Attention, Perception, & Psychophysics
2009, 71 (8), 1701-1715
doi:10.3758/APP.71.8.1701

Pitch perception

William A. Yost
Arizona State University, Tempe, Arizona

This article is a review of the psychophysical study of pitch perception. The history of the study of pitch has seen a continual competition between spectral and temporal theories of pitch perception. The pitch of complex stimuli is likely based on the temporal regularities in a sound's waveform, with the strongest pitches occurring for stimuli with low-frequency components. Thus, temporal models, especially those based on autocorrelation-like processes, appear to account for the majority of the data.

Pitch may be the most important perceptual feature of sound. Music without pitch would be drumbeats, speech without pitch processing would be whispers, and identifying sound sources without using pitch would be severely limited. The study of the perceptual attributes of pitch permeates the history of the study of sound, dating back almost to the beginnings of recorded time. For instance, Pythagoras established the existence of relationships between the length of plucked strings and the octave. The study of pitch perception is the study of the relationships among the physical properties of sound, its neural transforms, and the perception of pitch. The quest for a theory that establishes such a physical–perceptual (psychophysical) relationship is hundreds of years old, and there is still debate concerning what aspects of sound lead to the perception of pitch in the wide variety of contexts in which pitch is a major perceptual attribute of sound.

Sound, Auditory Periphery, and Pitch

Sound can be described in several ways; it is usually defined as comprising three physical properties: frequency, magnitude, and time/phase. The auditory periphery provides neural codes for each of these dimensions, so it would not seem to be very difficult to find a way to relate one or more of these properties to the perception of pitch. But it has been essentially impossible to do so in such a way as to establish a unified account of pitch perception for the wide variety of conditions leading to the perception of pitch. In the elemental case of a simple sound with a single frequency (i.e., a sinusoidal tonal sound), the frequency is its pitch. Even this simple sound has two representations: spectral and temporal. In the spectral domain, the sound is characterized as being a simple spectrum with a single spectral component at a given frequency and with a given magnitude and starting phase. The frequency of the spectral component is the sound's perceived pitch. The sound can also be equivalently represented by a sinusoidal time–pressure waveform. The reciprocal of the period of the waveform is also the pitch of such a simple sound. The time–pressure waveform and the spectrum are inverse functions of each other, in that the spectrum is the Fourier transform of the time–pressure waveform. Thus, one representation (e.g., the time–pressure waveform) can be transformed (via the Fourier transform) into the other representation (e.g., the spectral representation). So, it would appear that it would be difficult to decide between a spectral and a temporal explanation of pitch, in that one explanation is a simple transform of the other. Such a physical reality has complicated the ability to develop theories of pitch perception.

The description of the physical aspects of sounds is not the only basis for considering pitch processing. Sound must pass through the auditory system and, in so doing, is transformed significantly. The processing of sound by auditory mechanisms, especially peripheral structures, alters the representation of sound, and, as a consequence, these alterations affect the ways in which spectra and time–pressure waveforms contribute to pitch perception. At present, theories of pitch processing are based more on the possible neural representation of sound at the output of the auditory periphery than on the purely physical properties of sound. Even so, there remain two classes of theories: spectral and temporal. Testing one type of theory against the other is always complicated by the equivalence of the two views of sound. It is important, therefore, to carefully consider how the transformation of sound as it passes through the auditory periphery affects the neural representation of sound. The conflict between temporal and spectral accounts of pitch can be found in several reviews of pitch perception (Cohen, Grossberg, & Wyse, 1995; de Boer, 1976; Meddis & Hewitt, 1991; Plack, Oxenham, Fay, & Popper, 2005; Plomp, 1976; Yost, 2007).

To appreciate the crucial aspects of the neural peripheral code, consider a sound comprising a series of tones such that each tone has the same magnitude and starting phase and falls in a range from 200 to 8000 Hz in 200-Hz steps (i.e., 200, 400, 600, 800, . . . , 8000 Hz). Figure 1 shows the spectrum and the time–pressure waveform for

W. A. Yost, william.yost@asu.edu

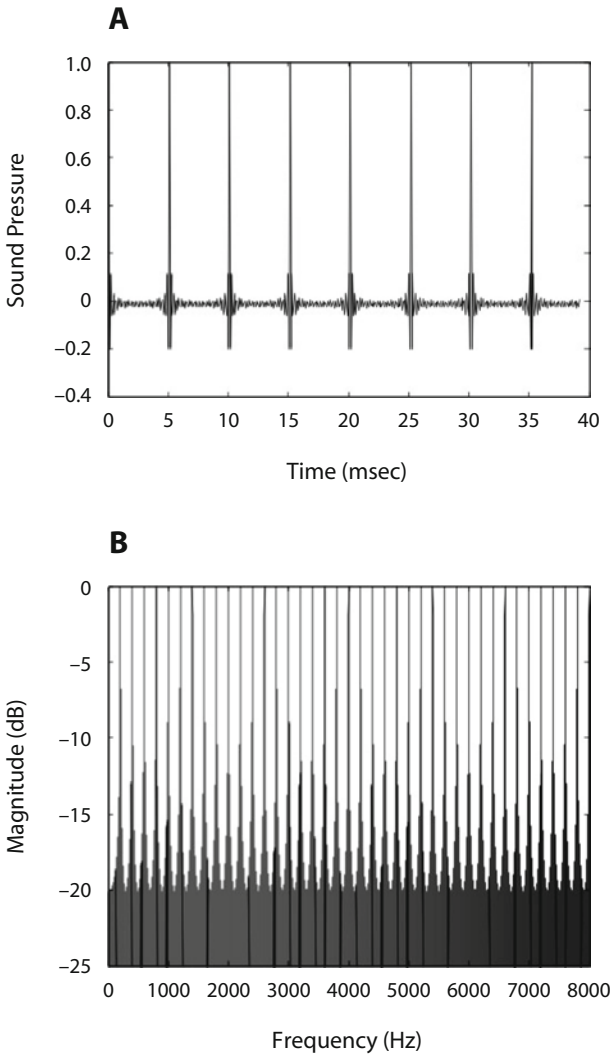

Figure 1. Illustration of the time-domain sound pressure waveform (A) and the amplitude spectrum of a harmonic complex with a 200-Hz fundamental frequency and the first 40 harmonics of 200 Hz (B).

this harmonic complex, with a fundamental frequency (i.e., the first harmonic) of 200 Hz and harmonics 2–40 (2–40 times the fundamental frequency of 200 Hz). The pitch of this sound is 200 Hz. Such harmonic complexes are often used in the study of pitch perception because they represent many everyday sounds, such as speech and musical sounds. That is, an individual sound produced by speech or a single musical tone consists of a fundamental and many harmonics, and the pitch of such everyday sounds is often the frequency of the fundamental (e.g., striking the middle A on a piano keyboard that is tuned to standard concert tuning produces a complex of harmonics with a fundamental frequency of 440 Hz, which is this note's pitch).

A simulation (Patterson, Allerhand, & Giguère, 1995) of the response of auditory nerve fibers to this harmonic complex is shown in Figure 2. The simulation is based on an outer- and middle-ear transform, a gamma-tone filter bank simulating the biomechanical action of the cochlea, and the Meddis hair cell (Meddis, 1986) simulating the transduction of the output of the biomechanical vibration of the cochlea into neural discharges in the auditory nerve. The display shows 100 msec of the neural response along the *x*-axis to a harmonic complex; the *y*-axis represents the output of different auditory nerve fibers as they are arranged from the base (top of the figure) to the apex (bottom of the figure) of the cochlea. Each horizontal line is a simulation of a small group of auditory nerve fibers innervating a similar place along the cochlea. Increases in the height of a line represent the probability of neural firing of this small set of fibers in the same way that a poststimulus time histogram indicates the probability of neural output. The auditory periphery contains a set of elaborate biomechanical–neural processes that function somewhat like a bank of bandpass filters; as a result, each group of nerve fibers, depicted along the *y*-axis, responds selectively to the frequency content of sound. That is, each small group of fibers depicted by a line in Figure 2 is tuned (responds best) to a narrow range of frequencies, and the center frequency of this narrow range increases from base to apex of the cochlea. As a consequence, a large probability of firing in a particular cochlear region signals the

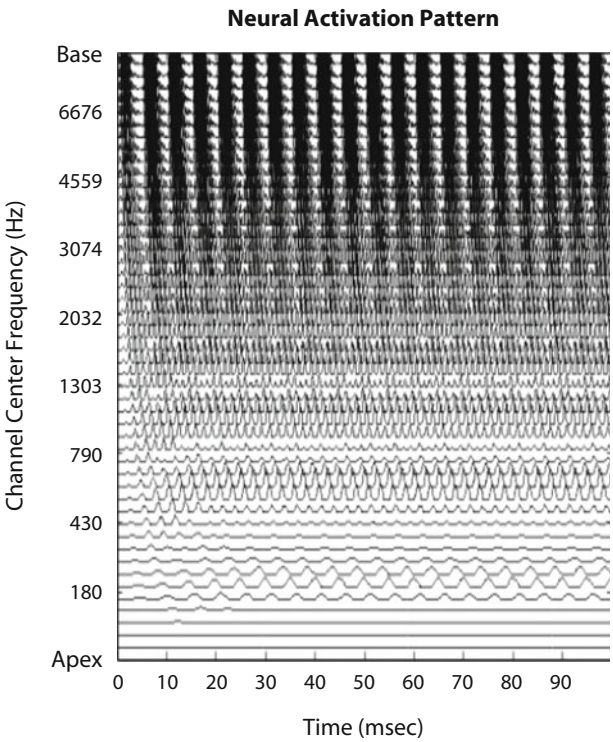

Figure 2. The neural activation pattern from the model of Patterson, Allerhand, and Giguère (1995), depicting a computational simulation of neural excitation in the auditory nerve to a harmonic complex with a 200-Hz fundamental (see Figure 1) low-pass filtered at 8000 Hz. Each line represents 100 msec of the neural firing rate of a small set of auditory fibers tuned to a narrow region of the spectrum. Responses of fibers represented at the bottom of the figure are fibers coming from the apex of the cochlea, and those toward the top are from the base.

presence of a particular range of frequency components in the stimulus (e.g., if most of the neural activity occurs for fibers at the base, the sound contains primarily high frequencies). This is a spectral code for frequency, in that, which nerve fibers are active indicates the spectral content of the stimulating sound.

Figure 3 depicts what is often referred to as a *neural excitation pattern*, or *auditory pattern*. The pattern is calculated by summing the activity for each group of nerve fibers over time and plotting the summed neural activity as a function of the frequency to which each nerve fiber is tuned (i.e., the frequency at which each nerve fiber responds best). Note that, for the lower frequencies, the magnitude of the neural output is modulated according to the frequency to which the fiber is tuned. Fibers tuned to 200, 400, 600, 800, 1000, 1200, and 1400 Hz respond best; fibers tuned to frequencies between these peaks respond less. This pattern indicates that auditory nerve fibers respond selectively to the frequency components of the 200-Hz harmonic complex, as shown in Figure 1B. The auditory nerve contains a spectral code.

The resolution of this code is limited, in that the auditory nerve does not continue to preserve the delineation of the spectral information above about 2000 Hz for this 200-Hz harmonic complex. This is because the resolution of the biomechanical processors decreases with increasing frequency (the width of the bandpass filters increase with increasing center frequency). In the low-frequency region where resolution is high, a single low-frequency auditory nerve fiber responds to only a few spectral components. At higher frequencies, a single high-frequency fiber responds almost equally to a fairly wide range of frequencies in a harmonic complex. Low-frequency fibers (toward the apex) can resolve small spectral differences, and high-frequency fibers (toward the base) cannot. What constitutes low and high frequencies is relative, since the spectral resolution of the auditory periphery is proportional to frequency: The higher the frequency, the poorer the spectral resolution. For a harmonic complex, harmonics above about the 10th are no longer resolved by the auditory periphery. Thus, 2000 Hz is the 10th harmonic of a 200-Hz fundamental, and the difference between 2000 Hz (10th harmonic) and 2200 Hz (11th harmonic) would probably not be resolved by the auditory periphery. However, 2000 Hz is the 4th harmonic of a 500-Hz fundamental, and the difference between 2000 Hz (4th harmonic) and 2500 Hz (5th harmonic) could probably be resolved by the auditory periphery; but the difference between 5000 Hz (10th harmonic of 500 Hz) and 5500 Hz (11th harmonic) could not be resolved.

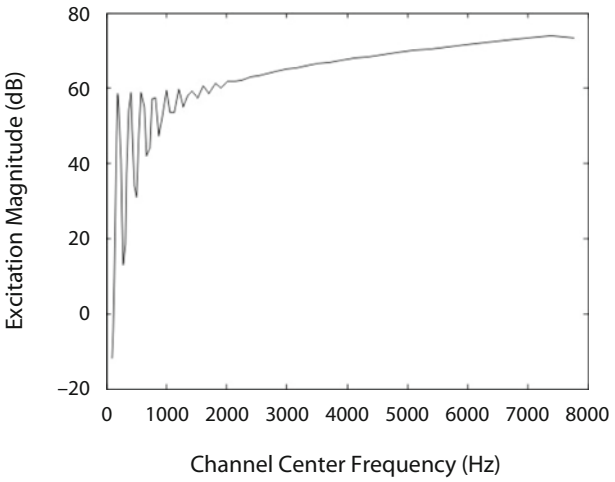

Figure 3. The neural excitation pattern or auditory spectrum of the output of the auditory periphery plotting summed neural excitation (in dB) as a function of the center frequency of the auditory tuned channels. The pattern is computed by summing across time for each channel shown in Figure 2. The neural excitation in the low-frequency channels (fibers) reflects the neurally resolved spectral peaks in the 200-Hz fundamental frequency harmonic complex.

A sound's amplitude varies over time in two general ways: There are relatively fast, cycle-by-cycle amplitude changes that constitute the fine structure of the waveform, and these fine-structure changes can also have slow overall changes in amplitude referred to as the *amplitude-modulated envelope of the waveform*. The entire sound pressure waveform can be expressed as the product of the fine-structure and envelope terms. The neural output also follows the temporal structure of the time–pressure waveform in two ways. Auditory nerve fibers can discharge in a phase-locked or synchronous manner to the cycle-by-cycle fluctuations (fine structure) in the time–pressure waveform. As the frequency to which the fibers respond best increases, the stimulus's period decreases, as does the period of the neural synchronous firing. For the 200-Hz harmonic complex, the periods of these periodic synchronous neural responses at the first four harmonics are integer divisions of 5 msec (e.g., 5 msec = 1/200 Hz; 2.5 msec = 1/400 Hz; 1.25 msec = 1/800 Hz; 0.625 msec = 1/1600 Hz). However, several properties of neural function (e.g., refractory period, and the low-pass property of neural transduction) limit the frequencies to which the auditory nerve fibers can reliability synchronize to the temporal fine structure of the input sound. If the fine structure varies more than about 5,000 times per second (5000 Hz, or a period of 0.2 msec), the auditory nerve cannot "keep up" and does not preserve a code for the cycle-by-cycle fine structure in the time–pressure waveform. Thus, the peripheral neural code captures the fine-structure fluctuations of an input sound, but the upper limit of the ability of auditory nerve fibers in humans to synchronize to the temporal fine-structure period of the sound is about 5000 Hz.

It was mentioned above that, for the 200-Hz harmonic complex, the high-frequency fibers respond to several of the spectral components (harmonics). If one investigates the sum of several harmonics of 200 Hz, one would determine that the time–pressure waveform of this sum has an overall amplitude that is modulated at 200 Hz (i.e., with a 5-msec period). This amplitude modulation pattern is the temporal envelope of the sound (see Figure 1A, and note that the overall neural magnitude at high frequencies has an envelope with a period of 5 msec). Thus, the output of a filter that would pass several harmonics of 200 Hz would

appear to have a 200-Hz amplitude-modulated envelope with a period of 5 msec. That is, these high-frequency fibers respond to a sum of the components of the harmonic complex that fall within the broad spectral bandwidth of such high-frequency fibers. Modulation in neural firing rate can code for a sound's temporal envelope, and the reciprocal of the period of this envelope is often the perceived pitch of the stimulus.

Thus, Figure 2 suggests three codes that might provide information about pitch: a spectral code indicating the ability of the auditory periphery to resolve a sound's spectral components, a synchronous firing code associated with the fine structure of low-frequency components, and a code related to the modulation in firing rate associated with the amplitude-modulated envelope fluctuations of a stimulus. Current models of pitch perception are based on one or more of these peripheral codes.

Measuring Pitch

In addition to having an appreciation of the biomechanical and neural transforms that a sound undergoes, it is also useful to note the various ways in which pitch is defined and measured. The national and international standard definition of *pitch* (American National Standards Institute, 1978) is "that subject dimension of sound that orders sound from low to high." This definition has been measured operationally in several ways. The most general means of measuring the pitch of a test sound is by use of a pitch-matching procedure in which the frequency or repetition rate of a comparison sound (e.g., a sinusoidal tone or a periodic train of clicks) is varied so that the listener judges the pitch of the comparison stimulus to be the same as that of the test stimulus. The frequency of the tone or the reciprocal of the period of repetition of a train of pulses (either expressed in Hz) is used as the *matched pitch* of the test stimulus. That is, if a listener adjusts the frequency of a comparison tone to be 200 Hz when this tone has the same perceived pitch of a test stimulus, the test stimulus has a 200-Hz pitch.

Pitch is also measured in a musical context. In standard equal-temperament tuning, 12 notes are arranged in a logarithmic spacing within a musical octave (i.e., an interval bounded by two pitches, the higher of which is twice the frequency of the lower pitch), in what a musician would refer to as a *chromatic scale*, with each note given a letter name (e.g., A, B♭, B, C, . . .). The logarithmic interval between any two adjacent notes (e.g., A–B♭) is called a *semitone*, and there are 12 semitone intervals in an octave. Each semitone interval comprises 100 cents (with 1,200 cents per octave). Thus, an individual note is measured in terms of its frequency (Hz), and the difference between two notes (i.e., an interval) can be reported as a frequency difference or in terms of octaves, semitones, and cents. The perception of pitch that supports melody recognition and the identification of musical intervals is an aspect of a person's sense of musical pitch, often considered the strongest form of pitch perception. That is, a sound might have a matched pitch obtained in a matching procedure, but variations from the matched pitch of that sound may prevent a listener from reliably judging a musical melody or interval.

The *mel scale* uses magnitude estimation scaling to measure pitch (see Stevens, Volkmann, & Newman, 1937). A standard 1000-Hz tone presented at a sound pressure level (SPL) of 40 dB is defined as having a pitch of 100 mels. A sound judged to have a pitch that is twice that of this standard has a pitch of 200 mels, a pitch of one half that of the standard has a pitch of 50 mels, and so forth. Currently, the mel scale is rarely used in the study of pitch perception.

Most of the work on pitch perception centers on complex sounds. As is indicated above, the pitch of a simple sound (e.g., a sinusoidal sound) is its frequency. It might be that a spectral code is responsible for the pitch of high-frequency tonal signals and that the period of the phase-locked synchronous firing to the period of the tone is the code used for pitch at low frequencies (Moore, 1993). Only a spectral code could be used to account for the pitch of tones with frequencies above 5000 Hz, since there is no phase-locked neural response to the tone's periodic vibrations above 5000 Hz. A simple tonal sound has a flat envelope with no amplitude modulation, so envelope cues are not applicable for simple tonal sounds. On the other hand, the tuning of auditory nerve fibers to low frequencies is not narrow enough to account for the small difference thresholds for discriminating a change in frequency (pitch). That is, at 250 Hz, a 1-Hz difference is discriminable, but neural tuning is not sharp enough to discern a 1-Hz difference. It is possible to imagine that a difference of 0.016 msec, which is the difference in the period between a 250-and a 249-Hz tone, could form the basis for determining a pitch difference. This argument and other data have suggested to several investigators (see Moore, 1993) that the pitch of a simple sound is based on two mechanisms: spectral processes at high frequencies and temporal processes at low frequencies.

A Historical Perspective

At this point, it is important to gain a historical perspective of the study of pitch perception. In the mid-19th century, the notion that the auditory system performed a limited Fourier series transform of sound (Ohm's acoustic law) suggested that the pitch of a sound would most likely be that of the lowest (because of spectral resolution) or most prominent spectral component. However, a stimulus such as the 200-Hz harmonic complex described above produces a 200-Hz pitch, even when the 200-Hz fundamental is removed from the stimulus. That is, the pitch of a missing-fundamental harmonic complex is still that of the fundamental. In fact, the pitch of the 200-Hz harmonic complex remains at 200 Hz, even when many of the lower (resolved) harmonics are removed. This is at odds with a spectral explanation based on the frequency of the lowest component, since the perceived pitch is not that of the lowest component for these missing-fundamental-type stimuli. Thus, the pitch of the missing-fundamental stimulus generated a significant challenge to the prevailing spectral view of pitch processing.

Nonlinear processing occurring as a result of cochlear transduction offers one possibility for the pitch of missing-fundamental stimuli being the fundamental, even when it is absent. A nonlinear system, such as the auditory periphery, produces a distortion product at a frequency equal to the difference in the frequencies of the spectral components of the originating sound. For instance, a nonlinear system would respond to a stimulus consisting of the second through sixth harmonics of a fundamental frequency of 200 Hz (i.e., 400, 600, 800, 1000, and 1200 Hz) by producing a distortion product (which could be audible) at 200 Hz (e.g., the frequency difference between adjacent harmonics). This 200-Hz distortion product could be the basis of the perceived 200-Hz pitch for the missing-fundamental stimulus. Over the years, several experiments (e.g., Licklider, 1954; Patterson, 1969) have shown that the pitch of the missing fundamental is not due to nonlinearities associated with cochlear transduction. Although the pitch of the missing fundamental is most likely not a result of nonlinear peripheral processing, nonlinear distortion products may contribute to the perceived pitch of other complex sounds, and, therefore, controls are often required to eliminate the perception of distortion products as the basis of pitch processing (see Pressnitzer & Patterson, 2001, for a recent study of issues related to nonlinear distortion and pitch perception).

Although the spectrum of the 200-Hz missing-fundamental stimulus has no energy at 200 Hz, the envelope of the sound has a 5-msec (reciprocal of 200 Hz) periodicity. Schouten (1938, 1940) pointed out that high-frequency fibers that could not resolve the spectral structure do produce periodic modulation of spike rate, with a period equal to the fundamental frequency (as shown in Figure 2). Thus, Schouten suggested that, for low frequencies at which the auditory periphery could resolve the harmonics of a missing-fundamental stimulus, a spectral account might be sufficient, but otherwise, one might need to use some form of temporal analysis for the high-frequency fibers to account for the pitch of harmonic sounds. Schouten referred to the pitch associated with unresolved spectral components (i.e., in the higher frequency regions) as a *residue pitch*.

A variation of the missing-fundamental stimulus introduced another important aspect of the pitch of complex sounds. This complex stimulus, first described by Schouten (1940) and later by de Boer (1956, 1961), is generated by shifting each component of a harmonic complex by a constant frequency (e.g., for the 200-Hz harmonic complex shifting each component by, for example, 40 Hz, producing spectral components at 440, 640, 840, 1040 Hz, etc.). The matched pitch of this sound is approximately 208 Hz, and the pitch of such frequency-shifted harmonic complexes is often referred to as the *pitch shift of the residue*. Although the frequency spacing between each component is 200 Hz, the components are harmonics of 40 Hz (but the components are the 6th, 11th, 16th, etc., harmonics of 40 Hz), and the envelope does not have a period equal to the reciprocal of 208 Hz. Figure 4 shows the matched pitch of various harmonic complexes with 200-Hz spacing of the components as a function of the frequency of the lowest component in the sound (in all cases, the other components in the sound are 200-Hz increments of the lowest component shown in Figure 4). The matched pitch is 200 Hz when the lowest component is a harmonic of 200 Hz (the complexes with the lowest components of 600, 800, and 1000 Hz are missing-fundamental cases). When the lowest com-

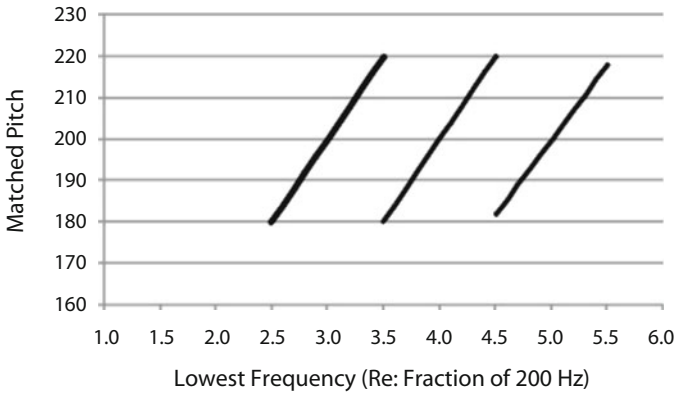

Figure 4. The matched pitch of 12-tone complexes producing the pitch shift of the residue. The lowest frequency of the tonal complex is shown on the *x*-axis in terms of fractions of 200 Hz (i.e., the frequency of the lowest component in the complex is the *x*-axis value multiplied times 200 Hz). The spacing between the 12 components in the complexes is always 200 Hz. The data are plotted using the slope estimates provided by Patterson (1973). Thus, the first point on the left represents the stimulus condition in which the lowest frequency component was 500 Hz (2.5 times 200 Hz), and the other 11 components ranged from 700 to 2700 Hz, in 200-Hz increments. This stimulus produces, on average, a pitch shift of the residue of 180 Hz.

ponent is an odd harmonic of 100 Hz (e.g., 700, 900, or 1100 Hz), two pitches may be matched to this pitch shift of the residue stimulus. The pitch shift of the residue was and is a difficult pitch for any of the current theories of pitch to account for, regardless of whether they are spectral or temporal theories.

There is one more concept of significant historical importance. Ritsma (1962) and Plomp (1967) conducted experiments indicating that the harmonics contributing the most to the pitch of complex harmonic sounds were in the region of the second to fifth harmonic. These second to fifth harmonics of a harmonic complex were dominant in determining the pitch of complex harmonic sounds, and this dominance region plays an important role in the development of theories of pitch perception. Different data sets (Moore, Glasberg, & Peters, 1985; Patterson & Wightman, 1976) suggest that which of the lower five or so harmonics are dominant depends on several stimulus conditions. Thus, the dominance region is in the general spectral region of the fifth harmonic or lower.

Models of Pitch Perception

The missing-fundamental stimulus, the pitch shift of the residue stimulus, and the existence of a dominance region provide enough information to enable a general description of the key features of the two types of current models of pitch perception. There is a spectral approach and a temporal approach to models of pitch perception, and each uses as its input the output of the auditory periphery, as Figure 2 shows.

Spectral modeling of pitch perception. The main proponents of the spectral approach are Goldstein (1973); to some extent Wightman (1973); Terhardt (1974); Cohen, Grossberg, and Wyse (1995); and Shamma and Klein (2000). Several of these models propose that pitch is extracted from the spectral components resolved by the auditory periphery (as shown in Figure 3). Most of these models suggest that some sort of process "fits" a harmonic structure to the resolved harmonics and that the fundamental of this fitted harmonic structure is the predicted pitch of the complex. This idea is depicted in Figure 5A. The sound is a 200-Hz fundamental harmonic complex, and Figure 5A depicts the neural excitation pattern for this stimulus (see Figure 3). The solid vertical lines represent the best-fitting harmonic structure to this excitation pattern. The solid lines are a representation of a 200-Hz fundamental structure, so the predicted pitch is 200 Hz and remains 200 Hz, even if many lower frequency components of the 200-Hz stimulus are missing. That is, this form of a spectral approach correctly predicts the pitch of missing-fundamental-type stimuli.

Figure 5B depicts the excitation pattern of a pitch shift of the residue stimulus condition, in which the stimulus components were 440, 640, 840, 1040 Hz, and so forth. The dominance region for this stimulus would be the excitation peaks at 440, 640, 840, and 1040 Hz (in the approximate spectral region of 2–5 times the 200-Hz fundamental shown in the circle). These dominant region peaks are shown as the dotted lines of the excitation pattern. The dotted line is the harmonic structure best fitting the peaks at 440, 640, 840, and 1040 Hz (i.e., in the dominance region). The fundamental of this best-fitting dotted line is about 208 Hz (i.e., the matched pitch of this pitch shift of the residue stimulus). The spectral template matching models each provide quantitative predictions for the pitch of complex sounds, and each is based on several additional assumptions and model components.

A series of experiments involving mistuning one harmonic in an otherwise tuned harmonic complex has been used to measure the resolution of the harmonic structure (*harmonic sieve*) required to determine the pitch of a complex stimulus from its resolved harmonics (see Lin & Hartmann, 1998; Moore, Peters, & Glasberg, 1986). If one of the harmonics of a harmonic complex is mistuned by 8% or more (e.g., the third harmonic of 200 Hz is mistuned to 648 Hz rather than to 600 Hz), the mistuned harmonic is perceived as a separate tone (pitch) from the complex pitch determined by the remaining tones. The mistuned harmonic "pops out" as a separate pitch from the complex pitch of the harmonic series. In the example, listeners would perceive a 648-Hz pitch and a pitch near 200 Hz. The complex pitch attributed to the remaining harmonic components (of 200 Hz in the example) is slightly altered when a harmonic is mistuned in this way (the pitch could be approximately 205 Hz). Thus, if the harmonics are within 8% of an integer multiple of the fundamental, the harmonics are fused as part of the complex sound whose spectral structure may be used to account for the sound's pitch. Thus, the resolution of the harmonic structure of resolved harmonics is approximately 8% of the fundamental.

Temporal modeling of pitch perception. Spectral modeling that is based on the peripherally resolved peaks in the neural excitation pattern that are in the dominance region for pitch can account for some of the key pitch perception data. There is a temporal explanation as well. The modern-day version of the temporal approach uses autocorrelation, as was originally suggested by Licklider (1951). Meddis and colleagues (Meddis & Hewitt, 1991; Meddis & O'Mard, 1997) have proposed the most fully developed versions of autocorrelation to account for the pitch of complex stimuli, and researchers, such as Slaney and Lyon (1993); Yost, Patterson, and colleagues (e.g., Yost, Patterson, & Sheft, 1996); and Bernstein and Oxenham (2005, 2008), have added refinements to autocorrelation-like models.

Autocorrelation, $A(\tau)$, is a transform of the spectrum (autocorrelation is the Fourier transform of the power spectrum), but can also be defined or described in the time domain as $A(\tau) = \int x(t)x(t + \tau)\, dt$, where $x(t)$ is a time-domain waveform and τ is temporal lag. That is, an original time pattern, $x(t)$, is time shifted (by τ) and multiplied times the original pattern, and the products are integrated (summed). The integrated product is determined as a function of the time shift (τ, lag) between the original and time-shifted pattern, and this forms the autocorrelation function. The normalized autocorrelation function is the integrated products divided by the autocorrelation value obtained when the lag is 0 (i.e., multiplying the original

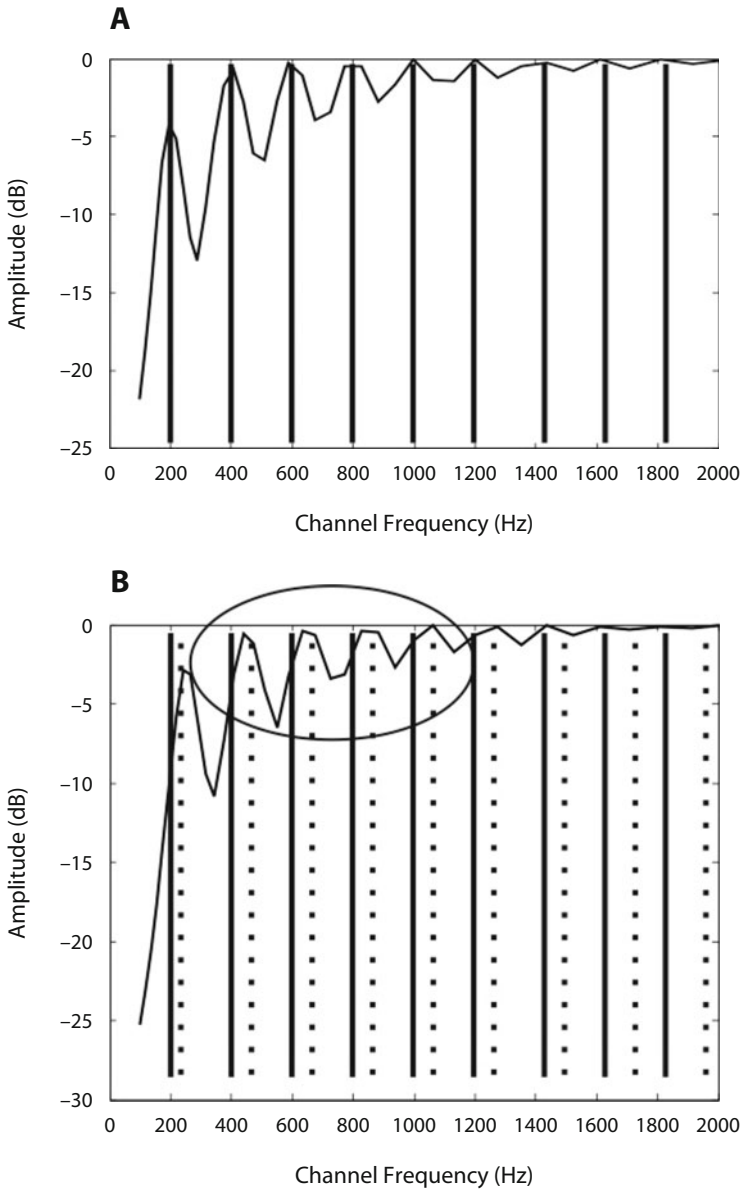

Figure 5. Examples of how spectral models use the possible harmonic structure in a complex stimulus to account for the pitch of complex sounds. The neural excitation pattern (see Figure 3) of a 200-Hz complex is shown in panel A, and a harmonic sequence with a 200-Hz fundamental (the vertical lines) fits the peaks in the excitation pattern; thus, the predicted pitch would be 200 Hz. In panel B, the neural excitation pattern of a pitch shift of the residue stimulus consisting of a 200-Hz fundamental harmonic complex with each component shifted 40 Hz is shown. The solid lines are the fit of a harmonic sequence with a 200-Hz fundamental, and the dotted lines are the fit of a 208-Hz-fundamental harmonic sequence. In the dominance region for pitch, shown in the circle (second to fifth harmonics), the 208-Hz-fundamental sequence provides the best fit to the peaks in the neural excitation pattern, leading to the prediction that this pitch shift of the residue stimulus would have a 208-Hz pitch.

pattern times itself). This temporal description of the autocorrelation of the timing pattern of auditory nerve fibers is used as a temporal model.

Figure 6A shows an autocorrelogram of the data of Figure 2. That is, an autocorrelation transform was applied to each frequency channel shown in Figure 2. The result produces a pattern of activity in which there is a high correlation at 5 msec (the reciprocal of the 200-Hz fundamental) and its integer multiples. Figure 6B shows a summary autocorrelogram obtained by summing the autocorrelogram (Figure 6A) across frequency channels. Note that the first major peak (major peak at the shortest lag that is not 0 msec) in the

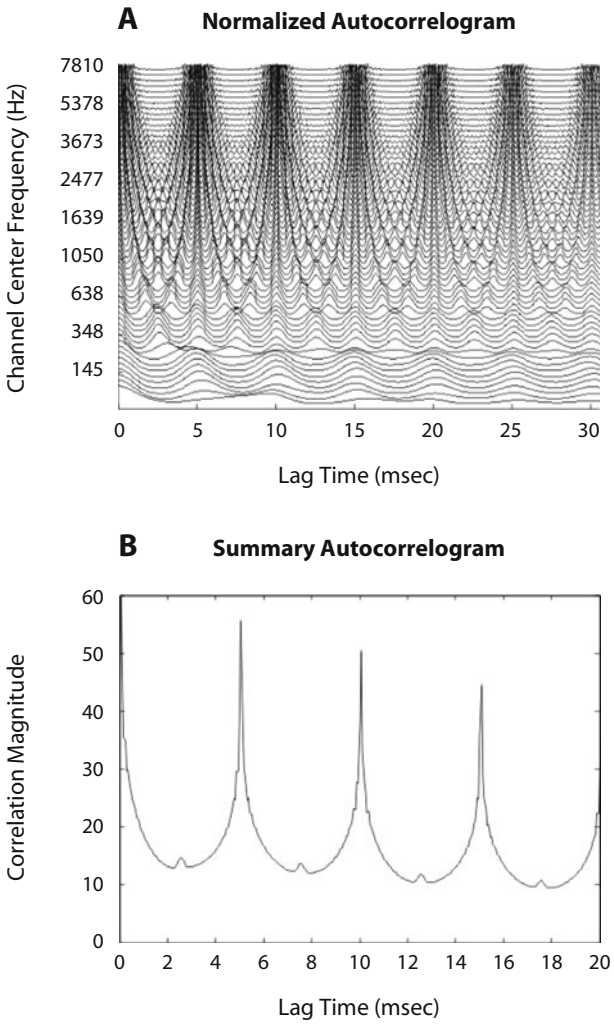

Figure 6. The autocorrelogram (A) for a 200-Hz-fundamental harmonic complex computed by taking the autocorrelation in each channel of the neural activation pattern shown in Figure 2. The high correlation every 5 msec is used to predict a 200-Hz pitch (reciprocal of 5 msec) for this stimulus. Panel B depicts the summary autocorrelogram for the 200-Hz-fundamental harmonic complex obtained by summing the correlations across channels for the autocorrelogram shown in panel A.

summary autocorrelogram is at a lag of 5 msec, the reciprocal of the pitch of the harmonic complex. This peak remains at 5 msec, even when the lower harmonics of the harmonic complex are removed. Thus, the autocorrelation model accounts for the pitch of missing-fundamental stimuli.

Figure 7 shows a summary autocorrelogram for a pitch shift of the residue stimulus consisting of 240, 440, 640, 840, 1040, 1240 Hz, and so forth, and this stimulus was filtered to emphasize the dominance region of 440–1040 Hz (i.e., the region of the second to fifth harmonic). Note that the first peak (at the dotted line) in the summary autocorrelogram is at 4.8 msec, approximately the period of 208 Hz, the reported pitch of this pitch shift of the residue stimulus. Thus, an autocorrelation model of temporal processing can account for the basic pitch perception data about as well as a spectral pattern model can.

Challenges to both the spectral and temporal models of pitch perception. The spectral pattern models predict a reduced ability to process pitch for conditions in which stimuli are high-pass filtered, resulting in only high-frequency stimuli. When a stimulus contains only high frequencies such that the spectral structure of the complex sound would be unresolved by the auditory periphery, spectral models would predict that such stimuli would have no perceived pitch. Temporal models, such as autocorrelation, can still provide pitch predictions for such high-frequency stimuli. Such high-frequency stimuli do have a weak but measurable pitch, one that can even allow one to recognize musical melodies. So, the pitch of high-frequency stimuli, for which the spectral structure of the stimulus is unresolved by the auditory periphery, can be accounted for only by temporal-based models, such as autocorrelation. However, it is important to recognize that pitch strength for such high-frequency stimuli is weak when they are compared with low-frequency conditions in which the spectral structure of the stimulus is spectrally resolved.

Pitch strength (also called *pitch saliency*) refers to the strength of the perceived pitch of a complex sound as compared with the overall perceived quality (timbre) of the sound (see Patterson, Yost, Handel, & Datta, 2000). Pitch strength is a relative term and is not the same as the loudness of the sound. A harmonic complex is usually perceived as having a "tinny" timbre along with a pitch that is often the same as the fundamental frequency of the complex. Pitch strength can be considered to be the perceived difference in the strength of the pitch sensation relative to that of the timbre of the sound. Pitch strength for a harmonic com-

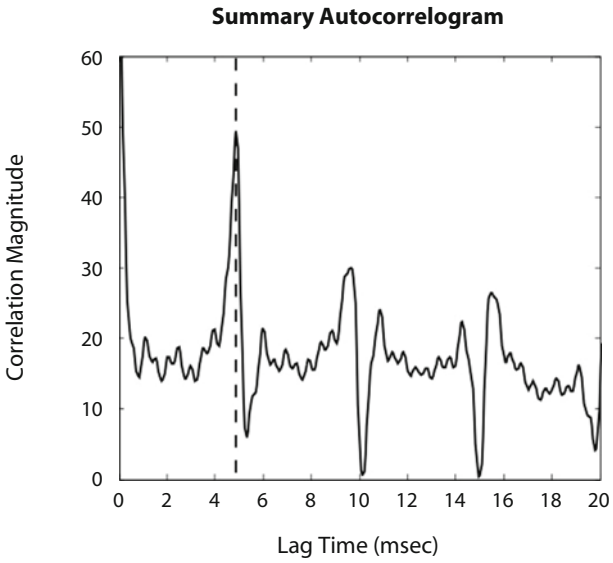

Figure 7. An estimated summary autocorrelogram (see Figure 6A) for the pitch shift of the residue stimulus used for Figure 5. Before the autocorrelogram was generated, the complex stimulus was bandpass filtered between 200 and 1000 Hz, to emphasize the dominance region. The peak in the summary autocorrelogram at a lag of 4.8 msec (dotted vertical line) would lead to a prediction of a 208-Hz pitch (208 Hz is the approximate reciprocal of 4.8 msec) for this pitch shift of the residue stimulus.

plex remains relatively invariant with changes in overall sound level (loudness), but weakens as more and more of the lower harmonics are removed (e.g., filtered) from the stimulus. That is, the pitch of a harmonic stimulus that is high-pass filtered is weak, compared with the overall timbral quality of the high-pass-filtered sound.

The pitch associated with amplitude-modulated noise is another example of the difficulty that a spectral model has in accounting for the pitch of complex sounds. A broadband amplitude-modulated noise has a random spectral structure, without any spectral features that can produce a pitch. Yet, sinusoidally amplitude-modulated noise has a weak pitch that can support melody recognition (Burns & Viemeister, 1981). The perceived pitch is the reciprocal of the period of the amplitude modulation. If amplitude-modulated noise is first passed through a nonlinear process as exists in the auditory periphery, the autocorrelation function of such modulated noises would contain a peak at a lag corresponding to the reciprocal of the perceived pitch (reciprocal of the period of amplitude modulation). Thus, an autocorrelation model can account for the pitch of amplitude-modulated noise.

The envelope of a sound cannot be used by itself to account for the different pitch shifts of the residue. The envelope is the same for different pitch shifts of the residue stimuli, but their perceived pitches differ (e.g., the 200-Hz pitch of a 400-, 600-, 800-, 1000-, 1200-Hz complex and the 208-Hz pitch of a 440-, 640-, 840-, 1040-, 1240-Hz complex have the same envelope). If a sound has only frequencies above the limit where phase locking no longer occurs (above approximately 5000 Hz), amplitude-modulated complex sounds can have a weak pitch, but the pitch shift of the residue does not change as a function of the same conditions that yield pitch differences at low frequencies (see Yost, Patterson, & Sheft, 1998). This result is consistent with the pitch of these high-frequency sounds being determined by the envelope, but, because the pitch of high harmonics is based on the envelope, a listener no longer perceives the pitch shift of the residue. It is also possible to generate low-frequency complex sounds with very low pitches (e.g., 50 Hz), such that the resolved spectral components of the complex sound are less than 5000 Hz (i.e., the sound's energy is in a region where phase locking of auditory nerve fibers can occur), but the spectral structure of the complex sound is unresolved (i.e., the 10th harmonic of 50 Hz is 500 Hz, so, for a 50-Hz fundamental complex, spectral differences above 500 Hz would not be resolved). In this case, it is still possible to perceive the pitch shift of the residue. This is consistent with the pitch of these sounds being determined by temporal fine structure information (as coded by the phased-lock responses of auditory nerve fibers). Autocorrelation-like approaches can describe both the pitch (including the pitch shift of the residue) due to fine-structure processing and the pitch of amplitude-modulated, high-frequency sounds when the pitch shift of the residue phenomena do not occur (Yost et al., 1998). Thus, temporal models, such as those based on autocorrelation, can account for more of the pitch data than can any other form of the proposed models of pitch perception.

Although autocorrelation has been successful in accounting for a large set of pitch perception data, there have been challenges to this approach. Autocorrelation is temporally symmetric—that is, a stimulus with a rising level (ramped; see Figure 8A) would have the same long-term autocorrelation function as would the same stimulus but with a declining level (damped; see Figure 8B). Yet, several aspects of pitch perception and other aspects of auditory perception differ significantly for ramped versus damped stimuli. Irino and Patterson (1996) and Patterson and Irino (1998) have argued that, although autocorrelation cannot account for data using ramped/damped stimuli, other mechanisms that extract the temporal fine structure (temporal regularity) of these stimuli can account for the results.

Kaernbach and Demany (1998) also argued that autocorrelation could not account for the perception of click-train stimuli arranged with particular temporal regulari-

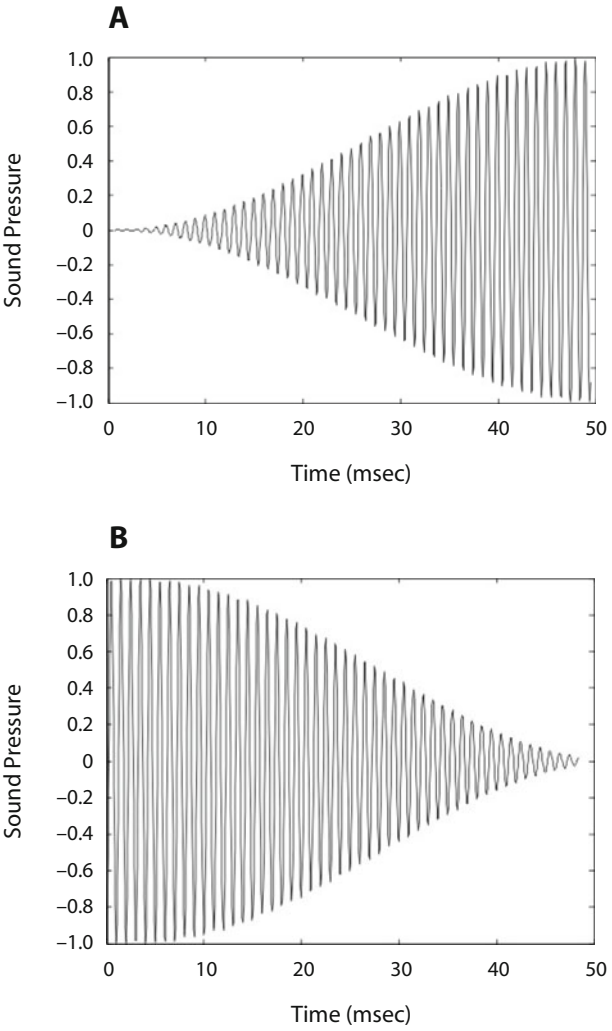

Figure 8. Ramped (A) and damped (B) sinusoids. The pitch percepts of these stimuli differ noticeably from one another. An autocorrelation analysis of the entire waveforms would be the same for both ramped and damped stimuli.

ties. They set up click trains arranged in three-, four-, or five-click sequences, such that the interval between the first two clicks in each three-click sequence was fixed at a constant duration, and the duration of each remaining interval was random. A three-click sequence is referred to as a *kxx stimulus*, where *k* indicates the interval with fixed duration and *x* indicates an interval of random duration (e.g., for a *kxx* click train, |10|7|15|10|12|9|10|4|11| . . . , the *k* interval is 10 msec [underlined] and the *x* intervals are of random durations). In a different arrangement of three-click sequences, which Kaernbach and Demany called an *abx sequence*, the duration of interval *a* was random, the sum of the durations of intervals $a + b$ was fixed, and the duration of the third interval, *x*, was random (e.g., |2|8|6|3|7|13|9|1|15| . . . , where the summed duration of the first two intervals, $a + b$ [underlined in the example], of each sequence equals 10 msec). The *abx* stimulus has the same autocorrelation function as the *kxx* stimulus, yet the perception of the pitch of the *abx* stimulus is considerably weaker than that of the *kxx* stimulus. This significant difference in pitch strength provides an argument against a strict autocorrelation model. However, a model that is autocorrelation-like (but not a full autocorrelator) and is based on temporal fine structure can account for the *kxx* versus *abx* pitch-strength differences (Pressnitzer, de Cheveigné, & Winter, 2004; Yost, Mapes-Riordan, Dye, Sheft, & Shofner, 2005).

Pitch of Continuous Spectra and Noisy Stimuli

In addition to amplitude-modulated noise, there is another noise-like stimulus used to study pitch perception. This stimulus has a continuous spectrum as compared with the discrete spectrum of a harmonic complex, and the spectral (and temporal) features of this stimulus are highly variable. The stimulus is iterated rippled noise (IRN; see Yost et al., 1996). Such noises are generated by delaying a random noise by some amount, *d*, and adding the delayed noise back to the undelayed noise in an iterative process. IRN stimuli have a noisy spectrum, with spectral peaks spaced at the reciprocal of the value of *d* (e.g., if *d* is 5 msec, the IRN spectrum has variable amplitude spectral peaks at 200, 400, 600 Hz, etc.). IRN stimuli also have a temporal fine structure, with the dominant temporal interval being *d* msec. The temporal envelope of IRN stimuli is noisy but on average flat. Figure 9 portrays the neural activation pattern (see Figure 2) of a three-iteration stage, high-pass-filtered (above 2000 Hz) IRN stimulus generated with a 5-msec delay. Figure 10 displays the auditory spectrum of this IRN stimulus. In both figures, there is little spectral or temporal structure in the displays that appear related to the perceived 200-Hz pitch of this IRN stimulus. Figure 11 displays the autocorrelogram and summary autocorrelogram of the IRN stimulus, and the peak at a lag of 5 msec is clearly discernable. The reciprocal of the lag at which this peak occurs and the relative height of this peak have been used in several studies to accurately account for the pitch, pitch strength, and other attributes of the pitch of IRN stimuli. IRN stimuli can be generated to represent all of the spectral and most of the temporal characteristics of the discrete stimuli (e.g., harmonic complexes) used to study complex pitch, but IRN is also a highly random stimulus. Again, the best approach to model these IRN data has been autocorrelation.

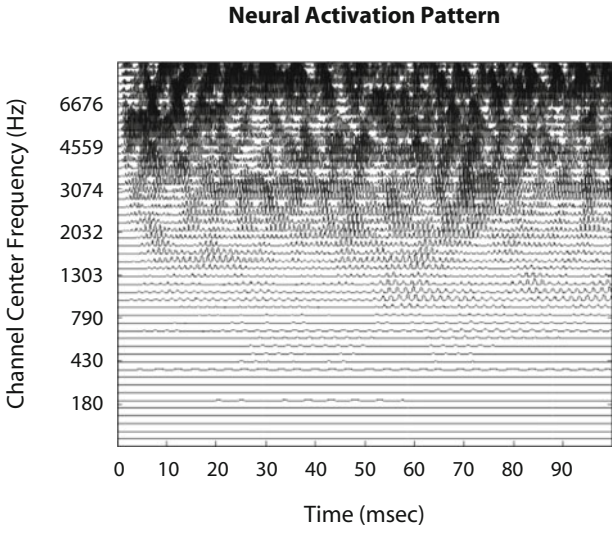

Figure 9. The neural activation pattern (see Figure 2) for an iterated ripple noise (IRN) stimulus generated with a delay of 5 msec and three stages of iteration. The IRN stimulus was band-pass filtered between 2000 and 8000 Hz and produced a strong 200-Hz pitch. The simulated neural activity is very noisy, making it difficult to identify a neural pattern that might be useable for predicting the pitch of IRN stimuli.

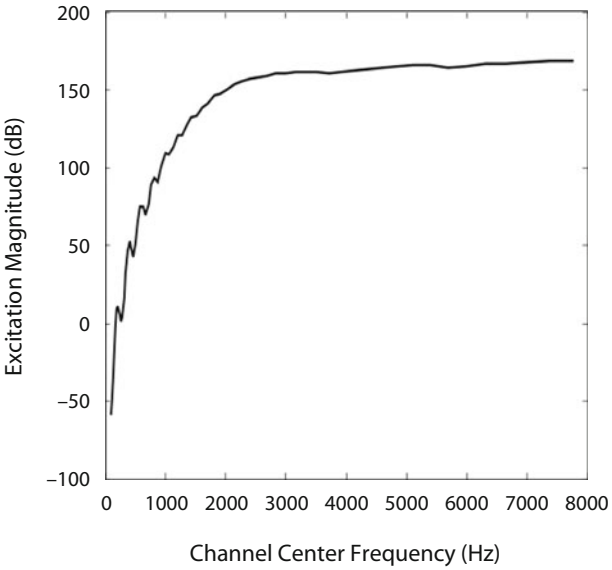

Figure 10. The neural excitation pattern for the stimulus condition shown in Figure 9. There is very little spectral variation (a decibel or less) upon which to base a pitch prediction (see Figure 3 for an description of neural activation pattern).

Spectrally Resolved Components

Although the perception of complex pitch exists when stimuli are high-pass filtered, the pitch is often very weak.

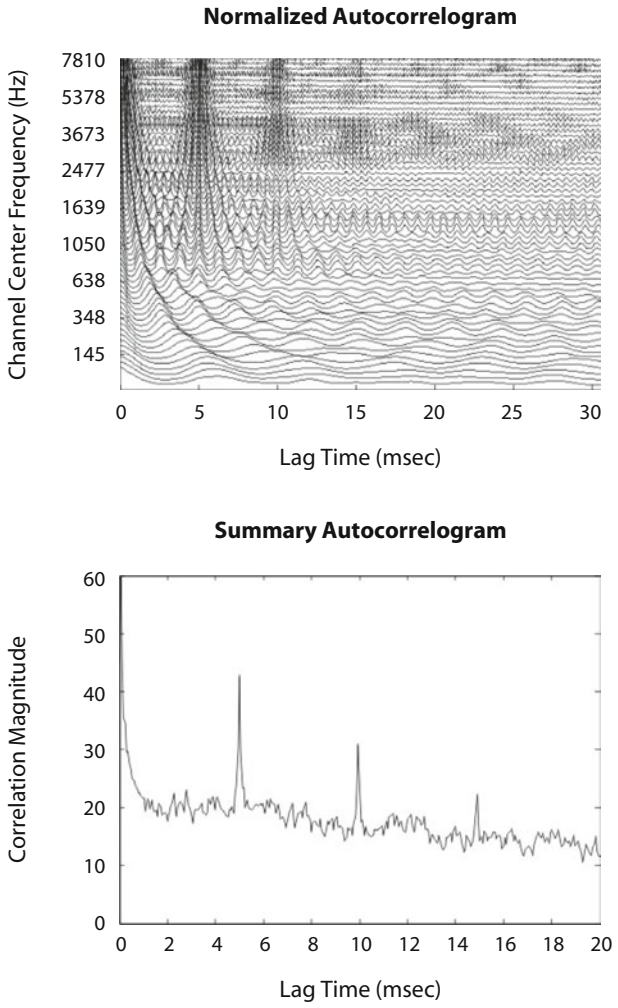

Figure 11. The autocorrelogram (Figure 6A) and summary autocorrelogram (Figure 6B) for the IRN stimulus condition used in Figures 9 and 10 are shown. There is a clear correlation at lags of 5 and 10 msec, with the largest correlation at 5 msec, which leads to a pitch prediction of 200-Hz (200 Hz is the reciprocal of 5 msec) for IRN stimulus condition.

Some data (e.g., Carlyon, 1998) suggest two spectral regions for the existence of pitch: a low-frequency region in which complex pitch is strong and a high-frequency region where the pitch is weak. One method used to measure the pitch strength of complex harmonic stimuli is to measure the ability of listeners to discriminate a change in the fundamental frequency ($f0$) of harmonic complexes. If listeners have low difference thresholds for discerning a difference in $f0$ (difference in pitch), then one might assume that the pitch was strong. Figure 12 shows the data from an experiment (Bernstein & Oxenham, 2005) in which listeners were asked to discriminate a change in $f0$ in a harmonic complex as a function of high-pass filtering the complex. As can be seen, as long as the complex had spectral components below about the 10th harmonic, $f0$ discrimination was very good. There is a steep transition to poor $f0$ discrimination above about the 10th harmonic. Thus, pitch strength could be strong when stimuli contain harmonics below the 10th harmonic and weak when there are no harmonics below the 10th. Recall that peripheral spectral resolution is very poor for harmonics above the 10th harmonic, so one interpretation of the data of Figure 12 is that the pitch strength of a harmonic complex decreases when the harmonics are not resolved by the auditory periphery.

Bernstein and Oxenham (2003) conducted an experiment that suggests that the stronger pitches for low-frequency stimuli may not be due only to the resolvability of low-frequency spectral structure. That is, resolvability may not be solely responsible for stronger, more salient pitches existing for low-frequency stimuli. These investigators provided a harmonic complex of 200 Hz to both ears in a diotic condition, and then a 200-Hz harmonic structure with every other harmonic delivered to one ear (200, 600, 1000, 1400 Hz, etc.) and the other harmonics delivered to the other ear (400, 800, 1200, 1600 Hz, etc.) in a dichotic condition. In the dichotic condition, the tonal components are further apart in frequency at each ear, and those components should be resolved at a higher cutoff frequency than the diotic stimulus would. That is, if the 10th harmonic is unresolved, this would occur at 4000 Hz in each ear in the dichotic case, but at 2000 Hz in the diotic case (i.e., 4000 Hz is the 20th harmonic of the diotic case). Despite the fact that components above 2000 Hz for the dichotic case were resolved in each ear, $f0$ discrimination was the same for the diotic and dichotic conditions, becoming worse at about 2000 Hz. Thus, even when components of a harmonic complex can be resolved at the periphery of both ears, pitch discrimination as an indication of pitch strength is still limited to about the 10th harmonic.

Ives and Patterson (2008) recently demonstrated a similar outcome for monaural harmonic complexes. That

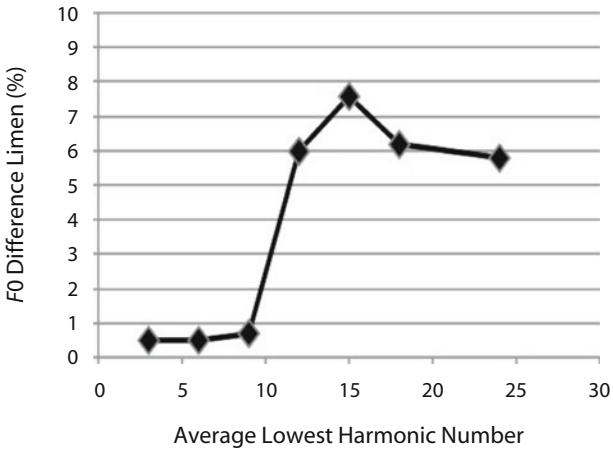

Figure 12. The threshold change in fundamental frequency (*F*0 difference limen) in percentage of $f0$, required to discriminate a change in *F*0 as a function of the lowest component (lowest harmonic) in a 200-Hz-fundamental complex harmonic sequence. There is a large decrement in performance when the harmonic number is approximately the 10th. This loss in performance might be due to a lack of resolvable harmonics above the 10th harmonic. These data are from Bernstein and Oxenham (2005).

is, harmonic complexes with high fundamental frequencies, but without many of the lower harmonics, appear to have a weaker pitch strength than do harmonic complexes with lower fundamentals and similar resolved harmonics. This suggests that some process other than spectral resolvability alone is responsible for the strong pitch of low-frequency stimuli. Bernstein and Oxenham (2005) suggested a refinement to the autocorrelation model of Meddis and O'Mard (1997) that can account for some of the *F*0 discrimination data below the 10th harmonic, even when spectral components can be resolved at one ear or the other.

Neural Correlates of Pitch Perception

There are no purely neural models or theories of pitch processing, but there are numerous neural studies investigating three primary areas related to pitch perception: (1) the role neural spectral tuning might play in pitch perception, (2) descriptions of possible temporal codes of the distribution of intervals between and among neural discharges that might be involved with pitch processing, and (3) the determination of neural pathways (as measured in animal models and in human subjects) that might participate in pitch processing.

The tuning of auditory nerve fibers and their role in frequency coding is described in the Sound, Auditory Periphery, and Pitch section. The main tenet of spectral theories of pitch processing is the exquisite tuning of auditory nerve fibers to the spectral content of sound. Such neural tuning indicates that fine details of the spectral structure of sound are limited to low frequencies and, in most cases, pitch is strongest when the spectral structure of sound can be captured by neural tuning (i.e., when the spectral structure can be resolved neurally). No neural circuits or coding mechanism has been proposed as a means for extracting information about the harmonic structure of sound, such as might be consistent with the perceptual spectral models of pitch perception (however, see Shamma & Klein, 2000).

A few investigators (Langner & Schreiner, 1988; Langner, Schreiner, & Biebel, 1998; Schreiner & Urbas, 1986) have suggested that neural pathways, especially those in auditory cortex and perhaps in the inferior colliculus (IC), are "tuned" to the envelope of amplitude-modulated sounds and may play a role in pitch perception. Some cortical units and, to some extent, some IC units respond selectively to the rate of amplitude modulation. The modulation rates to which these central neural units are most receptive are those in the range in which pitch is strongest (i.e., below approximately 500 Hz). However, as discussed previously, modulated envelopes alone cannot account for the various pitch percepts associated with the pitch shift of the residue, so any role that neural units tuned only to envelope modulation might play in pitch processing is likely to be limited.

Several investigators have studied the temporal structure of neural discharges of auditory nerve fibers (see Cariani & Delgutte, 1996) and of several different types of neural units in the cochlear nucleus and a few in the IC (see Sayles & Winter, 2008, for a review). These neural units respond to a wide variety of stimuli that produce the perception of pitch. The primary neural cue proposed for pitch processing is the interval between and among neural discharges. Figure 13 indicates a hypothetical neural discharge pattern to a periodic stimulus with a period *P*, such as might be generated by a harmonic complex whose fundamental frequency is the reciprocal of *P*. A common physiological metric used to describe the temporal response of neural units to sound is the interval histogram or interstimulus interval (ISI) histogram, which is a histogram of the intervals between successive neural discharges (i.e., first-order intervals). An all-interval histogram is a histogram of all orders of intervals among neural spikes (first order, plus second order, plus third order, etc.; see Figure 13); and an all-interval histogram is

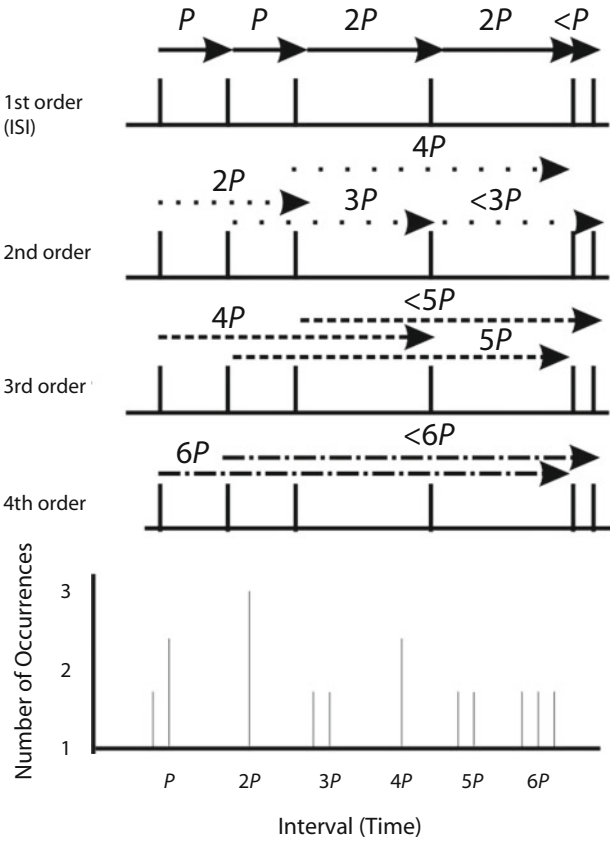

Figure 13. A depiction of a neural discharge pattern of six action potentials that could have resulted from the presentation of a periodic pulse train with a period of *P* msec. Each row indicates the order of intervals from first to fourth order (the one fifth-order interval is not shown). The first-order intervals indicate the same statistics used for an interstimulus interval histogram often computed in neural studies. There is one action potential at the end of the pattern that occurs less than *P* msec after the next to last action potential, but all other intervals between action potentials are integer multiples of *P*. The all-order interval histogram shown at the bottom is a histogram of all intervals and is equivalent to the autocorrelation function of the neural discharge pattern.

equivalent to the autocorrelation function of the temporal neural discharge pattern as indicated in Figure 13.

In many parts of the auditory system (e.g., in the auditory nerve), ISI histograms are sensitive to sound level, although neither the pitch nor the pitch strength of most sounds is very dependent on overall sound level. For this reason (and others) investigators (e.g., Cariani & Delgutte, 1996) usually have used an all-interval histogram, rather than a first-order-interval or ISI histogram, to measure the responses of auditory neural units to sounds that produce pitch. However, recently Sayles and Winter (2008; see also Wiegrebe & Winter, 2001) have analyzed the responses of fibers in the cochlear nucleus and have shown that a first-order interval analysis of the output of these brainstem neural fibers as is obtained for an ISI histogram can account for several perceptual pitch phenomena. They also continue to use an all-interval analysis of their brainstem single-unit data as a way to characterize the ability of these neural units to respond to the temporal fine structure information, because it might reveal useful information regarding pitch processing.

An all-interval histogram or autocorrelogram of neural discharges has been shown to preserve a great deal of information that is consistent with a wide range of pitch-perception phenomena. This is true for auditory nerve fibers (see Cariani & Delgutte, 1996) and different fiber types in the cochlear nucleus (e.g., Sayles & Winter, 2008; Shofner, 1999). Much of the physiological data is consistent with an autocorrelation model, but no physiological autocorrelation process has been described. Although it may not be the case that an all-order interval analysis (i.e., autocorrelation function) is necessary to account for pitch perception (i.e., maybe only the very lowest-order intervals, such as the ISI histogram, are required; see Wiegrebe & Winter, 2001), the physiological data suggest that the temporal regularity that is preserved in the interval statistics from many types of fibers in the auditory pathway is crucial for explaining pitch perception at the neural level.

Recent work with monkeys using single-unit recordings and with humans using neural imaging techniques (PET, fMRI, and MEG) has implicated cortical regions in and around Heschl's gyrus in humans (see Griffiths, Buechel, Frackowiak, & Patterson, 1998; Gutschalk, Patterson, Scherg, Uppenkamp, & Rupp, 2004; Hall & Plack, 2007; and Patterson & Johnsrude, 2008) and its homologue in monkeys (Bendor & Wang, 2005) as cortical centers that appear to process complex pitch. These studies usually use a wide variety of stimuli that produce the same pitch perception in human subjects, and the results indicate that these cortical pathways appear to be "tuned" selectively to pitch rather than to other aspects of the sounds. Although there is not yet a hypothesis as to how these cortical pathways process pitch, the evidence is growing that these pathways are in some way involved with pitch processing.

Summary and Conclusions

This review has described a historical seesaw between spectral and temporal accounts of the relationship between the temporal–spectral peripheral code for sound and pitch perception. Three aspects of the temporal–spectral code have been used in various ways to account for pitch perception: the tuning of auditory nerve fibers to the spectral structure of sound, phased-lock neural activity of low-frequency auditory nerve fibers representing the temporal fine structure of sound, and slow changes in auditory nerve fiber discharge rate resulting from the slow modulation of amplitude that can occur in a sound's envelope. Spectral accounts of pitch processing based on the tuning of auditory nerve fibers cannot account for the pitch of complex sounds with unresolved spectral structure. Nor can a spectral account handle the pitch of amplitude-modulated noise. The use of only the envelope of sound cannot account for the pitch shift of the residue. The phased-locked activity of auditory nerve fibers cannot account for the pitch of high-frequency narrow-band stimuli, such as tones, nor can it account for the pitch of amplitude-modulated noise. Thus, none of these three types of efforts based on peripheral processing alone can account for all of the phenomena associated with the perception of pitch. However, autocorrelation or similar processes with various modifications can account for the perceived pitch associated with resolved and unresolved harmonics, missing-fundamental pitch, pitch shift of the residue, pitch of narrow-band stimuli, and pitch of amplitude-modulated noise. On the other hand, there have been several challenges to a full autocorrelation approach as a model of pitch perception. Thus, even autocorrelation comes up short of being a complete model of pitch processing. There have been several suggestions for pitch perception processes that are not autocorrelation, but those share many of the properties of autocorrelation. These autocorrelation-like approaches include strobe-temporal integration (Patterson et al., 1995), the use of first-order interval calculations (Pressnitzer et al., 2004; Sayles & Winter, 2008; Wiegrebe & Winter, 2001), and cancellation correlation (de Cheveigné, 1998). These models or operations often help overcome some of the failures of a full autocorrelation model, but none have solved all of the problems of autocorrelation, nor have they been able to account for all of the major data sets described in this review.

One can view an autocorrelation process or many of the autocorrelation-like processes as determining the regular temporal intervals in a sound's waveform, sometimes arising from the sound's temporal fine structure and sometimes from its envelope. When such regular intervals are prevalent in sound, the sound usually has a perceived pitch associated with the reciprocal of the regular interval. Thus, although autocorrelation may or may not be the best process to determine these regular intervals, the evidence is strong that regular temporal intervals of simple and complex sounds provide a crucial base for explaining pitch. What is needed is a description of an actual central neural process that uses this temporal regularity for determining pitch and a description of how that process operates. It is likewise clear that an actual autocorrelation process does not exist neurally, but something that has many of the operations of autocorrelation would seem to be a good

starting point. In any case, the quest for a unified theory of pitch perception continues.

AUTHOR NOTE

Work on this article was supported by a grant from the NIDCD. I thank my colleagues at ASU for stimulating conversations that aided my writing of this review: Sid Bacon, Chris Brown, Michael Dorman, Tony Spahr, and Farris Wailing. Correspondence should be addressed to W. A.Yost, Speech and Hearing Science, Arizona State University, P.O. Box 870102, Tempe, AZ 85287-0102 (e-mail: william.yost@asu.edu).

REFERENCES

American National Standards Institute (1978). ANSI, S3.20-R1978-American National Standard on Bioacoustical Terminology. Acoustical Society of America.

Bendor, D. A., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, **436**, 1161-1165.

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. (2005). An autocorrelation model with place dependence to account for the effect of harmonic number on fundamental frequency discrimination. *Journal of the Acoustical Society of America*, **117**, 3816-3831.

Bernstein, J. G., & Oxenham, A. J. (2008). Harmonic segregation through mistuning can improve fundamental frequency discrimination. *Journal of the Acoustical Society of America*, **124**, 1653-1667.

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.

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. (1998). Comments on "A unitary model of pitch perception" [J. Acoust. Soc. Am. 102, 1811-1820 (1997)]. *Journal of the Acoustical Society of America*, **104**, 1118-1121.

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.

de Boer, E. (1956). *On the "residue" in hearing*. Unpublished doctoral dissertation, University of Amsterdam, Amsterdam.

de Boer, E. (1961). A note on phase distortion and hearing. *Acoustica*, **11**, 182-184.

de Boer, E. (1976). On the "residue" and auditory pitch perception. In W. D. Keidel & W. D. Neff (Eds.), *Handbook of sensory physiology* (pp. 479-583). New York: Springer.

de Cheveigné, A. (1998). Cancellation model of pitch perception. *Journal of the Acoustical Society of America*, **103**, 1261-1271.

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.

Griffiths, T. D., Buechel, C., Frackowiak, R. S. J., & Patterson, R. D. (1998). Analysis of temporal structure in sound by the brain. *Nature Neuroscience*, **1**, 422-427.

Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S., & Rupp, A. (2004). Temporal dynamics of pitch in human auditory cortex. *NeuroImage*, **22**, 755-766.

Hall, D., & Plack, C. (2007). Searching for a pitch centre in human auditory cortex. In B. Kollmeier et al. (Eds.), *Hearing: From sensory processing to perception* (pp. 83-94). Berlin: Springer.

Irino, T., & Patterson, R. D. (1996). Temporal asymmetry in the auditory system. *Journal of the Acoustical Society of America*, **99**, 2316-2331.

Ives, D. T., & Patterson, R. D. (2008). Pitch strength decreases as *F*0 and harmonic resolution increase in complex tines composed exclusively of high harmonics. *Journal of the Acoustical Society of America*, **123**, 2670-2679.

Kaernbach, C., & Demany, L. (1998). Psychophysical evidence against the autocorrelation theory of auditory temporal processing. *Journal of the Acoustical Society of America*, **104**, 2298-2306.

Langner, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat: I. Neuronal mechanisms. *Journal of Neurophysiology*, **60**, 1799-1822.

Langner, G., Schreiner, C. E., & Biebel, U. W. (1998). Functional implications of frequency and periodicity pitch in the auditory system. In A. R. Palmer, A. Rees, A. Q. Summersfield, & R. Meddis (Eds.), *Psychophysical and physiological advances in hearing* (pp. 277-285). London: Whurr.

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**, 945-950.

Lin, J.-Y., & Hartmann, W. M. (1998). The pitch of a mistuned harmonic: Evidence for a template model. *Journal of the Acoustical Society of America*, **103**, 2608-2613.

Meddis, R. (1986). Simulation of mechanical to neural transduction in the auditory receptor. *Journal of the Acoustical Society of America*, **79**, 702-711.

Meddis, R., & Hewitt, M. J. (1991). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America*, **89**, 2866-2882.

Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America*, **102**, 1811-1820.

Moore, B. C. J. (1993). Frequency analysis and pitch perception. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), *Human psychophysics* (pp. 56-116). New York: Springer.

Moore, B. C. J., Glasberg, B. J., & 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., & Glasberg, B. R. (1986). Thresholds for hearing mistuned partials as separate tones in harmonic complexes. *Journal of the Acoustical Society of America*, **80**, 479-483.

Patterson, R. D. (1969). Noise masking of a change in residue pitch. *Journal of the Acoustical Society of America*, **45**, 1520-1524.

Patterson, R. D. (1973). The effects of relative phase and the number of components on residue pitch. *Journal of the Acoustical Society of America*, **53**, 1565-1572.

Patterson, R. D., Allerhand, M. H., & Giguère, C. (1995). Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *Journal of the Acoustical Society of America*, **98**, 1890-1894.

Patterson, R. D., & Irino, T. (1998). Auditory temporal asymmetry and autocorrelation. In A. Palmer, A. Rees, Q. Summersfield, & R. Meddis (Eds.), *Psychological and physiological advances in hearing* (pp. 554-562). London: Whurr.

Patterson, R. D., & Johnsrude, I. S. (2008). Functional imaging of the auditory processing applied to speech sounds. *Philosophical Transactions of the Royal Society B*, **363**, 1012-1035.

Patterson, R. D., & Wightman, F. L. (1976). Residue pitch as a function of component spacing. *Journal of the Acoustical Society of America*, **59**, 1450-1459.

Patterson, R. D., Yost, W. A., Handel, S., & Datta, A. J. (2000). The perceptual tone/noise ratio of merged iterated rippled noises. *Journal of the Acoustical Society of America*, **107**, 1578-1588.

Plack, C. J., Oxenham, A. A., Fay, R. R., & Popper, A. N. (Eds.) (2005). *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.

Plomp, R. (1976). *Aspects of tone sensation: A psychophysical study*. London: Academic Press.

Pressnitzer, D., de Cheveigné, A., & Winter, I. M. (2004). Physiological correlates of the perceptual pitch shift for sounds with similar autocorrelation. *Acoustics Research Letters Online*, **5**, 1-6.

Pressnitzer, D., & Patterson, R. D. (2001). Distortion products and the perceived pitch of harmonic complex tones. In D. Breebaart, A. Houtsma, A. Kohlrausch, V. Prijs, & R. Schoonhovem (Eds.), *Physiological and psychophysical bases of auditory function* (pp. 97-104). Maastricht: Shaker.

Ritsma, R. J. (1962). Existence region of the tonal residue. *Journal of the Acoustical Society of America*, **34**, 1224-1229.

Sayles, M., & Winter, I. M. (2008). Ambiguous pitch and the temporal representation of inharmonic iterated rippled noise in the ventral cochlear nucleus. *Journal of Neuroscience*, **28**, 11925-11938.

SCHOUTEN, J. F. (1938). The perception of subjective tones. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, **41**, 1086-1093.

SCHOUTEN, J. F. (1940). The residue, a new component in subjective sound analysis. *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen*, **43**, 356-365.

SCHREINER, C. E., & URBAS, J. V. (1986). Representation of amplitude modulation in the auditory cortex of the cat: I. The anterior auditory field (AAF). *Hearing Research*, **21**, 227-241.

SHAMMA, S., & 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.

SHOFNER, W. P. (1999). Responses of cochlear nucleus units in the chinchilla to iterated rippled noises: Analysis of neural autocorrelograms. *Journal of Neurophysiology*, **81**, 2662-2674.

SLANEY, M., & LYON, R. F. (1993). On the importance of time: A temporal representation of sound. In M. Cooke, S. Beet, & M. Crawford (Eds.), *Visual representations of speech signals* (pp. 95-116). Chichester, U.K.: Wiley.

STEVENS, S. S., VOLKMANN, J., & NEWMAN, B. N. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, **8**, 185-190.

TERHARDT, E. (1974). Pitch, consonance, and harmony. *Journal of the Acoustical Society of America*, **55**, 1061-1069.

WIEGREBE, L., & WINTER, I. M. (2001). Psychophysics and physiology of regular interval noise: Critical experiments for current pitch models and evidence for a 1st-order temporal pitch code. In D. Breebaart, A. Houtsma, A. Kohlrausch, V. Prijs, & R. Schoonhoven (Eds.), *Physiological and psychophysical bases of auditory function* (pp. 121-128). Maastricht: Shaker.

WIGHTMAN, F. L. (1973). The pattern-transformation model of pitch. *Journal of the Acoustical Society of America*, **54**, 407-416.

YOST, W. A. (2007). Pitch perception. In P. Dallos, D. Oretel, & R. Hoy (Eds.), *The senses: A comprehensive reference—Audition* (Vol. 3, pp. 1023-1057). London: Academic Press.

YOST, W. A., MAPES-RIORDAN, D., DYE, R., SHEFT, S., & SHOFNER, W. (2005). Discrimination of first- and second-order regular intervals from random intervals as a function of high-pass filter cutoff frequency. *Journal of the Acoustical Society of America*, **117**, 59-62.

YOST, W. A., PATTERSON, R. D., & SHEFT, S. (1996). A time-domain description for the pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America*, **99**, 1066-1078.

YOST, W. A., PATTERSON, R. D., & SHEFT, S. (1998). The role of the envelope in auditory processing of regular interval stimuli. *Journal of the Acoustical Society of America*, **104**, 2349-2361.

(Manuscript received November 19, 2008;
revision accepted for publication July 5, 2009.)