

Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants

Andrew J. Oxenham, PhD

Pitch is important for speech and music perception, and may also play a crucial role in our ability to segregate sounds that arrive from different sources. This article reviews some basic aspects of pitch coding in the normal auditory system and explores the implications for pitch perception in people with hearing impairments and cochlear implants. Data from normal-hearing listeners suggest that the low-frequency, low-numbered harmonics within complex tones are of prime importance in pitch perception and in the perceptual segregation of competing sounds. The poorer frequency selectivity experienced by many hearing-impaired listeners leads to

less access to individual harmonics, and the coding schemes currently employed in cochlear implants provide little or no representation of individual harmonics. These deficits in the coding of harmonic sounds may underlie some of the difficulties experienced by people with hearing loss and cochlear implants, and may point to future areas where sound representation in auditory prostheses could be improved.

Keywords: Pitch; Hearing Loss; Cochlear Implants; Auditory Streaming

People with normal hearing can pick out the violin melody in an orchestral piece, or listen to someone speaking next to them in a crowded room, with relative ease. Despite the fact that we achieve these tasks with little or no effort, we know surprisingly little about how the brain and the ears achieve them. The process by which successive sounds from one source (such as a violin or a person talking) are perceptually grouped together and separated from other competing sounds is known as stream segregation, or simply “streaming” (Bregman, 1990; Darwin & Carlyon, 1995). Streaming involves at least two stages, which in some cases can influence and interact with each other. One stage is to bind together across frequency those elements emitted from one source and to separate them from other sounds that are present at the same time. This process is known as *simultaneous grouping* or *segregation*. There are a number of acoustic cues that influence the perceptual organization of

simultaneous sounds, such as onset and offset asynchrony (whether components start and stop at the same time) and harmonicity (whether components share the same fundamental frequency or F0); sounds that start or stop at the same time and share the same F0 are more likely to be heard as a single object. The other stage is to bind together across time those sound elements that belong to the same source and to avoid combining sequential sound elements that do not—a process known as sequential grouping or segregation. Again, a number of different acoustic cues play a role, such as the temporal proximity of successive sounds, as well as their similarity in F0 and spectrum, with more similar and more proximal sounds being more likely to be heard as belonging to a single stream. Beyond these acoustic, or bottom-up, cues, there are many top-down influences on perceptual organization, including attention, expectations, and prior exposure, all of which can also play an important role (e.g., Marrone, Mason, & Kidd, in press; Shinn-Cunningham & Best, in press).

Returning to acoustic cues, F0 and its perceptual correlate—pitch—have been shown to influence both sequential and simultaneous grouping; in fact, poorer pitch perception in people with hearing impairment and users of cochlear implants may underlie some of

From University of Minnesota, Minneapolis, Minnesota.

Address correspondence to: Andrew J Oxenham, PhD, Department of Psychology, University of Minnesota, 75 E. River Parkway, N218, Minneapolis, MN 55455; e-mail: oxenham@umn.edu.

the difficulties they face in complex everyday acoustic environments (Qin & Oxenham, 2003; Summers & Leek, 1998). This article reviews some recent findings on the basic mechanisms of pitch perception, and explores their implications for understanding auditory streaming in normal hearing, as well as their implications for hearing-impaired people and cochlear-implant users.

Fundamentals of Pitch Perception

Pitch of Pure Tones

The pitch of a pure tone is determined primarily by its frequency, with low frequencies eliciting low-pitch sensations and high frequencies eliciting high-pitch sensations. Changes in the intensity of a tone can affect pitch, but the influence tends to be relatively small and inconsistent (for a brief review, see Plack & Oxenham, 2005). Despite the apparently straightforward mapping from frequency to pitch, it is still not clear exactly what information the brain uses to represent pitch. The two most commonly cited possibilities are known as “place” and “temporal” codes. Both refer to the representation of sound within the auditory nerve. According to the place theory of pitch, the pitch of a tone is determined by which auditory nerve fibers it excites most. A tone produces a pattern of activity along the basilar membrane in the cochlea that varies according to the frequency of the tone—high frequencies produce maximal activity near the base of the cochlea, whereas low frequencies produce maximal activity near the apex of the cochlea. Because neural activity in each auditory nerve fiber reflects mechanical activity at a certain point along the basilar membrane, changes in the pattern of basilar membrane vibration are reflected by changes in which auditory nerve fibers respond most. This frequency-to-place, or frequency-to-fiber, mapping is known as “tonotopic representation.”

According to the temporal theory, pitch is represented by the timing of action potentials, or spikes, in the auditory nerve. At low frequencies (below about 4 kHz in mammals that have been studied so far), spikes are more likely to occur at one phase in the cycle of a sinusoid than at another. This property, known as phase locking, means that the time interval between pairs of spikes is likely to be a multiple of the period of the sinusoid (Rose, Brugge, Anderson, & Hind, 1967). By pooling information across multiple auditory nerve fibers, the time intervals

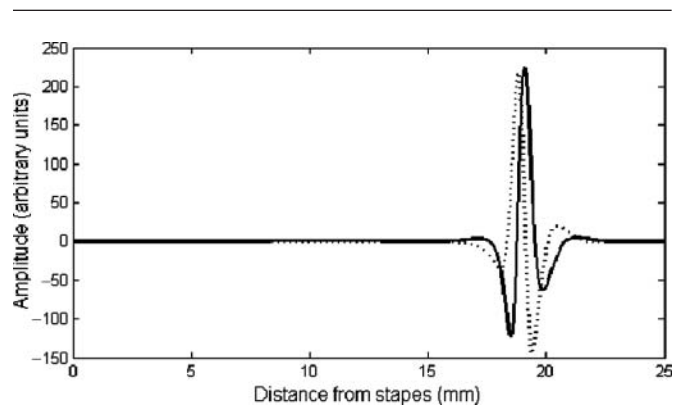


Figure 1. Schematic diagram of a traveling wave along the basilar membrane. Snapshots are shown of the same traveling wave at two points in time, spaced one fourth of a cycle apart.

between spikes can be used to derive information about the frequency of the pure tone that is generating them.

A third potential mechanism for coding pure tones involves both place and timing information. According to this place–time scheme, the timing information in the auditory nerve is used to derive the pitch, but it must be presented to the appropriate place (or tonotopic location) along the cochlea. A schematic representation of a snapshot in time of the cochlear traveling wave is shown in Figure 1. At any one point in time, different parts of the basilar membrane are at different phases in the sinusoidal cycle. The rate of change in phase is particularly rapid near the peak of stimulation where close-by points can have a phase shift between them of 180 degrees (π radians) or more. The points along the basilar membrane that are in or out of phase with each other will depend on the frequency of the stimulating sound. Thus, the patterns of phase differences along the basilar membrane could in principle be used by the auditory system to derive the frequency of a pure tone (Loeb, White, & Merzenich, 1983; Shamma, 1985).

All three potential codes (place, time, and place–time) have some evidence in their favor, and it has been difficult to distinguish between them in a decisive way. The place code would suggest that pitch discrimination of pure tones should depend on the frequency selectivity of cochlear filtering, with sharp tuning leading to better discrimination. This prediction is not well matched by data in normal-hearing listeners (Moore, 1973), which show that pitch discrimination is best at medium frequencies

(between about 500 and 2000 Hz) and is considerably worse at frequencies above about 4000 Hz; in contrast, frequency selectivity appears to stay quite constant (Glasberg & Moore, 1990) or may even improve (Shera, Guinan, & Oxenham, 2002) at high frequencies. The time code relies on phase-locking in the auditory nerve. Not much is known about phase-locking in the human auditory nerve, but data from other mammalian species, such as guinea pig and cat, suggest that phase locking degenerates above about 2 kHz, and is very poor beyond about 4 kHz. The fact that musical pitch perception (i.e., the ability to recognize and discriminate melodies and musical intervals) degrades for pure tones above about 4 kHz may be consistent with the supposed limits of phase locking in humans, based on studies in other mammals (Kim & Molnar, 1979). On the other hand, if phase locking is effectively absent above about 4 kHz, another mechanism must be posited to explain how pitch judgments remain possible, even up to very high frequencies of 16 kHz, albeit with severely reduced accuracy. In fact, a recent study using a computational model of the auditory nerve has suggested that even though phase locking is degraded, there may be enough information in the temporal firing patterns to explain human pitch perception even at very high frequencies (Heinz, Colburn, & Carney, 2001). Place-time models have not been tested quantitatively in their ability to predict pitch perception of pure tones, but it is likely that they would encounter the same limitations associated with both place and time models.

The question of what information in the auditory periphery is used by the central auditory system to code the frequency of tones is one of basic scientific interest, but it also has practical implications for better understanding hearing impairment and for coding strategies in cochlear implants. Nevertheless, there are limits to what the coding of pure tones can tell us about how more complex stimuli are processed. The following section reviews the pitch perception of harmonic complex tones, which form an important class of sounds in our everyday environment.

Pitch of Harmonic Complex Tones

Voiced speech, the sounds from most musical instruments, and many animal vocalizations are all formed by harmonic complex tones. These tones are made up of many pure tones, called harmonics, all of which have frequencies that are multiples of a

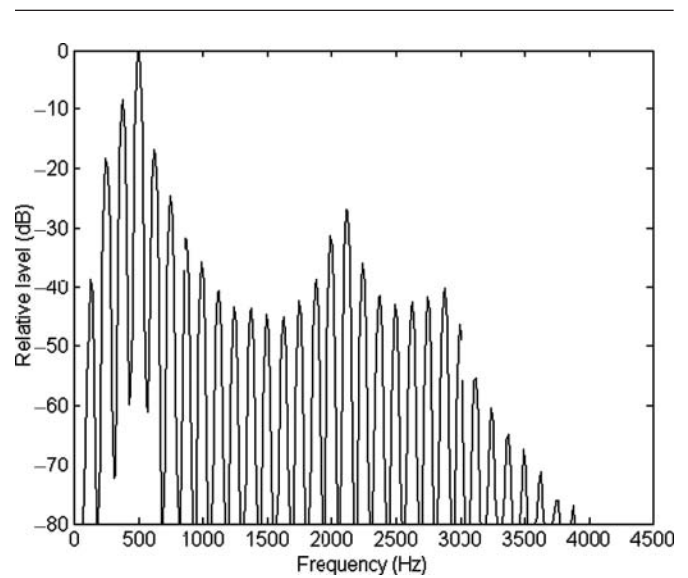


Figure 2. Spectrum of a synthesized vowel sound / ϵ / (as in “bet”), with an F0 of 125 Hz. Note the locations of the individual harmonics at integer multiples of 125 Hz, which determine the F0, and the locations of the spectral peaks, or formants, at around 500, 2100, and 2800 Hz, which determine the vowel’s identity.

single F0. Figure 2 shows the power spectrum of a short synthesized vowel (/ ϵ / as in “bet”), with energy at frequencies that are multiples of 125 Hz—the vowel’s F0. In most situations, we perceive a harmonic complex tone like that shown in Figure 2 as a single sound, and not as multiple sounds with pitches corresponding to the frequencies of the individual harmonics. Typically the pitch we hear corresponds to the F0 of the harmonic complex tone, even when there is no energy at the F0 itself. This phenomenon, known as the “pitch of the missing fundamental” (Licklider, 1954; Schouten, 1940), is important because it helps to ensure that sounds do not lose or change their identity when they are partially masked (McDermott & Oxenham, 2008).

Figure 3 shows how such a harmonic complex tone might be represented in the auditory periphery. The top panel shows the time waveform of a prototypical harmonic complex tone with an F0 of 100 Hz (similar to an average male talker). The second panel shows the frequency spectrum of the complex, illustrating that the sound has energy at integer multiples of the F0 - 200, 300, 400, 500 Hz, and so on. The third panel represents the filtering that takes place in the cochlea. Each point along the cochlea’s basilar membrane responds to a limited range of frequencies and so can be represented as a band-pass filter. In this way, the whole length of the basilar

membrane can be represented as a bank of overlapping band-pass filters. The bandwidths of these so-called auditory filters have been estimated in humans using masking and other psychoacoustic techniques (Fletcher, 1940; Glasberg & Moore, 1990; Zwicker, Flottorp, & Stevens, 1957) as well as with otoacoustic emissions (Shera et al., 2002). More direct measures of basilar-membrane (Rhode, 1971; Ruggero, Rich, Recio, Narayan, & Robles, 1997) and auditory-nerve (Lieberman, 1978) responses, or both (Narayan, Temchin, Recio, & Ruggero, 1998), have also been undertaken in various mammalian species. The results from these studies are broadly consistent in showing that the relative bandwidths of the filters (i.e., the bandwidth, as a proportion of the filter's center frequency) decrease with increasing center frequency up to about 1 kHz, and then either remain roughly constant (Evans, 2001; Glasberg & Moore, 1990), or continue to decrease (Shera et al., 2002; Tsuji & Liberman, 1997), depending on how the filters are measured. However, because the spacing of harmonics within a complex tone remains constant on an absolute frequency scale, it is more important to consider how the *absolute* bandwidth (in Hertz) of the auditory filters varies as a function of center frequency. The second panel of Figure 3 illustrates how the absolute bandwidths of the filters increase with increasing center frequency.

The effects of the relationship between harmonic spacing (Panel 2) and filter bandwidth (Panel 3) are shown in the lower two panels of Figure 3. The fourth panel shows the output, averaged over time, of each auditory filter, represented on the x-axis by its center frequency. For instance, the filter centered at 100 Hz has a high output because it responds strongly to the frequency component at 100 Hz, whereas the filter centered at 150 Hz has a lower output because it responds strongly neither to the 100-Hz nor the 200-Hz component. This type of representation is known as an excitation pattern. The excitation pattern displays a series of marked peaks at filters with center frequencies corresponding to the lower harmonic frequencies. Harmonics that produce such peaks in the excitation pattern are said to be resolved. At higher frequencies, the bandwidth of the filters begins to exceed the spacing between adjacent harmonics, so that filters centered between two adjacent harmonics begin to respond as strongly as filters centered on a harmonic. At this point, the peaks in response to individual harmonics become less distinct and eventually disappear, resulting in unresolved harmonics. According to the representation shown in Figure 3, which is based on the excitation pattern model of Glasberg and Moore (1990), at least the first five,

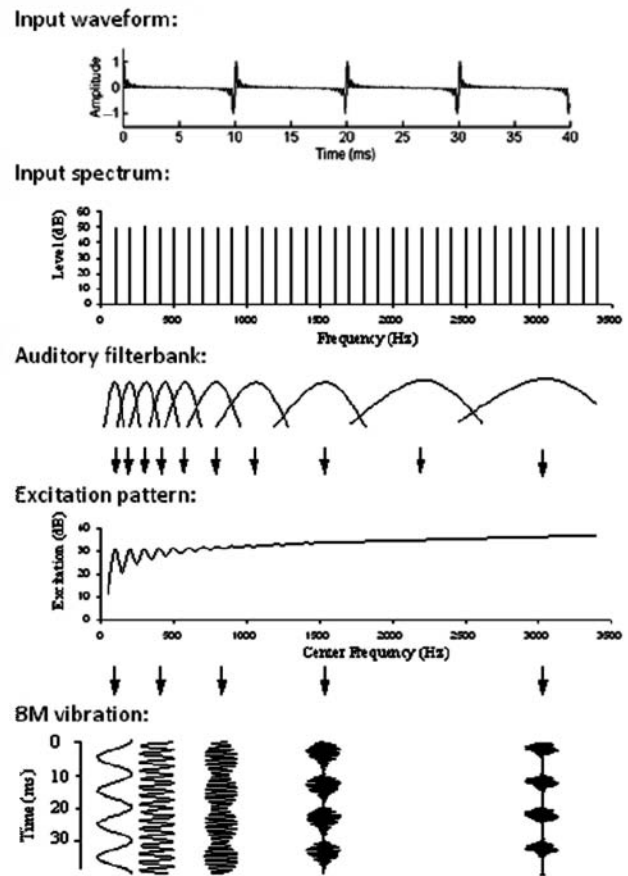


Figure 3. Peripheral representations of a harmonic complex tone. The upper panel shows the time waveform of a harmonic complex tone with an F0 of 100 Hz. The second panel shows the frequency spectrum of the same complex. The third panel shows the band-pass characteristics of the auditory filters, which have wider absolute bandwidths at higher center frequencies. The fourth panel shows the time-averaged output, or excitation, of these filters, as a function of the filter center frequency; this is known as the excitation pattern. The fifth panel shows the waveforms at the outputs of some sample auditory filters. Some lower filters respond only to one component of the complex, resulting in a sinusoidal output; higher-frequency filters respond to multiple components, producing a more complex output with a temporal envelope that repeats at the F0. The figure is adapted from Plack and Oxenham (2005).

and possibly the first eight, harmonics could be considered resolved. A number of models of complex pitch perception are based on the excitation pattern concept (Cohen, Grossberg, & Wyse, 1995; Goldstein, 1973; Terhardt, Stoll, & Seewann, 1982; Wightman, 1973). The general idea is that—just as with the place theory of pure-tone pitch—the peaks in the excitation pattern provide information about which frequencies are present, and then this information is combined to calculate the underlying F0, perhaps

via preformed “harmonic templates” (e.g., Goldstein, 1973).

Another representation to consider involves the temporal waveforms at the output of the auditory filters. Some examples are shown in the bottom panel of Figure 3. For the low-numbered resolved harmonics, filters centered at or near one of the harmonic frequencies respond primarily to that single component, producing a response that approximates a pure tone. As the harmonics become more unresolved, they interact within each filter producing a complex waveform with a temporal envelope that repeats at a rate corresponding to the F0. Temporal models of pitch use different aspects of these temporal waveforms to extract the pitch. The earliest temporal theories were based on the regular time intervals between high-amplitude peaks of the temporal fine structure in the waveform produced by unresolved harmonics (de Boer, 1956; Schouten, Ritsma, & Cardozo, 1962). Other temporal models use the timing information from resolved harmonics (Srulovicz & Goldstein, 1983) to extract the F0, but perhaps the most influential model, based on the autocorrelation function, extracts periodicities from the temporal fine structure of resolved harmonics, and the temporal envelope from high-frequency unresolved harmonics, by pooling temporal information from all channels to derive the most dominant periodicity, which in most cases corresponds to the F0 (Cariani & Delgutte, 1996; Licklider, 1951; Meddis & Hewitt, 1991a, 1991b; Meddis & O’Mard, 1997).

Just as with pure tones, complex-tone pitch perception has been explained with a place–time model, in addition to place or time models (Shamma & Klein, 2000), and recent physiological studies have explored this possibility (Cedolin & Delgutte, 2007; Larsen, Cedolin, & Delgutte, 2008). However, no studies have yet been able to definitively rule out or confirm any particular model in the auditory periphery. More centrally, Bendor and Wang (2005, 2006) have identified neurons in the auditory cortex of the marmoset, which seem to respond selectively to certain F0s, as opposed to frequencies or spectral regions, and roughly homologous regions have been identified in humans from imaging studies (Penagos, Melcher, & Oxenham, 2004). However, these higher-level responses do not provide much of a clue regarding the peripheral code that produces the cortical responses.

What can data from human perception experiments tell us about the validity of the various models? At present, there are no conclusive arguments in favor of any one theory over others. Many studies have found that low-numbered harmonics produce a more salient

and more accurate pitch than high-numbered harmonics, suggesting that the place or time information provided by resolved harmonics is more important than the temporal envelope information provided by unresolved harmonics (Bernstein & Oxenham, 2003; Dai, 2000; Houtsma & Smurzynski, 1990; Moore, Glasberg, & Peters, 1985; Plomp, 1967; Ritsma, 1967; Shackleton & Carlyon, 1994). Nevertheless, some pitch information is conveyed via the temporal envelope of unresolved harmonics (Kaernbach & Bering, 2001), as well as temporal-envelope fluctuations in random noise (Burns & Viemeister, 1981), suggesting that at least some aspects of pitch are derived from the timing information in the auditory nerve, and not solely place information. It is interesting to note that, when only unresolved harmonics are present, the phase relationships between the harmonics affect pitch perception discrimination, as one might expect if performance were based on the temporal envelope. The filter outputs shown in the bottom panel of Figure 3 illustrate the well-defined and temporally compact envelope that is produced when components are in sine phase with one another. When the phase relationships are changed, the envelope can become much less modulated, which usually leads to poorer pitch perception and discrimination, but only when purely unresolved harmonics are present in the stimulus (Bernstein & Oxenham, 2006a; Houtsma & Smurzynski, 1990; Shackleton & Carlyon, 1994).

Despite the ability of temporal models to explain many aspects of pitch perception, timing does not seem to be the whole story either: When the temporal information that would normally be available via low-frequency resolved harmonics is presented to a high-frequency region of the cochlea, it does not produce a pitch of the missing fundamental, suggesting that place information may be of importance (Oxenham, Bernstein, & Penagos, 2004). Along the same lines, auditory frequency selectivity and the resolvability of harmonics can predict pitch discrimination accuracy to some extent, suggesting that peripheral filtering is important in understanding pitch coding (Bernstein & Oxenham, 2006a, 2006b). On the other hand, artificially improving the resolvability of harmonics by presenting alternating harmonics to opposite ears does not lead to an improvement in pitch perception (Bernstein & Oxenham, 2003). It may be that, in normal hearing, all the information in the auditory periphery (place, time, and place–time) is used to some degree to increase the redundancy, and hence, the robustness of the pitch code.

Pitch Perception of Multiple Sounds

Surprisingly little research has been done on the pitch perception of multiple simultaneous harmonic complex tones. The little work that exists also points to a potentially important role for resolved harmonics. Beerends and Houtsma (1989) presented listeners with two simultaneous complex tones, with each complex comprised of only two adjacent harmonics, giving a total of four components. They found that listeners were able to identify the pitches of both complexes if at least one of the four components was resolved. Carlyon (1996) found that when two complexes, each containing many harmonics, were in the same spectral region and were presented at the same overall level, it was only possible to hear the pitch of one when the complexes contained some resolved harmonics; when the stimuli were filtered so that only unresolved harmonics were present, the mixture sounded more like a crackle than two pitches, suggesting that listeners are not able to extract two periodicities from a single temporal envelope.

Pitch Perception in Hearing Impairment

Pitch perception—both with pure tones and complex tones—is often poorer than normal in listeners with hearing impairment but, as with many auditory capabilities, there is large variability between different listeners (Moore & Carlyon, 2005). Moore and Peters (1992) measured pitch discrimination in young and old normal-hearing and hearing-impaired listeners using pure tones and complex tones filtered to contain the first 12 harmonics, only harmonics 1-5, or harmonics 4-12 or 6-12. They found a wide range of performance among their hearing-impaired listeners, with some showing near-normal ability to discriminate small differences in the frequency of pure tones, whereas others were severely impaired. They also found poorer-than-normal frequency discrimination in their older listeners with near-normal hearing, suggesting a possible age effect that is independent of hearing loss. Hearing impairment is often accompanied by poorer frequency selectivity, in the form of broader auditory filters. According to the place theory of pitch, poorer frequency selectivity should lead to poorer pitch perception and larger (worse) frequency difference limens (FDLs). Interestingly, however, FDLs were only weakly correlated with auditory filter bandwidths (Glasberg & Moore, 1990; Patterson, 1976), suggesting perhaps that place coding cannot fully account for pure-tone pitch perception. These and other results (Buss, Hall,

& Grose, 2004) have led to the conclusion that at least some hearing-impaired listeners may suffer from a deficit in temporal coding. An extreme example of deficits in temporal coding seem to be exhibited by subjects diagnosed with so-called auditory neuropathy, based on normal otoacoustic emissions but abnormal or absent auditory brain-stem responses (Sininger & Starr, 2001). Both low-frequency pure-tone frequency discrimination and modulation perception are typically highly impaired in people diagnosed with auditory neuropathy (Zeng, Kong, Michalewski, & Starr, 2005).

When it came to complex tones, Moore and Peters (1992) also found quite variable results, with one interesting exception: No listeners with broader-than-normal auditory filters showed normal F0 discrimination of complex tones. Bernstein and Oxenham (2006b) also investigated the link between auditory filter bandwidths and complex pitch perception. They specifically targeted the abrupt transition from good to poor pitch perception that is typically observed in normal-hearing listeners as the lower (resolved) harmonics are successively removed from a harmonic complex tone (Houtsma & Smurzynski, 1990). The main prediction was that poorer frequency selectivity should lead to fewer resolved harmonics. If resolved harmonics are important for good pitch perception, then pitch perception should deteriorate more rapidly in people with broader auditory filters as the lowest harmonics in each stimulus are successively removed. In other words, the point at which there are no resolved harmonics left in the stimulus should be reached more quickly with hearing-impaired listeners than with normal-hearing listeners, and that should be reflected in the pitch percept elicited by the stimuli. Bernstein and Oxenham found a significant correlation between the lowest harmonic number present at the point where the F0 difference limens (F0DLs) became markedly poorer and the bandwidth of the auditory filters in individual hearing-impaired subjects. The point at which F0DLs became poorer also correlated well with the point at which the phase relations between harmonics started to have an effect on F0DLs. Remember that phase effects are thought to occur only when harmonics are unresolved and interact within single auditory filters. Overall, the data were consistent with the idea that peripheral frequency selectivity affects the pitch perception of complex tones and that hearing-impaired listeners with poorer frequency selectivity suffer from poorer complex pitch perception.

Pitch Perception in Cochlear-Implant Users

Most cochlear implants today operate by filtering the incoming sound into a number of contiguous frequency bands, extracting the temporal envelope in each band, and presenting fixed-rate electrical pulses, amplitude-modulated by the temporal envelope of each band, to different electrodes along the array in a rough approximation of the tonotopic representation found in normal hearing.

Understanding the basic mechanisms by which pitch is coded in normal-hearing listeners has important implications for the design of cochlear implants. For instance, given the technical limitations of cochlear implants, should the emphasis be on providing a better place representation by increasing the number and selectivity of electrodes in an implant, or should designers focus on temporal representations by more closely imitating the firing patterns in the normal auditory nerve in the electrical pulses fed to the electrodes? If it emerges that a place–time code is required, further work will be required to determine whether such patterns can feasibly be presented via cochlear implants at all.

Research in cochlear implant users suggests that some use can be made of both place and timing cues. For instance, studies that ask subjects to rank the pitch percept associated with each electrode have generally found that the stimulation of electrodes near the base of the cochlea produce higher pitch percepts than electrodes nearer the apex (Nelson, van Tasell, Schroder, Soli, & Levine, 1995; Townsend, Cotter, van Compernelle, & White, 1987). More recent studies in individuals who have a cochlear implant in one ear but some residual hearing in the other ear have also suggested that pitch generally increases with the stimulation of more basally placed electrodes, in line with expectations based on the tonotopic representation of the cochlea (Boex et al., 2006; Dorman et al., 2007). However, it is difficult to know for sure what percepts the cochlear-implant users are experiencing—whether it is true pitch, or perhaps something more similar to the change in brightness or sharpness experienced by normal-hearing listeners when the centroid of energy in a broad stimulus is shifted to higher frequencies (von Bismark, 1974); for instance, when music is passed through a high-pass filter it sounds brighter, tinnier, perhaps even higher, but the notes in the music do not change their pitch in the traditional sense.

Some of the earliest studies of pitch perception in cochlear implant users were done by manipulating the timing cues presented to the subjects. For

instance, in studies where melodies were presented to a single electrode, thereby ruling out place cues, subjects were able to recognize and discriminate melodies (Moore & Rosen, 1979; Pijl & Schwarz, 1995). Simpler psychophysical studies measuring pulse-rate discrimination thresholds have found that rate changes can often be discriminated with base pulse rates between about 50 and 200 Hz (Pfingst et al., 1994; Zeng, 2002). Beyond about 200-300 Hz, most cochlear implant users are no longer able to discriminate changes in the rate of stimulation. Studies that have varied both electrode (place) and rate (timing) cues independently have found evidence that these two cues lead to separable perceptual dimensions (McKay, McDermott, & Carlyon, 2000; Tong, Blamey, Dowell, & Clark, 1983), suggesting that place and timing cues do not map onto the same pitch dimension.

The general finding has been that place and rate coding in cochlear implants do not yield the same fine-grained pitch perception experienced by normal-hearing listeners. This may in part be due to a complete lack of resolved harmonics in cochlear implants, as well as a lack of temporal fine structure coding—most current implants convey only temporal envelope information and discard the temporal fine structure. Although temporal-envelope (and pulse-rate) cues can provide some pitch information for cochlear implant users, the resulting pitch percept seems to be generally weak. For instance, a survey of previous studies (Moore & Carlyon, 2005) showed that the average just-noticeable change in pulse rate was more than 7% for a 100-Hz pulse. This exceeds a musical half-step, or semitone (about 6%), and is about an order of magnitude worse than performance by normal-hearing listeners when low-order resolved harmonics are present.

The Role of Pitch in Source Segregation

As mentioned in the introduction, F0 and its perceptual correlate—pitch—have long been thought to play an important role in our ability to group sounds that come from the same source and to segregate sounds that come from different sources, whether it be organizing simultaneous sounds into objects or sequential sounds into streams. The pitch and temporal fine structure information associated with voiced speech sounds have also been credited with helping to improve speech perception in the presence of complex, fluctuating backgrounds, such as a competing talker or a modulated noise. The

basic idea is that envelope fluctuations are important for speech intelligibility (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), but pitch may help decide which fluctuations belong to the target speech and which belong to the interferer (Qin & Oxenham, 2003).

Pitch and Source Segregation in Normal Hearing

The effect of F0 differences between simultaneous sources has been examined most thoroughly using pairs of simultaneously presented synthetic vowels (Assmann & Summerfield, 1987; de Cheveigné, 1997; Parsons, 1976; Scheffers, 1983). The task of listeners is to identify both vowels present in the mixture. The general finding is that performance improves as the F0 separation between the two vowels increases, in line with expectations from F0-based perceptual segregation. However, a number of studies have pointed out other potential cues that enable listeners to hear out the two vowels, which are not directly related to perceived pitch differences between the two vowels, such as temporal envelope fluctuations (or beating) caused by closely spaced harmonics (de Cheveigné, 1999; Summerfield & Assmann, 1991).

The influence of F0 differences between voiced segments of speech has also been examined using longer samples, such as sentences. In one of the earliest such studies, Brokx and Nootboom (1982) measured sentence intelligibility in the presence of competing speech, spoken by the same talker and resynthesized on a monotone. They found that the number of errors in reporting the monotone target speech decreased steadily as the F0 difference between the target and interfering speech was increased from 0 to 3 semitones. A similar benefit of an overall F0 difference between the target and interferer was also observed when the speech contained the F0 fluctuations of normal intonation. More recently, Bird and Darwin (1998) revisited this issue and examined the relative contributions of F0 differences in low- and high-frequency regions. In their control conditions, they too found that increasing the F0 difference between two monotone utterances improved the intelligibility of the target, out to at least 10 semitones. In their second experiment, the target and interfering speech were divided into low-pass and high-pass segments, with a crossover frequency of 800 Hz. The F0s of the target and interferer were manipulated independently in the two spectral regions. For instance, the target might

have an F0 of 100 Hz in the low-frequency region, but an F0 of 106 Hz in the high-frequency region, whereas the interferer might have an F0 of 106 Hz in both regions. In this example, there is an F0 difference between the target and interferer in the low-frequency region, but not in the high-frequency region. In this way, it was possible to determine which frequency region was most important in providing a benefit of F0 difference between the target and interferer. The 800-Hz crossover frequency ensured that resolved harmonics were present in the lower region, but did not rule out the possibility that some resolved harmonics were present in the higher region.

Bird and Darwin's (1998) results showed that the low region was dominant in determining whether a benefit of F0 difference between the target and interferer was observed: When F0 differences were introduced only in the high region, listeners did not benefit from them; when the F0 differences were present only in the low region, the benefit was almost as great as when the F0 differences were present in both spectral regions. However, their results do not imply that F0 information in the high region had no effect. In fact, when the F0s in the low and high regions were swapped (so that the F0 of target in the low region was the same as that of the interferer in the high region, and vice versa), performance deteriorated for F0 differences of 5 semitones or more, suggesting that in normal situations F0 differences and similarities across frequency are used to segregate or group components together, at least when the F0 differences are large. It is not clear whether the same results would have been observed if the high region had been further restricted to contain only unresolved harmonics, which are known to produce a weaker pitch percept, and so may contribute less to grouping by F0.

For nonspeech stimuli, the role of pitch in sequential streaming has been long established. Some early studies used repeating sequences of two pure tones, termed A and B. The sequences were created either by presenting the A and B tones in alternation (ABAB; Miller & Heise, 1950) or in triplets (ABA ABA; van Noorden, 1975). When the frequency separation between the two tones is small, listeners typically hear a single stream of alternating tones (Figure 4, upper panel). When the frequency separation is larger, the percept changes, with the two tones splitting into two separate perceptual streams (Figure 4, lower panel). The exact frequency separation at which the percept changes from one to two streams depends on many

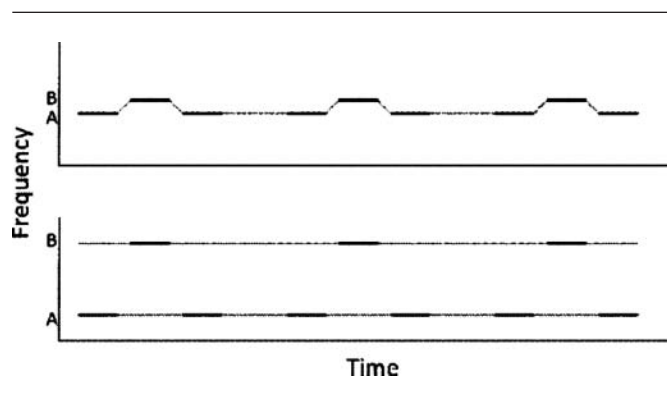


Figure 4. Schematic diagram of the alternating tone sequences used in many perceptual experiments of auditory stream segregation. Upper panel: At small frequency separations, the tones are heard as a single stream with a galloping rhythm. Lower panel: At larger frequency separations, the sequence splits into two auditory streams, each comprising one of the two tone frequencies, as shown by the dashed lines. When segregation occurs, the galloping rhythm is lost, and it becomes difficult to judge the timing between the A and B tones.

parameters, including stimulus and gap durations (Bregman, Ahad, Crum, & O'Reilly, 2000), total duration of the sequence (Anstis & Saida, 1985), exposure to previous sounds (Snyder, Carter, Lee, Hannon, & Alain, 2008), listener attention (Carlyon, Cusack, Foxton, & Robertson, 2001; Snyder, Alain, & Picton, 2006), and experimenter instructions (van Noorden, 1975).

Later work showed that the pattern of results for pure tones is also found for complex tones (Singh, 1987), even if the complex tones are filtered to contain only unresolved harmonics (Vliegen, Moore, & Oxenham, 1999; Vliegen & Oxenham, 1999). This suggests that streaming cannot be based solely on place coding, but that higher-level dimensions, such as pitch and timbre, can also affect perceptual organization—a conclusion that has also been supported by brain-imaging studies of stream segregation (Gutschalk, Oxenham, Micheyl, Wilson, & Melcher, 2007).

Pitch and Source Segregation in Hearing-Impaired People

Many of the paradigms described in the previous section have also been tested in people with hearing impairment. As discussed above, people with hearing impairment often exhibit deficits in pitch perception. Therefore, hearing-impaired people might be expected to experience difficulties with at least

those aspects of source segregation that depend on pitch.

The hypothesis that pitch perception affects simultaneous source segregation was tested directly by Summers and Leek (1998) who measured F0 discrimination for five synthetic vowels, and then measured listeners' ability to identify simultaneous pairs of those same vowels, as well as sentence recognition, as a function of their F0 separation. They found that hearing-impaired subjects with the poorest F0 discrimination thresholds also showed the least benefit of F0 differences in simultaneous-vowel identification. As a group, the hearing-impaired listeners also performed more poorly, and showed less benefit of an F0 difference, than did the normal-hearing group in the competing sentence task. However, the link between F0 discrimination and sentence recognition was not strong on an individual basis. Furthermore, there was some evidence that age played a role, with poorer performance correlating with higher age in both normal-hearing and hearing-impaired groups. A more recent study of double-vowel identification by Arehart, Rossi-Katz, and Swensson-Pruttsman (2005) also found poorer performance in hearing-impaired listeners, and further noted that hearing-impaired listeners were more likely to report hearing just one vowel, again suggesting deficits in their segregation abilities.

Nonspeech streaming in hearing-impaired listeners has received less attention, but some findings have been established. Using pure-tone stimuli, Rose and Moore (1997) found that the two pure tones (A and B) in repeating ABA triplet sequences had to be separated further in frequency for hearing-impaired than for normal-hearing listeners in order for the A and B sequences to be perceived as segregated. This might be attributed either to less salient pitch differences between the tones, or to poorer frequency selectivity, which would lead to less tonotopic (place-based) separation between the peripheral representations of the two tones. Arguing against the hypothesis that the results reflected peripheral frequency selectivity was the finding that listeners with unilateral hearing loss reported similar percepts in both their normal and impaired ears, despite the presumably different degrees of frequency selectivity in the two ears. Studies by Mackersie and colleagues (Mackersie, 2003; Mackersie, Prida, & Stiles, 2001) have established a link between auditory stream segregation of repeating triplet tone sequences and hearing-impaired listeners' performance in a sentence recognition task in the presence of an opposite-sex interfering talker. The results suggest that the ability to

perceive two streams at small frequency differences may correlate with the ability to make use of F0 differences between the male and female talkers.

Finally, just as in normal-hearing listeners (Roberts, Glasberg, & Moore, 2002; Vliegen et al., 1999; Vliegen & Oxenham, 1999), auditory stream segregation has been found to occur in hearing-impaired listeners for stimuli that do not differ in their spectral envelope but do differ in their temporal envelope (Stainsby, Moore, & Glasberg, 2004). This suggests that hearing-impaired listeners are able to use the temporal envelope pitch cues associated with unresolved harmonics in auditory stream segregation.

Because of the reduced availability and poorer coding of resolved harmonics, hearing-impaired people may well rely more on the temporal-envelope cues for pitch perception than do people with normal hearing. Although the evidence suggests that temporal-envelope coding is not necessarily impaired in hearing-impaired listeners, the reliance on such cues may negatively impact hearing-impaired people's ability to segregate sounds in natural environments for a number of reasons. First, temporal-envelope cues provide a weaker pitch cue than resolved harmonics, even under ideal conditions. Second, in real situations, such as in rooms or halls, reverberation smears rapidly fluctuating temporal-envelope pitch cues, meaning that performance will be worse than measured under anechoic laboratory conditions (Qin & Oxenham, 2005). Third, as mentioned above, normal-hearing listeners do not seem able to extract multiple pitches when two stimuli containing unresolved harmonics are presented to the same spectral region (Carlyon, 1996; Micheyl, Bernstein, & Oxenham, 2006), suggesting that hearing-impaired people may experience problems hearing more than one pitch at a time if they are relying solely on unresolved harmonics. This presumed inability to extract multiple pitches (which has not been directly tested yet), may explain at least some of the difficulties experienced in complex environments by people with hearing impairment. This question is addressed further in the section below on masking release.

Findings and Implications for Cochlear Implant Users

To our knowledge there have been no studies that have tested double-vowel recognition in cochlear-implant users. One study has, however, used a noise-excited envelope-vocoder technique, which simulates certain

aspects of cochlear-implant processing in normal-hearing listeners, to investigate the link between F0 perception and the segregation of simultaneous vowels (Qin & Oxenham, 2005). In the noise-vocoder technique, sound is split up into a number of contiguous frequency channels, and the temporal envelope is extracted from each channel, as in a cochlear implant. However, instead of using the channel envelopes to modulate electrical pulses, the envelopes are used to modulate the amplitudes of band-limited noise bands, each with the same center frequency and bandwidth as the original channel filter. After modulation, the noise bands are added together and presented acoustically to normal-hearing listeners. In this way, listeners are presented with an approximation of the original temporal envelopes, but with the spectral resolution degraded (depending on the number and bandwidths of the channels) and the original temporal fine structure replaced by the fine structure of the noise bands. The presentation of temporal-envelope information, without the original temporal fine structure, approximates the processing used in most current cochlear implants.

Qin and Oxenham (2005) found that F0 discrimination was considerably poorer with harmonic complex tones that were processed through an envelope vocoder. More importantly, performance was worsened further still by adding reverberation that approximated either a classroom or a large hall, in line with expectations based on the smearing effects of reverberation on the temporal envelope (Houtgast, Steeneken, & Plomp, 1980). However, in the absence of reverberation, or when the number of channels was relatively large, F0DLs were still less than 6%, corresponding to one musical semitone. Is such sensitivity enough to allow the use of F0 cues in segregating competing speech sounds? Qin and Oxenham tested this question by measuring the ability of normal-hearing listeners to identify of pairs of simultaneous vowels with or without envelope-vocoder processing. They found that no benefits were found for increasing the F0 difference between pairs of vowels, even with 24 channels of spectral information (far greater resolution than that found in current cochlear implants). This suggests that even in the best situations (with high spectral resolution and no reverberation), temporal envelope cues may not be sufficient to allow the perceptual segregation of simultaneous sounds based on F0 cues.

Similar conclusions have been reached using more natural speech stimuli in actual cochlear-implant users, as well as envelope-vocoder simulations. Deeks

and Carlyon (2004) used a six-channel envelope vocoder. Rather than modulating noise bands with the speech envelopes, they modulated periodic pulse trains to more closely simulate cochlear-implant processing. When two competing speech stimuli were presented, performance was better when the target was presented at the higher of two pulse rates (140 Hz, rather than 80 Hz), but there was no evidence that a difference in pulse rates between the masker and target improved the ability of listeners to segregate the two, whether the target and masker were presented in all six channels, or whether the masker and target were presented in alternating channels. Similarly, Stickney et al. (2007) found that F0 differences between competing sentences did not improve performance in real and simulated cochlear-implant processing. Interestingly, they found that reintroducing some temporal fine structure information into the vocoder-processed stimuli led to improved performance overall, as well as a benefit for F0 differences between the competing sentences. Restoring temporal fine structure has the effect of reintroducing some of the resolved harmonics. Although the results are encouraging, it is far from clear how these resolved harmonics (or temporal fine structure in general) can be best presented to real cochlear-implant users (Sit, Simonson, Oxenham, Faltys, & Sarpeshkar, 2007).

A few studies have now investigated perceptual organization in cochlear-implant users using non-speech stimuli. Using repeating sequences of ABA triplets, Chatterjee, Sarampalis, and Oba (2006) found some, albeit weak, evidence that differences in electrode location between A and B, as well as differences in temporal envelope, may be used by cochlear-implant users to segregate sequences of nonoverlapping sounds. In contrast, Cooper and Roberts (2007) suggested that self-reported stream segregation in cochlear-implant users may in fact just reflect their ability to discriminate differences in pitch between electrodes, rather than the automatic streaming experienced by normal-hearing listeners when alternating tones of very different frequencies are presented at high rates. To overcome the potential limitations of subjective reports of perception, Hong and Turner (2006) used a measure of temporal discrimination that becomes more difficult when the tones fall into different perceptual streams. They found a wide range of performance among their cochlear-implant users. Some showed near-normal performance, with thresholds in the discrimination task increasing (indicating worse performance) as the frequency separation between tones increased.

Others showed a very different pattern of results, with performance hardly being affected by the frequency separation between tones. This pattern of results is particularly interesting, because the implant users actually performed relatively better than normal when the tones were widely separated in frequency. Thus, it appears that at least some implant listeners experience streaming, at least with pure tones, whereas others may not. The difference may relate to differences in effective frequency discrimination—if a frequency difference between two tones is not perceived, then no streaming would be expected. Unfortunately, Hong and Turner did not measure frequency discrimination in their subjects. However, they did find a correlation between speech perception in noise or multitalker babble and performance on their streaming task, with better speech perception being related to stream segregation results that were closer to normal. To our knowledge, no streaming studies have been carried out with acoustic stimuli other than pure tones, so it is not clear to what extent F0 differences in everyday complex stimuli (such as voiced speech or music) lead to stream segregation in cochlear-implant users.

Masking Release in Normal, Impaired, and Electric Hearing

The phenomenon of “masking release” in speech perception has been attributed, at least in part, to successful source segregation. Masking release is defined as the improvement in performance when a stationary (steady-state) noise masker is replaced by a different (usually modulated) masker. In normal-hearing listeners, performance almost always improves when temporal and/or spectral fluctuations are introduced into a masking stimulus, either by modulating the noise or by using a single talker as the masker (Festen & Plomp, 1990; Peters, Moore, & Baer, 1998). The ability to make use of brief periods of low masker energy in a temporally fluctuating masker has been termed “listening in the dips” or “glimpsing.” To make use of these dips, it is necessary to know which fluctuations belong to the masker and which belong to the target. This is where source segregation comes into play.

Hearing-impaired listeners generally show much less masking release than normal-hearing listeners (Festen & Plomp, 1990; Peters et al., 1998). This has often been ascribed to reduced spectral and/or temporal resolution in hearing-impaired listeners, leading to lower audibility in the spectral and

temporal dips of the masker. An alternative hypothesis is that the problem lies not with audibility, but with a failure to segregate the target information from that of the masker because of poorer F0 perception (Summers & Leek, 1998) and/or a reduction in the ability to process temporal fine structure cues (Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006). A similar explanation could underlie the general failure of cochlear-implant users to show masking release; in fact, studies of cochlear-implant users (Nelson & Jin, 2004; Nelson, Jin, Carney, & Nelson, 2003; Stickney, Zeng, Litovsky, & Assmann, 2004) and envelope-vocoder simulations (Qin & Oxenham, 2003) have shown that in many cases, performance is actually poorer in a fluctuating background than in a stationary noise, perhaps because the masker fluctuations mask those of the target speech (Kwon & Turner, 2001).

The idea that accurate F0 information from the temporal fine structure of resolved harmonics underlies masking release in normal-hearing listeners was tested by Oxenham and Simonson (2009). The rationale was that high-pass filtering the target and masker to remove all resolved harmonic components should lead to a greater reduction in masking release than low-pass filtering the same stimuli, which would reduce some speech information but would retain most of the information important for F0 perception. The cutoff frequencies for the high-pass and low-pass filters were selected to produce equal performance (proportion of sentence words correctly reported) in the presence of stationary noise. In other conditions, the stationary noise was replaced by a single-talker interferer.

The results did not match expectations. Both low-pass and high-pass conditions produced less masking release than was found in the broadband (unfiltered) conditions. However, contrary to the predictions based on pitch strength and temporal fine structure, the amount of masking release was the same for the low-pass as for the high-pass condition, with no evidence found for superior performance in the low-pass condition with any of the different fluctuating maskers tested. The results of Oxenham and Simonson (2009) suggest that the temporal fine structure information associated with resolved harmonics may not play the crucial role in masking release that was suggested by studies in cochlear-implant users and vocoder simulations. This is not to say that temporal fine structure and accurate pitch coding are not important; they may

represent just one facet of the multiple forms of information, including other bottom-up and top-down segregation mechanisms, that are used by normal-hearing people when listening to speech in complex backgrounds.

Summary

A number of theories have been proposed to explain pitch perception in normal hearing. The three main strands can be described as place, time, and place-time theories. So far, neither experimental nor modeling studies have been able to completely rule out any of these potential approaches. Peripheral frequency selectivity has been shown to be linked to aspects of pitch perception, in line with the expectations of place-based models; however, there are many other aspects of pitch perception that are better understood in terms of timing-based mechanisms. The question is of more than just theoretical value—understanding what aspects of the peripheral auditory code are used by the brain to extract pitch could help in selecting the optimum processing strategy for cochlear implants.

Relatively little is known about our ability to perceive multiple pitches at once, even though this skill is clearly important in music appreciation, and may also play a vital role in the perceptual segregation of competing sounds, such as two people speaking at once. Based on the finding that resolved harmonics seem to be important in hearing more than one pitch from within a single spectral region, it seems likely that both hearing-impaired people and people with cochlear implants would experience great difficulties in such situations. More generally, pitch seems to play an important role in the perceptual organization of both sequential and simultaneous sounds. Better representation of pitch and temporal fine structure in cochlear implants and hearing aids remain an important, if elusive, goal.

Acknowledgments

Dr. Steven Neely provided the code to generate Figure 1. Dr. Marjorie Leek, one anonymous reviewer and the editor, Dr. Arlene Neuman, provided helpful comments on an earlier version of this article. The research described from the Auditory Perception and Cognition Lab was undertaken in collaboration with Drs. Joshua Bernstein, Christophe Micheyl, Michael Qin, and Andrea Simonson, and was supported by the National Institute

on Deafness and other Communication Disorders (NIDCD grant R01 DC 05216).

References

- Anstis, S., & Saida, S. (1985). Adaptation to auditory streaming of frequency-modulated tones. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 257-271.
- Arehart, K. H., Rossi-Katz, J., & Swensson-Prutsman, J. (2005). Double-vowel perception in listeners with cochlear hearing loss: Differences in fundamental frequency, ear of presentation, and relative amplitude. *Journal of Speech, Language, and Hearing Research*, *48*, 236-252.
- Assmann, P. F., & Summerfield, Q. (1987). Perceptual segregation of concurrent vowels. *Journal of the Acoustical Society of America*, *82*, S120.
- Beerends, J. G., & Houtsma, A. J. M. (1989). Pitch identification of simultaneous diotic and dichotic two-tone complexes. *Journal of the Acoustical Society of America*, *85*, 813-819.
- Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*, 1161-1165.
- Bendor, D., & Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology*, *16*, 391-399.
- Bernstein, J. G., & Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *Journal of the Acoustical Society of America*, *113*, 3323-3334.
- Bernstein, J. G., & Oxenham, A. J. (2006a). The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *Journal of the Acoustical Society of America*, *120*, 3916-3928.
- Bernstein, J. G., & Oxenham, A. J. (2006b). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *Journal of the Acoustical Society of America*, *120*, 3929-3945.
- Bird, J., & Darwin, C. J. (1998). Effects of a difference in fundamental frequency in separating two sentences. In A. R. Palmer, A. Rees, A. Q. Summerfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 263-269). London: Whurr.
- Boex, C., Baud, L., Cosendai, G., Sigrist, A., Kos, M. I., & Pelizzone, M. (2006). Acoustic to electric pitch comparisons in cochlear implant subjects with residual hearing. *Journal of the Association for Research in Otolaryngology*, *7*, 110-124.
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.
- Bregman, A. S., Ahad, P., Crum, P. A., & O'Reilly, J. (2000). Effects of time intervals and tone durations on auditory stream segregation. *Perception and Psychophysics*, *62*, 626-636.
- Brox, J. P., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*, 23-36.
- Burns, E. M., & Viemeister, N. F. (1981). Played again SAM: Further observations on the pitch of amplitude-modulated noise. *Journal of the Acoustical Society of America*, *70*, 1655-1660.
- Buss, E., Hall, J. W., & Grose, J. H. (2004). Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. *Ear and Hearing*, *25*, 242-250.
- Cariani, P. A., & Delgutte, B. (1996). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology*, *76*, 1698-1716.
- Carlyon, R. P. (1996). Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *Journal of the Acoustical Society of America*, *99*, 517-524.
- Carlyon, R. P., Cusack, R., Foxton, J. M., & Robertson, I. H. (2001). Effects of attention and unilateral neglect on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 115-127.
- Cedolin, L., & Delgutte, B. (2007). Spatio-temporal representation of the pitch of complex tones in the auditory nerve. In B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, et al. (Eds.), *Hearing: From basic research to applications* (pp. 61-70). New York/Berlin: Springer-Verlag.
- Chatterjee, M., Sarampalis, A., & Oba, S. I. (2006). Auditory stream segregation with cochlear implants: A preliminary report. *Hearing Research*, *222*, 100-107.
- Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *Journal of the Acoustical Society of America*, *98*, 862-879.
- Cooper, H. R., & Roberts, B. (2007). Auditory stream segregation of tone sequences in cochlear implant listeners. *Hearing Research*, *225*, 11-24.
- Dai, H. (2000). On the relative influence of individual harmonics on pitch judgment. *Journal of the Acoustical Society of America*, *107*, 953-959.
- Darwin, C. J., & Carlyon, R. P. (1995). Auditory organization and the formation of perceptual streams. In B. C. J. Moore (Ed.), *Handbook of perception and cognition: Vol. 6. Hearing*. San Diego: Academic.
- de Boer, E. (1956). *On the "residue" in hearing*. Unpublished doctoral dissertation submitted to University of Amsterdam, Netherlands.
- de Cheveigné, A. (1997). Concurrent vowel identification. III. A neural model of harmonic interference cancellation. *Journal of the Acoustical Society of America*, *101*, 2857-2865.
- de Cheveigné, A. (1999). Waveform interactions and the segregation of concurrent vowels. *Journal of the Acoustical Society of America*, *106*, 2959-2972.
- Deeks, J. M., & Carlyon, R. P. (2004). Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation.

- Journal of the Acoustical Society of America*, 115, 1736-1746.
- Dorman, M. F., Spahr, T., Gifford, R., Loisel, L., McKarns, S., Holden, T., et al. (2007). An electric frequency-to-place map for a cochlear implant patient with hearing in the nonimplanted ear. *Journal of the Association for Research in Otolaryngology*, 8, 234-240.
- Evans, E. F. (2001). Latest comparisons between physiological and behavioural frequency selectivity. In J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, & R. Schoonhoven (Eds.), *Physiological and psychophysical bases of auditory function* (pp. 382-387). Maastricht: Shaker.
- Festen, J. M., & Plomp, R. (1990). Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing. *Journal of the Acoustical Society of America*, 88, 1725-1736.
- Fletcher, H. (1940). Auditory patterns. *Reviews of Modern Physics*, 12, 47-65.
- Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, 47, 103-138.
- Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America*, 54, 1496-1516.
- Gutschalk, A., Oxenham, A. J., Micheyl, C., Wilson, E. C., & Melcher, J. R. (2007). Human cortical activity during streaming without spectral cues suggests a general neural substrate for auditory stream segregation. *Journal of Neuroscience*, 27, 13074-13081.
- Heinz, M. G., Colburn, H. S., & Carney, L. H. (2001). Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Computation*, 13, 2273-2316.
- Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *Journal of the Acoustical Society of America*, 120, 360-374.
- Houtgast, T., Steeneken, H. J. M., & Plomp, R. (1980). Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. *Acustica*, 46, 60-72.
- Houtsma, A. J. M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America*, 87, 304-310.
- Kaernbach, C., & Bering, C. (2001). Exploring the temporal mechanism involved in the pitch of unresolved harmonics. *Journal of the Acoustical Society of America*, 110, 1039-1048.
- Kim, D. O., & Molnar, C. E. (1979). A population study of cochlear nerve fibres: Comparison of spatial distributions of average-rate and phase-locking measures of responses to single tones. *Journal of Neurophysiology*, 42, 16-30.
- Kwon, B. J., & Turner, C. W. (2001). Consonant identification under maskers with sinusoidal modulation: Masking release or modulation interference? *Journal of the Acoustical Society of America*, 110, 1130-1140.
- Larsen, E., Cedolin, L., & Delgutte, B. (2008). Pitch representations in the auditory nerve: Two concurrent complex tones. *Journal of Neurophysiology*, 100, 1301-1319.
- Lieberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. *Journal of the Acoustical Society of America*, 63, 442-455.
- Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia*, 7, 128-133.
- Licklider, J. C. R. (1954). "Periodicity" pitch and "place" pitch. *Journal of the Acoustical Society of America*, 26, 945.
- Loeb, G. E., White, M. W., & Merzenich, M. M. (1983). Spatial cross correlation: A proposed mechanism for acoustic pitch perception. *Biological Cybernetics*, 47, 149-163.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences (USA)*, 103, 18866-18869.
- Mackersie, C. L. (2003). Talker separation and sequential stream segregation in listeners with hearing loss: Patterns associated with talker gender. *Journal of Speech, Language, and Hearing Research*, 46, 912-918.
- Mackersie, C. L., Prida, T. L., & Stiles, D. (2001). The role of sequential stream segregation and frequency selectivity in the perception of simultaneous sentences by listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, 44, 19-28.
- Marrone, N., Mason, C. R., & Kidd, G. D., Jr. (2008). Evaluating the benefit of hearing aids in solving the cocktail party problem. *Trends in Amplification*, 12, 300-315.
- McDermott, J. H., & Oxenham, A. J. (2008). Spectral completion of partially masked sounds. *Proceedings of the National Academy of Sciences (USA)*, 105, 5939-5944.
- McKay, C. M., McDermott, H. J., & Carlyon, R. P. (2000). Place and temporal cues in pitch perception: Are they truly independent? *Acoustics Research Letters Online*, 1, 25-30.
- Meddis, R., & Hewitt, M. (1991a). Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, 89, 2866-2882.
- Meddis, R., & Hewitt, M. (1991b). Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. II: Phase sensitivity. *Journal of the Acoustical Society of America*, 89, 2882-2894.
- Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, 102, 1811-1820.
- Micheyl, C., Bernstein, J. G., & Oxenham, A. J. (2006). Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise.

- Journal of the Acoustical Society of America*, 120, 1493-1505.
- Miller, G. A., & Heise, G. A. (1950). The trill threshold. *Journal of the Acoustical Society of America*, 22, 637-638.
- Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 54, 610-619.
- Moore, B. C. J., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In C. J. Plack, A. J. Oxenham, R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception*. New York: Springer.
- Moore, B. C. J., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America*, 77, 1853-1860.
- Moore, B. C. J., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *Journal of the Acoustical Society of America*, 91, 2881-2893.
- Moore, B. C. J., & Rosen, S. M. (1979). Tune recognition with reduced pitch and interval information. *Quarterly Journal of Experimental Psychology*, 31, 229-240.
- Narayan, S. S., Temchin, A. N., Recio, A., & Ruggero, M. A. (1998). Frequency tuning of basilar membrane and auditory nerve fibers in the same cochleae. *Science*, 282, 1882-1884.
- Nelson, P. B., & Jin, S. H. (2004). Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*, 115, 2286-2294.
- Nelson, P. B., Jin, S. H., Carney, A. E., & Nelson, D. A. (2003). Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*, 113, 961-968.
- Nelson, D. A., van Tasell, D. J., Schroder, A. C., Soli, S., & Levine, S. (1995). Electrode ranking of "place pitch" and speech recognition in electrical hearing. *Journal of the Acoustical Society of America*, 98, 1987-1999.
- Oxenham, A. J., Bernstein, J. G. W., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences (USA)*, 101, 1421-1425.
- Oxenham, A. J., & Simonson, A. M. (2009). Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference. *Journal of the Acoustical Society of America*.
- Parsons, T. W. (1976). Separation of speech from interfering speech by means of harmonic selection. *Journal of the Acoustical Society of America*, 60, 911-918.
- Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America*, 59, 640-654.
- Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in non-primary human auditory cortex revealed with fMRI. *Journal of Neuroscience*, 24, 6810-6815.
- Peters, R. W., Moore, B. C. J., & Baer, T. (1998). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. *Journal of the Acoustical Society of America*, 103, 577-587.
- Pfingst, B. E., Holloway, L. A., Poopat, N., Subramanya, A. R., Warren, M. F., & Zwolan, T. A. (1994). Effects of stimulus level on nonspectral frequency discrimination by human subjects. *Hearing Research*, 78, 197-209.
- Pijl, S., & Schwarz, D. W. (1995). Melody recognition and musical interval perception by deaf subjects stimulated with electrical pulse trains through single cochlear implant electrodes. *Journal of the Acoustical Society of America*, 98, 886-895.
- Plack, C. J., & Oxenham, A. J. (2005). Psychophysics of pitch. In C. J. Plack, A. J. Oxenham, A. N. Popper, & R. Fay (Eds.), *Pitch: Neural coding and perception*. New York: Springer.
- Plomp, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America*, 41, 1526-1533.
- Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America*, 114, 446-454.
- Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear and Hearing*, 26, 451-460.
- Rhode, W. S. (1971). Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. *Journal of the Acoustical Society of America*, 49, 1218-1231.
- Ritsma, R. J. (1967). Frequencies dominant in the perception of the pitch of complex sounds. *Journal of the Acoustical Society of America*, 42, 191-198.
- Roberts, B., Glasberg, B. R., & Moore, B. C. (2002). Primitive stream segregation of tone sequences without differences in fundamental frequency or passband. *Journal of the Acoustical Society of America*, 112, 2074-2085.
- Rose, J. E., Brugge, J. F., Anderson, D. J., & Hind, J. E. (1967). Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology*, 30, 769-793.
- Rose, M. M., & Moore, B. C. J. (1997). Perceptual grouping of tone sequences by normally hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 102, 1768-1778.
- Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., & Robles, L. (1997). Basilar-membrane responses to tones at the base of the chinchilla cochlea. *Journal of the Acoustical Society of America*, 101, 2151-2163.
- Scheffers, M. T. M. (1983). Simulation of auditory analysis of pitch: An elaboration on the DWS pitch meter. *Journal of the Acoustical Society of America*, 74, 1716-1725.
- Schouten, J. F. (1940). The residue and the mechanism of hearing. *Proc. Kon. Akad. Wetenschap*, 43, 991-999.

- Schouten, J. F., Ritsma, R. J., & Cardozo, B. L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America*, 34, 1418-1424.
- Shackleton, T. M., & Carlyon, R. P. (1994). The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *Journal of the Acoustical Society of America*, 95, 3529-3540.
- Shamma, S. A. (1985). Speech processing in the auditory system. I: The representation of speech sounds in the responses in the auditory nerve. *Journal of the Acoustical Society of America*, 78, 1612-1621.
- Shamma, S. A., & Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America*, 107, 2631-2644.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303-304.
- Shera, C. A., Guinan, J. J., & Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences Online (USA)*, 99, 3318-3323.
- Shinn-Cunningham, B., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, 12, 283-299.
- Singh, P. G. (1987). Perceptual organization of complex-tone sequences: A trade-off between pitch and timbre? *Proceedings of the National Academy of Sciences Online (USA)*, 82, 886-895.
- Sininger, Y., & Starr, A. (Eds.). (2001). *Auditory neuropathy*. San Diego, CA: Singular.
- Sit, J.-J., Simonson, A. M., Oxenham, A. J., Faltys, M. A., & Sarpeshkar, R. (2007). A low-power asynchronous interleaved sampling algorithm for cochlear implants that encodes envelope and phase information. *IEEE Transactions on Biomedical Engineering*, 54, 138-149.
- Snyder, J. S., Alain, C., & Picton, T. W. (2006). Effects of attention on neuroelectric correlates of auditory stream segregation. *Journal of Cognitive Neuroscience*, 18, 1-13.
- Snyder, J. S., Carter, O. L., Lee, S. K., Hannon, E. E., & Alain, C. (2008). Effects of context on auditory stream segregation. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 1007-1016.
- Srulovicz, P., & Goldstein, J. L. (1983). A central spectrum model: A synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum. *Journal of the Acoustical Society of America*, 73, 1266-1276.
- Stainsby, T. H., Moore, B. C., & Glasberg, B. R. (2004). Auditory streaming based on temporal structure in hearing-impaired listeners. *Hearing Research*, 192, 119-130.
- Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F. G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *Journal of the Acoustical Society of America*, 122, 1069-1078.
- Stickney, G. S., Zeng, F. G., Litovsky, R., & Assmann, P. (2004). Cochlear implant speech recognition with speech maskers. *Journal of the Acoustical Society of America*, 116, 1081-1091.
- Summerfield, A. Q., & Assmann, P. F. (1991). Perception of concurrent vowels: Effects of pitch-pulse asynchrony and harmonic misalignment. *Journal of the Acoustical Society of America*, 89, 1364-1377.
- Summers, V., & Leek, M. R. (1998). F0 processing and the separation of competing speech signals by listeners with normal hearing and with hearing loss. *Journal of Speech, Language, and Hearing Research*, 41, 1294-1306.
- Terhardt, E., Stoll, G., & Seewann, M. (1982). Algorithm for extraction of pitch and pitch salience from complex tonal signals. *Journal of the Acoustical Society of America*, 71, 679-688.
- Tong, Y. C., Blamey, P. J., Dowell, R. C., & Clark, G. M. (1983). Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant. *Journal of the Acoustical Society of America*, 74, 73-80.
- Townsend, B., Cotter, N., van Compernelle, D., & White, R. L. (1987). Pitch perception by cochlear implant subjects. *Journal of the Acoustical Society of America*, 82, 106-115.
- Tsuji, J., & Liberman, M. C. (1997). Intracellular labeling of auditory nerve fibers in guinea pig: Central and peripheral projections. *Journal of Comparative Neurology*, 381, 188-202.
- van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation submitted to Eindhoven University of Technology, Netherlands.
- Vliegen, J., Moore, B. C. J., & Oxenham, A. J. (1999). The role of spectral and periodicity cues in auditory stream segregation, measured using a temporal discrimination task. *Journal of the Acoustical Society of America*, 106, 938-945.
- Vliegen, J., & Oxenham, A. J. (1999). Sequential stream segregation in the absence of spectral cues. *Journal of the Acoustical Society of America*, 105, 339-346.
- von Bismark, G. (1974). Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30, 159-172.
- Wightman, F. L. (1973). The pattern-transformation model of pitch. *Journal of the Acoustical Society of America*, 54, 407-416.
- Zeng, F. G. (2002). Temporal pitch in electric hearing. *Hearing Research*, 174, 101-106.
- Zeng, F. G., Kong, Y. Y., Michalewski, H. J., & Starr, A. (2005). Perceptual consequences of disrupted auditory nerve activity. *Journal of Neurophysiology*, 93, 3050-3063.
- Zwicker, E., Flottorp, G., & Stevens, S. S. (1957). Critical bandwidth in loudness summation. *Journal of the Acoustical Society of America*, 29, 548-557.