hear a wide range of frequencies is the ability to analyze the frequency content of sounds. Both our sensitivity and our selectivity with respect to frequency originate in the cochlea of the inner ear.

Cochlear Tuning and Frequency Selectivity

The basilar membrane runs along the length of the cochlea and vibrates in response to the sounds that enter the cochlea via the vibrations of the eardrum and the middle ear bones. The action of the basilar membrane can be compared to that of a prism—the wide range of frequencies within a typical sound are dispersed to different locations along the basilar membrane within the cochlea. Every point along the basilar membrane responds best to a certain frequency, known as the best frequency or characteristic frequency (CF). In this way, the frequency content of a sound is represented along the length of the basilar membrane in a frequency-to-place map, providing tonotopic organization with a gradient from low to high frequencies from the apex to the base of the cochlea. This organization is maintained from the cochlea via the inner hair cells and the auditory nerve, through the brainstem and midbrain, to the primary auditory cortex. Place coding thus represents a primary organizational principle for both neural coding and perception.

Although the passive properties of the basilar membrane (e.g., its mass and stiffness gradients) provide the foundations for the tonotopic organization (von Békésy 1960), the separation of frequencies along the basilar membrane is enhanced by sharp tuning that is mediated by the action of the outer hair cells within the cochlea (Dallos et al. 2006). This sharp tuning has a profound impact on our perception of sound. There are many ways to measure our perceptual ability to separate sounds of different frequencies, or our frequency selectivity. One of the most common perceptual measures involves the masking of one sound by another. By parametrically varying the frequency relationship between the masking sound and a target and measuring the level of the masker and target sounds at the detection threshold, it is possible to determine the sharpness of

tuning (Patterson 1976). It has often been assumed that cochlear tuning determines the frequency selectivity measured behaviorally, so that the first stages of auditory processing limit the degree to which we are able to hear out different frequencies within a mixture. However, because of the inability to make direct measurements of the cochlea in humans, and because of the difficulty of deriving behavioral measures in animals, this assumption has rarely been tested. Another difficulty is posed by the highly nonlinear nature of cochlear processing, as shown directly through physiological measures in animals (Ruggero 1992) and indirectly through behavioral measures in humans (Oxenham & Plack 1997). The nonlinearity means that estimates of frequency selectivity will differ depending on the precise measurement technique used and the stimulus level at which the measurements are made.

One study measured both cochlear tuning and behavioral frequency selectivity in guinea pigs and found reasonably good correspondence between the two (Evans 2001). Because it has generally been assumed that the cochleae of humans and of mammals commonly used in laboratory experiments (such as guinea pigs, cats, and chinchillas) are similar, it has also been assumed that human cochlear tuning is similar to that of other mammals and that, therefore, human perceptual frequency selectivity is also limited by cochlear tuning. Indeed, a number of physiological studies have examined the representation of speech sounds in the auditory nerves of other species, making the explicit assumption that cochlear tuning is similar across species (Delgutte 1984, Young & Sachs 1979). As our understanding of otoacoustic emissions (OAEs)—sounds generated by the ear—has improved, it has been possible to probe the tuning properties of the human cochlea in a noninvasive manner (e.g., Bentsen et al. 2011). The combination of OAE measurements and behavioral masking studies in humans has led to confirmation of the idea that behavioral frequency selectivity reflects cochlear tuning, but also (and more surprisingly), that human tuning may be considerably sharper than that found in common laboratory animals, such as cats and guinea pigs (Shera et al. 2002). This conclusion was based on the fact that the latencies (or delays) of stimulus-frequency OAEs (SFOAEs) in humans are longer than those measured in other mammals and that latency is related to the sharpness of cochlear tuning (Shera et al. 2010).

The claim that human cochlear tuning is sharper than that in many other species has generated some controversy (Lopez-Poveda & Eustaquio-Martin 2013, Ruggero & Temchin 2005). Nevertheless, the initial claims have been supported by further studies in different rodent species (Shera et al. 2010), as well as in a species of old-world monkey, where cochlear tuning appears to be intermediate between that of rodent and human, suggesting a progression from small nonprimate mammals to small primates to humans (Joris et al. 2011). All of these studies have used a combination of (a) OAE measurements, (b) direct measurements of tuning in the auditory nerve, and (c) behavioral measurements of frequency selectivity. However, none of the earlier studies used all three methods in the same species; the auditory nerve measurements are too invasive to be carried out in humans, and the behavioral measurements have posed challenges in terms of animal training. More recently, a study was carried out in ferrets that included all three measurements. The results from this study reveal a good correspondence between all three types of measurement and confirm that tuning is, indeed, broader in ferrets than in humans (Sumner et al. 2014).

In summary, our current thinking is that the frequency tuning established in the cochlea determines our perceptual ability to separate sounds of different frequencies. In some ways, this is a remarkable finding, given the extensive and complex processing of neural signals between the cochlea and the auditory cortex: These multiple stages of neural processing neither enhance nor degrade the basic tuning patterns that are established in the cochlea. Another main conclusion is that human frequency tuning is sharper than that of many other mammals. This finding has important, and not yet fully explored, implications for understanding human hearing and acoustic communication in general. It may be that sharp tuning is a prerequisite for developing the fine

acoustic communication skills necessary for speech. However, this speculation is rendered less likely by the fact that speech is highly robust to spectral degradation and remains intelligible even under conditions of very poor spectral resolution (Shannon et al. 1995). It currently appears more likely that our sharp cochlear tuning underlies our fine pitch perception and discrimination abilities. As discussed in the next section, there appear to be some fundamental and qualitative differences in the way pitch is perceived by humans and by other species, which, in turn, may be related to the differences in frequency tuning found in the very first stages of auditory processing.

PITCH PERCEPTION AND NEURAL CODING

Pitch is a perceptual quality that relates most closely to the physical variable of frequency or repetition rate of a sound. Its technical definition, provided by the American National Standards Institute, is "that attribute of auditory sensation by which sounds are ordered on the scale used for melody in music" (ANSI 2013, p. 58). Pitch plays a crucial role in auditory perception. In music, sequences of pitch define melody, and simultaneous combinations of pitch define harmony and tonality. In speech, pitch contours provide information about prosody and speaker identity; in tone languages, such as Mandarin or Cantonese, pitch contours also provide lexical information. In addition, differences in pitch between sounds enable us to segregate competing sources, thereby helping solve the cocktail party problem (Darwin 2005).

Place and Time Theories

The questions of how pitch is extracted from acoustic waveforms and how it is represented in the auditory system have been debated for well over a century but remain topics of current investigation and some controversy. Two broad categories of theories addressing these questions can be identified, both of which have long histories: place theories and timing theories. As outlined in the previous section, the cochlea establishes a tonotopic representation that is maintained throughout the early auditory pathways. Broadly speaking, the premise of place theories is that the brain is able to extract the frequency content of sounds from this tonotopic representation to derive the percept of pitch. Timing theories, on the other hand, are based on the observation that action potentials, or spikes, in the auditory nerve tend to occur at a given phase in the cycle of a stimulating waveform, producing a precise relationship between the waveform and the timing of the spikes, known as phase locking (Rose et al. 1967). Auditory nerve phase locking enables the auditory system to extract timing differences between a sound arriving at each of the two ears, enabling us to localize sounds in space (Blauert 1997). The fact that humans can discriminate interaural time differences as small as 20 µs attests to the exquisite sensitivity of the auditory system to timing information. Timing theories of pitch postulate that the same exquisite sensitivity to timing can be harnessed by the auditory system to measure the time intervals between spikes, which are related to the period (i.e., the duration of one repetition) of the waveform.

A third category of theories could be termed place-time theories. According to these approaches, spike timing information is used by the auditory system not by comparing time intervals between successive spikes but rather by using the phase dispersion along the basilar membrane and utilizing coincident spikes from different cochlear locations to extract information about the frequency of a tone (Loeb et al. 1983, Shamma 1985). Various place, timing, and place-time theories have been postulated over the decades to account for pitch of both pure and complex tones. Some background and recent findings are reviewed in the following sections for both classes of stimuli.

Pitch of Pure Tones

Pure tones—sinusoidal variations in air pressure—produce a salient pitch sensation over a wide range of frequencies. In the range of greatest sensitivity to frequency changes (between approximately 1 and 2 kHz), humans can discriminate between two frequencies that differ by as little as 0.2% (Micheyl et al. 2012). At the low and high ends of the frequency spectrum (below approximately 500 Hz and above approximately 4,000 Hz), sensitivity deteriorates. At high frequencies, the deterioration is particularly dramatic, with increases in frequency discrimination thresholds by an order of magnitude between 2 and 8 kHz (Moore & Ernst 2012). Indeed, our ability to recognize musical intervals (such as an octave or a fifth), or even familiar melodies, essentially disappears at frequencies above 4–5 kHz (Attneave & Olson 1971).

There is reasonably good correspondence between the deterioration in our ability to discriminate between frequencies and the deterioration in the accuracy of auditory nerve phase locking in small mammals that occurs as frequency increases: The synchronization index for phase locking degrades to about half its maximum value by approximately 2–3 kHz, and significant phase locking is no longer observed above approximately 4–5 kHz, depending somewhat on the species (Heil & Peterson 2015). Phase locking has not been measured directly in the human auditory nerve due to the invasive nature of the measurements. On the one hand, usable phase locking may only extend up to approximately 1.5 kHz, as indicated by the fact that we cease to be able to detect timing differences between the two ears for pure tones above 1.5 kHz (Blauert 1997). On the other hand, the fact that frequency discrimination thresholds (as a proportion of the center frequency) continue to increase up to approximately 8 kHz and then remain roughly constant at even higher frequencies has been proposed as evidence for some residual phase locking up to 8 kHz (Moore & Ernst 2012). Thus, if it is agreed that human phase locking is at least qualitatively similar to that observed in other mammals, it seems reasonable to assume that its effects begin to degrade above 1 kHz and are no longer perceptually relevant above 8 kHz, placing the highest frequency at which phase locking is used at least within the range of the 4–5 kHz limit for musical pitch (Attneave & Olson 1971).

The fact that the breakdown in phase locking seems to occur at around the same frequency as the breakdown in musical pitch perception has led to the proposal that timing information from the auditory nerve is necessary for musical pitch perception; indeed, it is tempting to speculate that the highest note on current musical instruments (e.g., C_8 on the grand piano, with a frequency of 4,186 Hz) is determined by the coding limitations imposed at the earliest stages of the auditory system (Oxenham et al. 2011). In contrast, place theory provides no explanation for the fact that frequency discrimination and pitch perception both degrade at high frequencies—if anything, cochlear filters become sharper at high frequencies, suggesting more accurate place coding (Shera et al. 2010).

Another argument in favor of a timing theory of pitch for pure tones is the fact that our ability to discriminate between two frequencies is much finer than would be predicted by basic place theories of pitch. According to place theories, an increase in frequency is detected by a shift in the peak of response from a more apical to a more basal cochlear location, which, in turn, produces a decrease in response from cochlear locations apical to the peak and an increase in response from locations basal to the peak (**Figure 1a**). Place theories have contended that the frequency increase becomes detectable when the change in response at any given cochlear location exceeds some threshold. In this way, changes in frequency can be coded as changes in amplitude (Zwicker 1970; **Figure 1a**). The challenge for the place theory is that a just-detectable change in frequency produces a predicted change in cochlear response that is much smaller than that needed to detect a change in the amplitude of a tone (Heinz et al. 2001).

Although the current weight of evidence seems to favor timing theories of pure-tone pitch, some recent studies have led to a reconsideration of these earlier ideas (Micheyl et al. 2013b, Whiteford
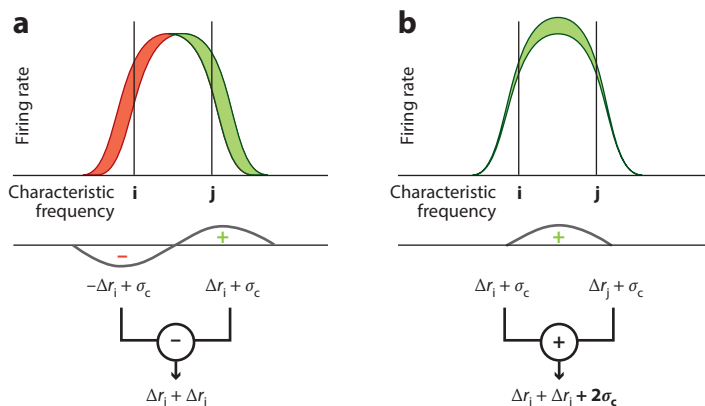
Figure 1

(*a*) Schematic of response (excitation pattern) for two tones close together in frequency. An increase in frequency leads to a shift in the peak of the response to the right, which, in turn, leads to a response decrease below the peak (r_i) and a response increase above the peak (r_j). If some of the neuronal noise is correlated, then the correlated portion of the noise (σ_c) will be canceled out when the two responses are compared by subtraction, leading to improved discrimination. (*b*) For intensity discrimination, correlated neuronal noise between i and j has a different effect because the increment in intensity is detected by adding (not subtracting) the neuronal responses, leading to an increased effect of the correlated noise and, thus, poorer discrimination. Overall, smaller differences in response (or excitation) patterns are required for the detection of a change in frequency than a change in intensity, in line with human perceptual data (Micheyl et al. 2013b).

& Oxenham 2015). One question is whether frequency discrimination, and pitch perception more generally, is really limited by peripheral constraints. In contrast to frequency selectivity (tuning), discussed in the section titled Cochlear Tuning and Frequency Selectivity, frequency discrimination is highly susceptible to training, with dramatic improvements often observed over fairly short periods of time. For instance, professional musicians have been found to have lower (better) frequency discrimination thresholds than nonmusicians by a factor of approximately 6, but nonmusicians can reach levels of performance similar to those of the professional musicians after only 4–8 hours of training (Micheyl et al. 2006). One interpretation of this extended perceptual learning is that discrimination is limited not by peripheral coding constraints (e.g., auditory nerve phase locking), but rather by more central, possibly cortical, coding constraints that are more likely to demonstrate rapid plasticity (e.g., Yin et al. 2014). If so, we may not expect perceptual performance to mirror peripheral limitations, such as auditory nerve phase locking; instead, performance may reflect higher-level constraints, perhaps shaped by passive exposure, with high-frequency tones being sparsely represented and poorly perceived due to the lack of exposure to them in everyday listening conditions. Indeed, the fact that some people have been reported to perceive musical intervals for pure tones of approximately 10 kHz (Burns & Feth 1983) suggests that the more usual limit of 4–5 kHz is not imposed by immutable peripheral coding constraints.

Another line of evidence suggesting that frequency discrimination is not limited by peripheral constraints comes from studies of frequency modulation (FM) and amplitude modulation (AM). According to timing-based theories, the detection of FM at slow modulation rates is mediated by phase locking to the temporal fine structure of the pure tone, whereas the detection of FM at fast modulation rates is mediated via the transformations of the FM to AM via cochlear filtering (Moore & Sek 1996). A recent study of individual differences in 100 young normal-hearing listeners found that slow-rate FM thresholds were significantly correlated with slow-rate changes in interaural

time differences, which are known to be mediated by phase locking. However, slow-rate FM detection thresholds were just as strongly correlated with fast-rate FM and AM detection thresholds, suggesting that the individual differences were not mediated by the peripheral coding constraints of phase locking, but rather by more central constraints (Whiteford & Oxenham 2015).

One remaining problem for the place theories of pure-tone pitch is the apparently large difference in sensitivity between frequency discrimination and intensity discrimination. A computational modeling study of cortical neural coding has provided one solution to this problem. Using simple assumptions about the properties of cortical neurons with tuning similar to that observed in the auditory nerve and in the cortex of primate species, Micheyl et al. (2013b) were able to resolve the apparent discrepancy between frequency and intensity discrimination abilities within a unified place-based code. They assumed some underlying correlation between the firing rates of neurons with similar CFs that is independent of the stimulus. The effect of this noise correlation (e.g., Cohen & Kohn 2011) is to limit the usefulness of integrating information across multiple neurons in the case of intensity discrimination, where the correlation decreases the independence of the information in each neuron. However, in the case of frequency discrimination, the effect of the noise correlation is less detrimental because it can be reduced by subtracting the responses of neurons with CFs above the stimulus frequency from the responses of neurons with CFs below the stimulus frequency, thereby enhancing the effects of a shift in frequency. In this way, the same model, with the same sensitivity, can account for observed human performance in both frequency and intensity discrimination tasks (Micheyl et al. 2013b; **Figure 1**).

Regardless of how pitch is extracted from information in the auditory nerve, these representations clearly involve some transformations between the cochlea and the cortex. Timing information becomes increasingly coarse at higher stages of the auditory pathways. In the cochlear nucleus (the first stage of processing beyond the cochlea), phase-locked information is maintained via primary-like neurons that seem to maintain the temporal properties of auditory nerve fibers (Rhode et al. 1983). However, already in the inferior colliculus of the midbrain, phase-locked responses are not normally observed above 1,000 Hz (Liu et al. 2006), and in the auditory cortex, phase locking is generally not observed above 100 Hz (e.g., Lu & Wang 2000). Therefore, any timing-based code in the auditory periphery must be transformed into a population rate or place code at higher stages of processing. In contrast, the place-based, or tonotopic, representation in the auditory periphery is maintained at least up to the primary auditory cortex (e.g., Moerel et al. 2014).

In summary, some aspects of pure-tone pitch perception and frequency discrimination are well accounted for by a timing theory. However, in most cases, a place-based or tonotopic theory can also be used to account for the available perceptual data. Questions surrounding the coding of pure tones in the auditory periphery are not only of basic scientific interest; they also have important implications for attempts to restore hearing via auditory prostheses, such as cochlear implants. This topic is addressed below (see the section titled Perceptual Consequences of Hearing Loss and Cochlear Implants). In any case, pure tones are a special case and are not a particularly ecologically relevant class of stimuli. For a more general case, we turn to harmonic complex tones, such as those we encounter in speech and music.

Pitch of Complex Tones

A complex tone is defined as any sound composed of more than one sinusoid or pure tone. Harmonic complex tones consist of a fundamental frequency (F0) and harmonics (also known as upper partials or overtones) with frequencies at integer multiples of the F0. For instance, a violin playing a note with a pitch corresponding to an orchestral A (440 Hz) produces a waveform that repeats 440 times per second (**Figure 2a**) but has energy not only at 440 Hz but also at
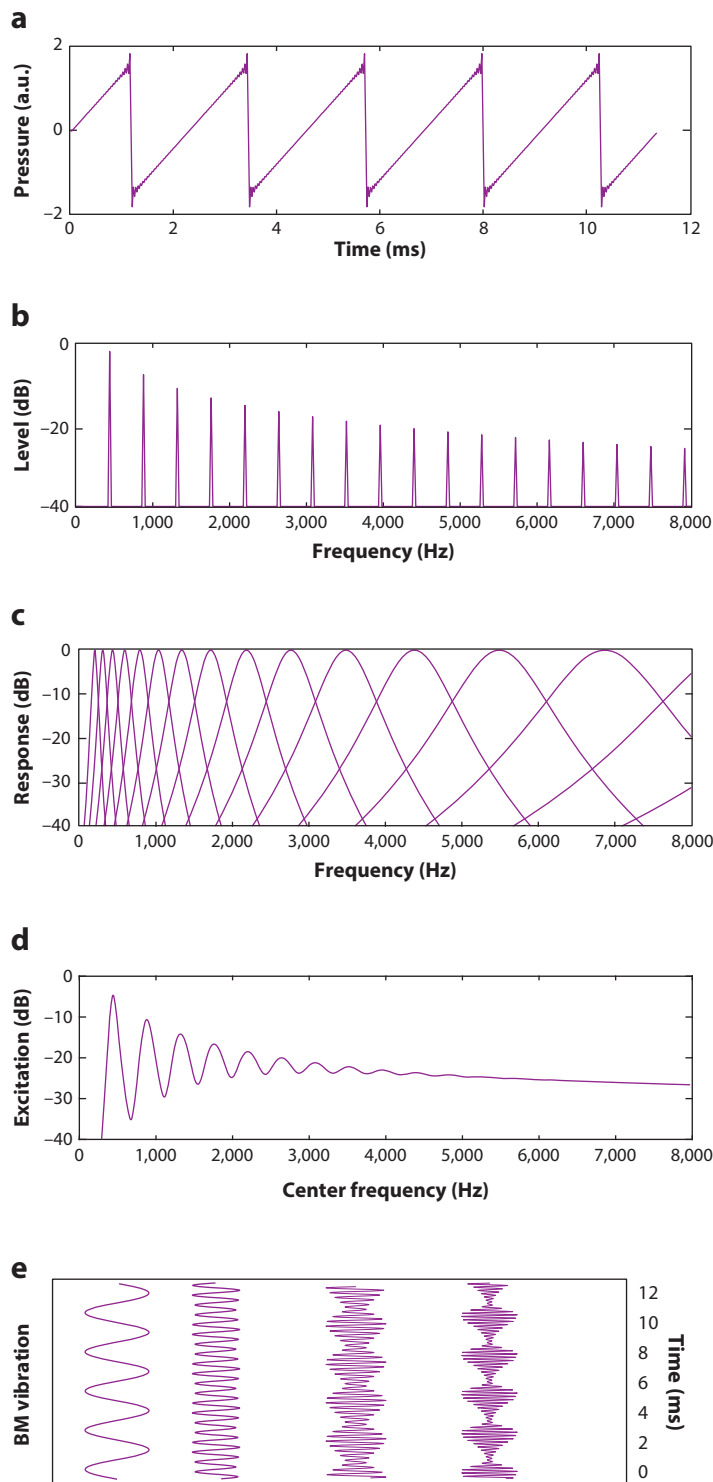
880 Hz, 1,320 Hz, 1,760 Hz, etc. (**Figure 2b**). Interestingly, we tend to hear a single sound with a single pitch, corresponding to the F0, despite the presence of many other frequencies. Indeed, the pitch continues to be heard at the F0 even if the energy at the F0 is removed or masked. This phenomenon is known as residue pitch, periodicity pitch, or the pitch of the missing fundamental. The constancy of the pitch in the presence of masking makes sense from an ecological perspective: We would expect the primary perceptual properties of a sound to remain invariant in the presence of other competing sounds in the environment, just as we expect perceptual constancy of visual objects under different lighting conditions, perspectives, and occlusions. But if it is not derived from the component at the F0 itself, how is pitch extracted from a complex waveform?

To better understand how pitch is extracted from a complex tone, it is useful to consider first how the tone is represented in the auditory periphery. **Figure 2c** illustrates the filtering process of the cochlea, represented as a bank of bandpass filters. Although the filters tend to sharpen somewhat with increasing CF in terms of their bandwidth relative to the CF (known as quality factor, or Q, in filter theory; Shera et al. 2010), their absolute bandwidths in Hz increase with increasing CF, as shown in **Figure 2c**. This means that the filters are narrow, relative to the spacing of the harmonics, for the low-numbered harmonics but become broader with increasing harmonic number. The implications of the relationship between filter bandwidth and harmonic spacing are illustrated in **Figure 2d**, which shows the excitation pattern produced when the harmonic complex is passed through the filter bank illustrated in **Figure 2c**: Low-numbered harmonics each produce distinct peaks in the excitation pattern and are, therefore, spectrally resolved, whereas multiple higher harmonics fall within the bandwidth of a single filter, meaning that they are no longer resolved. The putative time waveforms produced by the complex at different locations along the basilar membrane are shown in **Figure 2e**. For the resolved harmonics, the output resembles a pure tone, whereas for the higher, unresolved harmonics, the output of each filter is itself a complex waveform that repeats at a rate corresponding to the F0.

Numerous studies have shown that the overall pitch of a complex tone is dominated by the lower, resolved harmonics (Plomp 1967). These harmonics could be represented by their place (**Figure 2d**) or their time (**Figure 2e**) representations. In contrast, the unresolved harmonics do not produce clear place cues. Indeed, the fact that any pitch information can be transmitted via unresolved harmonics provides strong evidence that the auditory system is able to use timing information to extract pitch. However, the pitch strength of unresolved harmonics is much weaker than that produced by resolved harmonics, and F0 discrimination thresholds are generally much poorer (by up to an order of magnitude) than thresholds for complexes with resolved harmonics (Bernstein & Oxenham 2003, Houtsma & Smurzynski 1990, Shackleton & Carlyon 1994). The reliance on low-numbered harmonics for pitch may be due to the greater robustness of these harmonics to interference. For instance, the lower-numbered harmonics tend to be more intense and, therefore, less likely to be masked. Also, room acoustics and reverberation can scramble the phase relationships between harmonics. This has no effect on resolved harmonics, but it can severely degrade the temporal envelope information carried by the unresolved harmonics (Qin & Oxenham 2005, Sayles & Winter 2008). The most important question for natural pitch perception, therefore, is how pitch is extracted from the low-numbered resolved harmonics.

As with pure tones, the pitch of harmonic complex tones has been explained in terms of place, timing, and place-time information (Cedolin & Delgutte 2010, Shamma & Klein 2000). Most recent perceptual work has been based on the premise that timing information is extracted, and there has been a plethora of studies concentrating on the perceptual effects of temporal fine structure (TFS), a term that usually refers to the timing information extracted from resolved harmonics or similarly narrowband sounds (Lorenzi et al. 2006, Smith et al. 2002). However, for the same reasons that it is difficult to distinguish place and time codes for pure tones, it is difficult

to determine whether the TFS of resolved harmonics is being coded via an auditory nerve timing code or a via a place-based mechanism. Two studies have suggested that timing information from individual harmonics presented to the wrong locations in the cochlea cannot be used to extract pitch information corresponding to the missing F0 (Deeks et al. 2013, Oxenham et al. 2004), suggesting that timing information is not sufficient for the extraction of pitch information. In addition, one study has demonstrated that the pitch of the missing F0 can be extracted from resolved harmonics even when all the harmonics are above approximately 7.5 kHz (Oxenham et al. 2011). If one accepts that phase locking is unlikely to be effective at frequencies of 8 kHz and above, this result suggests that timing information is also not necessary for the perception of complex pitch. Finally, a recent study has found that the F0 discrimination found for these very high-frequency complexes is better than predicted based on optimal integration of the information from each individual harmonic, suggesting that performance is not limited by peripheral coding constraints, such as limited phase locking, and is instead limited at a more central processing stage, where the information from the individual harmonics has already been combined (Lau et al. 2017).

Some studies have attempted to limit place information by presenting stimuli at high levels, where frequency selectivity is poorer, and have found results that do not seem consistent with a purely place-based code (e.g., Marmel et al. 2015). In particular, the discrimination of harmonic from inharmonic complex tones was possible in situations where the changes in the frequencies of the tones produced no measurable change in the place of stimulation based on masking patterns. However, as discussed above (see **Figure 1**), relatively small changes in excitation may be sufficient to code changes in pitch even if they are too small to measure in a masking paradigm (Micheyl et al. 2013b).

Studies of pitch perception in other species have generally concluded that animals can perceive a pitch corresponding to the missing F0. However, some important differences between humans and other species have been identified. First, absolute pitch, or simply spectral similarity, seems to be particularly salient for other species, including other mammals (Yin et al. 2010) and songbirds (Bregman et al. 2016), whereas humans tend to focus on relative pitch relations. Second, the few studies that have attempted to determine the mechanisms of pitch perception in other species have found that judgments seem to be based on temporal envelope cues from unresolved harmonics rather than resolved harmonics, perhaps because the poorer frequency selectivity of other species means that fewer harmonics are resolved than in humans (Shofner & Chaney 2013).

As with pure tones, no matter how complex tones are represented in the auditory nerve, it seems likely that the code is transformed into some form of rate or rate-place population-based code at higher levels of the auditory system. Some behavioral studies have demonstrated that perceptual grouping effects (which are thought to be relatively high-level phenomena) can affect the perception of pitch; conversely, pitch and harmonicity can strongly affect perceptual grouping, suggesting that pitch itself is a relatively high-level, possibly cortical, phenomenon (Darwin 2005). Studies in nonhuman primates (marmosets) have identified small regions of the auditory cortex that seem to respond selectively to harmonic stimuli in ways that are either independent of or dependent on the harmonic numbers presented. Neurons in the former category have been termed pitch neurons (Bendor & Wang 2005), whereas the neurons in the latter category have been termed harmonic template neurons (Feng & Wang 2017). Such fine-grained analysis has not been possible in human neuroimaging studies, including positron emission tomography, functional magnetic resonance imaging, and, more recently, electrocorticography (ECoG); however, there exist a number of reports of anterolateral regions of the human auditory cortex, potentially homologous to the regions identified in marmoset monkeys, that seem to respond selectively to harmonic stimuli in ways that suggest that they are responsive to perceived pitch strength, rather than just stimulus regularity (e.g., Norman-Haignere et al. 2013, Penagos et al. 2004).

Despite these encouraging findings, it remains unclear to what extent such neurons extract pitch without regard to other aspects of the stimulus. For instance, a study in ferrets used stimuli that varied along three dimensions, F0 (corresponding to the perception of pitch), location, and spectral centroid (corresponding to the timbral dimension of brightness), and failed to find evidence for neurons that were sensitive to changes in one dimension but not the others (Walker et al. 2011). In particular, neurons that were modulated by changes in F0 were generally also sensitive to changes in spectral centroid. Interestingly, a human neuroimaging study that also covaried F0 and spectral centroid came to a similar conclusion (Allen et al. 2017). In fact, the failure to find a clear neural separation between dimensions relating to pitch and timbre is consistent with results from perceptual experiments that have demonstrated strong interactions and interference between the two dimensions (e.g., Allen & Oxenham 2014).

Combinations of Pitches: Consonance and Dissonance

Some combinations of pitches sound good or pleasing together (consonant), whereas other do not (dissonant). In this section, we consider only the very simple case of tones presented simultaneously in isolation from any surrounding musical context. The question of which combinations are consonant and why has intrigued scientists, musicians, and music theorists for over two millennia. Pythagoreans attributed the pleasing nature of some consonant musical intervals, such as the octave (2:1 frequency ratio) or the fifth (3:2 frequency ratio), to the inherent mathematical beauty of low-numbered ratios. Indeed, some combinations, such as the octave and the fifth, do seem to occur across multiple cultures and time periods, suggesting explanations that are more universal than simple acculturation (McDermott & Oxenham 2008). More recently, consonance has been attributed to an absence of acoustic beats—the amplitude fluctuations that occur when two tones are close but not identical in frequency (e.g., Fishman et al. 2001, Plomp & Levelt 1965). Another alternative is that a combination of harmonic tones is judged as being most consonant when the combined harmonics most closely resemble a single harmonic series (e.g., Tramo et al. 2001). It has been difficult to distinguish between theories based on acoustic beats and those based on harmonicity because the two properties generally covary: The less a combination of tones resembles a single harmonic series, the more likely it is to contain beating pairs of harmonics.

The question has recently been addressed by exploiting individual differences in preferences. McDermott et al. (2010) used artificial diagnostic stimuli to independently test preference ratings for stimuli in which acoustic beats were either present or absent and for stimuli that were either harmonic or inharmonic. They then correlated individual preference ratings for the diagnostic stimuli with preferences for consonant and dissonant musical intervals and chords using real musical sounds. The outcome was surprisingly clear cut: Preferences for harmonicity correlated strongly with preferences for consonant versus dissonant musical intervals, whereas preferences for (or antipathy to) acoustic beats did not. In addition, the number of years of musical training was found to correlate with harmonicity and musical consonance preferences but not with acoustic beat preferences. From this study, it seems that harmonicity, rather than acoustic beats, determines preferences for consonance, and that these preferences may be learned to some extent.

The suggestion that preferences for consonance may be learned was somewhat surprising at the time given the fact that some earlier studies in infants had suggested that a preference for consonance may be innate (Trainor & Heinmiller 1998, Zentner & Kagan 1996). However, a more recent study in infants questioned the findings from these earlier studies and found no preference for consonant over dissonant intervals (Plantinga & Trehub 2014); instead, this study found only a preference for music to which the infants had previously been exposed. The lack of any innate aspect of consonance judgments was supported by a recent cross-cultural study that compared

the judgments of members of a native Amazonian society with little or no exposure to Western culture or music to those of urban residents in Bolivia and the United States (McDermott et al. 2016). That study found that the members of the Amazonian society exhibited no clear distinctions in preference between musical intervals that are deemed consonant and those that are deemed dissonant in Western music.

In summary, studies in adults and infants, as well as studies across cultures, seem to be converging on the conclusion that Western judgments of consonance and dissonance for isolated simultaneous combinations of tones are driven by the harmonicity of the combined tones, rather than the presence of acoustic beats, and that these preferences are primarily learned through active or passive exposure to Western music.

SURVEYING THE AUDITORY SCENE

Acoustic Cues to Solve the Cocktail Party Problem

The auditory system makes use of regularities in the acoustic structure of sounds from individual sources to assist in parsing the auditory scene (Bregman 1990). The first stage of parsing occurs in the cochlea, where sounds are mapped along the basilar membrane according to their frequency content or spectrum. Thus, two sounds with very different spectra will activate different portions of the basilar membrane and will, therefore, stimulate different populations of auditory nerve fibers. In such cases, the perceptual segregation of sounds has a clear basis in the cochlea itself. This phenomenon forms the core of the peripheral channeling theory of stream segregation— sequences of sounds can be perceptually segregated only if they stimulate different populations of peripheral neurons (Hartmann & Johnson 1991). Although peripheral channeling remains the most robust form of perceptual segregation of competing sources, there have since been several instances reported in which streaming occurs even in the absence of peripheral channeling. For instance, by using harmonic complex tones containing only unresolved harmonics, Vliegen & Oxenham (1999) showed that differences in F0 or pitch could lead to perceptual segregation even when the complexes occupied exactly the same spectral region. Similar results have been reported using differences in wave shape, even when the same harmonic spectrum was used (Roberts et al. 2002). Indeed, it has been proposed that perceptual segregation can occur with differences along any perceptual dimension that can be discriminated (Moore & Gockel 2002).

One aspect of ongoing sound sequences (such as speech or music) that is important in binding together elements and features of sound is temporal coherence, i.e., a repeated synchronous relationship between elements. In addition to differences in features such as spectral content or F0, which are necessary to induce stream segregation of two sound sequences, another necessary component is some form of temporal incoherence between the two sequences. If the sequences are presented coherently and synchronously, they will tend to form a single stream, even if they differ along other dimensions (Micheyl et al. 2013a). Note that temporal coherence goes beyond simple synchrony: Although sound elements are generally perceived as belonging to a single source if they are gated synchronously (e.g., Bregman 1990), when synchronous sound elements are embedded in a longer sequence of similar sounds that are not presented synchronously or coherently, no grouping occurs, even between the elements that are synchronous (Christiansen & Oxenham 2014, Elhilali et al. 2009).

Some attempts have been made to identify the neural correlates of sensitivity to temporal coherence. Despite the perceptual difference between synchronous and alternating tones, an initial study found that responses to sequences of tone pairs in the primary auditory cortex of awake but passive ferrets did not depend on whether the two tones were synchronous or alternating

(Elhilali et al. 2009). This outcome led the authors to conclude that the neural correlates of the differences in perception elicited by synchronous and alternating tone sequences must emerge at a level higher than the primary auditory cortex. However, another study that compared the neural responses of ferrets when they were either passively listening or actively attending to the sounds found evidence supporting the theory of temporal coherence, with alternating tones producing suppression relative to the responses elicited by synchronous tones, but only when the ferrets were actively attending to the stimuli (Lu et al. 2017).

In addition to using differences in acoustic properties, the auditory system is able to make use of the regularities and repetitive natures of many natural sounds to help in the task of segregating competing sources. McDermott et al. (2011) found that listeners were able to segregate a repeating target sound from a background of varying sounds even when there were no acoustic cues with which to segregate the target sound. It seems that the repetitions themselves, against a varying background, allow the auditory system to extract the stable aspects of the sound. The authors proposed that this may be one way in which we are able to learn new sounds, even when they are never presented to us in complete isolation (McDermott et al. 2011).

Perceptual Multistability, Informational Masking, and the Neural Correlates of Auditory Attention and Awareness

As is the case with visual stimuli, the same acoustic stimulus can be perceived in more than one way, leading to perceptual ambiguity and, in some cases, multistability (Mehta et al. 2016). Alternating sound sequences provide one common example of such ambiguity: A rapidly alternating sequence of two tones is perceived as a single auditory stream if the frequency separation is small. However, the same sequence will be perceived as two separate streams (one high and one low) if the frequency separation between the two tones is large. In between, there exists a gray region where the percept can alternate between the two states and can depend on the attentional state of the listener. Studies comparing the dynamics of this bistability have found that it has a similar time course as analogous conditions in the visual domain but that the times at which switching occurs within each sensory modality are independent of the others even when the auditory and visual stimuli are presented at the same time (Pressnitzer & Hupe 2006). Neural correlates of such bistability have been identified in both auditory (Gutschalk et al. 2005) and nonauditory (Cusack 2005) regions of the cortex. Such stimuli are useful because they can, in principle, be used to distinguish between neural responses to the stimuli and neural correlates of perception.

Another approach to elucidating the neural correlates of auditory perception, attention, and awareness has been to use a phenomenon known as informational masking (Durlach et al. 2003). Informational masking is a term used to describe most kinds of masking that cannot be explained in terms of interactions or interference within the cochlea. Such peripherally based masking is known as energetic masking. Informational masking tends to occur when the masker and the target share some similarities (e.g., they both consist of pure tones and both emanate from the same spatial location) and when there is some uncertainty associated with the spectrotemporal properties of the masker or the target sound. Uncertainty can be produced by using randomly selected frequencies for tones within the masker. Informational masking shares some similarities with a visual phenomenon known as visual crowding (Whitney & Levi 2011) in that it, too, cannot be explained in terms of the limits of peripheral resolution. The term informational masking has been applied to both nonspeech sounds (Oxenham et al. 2003) and speech (Kidd et al. 2016), although it is not clear if the same mechanisms underlie both types of masking.

Because informational masking occurs when stimuli that are clearly represented in the auditory periphery are not heard, it provides an opportunity to probe the neural correlates of auditory

awareness or consciousness. An early study into the neural correlates of auditory awareness using informational masking in combination with magnetoencephalography found that the earliest cortical responses to sound, measured via the steady-state response to a 40-Hz modulation, provided a robust representation of the target sound that did not depend on whether the target was heard or not. In contrast, a later response (peaking approximately 100–150 ms after stimulus onset) was highly dependent on whether the target was heard, with no measurable response recorded when the target remained undetected (Gutschalk et al. 2008). This outcome suggests that informational masking does not occur in subcortical processing but already affects responses in the auditory cortex itself. However, it seems clear that the effects are not limited to the auditory cortex. For instance, the fact that visual stimuli can influence these responses suggests a feedback mechanism based on supramodal processing (Hausfeld et al. 2017).

Several recent studies have reported strong attentional modulation of auditory cortical responses using both speech (O'Sullivan et al. 2015) and nonspeech sounds (Chait et al. 2010). A study in patients using ECoG was able to accurately determine which of two talkers was attended based on cortical responses, to the extent that the neural response was wholly dominated by the sound of the attended talker (Mesgarani & Chang 2012).

PERCEPTUAL CONSEQUENCES OF HEARING LOSS AND COCHLEAR IMPLANTS

Importance of Hearing Loss

Hearing loss is a very common problem in industrial societies. In the United States alone, it is estimated that approximately 38 million adults have some form of bilateral hearing loss (Goman & Lin 2016). The problem worsens dramatically with age, so that more than 25% of people in their 60s suffer from hearing loss; for people in their 80s, the incidence rises to nearly 80% (Lin et al. 2011). If we take a stricter definition of a substantial or disabling hearing loss, meaning greater than 40 dB average loss between 500 and 4,000 Hz, the numbers are still very high, incorporating approximately one third of the world's adults aged 65 or older (WHO 2012). Hearing loss is defined as a loss of sensitivity to quiet sounds, but one of the most pressing problems associated with hearing loss is a reduced ability to hear out or segregate sounds, such as someone talking against a background of other sounds. This difficulty in understanding, and thus taking part in, conversations leads many people with hearing loss to avoid crowded situations, which, in turn, can lead to more social isolation, potential cognitive decline, and more general health problems (Kamil et al. 2016, Sung et al. 2016, Wayne & Johnsrude 2015). Understanding how hearing loss occurs, and how best to treat it, is a challenge of growing importance in our aging societies.

Cochlear Hearing Loss

By far the most common form of hearing loss is cochlear in origin. As discussed in the section titled Cochlear Tuning and Frequency Selectivity, the outer hair cells provide the cochlea with amplification of low-level sounds and sharp tuning. Strong amplification at low levels and little or no amplification at high levels produce a compressive input–output function, where a 100-dB range of sound levels is fitted into a much smaller range of vibration amplitudes in the cochlea (Ruggero 1992). A loss of function of the outer hair cells results in (*a*) a loss of sensitivity, (*b*) a loss of dynamic range compression, and (*c*) poorer frequency tuning. Each of these three factors has perceptual consequences for people with cochlear hearing loss (Oxenham & Bacon 2003). The loss of sensitivity is the classic symptom of hearing loss and the symptom that is measured

most frequently in clinical tests of hearing via the audiogram. Although the audibility of quiet sounds can be restored by amplification (e.g., with a hearing aid), simple linear amplification does not restore normal hearing because it does not address the remaining two factors of dynamic range and frequency tuning. The loss of dynamic range means that low-level sounds are no longer audible, but high-level sounds seem just as loud, leading to a smaller range of audible but tolerable sound levels. This phenomenon of loudness recruitment (Moore 2007) was known long before it was discovered that it could be explained by changes in the mechanics of the cochlea caused by a linearization of the basilar membrane response to sound in the absence of functioning outer hair cells (Ruggero 1992). Some aspects of loudness recruitment can be compensated for by introducing a compression circuit, which amplifies low-level sounds more than high-level sounds, into a hearing aid. However, this still leaves the consequences of poorer frequency tuning untreated.

The effects of poor cochlear frequency tuning can be measured behaviorally using the same masking methods that are employed to measure frequency selectivity in people with normal hearing, and such methods generally show poorer-than-normal frequency selectivity in people with hearing loss (Moore 2007). The loss of frequency selectivity may explain some of the difficulties faced by people with hearing loss in noisy environments: Poorer selectivity implies a reduced ability to segregate competing sounds.

Pitch perception is also generally poorer than normal in people with cochlear hearing loss. Again, this may be due in part to poorer frequency selectivity and a loss of spectrally resolved harmonics (Bernstein & Oxenham 2006, Bianchi et al. 2016). Relatively few studies have explored auditory stream segregation in hearing-impaired listeners, but those studies that exist also indicate that poorer frequency selectivity affects segregation abilities, which, in turn, is likely to explain some of the difficulties experienced by hearing-impaired listeners when trying to understand speech in complex acoustic environments (Mackersie 2003).

Unfortunately, hearing aids cannot restore sharp cochlear tuning. Because damage to the outer hair cells is currently irreversible, and because the consequences of hearing loss can be severe and wide ranging, it is particularly important to protect our hearing from overexposure to loud sounds. As outlined in the next section, even avoiding damage to the outer hair cells may not be sufficient to maintain acute hearing over the lifespan.

Hidden Hearing Loss

Most of us have experienced temporary threshold shift (TTS) at some time or other, such as after a very loud sporting event or rock concert. The phenomenon is often accompanied by a feeling of wooliness and, possibly, a sensation of ringing, but it usually resolves itself within 24 to 48 hours. However, recent physiological studies have suggested that the long-term consequences of TTS may not be as benign as previously thought. A landmark study by Kujawa & Liberman (2009) in mice revealed that noise exposure sufficient to cause TTS, but not sufficient to cause permanent threshold shifts, can result in a significant loss of the synapses between the inner hair cells in the cochlea and the auditory nerve. These synapses effectively connect the ear to the brain, so a 50% loss of synapses (as reported in many recent animal studies; e.g., Kujawa & Liberman 2009) is likely to have some important perceptual consequences. The surprising aspect of these results is that a 50% loss of synapses does not produce a measurable change in absolute thresholds, meaning that it would not be detected in a clinical hearing test, leading to the term hidden hearing loss (Schaette & McAlpine 2011).

The questions currently in need of urgent answers are: (*a*) Do humans suffer from hidden hearing loss? (*b*) If so, how prevalent is it? (*c*) What are the perceptual consequences in everyday life? Finally, (*d*) how can it best be diagnosed? A number of studies are currently under way

to provide answers to these questions. Indeed, studies have already suggested that some of the difficulties encountered by middle-aged and older people in understanding speech in noise may be related to hidden hearing loss (Bharadwaj et al. 2015, Ruggles et al. 2011). In addition, some consideration has gone into developing either behavioral or noninvasive physiological tests as indirect diagnostic tools to detect hidden hearing loss (Liberman et al. 2016, Plack et al. 2016, Stamper & Johnson 2015). Although it seems likely that people with more noise exposure would suffer from greater hidden hearing loss, the results from the first study with a larger sample of younger listeners (>100) have not yet revealed clear associations (Prendergast et al. 2017).

It may appear puzzling that a 50% loss of fibers leads to no measurable change in absolute thresholds for sound. There are at least three possible reasons for this, which are not mutually exclusive. First, further physiological studies have shown that the synapses most affected are those that connect to auditory nerve fibers with high thresholds and low spontaneous firing rates (Furman et al. 2013). These fibers are thought to be responsible for coding the features of sound that are well above absolute threshold, so a loss of these fibers may not affect sensitivity to very quiet sounds near absolute threshold. Second, higher levels of auditory processing, from the brainstem to the cortex, may compensate for the loss of stimulation by increasing neural gain (Chambers et al. 2016, Schaette & McAlpine 2011). Third, theoretical considerations based on signal detection theory have suggested that the perceptual consequences of synaptic loss may not be very dramatic until a large proportion of the synapses are lost (Oxenham 2016). In fact, with fairly simple and reasonable assumptions, it can be predicted that a 50% loss of synapses would result in only a 1.5-dB worsening of thresholds, which would be unmeasurable. Taken further, a 90% loss of fibers would be required to produce a 5-dB worsening of thresholds—still well below the 20-dB loss required for a diagnosis of hearing loss (Oxenham 2016). However, if the loss of fibers is concentrated in the small population of fibers with high thresholds and low spontaneous rates, then a loss of 90% or more is feasible, and may result in severe deficits for the processing of sounds that are well above absolute threshold—precisely the deficits that cause middle-aged and elderly people to have difficulty understanding speech in noisy backgrounds. In summary, hidden hearing loss remains a topic of considerable interest that has the potential to dramatically change the way hearing loss is diagnosed and treated.

Cochlear Implants

Cochlear implants represent by far the most successful sensory–neural prosthetic. They have enabled hundreds of thousands of people who would otherwise be deaf or severely hearing impaired to regain some auditory and speech capacities. Cochlear implants consist of an array of tiny electrodes that are surgically inserted into the turns of the cochlea with the aim of bypassing the ear and electrically stimulating the auditory nerve. Placing electrodes along the length of the array and stimulating them with different parts of the audio frequency spectrum are intended to recreate an approximation of the tonotopic mapping that occurs in the normal cochlea. Given that a crude array of 12–24 electrodes is used to replace the functioning of around 3,500 inner hair cells, perhaps the most surprising aspect of cochlear implants is that they work at all. However, many people with cochlear implants can understand speech in quiet conditions, even without the aid of lip reading.

One reason why cochlear implants have been so successful in transmitting speech information to their recipients is that speech is extremely robust to noise and distortion and requires very little in terms of spectral resolution (Shannon et al. 1995). Thus, even the limited number of spectral channels provided by a cochlear implant can be sufficient to convey speech. Pitch, on the other hand, requires much finer spectral resolution and, thus, remains a major challenge for cochlear

implants. Two main dimensions of pitch have been explored in cochlear implants. The first is referred to as place pitch and varies with the location of the stimulating electrode, with lower pitches reported as the place of stimulation moves further in toward the apex of the cochlea. The second is referred to as temporal, or rate, pitch and increases with an increasing rate of electrical pulses, at least up to approximately 300 Hz (McDermott 2004). It has generally been found that place pitch and temporal pitch in cochlear implant users are represented along independent dimensions (McKay et al. 2000) in much the same way as pitch and brightness are considered different dimensions (despite some interference) in acoustic hearing (Allen & Oxenham 2014). Thus, the place pitch in cochlear implant users may be more accurately described as a dimension of timbre (McDermott 2004).

In general, the pitch extracted from the pulse rate or envelope modulation rate by cochlear implant users is weak and inaccurate, with average thresholds often between 5% and 10%, or nearly 1–2 semitones (Kreft et al. 2013). Interestingly, similar thresholds are found in normal hearing listeners when they are restricted to just the temporal envelope cues provided by unresolved harmonics (Kreft et al. 2013, Shackleton & Carlyon 1994).

To restore accurate pitch sensations via cochlear implants would require the transmission of the information normally carried by resolved harmonics. Can this be achieved? A number of factors suggest that this may be challenging. First, the number of electrodes in current devices is limited to between 12 and 24, depending on the manufacturer. This would likely to be too few to provide an accurate representation of harmonic pitch. Second, even with a large number of channels, resolution is limited by the spread and interaction of current between adjacent electrodes and by possibly uneven neural survival along the length of the cochlea. For instance, in speech perception, the performance of cochlear implant users as the number of electrodes increases typically reaches a plateau at approximately 8 electrodes (Friesen et al. 2001) because the interference or crosstalk between electrodes limits the number of effectively independent channels (Bingabr et al. 2008). Third, the depth of insertion of an implant is limited by surgical constraints, meaning the implants generally do not reach the most apical portions of the cochlea, which, in turn, means that the auditory nerve fibers tuned to the lowest frequencies (and the ones most relevant for pitch) are not reached by the implant.

Some studies have used acoustic simulations to estimate the number of channels that might be needed to transmit accurate pitch information via cochlear implants (Crew et al. 2012, Kong et al. 2004). However, these studies allowed the use of temporal pitch cues, as well as cues based on the lowest frequency present in the stimulus, and so did not test the ability of listeners to extract information from resolved harmonics. A recent study limited listeners' access to the temporal envelope and spectral edge cues and found that at least 32 channels would be needed but, less encouragingly, that extremely narrow stimulation would be required from each channel. Simulating current spread with attenuation slopes as steep as 72 dB per octave was still not sufficient to elicit accurate pitch (Mehta & Oxenham 2017). To put that into context, current cochlear implants deal with spread that is equivalent to closer to 12–24 dB per octave (Bingabr et al. 2008, Oxenham & Kreft 2014). Even with recent developments in current focusing (Bierer & Litvak 2016), it is highly unlikely that sufficiently focused stimulation can be achieved using today's devices. This result suggests that novel interventions may be needed; these interventions may include neurotrophic agents that encourage neuronal growth toward the electrodes (Pinyon et al. 2014), optogenetic approaches that provide greater specificity of stimulation (Hight et al. 2015), or a different location of implantation, such as in the auditory nerve itself, where electrodes can achieve more direct contact with the targeted neurons (Middlebrooks & Snyder 2010).

Improving pitch perception via cochlear implants will not only provide the users with improved music perception, but should also improve many aspects of speech perception, especially for tone

languages, as well as the ability of cochlear implant recipients to hear out target sounds in the presence of interferers.

CONCLUSIONS

Auditory perception provides us with access to the acoustic environment and enables communication via speech and music. Some of the fundamental characteristics of auditory perception, such as frequency selectivity, are determined in the cochlea of the inner ear. Other aspects, such as pitch, are derived from higher-level representations, which are nonetheless affected by cochlear processing. More than 60 years since Cherry (1953) posed the famous cocktail party problem, work on human and animal behavior, work on human neuroimaging, and work on animal neurophysiology are being combined to answer the question of how the auditory brain is able to parse information in complex acoustic environments. The furthering of our knowledge of basic auditory processes has helped us to understand the causes of many types of hearing loss, but new findings on hidden hearing loss may signal a dramatic shift in how hearing loss is diagnosed and treated. Cochlear implants represent a highly successful intervention that provides speech understanding to many recipients, but they also highlight current limitations in technology and in our understanding of the underlying auditory processes.

FUTURE ISSUES

1. Is human frequency selectivity really much sharper than that found in other animals, and, if so, what differences in auditory perception between humans and other species can this variation explain?

2. Can we harness the knowledge gained from perceptual and neural studies of auditory scene analysis and source segregation to enhance automatic speech recognition and sound identification by computers?

3. Is cochlear synaptopathy, or hidden hearing loss, a common phenomenon in humans? If so, what are its consequences, and how can it best be diagnosed and treated?

4. How can we best restore pitch perception to recipients of cochlear implants? Will this restoration require a new electrode–neural interface or a completely different site of implantation?

DISCLOSURE STATEMENT

The author is not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

LITERATURE CITED

Allen EJ, Burton PC, Olman CA, Oxenham AJ. 2017. Representations of pitch and timbre variation in human auditory cortex. *J. Neurosci.* 37:1284–93

Allen EJ, Oxenham AJ. 2014. Symmetric interactions and interference between pitch and timbre. *J. Acoust. Soc. Am.* 135:1371–79

ANSI (Am. Nat. Stand. Inst.). 2013. *American National Standard: acoustical terminology*. Rep. S1.1-2013, Am. Nat. Stand. Inst./Accredit. Stand. Comm. Acoust., Acoust. Soc. Am., Washington, DC/Melville, NY

Attneave F, Olson RK. 1971. Pitch as a medium: a new approach to psychophysical scaling. *Am. J. Psychol.* 84:147–66

Bendor D, Wang X. 2005. The neuronal representation of pitch in primate auditory cortex. *Nature* 436:1161–65

Bentsen T, Harte JM, Dau T. 2011. Human cochlear tuning estimates from stimulus-frequency otoacoustic emissions. *J. Acoust. Soc. Am.* 129:3797–807

Bernstein JG, Oxenham AJ. 2003. Pitch discrimination of diotic and dichotic tone complexes: harmonic resolvability or harmonic number? *J. Acoust. Soc. Am.* 113:3323–34

Bernstein JG, Oxenham AJ. 2006. The relationship between frequency selectivity and pitch discrimination: sensorineural hearing loss. *J. Acoust. Soc. Am.* 120:3929–45

Bharadwaj HM, Masud S, Mehraei G, Verhulst S, Shinn-Cunningham BG. 2015. Individual differences reveal correlates of hidden hearing deficits. *J. Neurosci.* 35:2161–72

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

Bierer JA, Litvak L. 2016. Reducing channel interaction through cochlear implant programming may improve speech perception: current focusing and channel deactivation. *Trends Hear.* 20:2331216516653389

Bingabr M, Espinoza-Varas B, Loizou PC. 2008. Simulating the effect of spread of excitation in cochlear implants. *Hear. Res.* 241:73–79

Blauert J. 1997. *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press

Bregman AS. 1990. *Auditory Scene Analysis: The Perceptual Organisation of Sound*. Cambridge, MA: MIT Press

Bregman MR, Patel AD, Gentner TQ. 2016. Songbirds use spectral shape, not pitch, for sound pattern recognition. *PNAS* 113:1666–71

Burns EM, Feth LL. 1983. Pitch of sinusoids and complex tones above 10 kHz. In *Hearing—Physiological Bases and Psychophysics*, ed. R Klinke, R Hartmann, pp. 327–33. Berlin: Springer Verlag

Cedolin L, Delgutte B. 2010. Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. *J. Neurosci.* 30:12712–24

Chait M, de Cheveigne A, Poeppel D, Simon JZ. 2010. Neural dynamics of attending and ignoring in human auditory cortex. *Neuropsychologia* 48:3262–71

Chambers AR, Resnik J, Yuan Y, Whitton JP, Edge AS, et al. 2016. Central gain restores auditory processing following near-complete cochlear denervation. *Neuron* 89:867–79

Cherry EC. 1953. Some experiments on the recognition of speech, with one and two ears. *J. Acoust. Soc. Am.* 25:975–79

Christiansen SK, Oxenham AJ. 2014. Assessing the effects of temporal coherence on auditory stream formation through comodulation masking release. *J. Acoust. Soc. Am.* 135:3520–29

Cohen MR, Kohn A. 2011. Measuring and interpreting neuronal correlations. *Nat. Neurosci.* 14:811–19

Crew JD, Galvin JJ III, Fu QJ. 2012. Channel interaction limits melodic pitch perception in simulated cochlear implants. *J. Acoust. Soc. Am.* 132:EL429–35

Cusack R. 2005. The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* 17:641–51

Dallos P, Zheng J, Cheatham MA. 2006. Prestin and the cochlear amplifier. *J. Physiol.* 576:37–42

Darwin CJ. 2005. Pitch and auditory grouping. In *Pitch: Neural Coding and Perception*, ed. CJ Plack, AJ Oxenham, R Fay, AN Popper, pp. 278–305. Berlin: Springer Verlag

Deeks JM, Gockel HE, Carlyon RP. 2013. Further examination of complex pitch perception in the absence of a place-rate match. *J. Acoust. Soc. Am.* 133:377–88

Delgutte B. 1984. Speech coding in the auditory nerve: II. Processing schemes for vowel-like sounds. *J. Acoust. Soc. Am.* 75:879–86

Durlach NI, Mason CR, Kidd G Jr., Arbogast TL, Colburn HS, Shinn-Cunningham BG. 2003. Note on informational masking. *J. Acoust. Soc. Am.* 113:2984–87

Elhilali M, Ma L, Micheyl C, Oxenham AJ, Shamma SA. 2009. Temporal coherence in the perceptual organization and cortical representation of auditory scenes. *Neuron* 61:317–29

Evans EF. 2001. Latest comparisons between physiological and behavioural frequency selectivity. In *Physiological and Psychophysical Bases of Auditory Function*, ed. J Breebaart, AJM Houtsma, A Kohlrausch, VF Prijs, R Schoonhoven, pp. 382–87. Maastricht: Shaker

Feng L, Wang X. 2017. Harmonic template neurons in primate auditory cortex underlying complex sound processing. *PNAS* 114:E840–48

Fishman YI, Volkov IO, Noh MD, Garell PC, Bakken H, et al. 2001. Consonance and dissonance of musical chords: neural correlates in auditory cortex of monkeys and humans. *J. Neurophysiol.* 86:2761–88

Friesen LM, Shannon RV, Baskent D, Wang X. 2001. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J. Acoust. Soc. Am.* 110:1150–63

Furman AC, Kujawa SG, Liberman MC. 2013. Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *J. Neurophysiol.* 110:577–86

Goman AM, Lin FR. 2016. Prevalence of hearing loss by severity in the United States. *Am. J. Public Health* 106:1820–22

Gutschalk A, Micheyl C, Melcher JR, Rupp A, Scherg M, Oxenham AJ. 2005. Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* 25:5382–88

Gutschalk A, Micheyl C, Oxenham AJ. 2008. Neural correlates of auditory perceptual awareness under informational masking. *PLOS Biol.* 6:1156–65

Hartmann WM, Johnson D. 1991. Stream segregation and peripheral channeling. *Music Percept.* 9:155–84

Hausfeld L, Gutschalk A, Formisano E, Riecke L. 2017. Effects of cross-modal asynchrony on informational masking in human cortex. *J. Cogn. Neurosci.* 29(6):980–90

Heil P, Peterson AJ. 2015. Basic response properties of auditory nerve fibers: a review. *Cell Tissue Res.* 361:129–58

Heinz MG, Colburn HS, Carney LH. 2001. Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory. *Neural Comput.* 13:2273–316

Hight AE, Kozin ED, Darrow K, Lehmann A, Boyden E, et al. 2015. Superior temporal resolution of Chronos versus channelrhodopsin-2 in an optogenetic model of the auditory brainstem implant. *Hear Res.* 322:235–41

Houtsma AJM, Smurzynski J. 1990. Pitch identification and discrimination for complex tones with many harmonics. *J. Acoust. Soc. Am.* 87:304–10

Joris P, Bergevin C, Kalluri R, McLaughlin M, Michelet P, et al. 2011. Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. *PNAS* 108:17516–20

Kamil RJ, Betz J, Powers BB, Pratt S, Kritchevsky S, et al. 2016. Association of hearing impairment with incident frailty and falls in older adults. *J. Aging Health* 28:644–60

Kersten D, Mamassian P, Yuille A. 2004. Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55:271–304

Kidd G Jr., Mason CR, Swaminathan J, Roverud E, Clayton KK, Best V. 2016. Determining the energetic and informational components of speech-on-speech masking. *J. Acoust. Soc. Am.* 140:132–44

Kong YY, Cruz R, Jones JA, Zeng FG. 2004. Music perception with temporal cues in acoustic and electric hearing. *Ear Hear.* 25:173–85

Kreft HA, Nelson DA, Oxenham AJ. 2013. Modulation frequency discrimination with modulated and unmodulated interference in normal hearing and in cochlear-implant users. *J. Assoc. Res. Otolaryngol.* 14:591–601

Kujawa SG, Liberman MC. 2009. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J. Neurosci.* 29:14077–85

Lau BK, Mehta AH, Oxenham AJ. 2017. Superoptimal perceptual integration suggests a place-based representation of pitch at high frequencies. *J. Neurosci.* 37:9013–21

Liberman MC, Epstein MJ, Cleveland SS, Wang H, Maison SF. 2016. Toward a differential diagnosis of hidden hearing loss in humans. *PLOS ONE* 11:e0162726

Lin FR, Niparko JK, Ferrucci L. 2011. Hearing loss prevalence in the United States. *Arch. Intern. Med.* 171:1851–52

Liu LF, Palmer AR, Wallace MN. 2006. Phase-locked responses to pure tones in the inferior colliculus. *J. Neurophysiol.* 95:1926–35

Loeb GE, White MW, Merzenich MM. 1983. Spatial cross correlation: a proposed mechanism for acoustic pitch perception. *Biol. Cybern.* 47:149–63

Lopez-Poveda EA, Eustaquio-Martin A. 2013. On the controversy about the sharpness of human cochlear tuning. *J. Assoc. Res. Otolaryngol.* 14:673–86

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BCJ. 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *PNAS* 103:18866–69

Lu K, Xu Y, Yin P, Oxenham AJ, Fritz JB, Shamma SA. 2017. Temporal coherence structure rapidly shapes neuronal interactions. *Nat. Commun.* 8:13900

Lu T, Wang X. 2000. Temporal discharge patterns evoked by rapid sequences of wide-and narrowband clicks in the primary auditory cortex of cat. *J. Neurophysiol.* 84(1):236–46

Mackersie CL. 2003. Talker separation and sequential stream segregation in listeners with hearing loss: patterns associated with talker gender. *J. Speech Lang. Hear. Res.* 46:912–18

Marmel F, Plack CJ, Hopkins K, Carlyon RP, Gockel HE, Moore BC. 2015. The role of excitation-pattern cues in the detection of frequency shifts in bandpass-filtered complex tones. *J. Acoust. Soc. Am.* 137:2687–97

McDermott HJ. 2004. Music perception with cochlear implants: a review. *Trends Amplif.* 8:49–82

McDermott JH, Lehr AJ, Oxenham AJ. 2010. Individual differences reveal the basis of consonance. *Curr. Biol.* 20:1035–41

McDermott JH, Oxenham AJ. 2008. Music perception, pitch, and the auditory system. *Curr. Opin. Neurobiol.* 18:452–63

McDermott JH, Schultz AF, Undurraga EA, Godoy RA. 2016. Indifference to dissonance in native Amazonians reveals cultural variation in music perception. *Nature* 535:547–50

McDermott JH, Wrobleski D, Oxenham AJ. 2011. Recovering sound sources from embedded repetition. *PNAS* 108:1188–93

McKay CM, McDermott HJ, Carlyon RP. 2000. Place and temporal cues in pitch perception: Are they truly independent? *Acoust. Res. Lett. Online* 1:25–30

Mehta AH, Oxenham AJ. 2017. Vocoder simulations explain complex pitch perception limitations experienced by cochlear implant users. *J. Assoc. Res. Otolaryngol.* In press

Mehta AH, Yasin I, Oxenham AJ, Shamma S. 2016. Neural correlates of attention and streaming in a perceptually multistable auditory illusion. *J. Acoust. Soc. Am.* 140:2225

Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485:233–36

Micheyl C, Delhommeau K, Perrot X, Oxenham AJ. 2006. Influence of musical and psychoacoustical training on pitch discrimination. *Hear. Res.* 219:36–47

Micheyl C, Hanson C, Demany L, Shamma S, Oxenham AJ. 2013a. Auditory stream segregation for alternating and synchronous tones. *J. Exp. Psychol. Hum. Percept. Perform.* 39:1568–80

Micheyl C, Schrater PR, Oxenham AJ. 2013b. Auditory frequency and intensity discrimination explained using a cortical population rate code. *PLOS Comput. Biol.* 9:e1003336

Micheyl C, Xiao L, Oxenham AJ. 2012. Characterizing the dependence of pure-tone frequency difference limens on frequency, duration, and level. *Hear. Res.* 292:1–13

Middlebrooks JC, Snyder RL. 2010. Selective electrical stimulation of the auditory nerve activates a pathway specialized for high temporal acuity. *J. Neurosci.* 30:1937–46

Moerel M, De Martino F, Formisano E. 2014. An anatomical and functional topography of human auditory cortical areas. *Front. Neurosci.* 8:225

Moore BCJ. 2007. *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues.* Chichester, UK: Wiley

Moore BCJ, Ernst SM. 2012. Frequency difference limens at high frequencies: evidence for a transition from a temporal to a place code. *J. Acoust. Soc. Am.* 132:1542–47

Moore BCJ, Gockel H. 2002. Factors influencing sequential stream segregation. *Acta Acust. Unit. Acust.* 88:320–33

Moore BCJ, Sek A. 1996. Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. *J. Acoust. Soc. Am.* 100:2320–31

Norman-Haignere S, Kanwisher N, McDermott JH. 2013. Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *J. Neurosci.* 33:19451–69

O'Sullivan JA, Power AJ, Mesgarani N, Rajaram S, Foxe JJ, et al. 2015. Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25:1697–706

Oxenham AJ. 2012. Pitch perception. *J. Neurosci.* 32:13335–38

Oxenham AJ. 2016. Predicting the perceptual consequences of hidden hearing loss. *Trends Hear.* 20:2331216516686768

Oxenham AJ, Bacon SP. 2003. Cochlear compression: perceptual measures and implications for normal and impaired hearing. *Ear Hear.* 24:352–66

Oxenham AJ, Bernstein JGW, Penagos H. 2004. Correct tonotopic representation is necessary for complex pitch perception. *PNAS* 101:1421–25

Oxenham AJ, Fligor BJ, Mason CR, Kidd G Jr. 2003. Informational masking and musical training. *J. Acoust. Soc. Am.* 114:1543–49

Oxenham AJ, Kreft HA. 2014. Speech perception in tones and noise via cochlear implants reveals influence of spectral resolution on temporal processing. *Trends Hear.* 18:2331216514553783

Oxenham AJ, Micheyl C, Keebler MV, Loper A, Santurette S. 2011. Pitch perception beyond the traditional existence region of pitch. *PNAS* 108:7629–34

Oxenham AJ, Plack CJ. 1997. A behavioral measure of basilar-membrane nonlinearity in listeners with normal and impaired hearing. *J. Acoust. Soc. Am.* 101:3666–75

Patterson RD. 1976. Auditory filter shapes derived with noise stimuli. *J. Acoust. Soc. Am.* 59:640–54

Penagos H, Melcher JR, Oxenham AJ. 2004. A neural representation of pitch salience in non-primary human auditory cortex revealed with fMRI. *J. Neurosci.* 24:6810–15

Pinyon JL, Tadros SF, Froud KE, Wong ACY, Tompson IT, et al. 2014. Close-field electroporation gene delivery using the cochlear implant electrode array enhances the bionic ear. *Sci. Transl. Med.* 6:233ra54

Plack CJ, Leger A, Prendergast G, Kluk K, Guest H, Munro KJ. 2016. Toward a diagnostic test for hidden hearing loss. *Trends Hear.* 20:2331216516657466

Plantinga J, Trehub SE. 2014. Revisiting the innate preference for consonance. *J. Exp. Psychol. Hum. Percept. Perf.* 40:40–49

Plomp R. 1967. Pitch of complex tones. *J. Acoust. Soc. Am.* 41:1526–33

Plomp R, Levelt WJM. 1965. Tonal consonance and critical bandwidth. *J. Acoust. Soc. Am.* 38:548–60

Prendergast G, Guest H, Munro KJ, Kluk K, Leger A, et al. 2017. Effects of noise exposure on young adults with normal audiograms: I. Electrophysiology. *Hear. Res.* 344:68–81

Pressnitzer D, Hupe JM. 2006. Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* 16:1351–57

Qin MK, Oxenham AJ. 2005. Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear Hear.* 26:451–60

Rhode WS, Oertel D, Smith PH. 1983. Physiological response properties of cells labeled intracellularly with horseradish peroxidase in cat ventral cochlear nucleus. *J. Comp. Neurol.* 213:448–63

Roberts B, Glasberg BR, Moore BCJ. 2002. Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *J. Acoust. Soc. Am.* 112:2074–85

Rose JE, Brugge JF, Anderson DJ, Hind JE. 1967. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *J. Neurophysiol.* 30:769–93

Ruggero MA. 1992. Responses to sound of the basilar membrane of the mammalian cochlea. *Curr. Opin. Neurobiol.* 2:449–56

Ruggero MA, Temchin AN. 2005. Unexceptional sharpness of frequency tuning in the human cochlea. *PNAS* 102:18614–19

Ruggles D, Bharadwaj H, Shinn-Cunningham BG. 2011. Normal hearing is not enough to guarantee robust encoding of suprathreshold features important in everyday communication. *PNAS* 108:15516–21

Sayles M, Winter IM. 2008. Reverberation challenges the temporal representation of the pitch of complex sounds. *Neuron* 58:789–801

Schaette R, McAlpine D. 2011. Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *J. Neurosci.* 31:13452–57

Shackleton TM, Carlyon RP. 1994. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* 95:3529–40

Shamma S, Klein D. 2000. The case of the missing pitch templates: how harmonic templates emerge in the early auditory system. *J. Acoust. Soc. Am.* 107:2631–44

Shamma SA. 1985. Speech processing in the auditory system: II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J. Acoust. Soc. Am.* 78:1622–32

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. 1995. Speech recognition with primarily temporal cues. *Science* 270:303–4

Shera CA, Guinan JJ, Oxenham AJ. 2002. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *PNAS* 99:3318–23

Shera CA, Guinan JJ Jr., Oxenham AJ. 2010. Otoacoustic estimation of cochlear tuning: validation in the chinchilla. *J. Assoc. Res. Otolaryngol.* 11:343–65

Shofner WP, Chaney M. 2013. Processing pitch in a nonhuman mammal (*Chinchilla laniger*). *J. Comp. Psychol.* 127:142–53

Smith ZM, Delgutte B, Oxenham AJ. 2002. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416:87–90

Stamper GC, Johnson TA. 2015. Auditory function in normal-hearing, noise-exposed human ears. *Ear Hear.* 36:172–84

Sumner CJ, Wells T, Bergevin C, Palmer AR, Oxenham AJ, Shera CA. 2014. Comparing otoacoustic, auditory-nerve, and behavioral estimates of cochlear tuning in the ferret. *Proc. Midwinter Meet. Assoc. Res. Otolaryngol., San Diego, CA*, p. 123. San Diego, CA: Assoc. Res. Otolaryngol.

Sung YK, Li L, Blake C, Betz J, Lin FR. 2016. Association of hearing loss and loneliness in older adults. *J. Aging Health* 28:979–94

Trainor LJ, Heinmiller BM. 1998. The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behav. Dev.* 21:77–88

Tramo MJ, Cariani PA, Delgutte B, Braida LD. 2001. Neurobiological foundations for the theory of harmony in Western tonal music. *Ann. N. Y. Acad. Sci.* 930:92–116

Vliegen J, Oxenham AJ. 1999. Sequential stream segregation in the absence of spectral cues. *J. Acoust. Soc. Am.* 105:339–46

von Békésy G. 1960. *Experiments in Hearing*. New York: McGraw-Hill

Walker KM, Bizley JK, King AJ, Schnupp JW. 2011. Multiplexed and robust representations of sound features in auditory cortex. *J. Neurosci.* 31:14565–76

Wayne RV, Johnsrude IS. 2015. A review of causal mechanisms underlying the link between age-related hearing loss and cognitive decline. *Aging Res. Rev.* 23:154–66

Whiteford KL, Oxenham AJ. 2015. Using individual differences to test the role of temporal and place cues in coding frequency modulation. *J. Acoust. Soc. Am.* 138:3093–104

Whitney D, Levi DM. 2011. Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn. Sci.* 15:160–68

WHO (World Health Organ.). 2012. *Hearing loss in persons 65 years and older based on WHO global estimates on prevalence of hearing loss: mortality and burden of diseases and prevention of blindness and deafness*. Rep., World Health Organ., Geneva. **http://www.who.int/pbd/deafness/news/GE_65years.pdf**

Yin P, Fritz JB, Shamma SA. 2010. Do ferrets perceive relative pitch? *J. Acoust. Soc. Am.* 127:1673–80

Yin P, Fritz JB, Shamma SA. 2014. Rapid spectrotemporal plasticity in primary auditory cortex during behavior. *J. Neurosci.* 34:4396–408

Young ED, Sachs MB. 1979. Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibres. *J. Acoust. Soc. Am.* 66:1381–403

Zentner MR, Kagan J. 1996. Perception of music by infants. *Nature* 383:29

Zwicker E. 1970. Masking and psychological excitation as consequences of the ear's frequency analysis. In *Frequency Analysis and Periodicity Detection in Hearing*, ed. R Plomp, GF Smoorenburg, pp. 376–94. Leiden: Sijthoff

**Annual Review of
Psychology**

Volume 69, 2018

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Psychology* articles may be found at
http://www.annualreviews.org/errata/psych