

Chapter 4

Pitch perception

Alain de Cheveigné

4.1 Introduction

Pitch is the stuff of which music is made. Melody, harmony, and tonality are either built upon pitch, or else they depend upon similar properties of the physical stimulus. In speech, pitch is a vector of prosody, and for tonal languages it also carries syllabic information. Pitch (or its physical correlate, periodicity) is important to perceptually segregate competing sound sources. Pitch and harmony have fascinated thinkers since antiquity and, for many early authors, to explain pitch amounted to explaining auditory perception.

For the psychoacoustician, pitch is the perceptual correlate of fundamental frequency (F_0), that is, the rate at which a periodic waveform repeats itself. A periodic sound produces a pitch that depends on the period $T = 1/F_0$: the shorter the period, the higher the pitch. The quantitative relation between period of vibration and notes of the musical scale was established early in the 17th century by Mersenne and Galileo (see de Cheveigné (2005) for a review). Recent work has been invested in mapping out the properties and limits of pitch perception (Plack and Oxenham, 2005), and probing the mechanisms by which the pitch percept emerges within the auditory system (Winter, 2005).

Stimuli may differ in amplitude, duration, spatial position, and spectral content, and nevertheless evoke the same pitch. Pitch is a many-to-one mapping from a high dimensional set of sounds to a percept that is unidimensional (in first approximation). Thus, a trained listener may accurately match a piano note to a pure tone, to a complex tone with high-order partials, or to an exotic binaural stimulus that sounds like a featureless ‘shhhh’ when listened to with one ear, but is distinctly pitch-like when listening with both ears (Cramer and Huggins, 1958). Pitch is the abstract quality common to these sounds. To understand pitch perception, we must explain not only our exquisite sensitivity to small changes along the physical dimension of period (or fundamental frequency), but also our ability to ignore enormous differences along other dimensions.

It is customary in psychoacoustics to distinguish *pure tones*, with a sinusoidal waveform and a single-component spectrum, from *complex tones* with a waveform that is arbitrarily shaped but nevertheless periodic, and a spectrum with multiple components that are harmonically related (i.e. all multiples of the same F_0). Much past research on pitch has focused on pure tones under the belief that the percept that they evoke is somehow ‘elementary’. Here we treat the pure tone as one among the many stimuli that may evoke a pitch. Some examples of pitch-evoking stimuli are illustrated in Fig. 4.1.

4.2 What is pitch?

In 1960 the American Standards Association defined pitch as ‘that attribute of auditory sensation in terms of which sounds may be ordered on a musical scale’ (ASA, 1960), a definition that suggests a percept with extent along a linear perceptual dimension. The concept of ‘dimension’ carries the idea that diverse sounds may map to the same point along this dimension. Indeed, notes

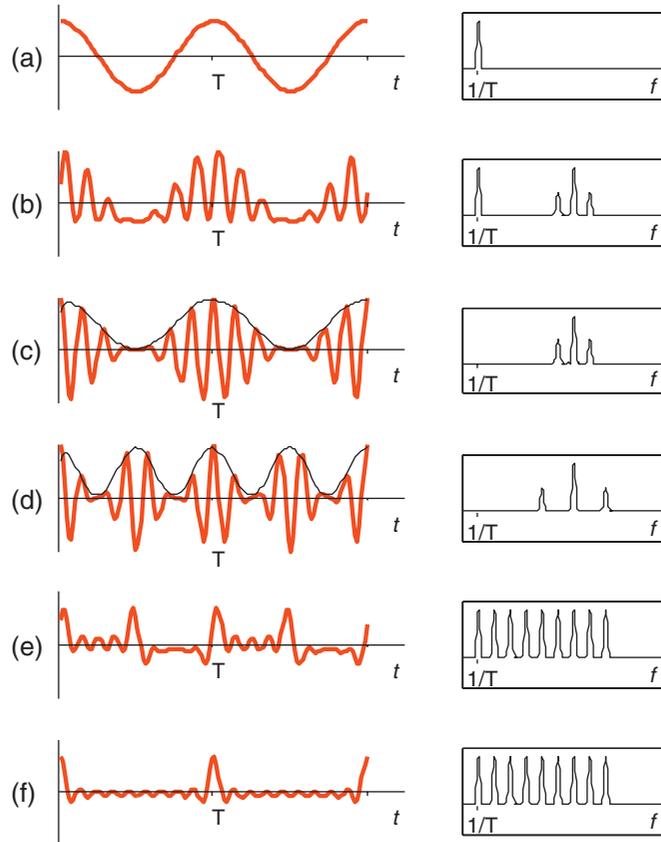


Fig. 4.1 Waveforms and spectra of various stimuli that evoke pitch. Stimuli with the same period tend to evoke the same pitch despite their different amplitude, duration, spectra, or spatial characteristics. (a) Pure tone; (b) complex tone made up of the fundamental and three higher harmonics; (c) same, without the fundamental; (d) same, with partials spaced by twice the F_0 ; (e) complex made up of 9 partials in alternating sine/cosine phase; (f) same, in cosine phase. The stimuli in (c) and (d) can be described as the result of modulating a carrier of frequency $f = 7/T$ by a more slowly varying *temporal envelope* (thin line). The frequency of the temporal envelope is equal to the spacing of the partials, whereas the fundamental frequency $F_0 = 1/T$ is equal to the largest common divisor of the partial frequencies. For some stimuli, pitch may follow the period of the temporal envelope, rather than the true fundamental, i.e. tones as in (d) (and also (e)) may sound an octave higher than expected based on their F_0 . In the spectral domain, the *spectral envelope* is a smooth function of frequency that describes the amplitude of the partials. The spectral envelope mainly determines the timbre.

produced on different instruments may have distinct timbres but the same pitch. A more recent version of the standard (ANSI, 1994) added that pitches are ordered from *low to high*, suggesting a vertical orientation for this dimension. Whether this orientation is universal or cultural is a matter of debate (e.g. Rusconi *et al.*, 2006), but in our culture it fits the vertical axis of a musical score or a spectrogram.

In music, pitch usually varies over time. Every new note of a melody evokes a percept that depends to some extent on the physical characteristics of that note, but also on the note that precedes it.

Indeed, it may seem that our perception of melodic pitch is determined by *intervals* between notes rather than, or in addition to, the notes themselves. More generally, the musical effect of each new note depends strongly on its context (Bigand and Tillmann, 2005), a property that is not quite captured by the ANSI definitions, or indeed, by most psychoacoustic accounts of pitch.

To a first approximation, equal *ratios* of frequency produce pitch steps of equal salience, as if pitch were a logarithmic function of frequency. However, the relation between notes on the scale is also governed by complex rules of harmony, also not captured by the standard definition of pitch. For example, notes an octave apart are perceptually similar, and in some cases interchangeable. They are said to share the same *chroma*. Chroma is an equivalence relation: multiple notes map to the same chroma. The similarity between two notes depends in part on their proximity along a logarithmic frequency scale, and in part on their chroma.

To capture this property, more complex geometrical models of pitch have been proposed, such as a helix with a linear axis that fits the standard 'low-to-high' dimension of pitch height, and a circular dimension of chroma (Bachem, 1950; Ueda and Ohgushi, 1987; Giangrande *et al.*, 2003). Yet more complex structures such as toroids have been proposed to incorporate additional tonal relations such as fifths (Shepard, 1982). However it has been argued that they may reflect less the perceptual structure of pitch than the harmonic spectra of most European instrumental sounds (Burns, 1981; Sethares, 1997). In any case, the fixed nodes of such a structure cannot capture the dynamic effects noted earlier, such as determined by the order in which two notes appear (Giangrande *et al.*, 2003; Bigand and Tillmann, 2005).

A fascinating aspect of pitch perception is inter-individual variability. Discrimination thresholds vary between individuals over several orders of magnitude. Thresholds improve considerably with training (Demany and Semal 2002; Micheyl *et al.*, 2006a), and thus experience may account for some differences between listeners. Genetic factors may also contribute (Drayna *et al.*, 2001; Douglas and Bilkey 2007), and there are hints that the phenomenology of pitch may actually differ among people. For example, Semal and Demany (2006) found that most subjects can judge the direction of a pitch change (high to low or vice versa) as soon as it is detectable, but other subjects found it impossible to say which note is higher, for a pitch difference that they nevertheless could detect with ease.

Absolute pitch is the relatively rare capability to assign labels to pitches regardless of context (Ward 1999; Zatorre, 2003; Levitin and Rogers 2005; Hsieh and Saberi 2007). Most listeners possess instead *relative pitch*, the ability to judge the pitch of a note relative to a preceding note. This question is interesting because the two forms of pitch imply rather different mechanisms. For example, most models of pitch easily account for absolute pitch but not relative pitch. Something else is required for relative pitch, and as both types of pitch exist, the brain must be capable of both.

To summarize, pitch is a very important aspect of sound perception. We can discriminate exquisitely small differences in pitch, while ignoring salient differences along other perceptual dimensions. Pitch has more to it than the simple, one-dimensional construct assumed by psychophysics, and yet we have few models to account for these complexities. That so much is yet unknown about pitch is sobering for those of us who have been working on it for years, and exhilarating for whoever sets out to search for more: there's lots more to discover!

4.3 The limits of pitch

Periodic stimuli evoke pitch over a very wide range of their parameters: F0, amplitude, duration, spectral envelope, etc. Pitch is exquisitely sensitive to the first parameter (F0), and yet remarkably *stable* over large variations of the others, variations that themselves may produce salient changes in loudness, subjective duration, or timbre. Stimuli that are only approximately periodic may also

evoke a pitch (inharmonic complexes, stimuli in noise), and the same may occur for stimuli that are, strictly speaking, not periodic at all, such as amplitude-modulated noise or binaurally correlated noise. Conversely, a periodic stimulus may fail to evoke a pitch if its parameters fall outside certain bounds that delimit the *region of existence* of pitch.

Musical pitch arises if the F0 is within a range of about 30 Hz to 5000 Hz. At the lower end of the scale, Pressnitzer *et al.*, (2001) found that subjects failed to detect a one-semitone mistuning in a four-note chromatic melody if its lowest note fell below 30 Hz (33 ms). Periodicity can be detected, and the period discriminated, for longer periods up to several seconds, but the percept is not ‘musical’ and discrimination thresholds are large (Warren *et al.*, 1980; Kaernbach, 1993). Thresholds improve by an order of magnitude as F0 increases from 16 Hz to 64 Hz (Krumbholz *et al.*, 2000). At the other end of the scale, stimuli lose their musical properties beyond about 5 kHz (Bachem, 1948; Semal and Demany, 1990). The limit is rather sharp (subjects report a ‘highest musical note’) but it is subject- (and even ear-) dependent. Burns (1983) nevertheless found that some subjects performed above chance on musical tasks at frequencies of 10 kHz or beyond.

Difference limens for pure tones (expressed as a proportional change in frequency) are smallest at around 1–2 kHz (about 0.2% for good subjects) but they increase abruptly as frequency exceeds 4 kHz (Moore, 1973) (see Fig. 4.2). The interval from 30 Hz to 4000 Hz spans about seven octaves and includes the range of most musical instruments.

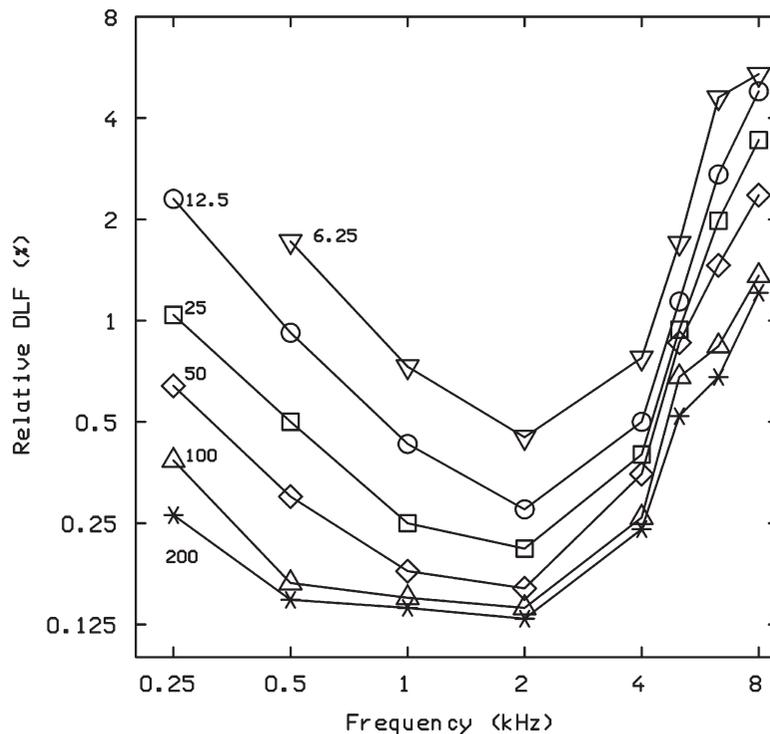


Fig. 4.2 Frequency difference limens (smallest detectable relative frequency difference) for pure tones. Each curve is for a different stimulus duration (in ms). Discrimination is best for frequencies near 2 kHz and degrades rapidly above 4 kHz. Discrimination is better for longer durations. From Moore (1973) with permission.

Pitch changes remarkably little with level. The frequency of a pure tone can be discriminated, presumably on the basis of pitch, as soon as it is detectable (Pollack, 1947; Cardozo, 1974; Gockel *et al.*, 2006). Indeed, pitch may be the cue that allows us to detect a tone in noise (Moore, 1981; Carney *et al.*, 2002). Frequency discrimination is less good at low levels (e.g. Wier *et al.*, 1977) or high levels (e.g. Bernstein and Oxenham, 2006), but the value of the pitch changes with level by at most a few percent for pure tones, and even less for complex tones (Hartmann, 1997). This is remarkable in the face of the strong level-dependency of several physiological responses to sound that have been considered as candidate substrates for pitch perception (Winter, 2005).

Frequency discrimination is possible for stimuli as short as one or two cycles (Mark and Rattay, 1990), although at such short durations it is uncertain whether discrimination is based on pitch or timbre changes (Hartmann *et al.*, 1985). A clear tonal percept requires a longer stimulus (Mark and Rattay, 1990; Robinson and Patterson, 1995). For very short stimuli, the value of the pitch may differ according to the duration, or the shape of the temporal envelope, but the differences are no more than a few percent (Hartmann, 1978; Hartmann *et al.*, 1985). As stimuli are made longer, frequency discrimination becomes more accurate (Moore, 1973; White and Plack, 2003; Gockel *et al.*, 2007; Hsieh and Saberi, 2007) (Fig. 4.2).

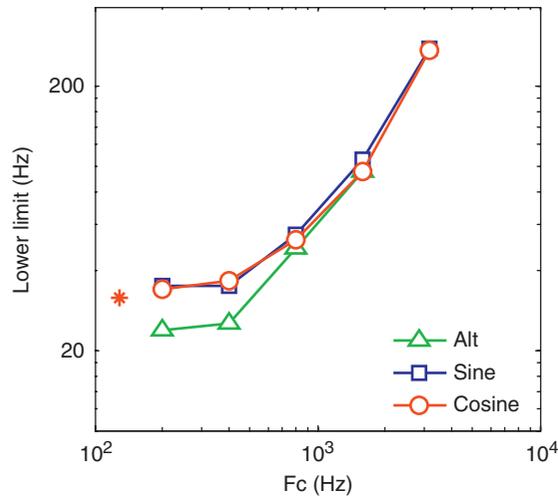
For stimuli with only low-order harmonics, differences in their relative phase are imperceptible, as stated by Ohm's acoustic law. With higher harmonics (closer spaced relative to their frequencies), phase may affect the *timbre* of the stimulus, but usually not the value of its pitch. Phase may, however, affect the *salience* of the pitch, and it may also change the relative weight of competing pitch candidates in stimuli with ambiguous pitch. For example, the pitch of a stimulus with closely spaced partials may be one octave higher if phases alternate between sine and cosine (as illustrated in Fig. 4.1(e)), than if they are all sine or all cosine (as illustrated in Fig. 4.1(f)). In a few rare situations one may observe small phase-dependent shifts of the value of the pitch (Plomp, 1967*b*; Pressnitzer *et al.*, 2002).

Pitch is evoked by many periodic stimuli, with very different spectra, but not all: some periodic stimuli evoke a pitch that is weak or absent. The stimulus parameter space has many dimensions, and therefore it is not straightforward to map the limits of the existence region. Roughly speaking, a complex tone may fail to evoke a pitch if: (1) its period is too long (Pressnitzer *et al.*, 2001), or (2) the *rank* of its lowest harmonic is too high, or (3), the *frequency* of its lowest harmonic is too high. These limits depend somewhat on the total number of harmonics within the stimulus: they are narrower for two adjacent harmonics (Smooenburg, 1970) than for three harmonics (Ritsma, 1962, 1963) or more. There are also interactions between parameters, for example musical pitch extends to a lower F0 (30 Hz) for a wideband stimulus than for a 3200-Hz, high-pass stimulus (270 Hz) (Fig. 4.3) (Pressnitzer *et al.*, 2001).

Tones with the same period but different spectral envelopes usually evoke the same pitch, despite large timbre differences. It is nevertheless more difficult to match the pitches of stimuli that occupy distinct rather than overlapping spectral regions (Micheyl and Oxenham, 2004). Such is notably the case when a pure tone is matched to a complex tone (Moore *et al.*, 1992). This difficulty might be due to perceptual interference from the salient difference in timbre, or it might result from the lack of overlap within an internal tonotopically organized representation of stimuli, or both.

For wideband stimuli, with partials spread over a wide frequency range, the various spectral regions carry unequal weight. One way to reveal this is to mistune the frequency of one partial of the complex and observe the shift in the overall pitch. This effect is appreciable only if the partial falls within the *dominant region*. The extent of this region depends on the F0: below 50 Hz partials beyond the 6th are dominant (Moore *et al.*, 2007), above 1400 Hz the fundamental is dominant, and in the intermediate range the dominant partials are usually between 2 and 6 but with considerable interindividual differences (Plomp, 1967*a*; Moore *et al.*, 1985*b*).

Fig. 4.3 The lower limit of melodic pitch as a function of the frequency of the lowest partial (F_c) and for different phase relationships: sine, cosine, and alternating sine and cosine (alt). The asterisk is for wideband cosine phase stimuli (click trains). Mean of three subjects. Replotted from Pressnitzer *et al.* (2001).



Pitch may be evoked by stimuli that are only imperfectly periodic such as ‘iterated rippled noise’, IRN. IRN is obtained by delaying a segment of white noise repeatedly, and adding together the delayed and non-delayed waveforms (Yost, 1996; Hartmann, 1997—Chapter 15). Pitch may also arise with stimuli that are, strictly speaking, aperiodic such as amplitude-modulated white noise (Burns and Viemeister, 1976). It is as if the auditory system searches for the best periodic approximation to the stimulus, according to some metric that tolerates various forms of mismatch. The same stimulus might allow multiple matches, in which case its pitch may be ambiguous. For example, a complex tone with a narrow spectral peak or ‘formant’ (Fig. 4.1(c)) may evoke a pitch that fits that peak. This high pitch, corresponding to the formant, may compete with the low pitch corresponding to the F_0 , particularly if the F_0 is relatively low and constant, and the frequency of the spectral peak varies. This effect is exploited in overtone singing (Bloothoof *et al.*, 1992).

To summarize, pitch may be evoked by a very wide range of stimuli. Pitch depends mainly on the period ($1/F_0$) and is remarkably insensitive to changes along other stimulus dimensions that provoke salient changes in qualities other than pitch. A very wide range of different stimuli map to the same pitch. This is possibly the hardest to explain: how does the auditory system perform tasks that require accurate discrimination using pitch, while ignoring the very salient effects of differences along other stimulus dimensions? Pitch theories need to explain both the accuracy and the constancy of pitch perception.

4.4 The pure tone

Much of the psychoacoustics of pitch has been established using pure-tone stimuli with sinusoidal waveforms (Fig. 4.1(a)). A pure tone evokes a pitch similar to other periodic stimuli of the same period (e.g. Fig. 4.1(b–f)). However, two things make pure tones ‘special’, and set them apart from stimuli with complex spectra. The first is their special status for the physics and mathematics of sounds. The second is that a pure tone produces essentially the same shape of vibration at every point of the cochlea, albeit with different amplitude and phase.

Sinusoids (more precisely: complex exponentials) are eigenvectors of linear transforms. This means that a pure tone remains a pure tone after propagation through air, reflection from obstacles, or mechanical transmission within the ear. The amplitude and phase of the waveform may be affected by the filtering involved, but it retains a sinusoidal shape, and its frequency remains

the same. Furthermore, according to Fourier's theorem any waveform may be decomposed into a sum of such sinusoids. Sinusoids are not the only functions to allow such a decomposition, but their mathematical properties make them a good basis of 'elementary waveforms' from which other waveforms can be built.

It is tempting to assume that, just like a complex stimulus is a sum of sinusoids, the percept that it evokes too is the sum of *elementary percepts* evoked by its sinusoidal components. If this were the case psychophysics would be very much simplified indeed: we would only need to study effects of pure tones to predict the effect of a stimulus of arbitrary complexity. To some extent, psychoacoustics has progressed on the basis of this assumption, witness the large proportion of studies involving pure tones. The idea was inspired by the intense development of harmonic analysis between the 17th and 19th centuries that culminated in Fourier's theorem, that Ohm (1843) and Helmholtz (1877) extended to the sensory domain (de Cheveigné, 2005; Darrigol, 2007). The same idea is embodied in the 'virtual pitch' theory of Terhardt (1974, 1979), according to which the pitch of a complex tone is composed from 'spectral pitches' evoked by its partials.

Unfortunately, there is little to support this idea. Introspection tells us that the percept evoked by a complex tone differs radically from the percepts evoked by its sinusoidal parts in isolation. It requires some faith to believe that one is *composed* of the others. True, we can sometimes focus our attention and 'hear out' an individual partial, but this requires skill and training, and it succeeds only in particular situations. The issue of hearing out multiple pitches is addressed in Section 4.6.

Anticipating that discussion, note that a complex tone causes different parts of the cochlea to vibrate with different waveforms depending on which stimulus components are reinforced by cochlear filtering (see Fig. 4.4 in Section 4.7 below). In particular, some channels may respond mostly to a single resolved partial (e.g. 3rd harmonic in Fig. 4.4). If we suppose that attention can be focused on that subset of cochlear channels, it may be possible to hear out the partial as originally reported by Mersenne (1636), or instead possibly a 'residue' of unresolved partials described by Schouten (1940). In the case of a pure tone, however, all parts of the cochlea respond with the same waveform, so attending to a subset of channels should not produce a different percept. Pure tones are 'pure' in the sense that they cannot be partitioned in this way, not in the sense that the percepts that they evoke compose those of complex sounds. Studies that use pure-tone stimuli are informative for those particular stimuli, and by extension for the wider class of periodic stimuli that they belong to, but they do not really probe the 'elements' of the perception of sound.

The value of the pitch of an arbitrary stimulus has been defined as the frequency of the pure tone to which it can be matched (Hartmann, 1997). This provides a convenient means to quantify pitch. However, the pitch of a pure tone varies with sound level and across ears (Burns, 1982), at low frequencies a pure tone must have a high amplitude to compensate for the high-pass characteristics of the middle ear (this may introduce distortion products), and for F0s below 2 kHz frequency discrimination is less accurate for pure than for complex tones (Henning and Grosberg, 1968). The presence of noise induces pitch shifts that tend to be larger for pure tones than complex tones (Houtsma, 1981). For all these reasons, it would make sense to replace the pure-tone standard by a complex tone standard, for example a click train.

Studies that use only pure tones do not probe the property of invariance across stimuli with different spectral content. This is a concern for physiological and brain imaging studies, as it may be uncertain that a response reflects pitch rather than some other correlate of stimulus manipulation.

To summarize, pure-tone stimuli are illustrative of a wider class of pitch-evoking stimuli, but there is little reason to think that the percept evoked by a complex sound is *composed* of the percepts evoked by each of its sinusoidal components. Studies that aim to establish that a response reflects pitch need to use a wider range of stimuli.

4.5 The missing fundamental

Rarely has a paradox provoked such a long-lasting and heated debate. The pitch evoked by a pure tone remains the same if we add additional tones with frequencies that are integer multiples of that of the original pure tone (harmonics). It also does not change if we then remove the original pure tone (the fundamental): this is the ‘paradox’ of the missing fundamental. At issue is whether, and how, a periodic tone that lacks a sinusoidal component at its F_0 can evoke a pitch. Concerning the stimulus itself there is no paradox: Fourier’s theorem states that a periodic waveform is composed of sinusoids with frequencies that are integer multiples of F_0 , but it does not say that they must *all* be present. Compare, for example, the waveforms of Fig. 4.1(b) and Fig. 4.1(c): both are clearly periodic. Concerning the percept, the paradox vanishes if one accepts that pitch is associated with the *periodicity* of the stimulus. The paradox appears only if we insist that pitch requires the presence of a sinusoidal fundamental component.

The psychophysics is quite clear: the presence of a fundamental component is not required for pitch. This was established already in the 19th century by Seebeck, who synthesized stimuli where the fundamental was weak or absent (de Boer, 1976; Turner, 1977). It was confirmed by Schouten (1938) who addressed the issue of a possible distortion product (created within the apparatus or within the ear) by adding to the stimulus a sinusoidal component with carefully controlled amplitude and phase to cancel any remaining energy at the F_0 . Licklider (1954) corroborated his conclusion by adding low-pass noise to mask any distortion products, and this has since been replicated in hundreds of studies. There is little support for Ohm’s dogma, according to which pitch requires a sinusoidal component at the fundamental.

In spite of all this evidence, there is a reluctance to abandon this idea. The observation of fundamental components in the analysis of physiological recordings (where they arise naturally as the result of non-linearities), or the demonstration that relatively high-amplitude distortion products may arise in the ear (Pressnitzer and Patterson, 2001), keep alive the suspicion that the fundamental might sneak into the ear unnoticed. However the most potent reason to retain Helmholtz’s picture of the ear as a Fourier analyzer is that it is too attractive to abandon. The missing fundamental will be with us for some time.

4.6 Hearing out pitch

We are adept at *hearing out* sounds, for example a faint voice among the sounds of a forest, or the melodic line of an instrument within the orchestra. We usually study pitch in relation to the acoustic waveform, but in real life it often emerges from a fraction of the stimulus, the rest constituting a masker to be ignored. The masker itself may include one or more pitch-evoking sources, as in music when several instruments play at the same time. In order to hear each pitch, the ear must overcome masking from the energy of the competing sounds (‘energetic masking’), and also ignore the percepts that those sounds evoke (‘informational masking’). This is an example of the process of Auditory Scene Analysis (Bregman, 1990) by which we parse an acoustic scene and attend to its parts. There are large interindividual differences in the susceptibility to informational masking, musicians being more adept at focusing on a frequency range than non-musicians (Oxenham *et al.*, 2003).

Despite its obvious musical relevance, the psychoacoustics of competing pitches is surprisingly sparse (Beerends and Houtsma, 1989; Carlyon, 1996a; Assmann and Paschall, 1998; Micheyl *et al.*, 2006b). If several partials of a tone dominate part of the spectrum, such that they are resolved by at least some cochlear channels, the ear may be able to focus on that tone and ignore the others. For example, in polyphonic music, the spectrotemporal envelope of one voice may

have ‘windows’ of low energy within which the other voices may be glimpsed in this way. However, experiments have shown that concurrent tone pairs that overlap in both frequency and time may also evoke salient pitches, as long as their partials are sufficiently spaced to be *resolvable* by cochlear filtering (Carlyon, 1996a; Micheyl *et al.*, 2006b). In contrast, if two tones contain only unresolved partials, mixing them gives rise to a noise-like sound or ‘crackle’ (Carlyon, 1996a) unless one is stronger than the other, in which case only one pitch is heard (Micheyl *et al.*, 2006b). The concept of resolvability is discussed in more detail in Section 4.7. In summary, it is often possible to hear out a pitch from a background of interfering sounds, some of which may themselves evoke a pitch.

Pitch is nevertheless degraded by sounds presented simultaneously, or even sequentially in close temporal proximity. Presenting temporal ‘fringes’ before or after a stimulus degrades discrimination (Carlyon, 1996b; Micheyl and Carlyon, 1998), apparently because information from the fringes leaks into the integration window that sums pitch information over time. Likewise, a distractor tone presented simultaneously in a remote frequency region produces what is known as ‘pitch discrimination interference’ (Gockel *et al.*, 2004; Micheyl and Oxenham, 2007). The degree of interference depends on both *spectral proximity* and *pitch similarity*, suggesting that interference may occur at two stages: within a tonotopically organized, low-level representation of the signal, and within a higher level representation of pitch (or yet higher levels such as attention, etc.).

If the masker is a noise band that overlaps the target in time and frequency region, the pitch of the target may be discriminable as soon as the target is detectable (Moore and Glasberg, 1991; Gockel *et al.*, 2006; Micheyl *et al.*, 2006b). It is as if the cue for detecting the tone within noise were pitch. In contrast, if the masker is a complex tone, the detection threshold is about 15–25 dB *below* the level at which the target’s pitch can be discriminated (Micheyl *et al.*, 2006b). The detection cue here is more likely the disruption of the tonal percept evoked by the masker. These phenomena are of interest because they may shed light on how pitch is extracted and represented within the auditory system.

Perhaps the earliest report of multiple pitches within the same sound is that of Mersenne (1636), who heard, within the sound of a plucked string, ‘at least five sounds’ corresponding to the fundamental and first few harmonics. Hearing out partials requires concentration and training: Sauveur (1701) recommended listening at night, while Helmholtz (1877) relied on special resonators that he designed to enhance the partials. The task is easier for harmonics of low rather than high rank. Estimates of the highest audible partial vary according to the method employed, ranging from the 5th to 8th (Plomp, 1964) or the 9th to 11th (Bernstein and Oxenham, 2003). They are roughly consistent with the hypothesis that a partial may be heard out if the distance from its closest neighbor is greater than about 1.25 ERB (Moore and Ohgushi, 1993; Moore *et al.*, 2006) (the equivalent rectangular bandwidth, ERB, is a measure of cochlear filter width, Moore and Glasberg, 1983). Hearing out partials is easier for tones of long rather than short duration (Gockel *et al.*, 2007). It is also easier if the partial is mistuned from the harmonic series, or amplitude-modulated, or turned on later than the rest of the complex tone (Peters *et al.*, 1983; Moore *et al.*, 1985a; Hartmann and Doty, 1996; Bernstein and Oxenham, 2003; Hartmann and Goupell, 2006).

Although partials of a complex may be heard out with effort and attention, they are rarely salient spontaneously unless enhanced by some spectral or temporal irregularity (Bernstein and Oxenham, 2003; Hartmann and Goupell, 2006). Listeners differ in their propensity to listen to a complex tone as a whole (‘synthetic listening’) or as composed of parts (‘analytic listening’) (Smoorenburg, 1970; Laguitton *et al.*, 1998). Audibility of partials inspired the doctrine according to which the percept of a complex tone is composed of the percepts of its partials, but an

alternative interpretation is that the ear, being adept at hearing out weak sounds within a background, can sometimes perform this feat with the partials that compose a complex tone.

4.7 Resolvability

An important concept, already mentioned, is *resolvability*, the ability of cochlear filtering to isolate individual partials of a complex sound. A partial is resolved if it is sufficiently remote in frequency from other partials so that it dominates the response of at least some cochlear filters. By extension, a complex stimulus is said to be ‘resolved’ if it contains at least some resolved partials (and ‘unresolved’ if it contains none). Figure 4.4 illustrates the concept for a 30-component complex tone with a 100 Hz fundamental. Each of the lowest 3–5 partials dominates the output of a few filters, near the apex of the cochlea. The output of these filters is quasi-sinusoidal at the frequency of the partial (Fig. 4.4, bottom left): these partials are resolved. Partial of higher rank are less well isolated: the best filter contains a large proportion of power from other components, and the output waveform is more strongly modulated and less sinusoidal (Fig. 4.4, bottom right). Partial such as these are unresolved. As an aside, we note that the uppermost partial is comparatively well isolated, which might account for the relatively salient pitch of the highest partial of such a complex (Moore *et al.*, 2006).

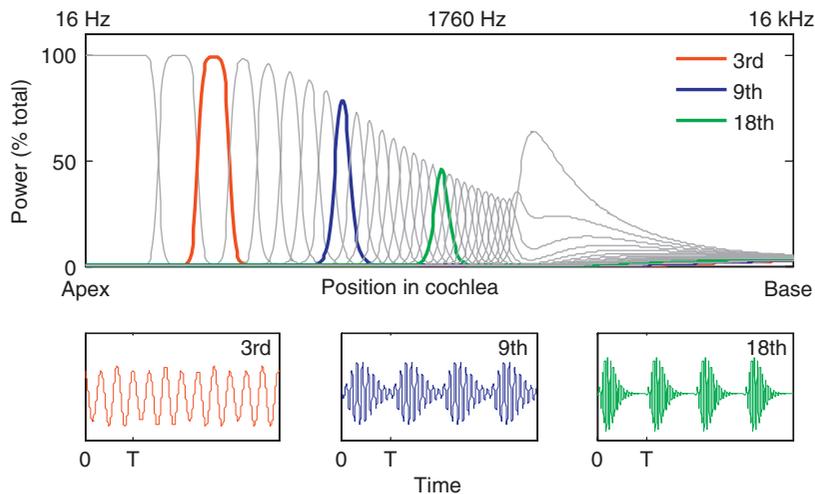


Fig. 4.4 Resolvability of harmonics of a complex tone. Top: percentage of power for each harmonic of a 30-component complex tone at the output of cochlear filters, plotted as a function of their position along the cochlea. The apex of the cochlea is tuned to low frequencies (left) and the base to high frequencies (right). Bottom: waveforms at the output of filters tuned to the 3rd harmonic, 9th harmonic, and 18th harmonic. Each of the lower harmonics is isolated within a set of filters near the apex of the cochlea (left of upper plot). The output of these filters is quasi-sinusoidal (bottom left): the lower harmonics are ‘resolved’. Each of the higher harmonics excites a narrower range of filters (middle and right of upper plot). Outputs of these filters are ‘pulsatile’, indicating that the filters respond to more than one harmonic: higher harmonics are ‘unresolved’. The cochlea is modeled here as a bank of linear 4th-order gammatone filters uniformly spaced in terms of equivalent rectangular bandwidth (ERB; Moore and Glasberg, 1983), which is roughly equivalent to uniform spacing along the cochlea.

Resolvability determines, in part, whether partials of a complex tone can be ‘heard out’. It also seems to play a role in the pitch of the complex as a whole: stimuli with one or more resolved partials tend to have a strong pitch, while those with only unresolved partials have a weak pitch. This is puzzling because unresolved partials produce beats along the basilar membrane at the F0, and one might expect this to be a clear cue to pitch. Instead, partials of low rank dominate the pitch of a complex. For example, mistuning those partials affects the pitch of the complex (Plomp, 1967a; Moore *et al.*, 1985b), and F0 discrimination thresholds are an order of magnitude smaller if the stimulus contain partials below about the 10th (Fig. 4.5) (Houtsma and Smurzynski, 1990; Shackleton and Carlyon, 1994; Plack and Carlyon, 1995; Bernstein and Oxenham, 2003) than if it does not. Accurate discrimination requires longer stimuli for unresolved than resolved partials (White and Plack, 2003). The ability to hear the pitches of concurrent complex tones is limited to tones with resolved partials: mixtures of unresolved tones evoke a crackling sound (Carlyon, 1996a). All these phenomena hint at a role for resolvability in pitch.

The contrast between performance for resolved- and unresolved-partial stimuli has led to the hypothesis that their pitches are processed by different mechanisms: *pattern matching* for resolved, and *autocorrelation* for unresolved (see Section 4.9). This necessarily implies also a third mechanism to translate between the two. Indeed, degraded performance in comparing pitch between resolved and unresolved stimuli has been taken as evidence of a translation cost between pitch mechanisms (Carlyon and Shackleton, 1994).

However, several results do not fit with this interpretation: (1) Moore *et al.* (2007) found for F0s of 35 and 50 Hz that partials with ranks greater than 6 (presumably *unresolved* at those F0s)

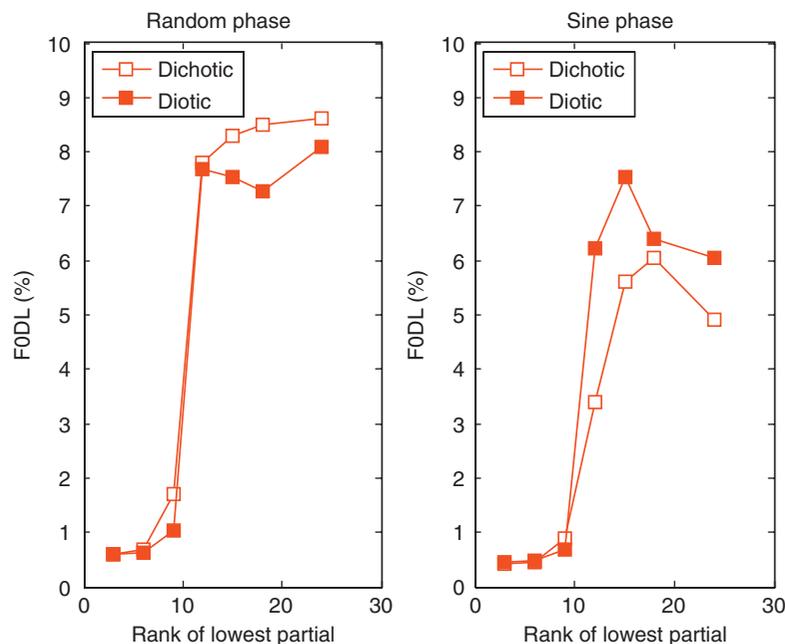


Fig. 4.5 F0 difference limens for a 200-Hz complex tone as a function of the rank of the lowest partial. Phases were either random (left) or sine (right). In the ‘diotic’ condition (full markers) all partials were presented to both ears. In the ‘dichotic’ condition (open markers) partials of odd rank were presented to one ear, and partials of even rank to the other ear. Average of four subjects, replotted from Bernstein and Oxenham (2003).

dominated the pitch of the complex. This suggests that dominance may depend on a factor other than resolvability. (2) The lower limit of melodic pitch does not coincide with the limit of resolvability: tones that lack any resolved partial may nevertheless evoke a pitch that supports melody (Pressnitzer *et al.*, 2001). (3) Bernstein and Oxenham (2003, 2008) found that the deterioration of discrimination thresholds beyond the 10th partial occurred regardless of whether all partials were presented to the same ear (in which case they were too closely spaced to be resolved) or else partials of even and odd rank were distributed to opposite ears (in which case their spacing is doubled, so that partials up to the 20th should be resolved according to their criteria) (Fig. 4.5). Peripheral resolvability is therefore not the factor that determines accurate pitch. (4) It was said earlier that complex tones that are mixed together each evoke a pitch only if they contain resolved partials. However, the meaning of ‘resolved’ in that context applied to the complex tones *in isolation*, before mixing. Obviously after mixing, the partials may no longer be resolved. The pitches may nevertheless be salient, so resolvability of partials *within the stimulus* is not crucial (Micheyl *et al.*, 2006b). (5) Finally, Micheyl and Oxenham (2004) reexamined the issue of a putative translation cost between distinct pitch mechanisms for resolved and unresolved stimuli, and concluded that there is none (see also Gockel *et al.*, 2004 and Micheyl and Oxenham, 2005). All these results suggest that peripheral resolvability *per se* is not what determines the salience of the pitch of a complex tone.

To summarize, stimuli with partials of low rank may evoke a pitch that is more salient and accurately discriminable than stimuli that only contain partials of high rank. This is often attributed to differences in resolvability of partials within the complex tones, but this interpretation does not fit some aspects of the data, and other interpretations have been proposed, such as that the duration of internal delays varies according to cochlear filter characteristic frequency (Moore, 1982; Bernstein and Oxenham, 2005; de Cheveigné and Pressnitzer, 2006). The debate is of importance to decide which strategies are used to hear pitch (Section 4.9).

4.8 Binaural pitch

Our ears sample the acoustic field in two points, and this helps us to localize sounds and make sense of complex acoustic scenes. Certain pitch phenomena require binaural interaction, and binaural hearing and pitch may actually have something in common. Both seem to be based on temporal cues analyzed within the brain, and two influential models of sound localization (Jeffress, 1948) and pitch (Licklider, 1951) both postulate neural processing based on time delays and coincidence counting. Experimentally, binaural stimulation adds a degree of freedom that may help us locate the site of pitch extraction within the brain.

Fascinating among auditory phenomena is *Huggins pitch*, which arises while listening with two ears to a binaural stimulus that sounds like noise when listened to with either ear alone (Cramer and Huggins, 1958; Culling, 1999). Huggins pitch is one of several binaurally created pitch phenomena (see Culling *et al.*, 1998 for a review). The stimulus consists of white noise that is identical at the two ears except for a narrow frequency region for which noise is decorrelated between the two ears (for example, the noise in one ear may undergo a phase transition of 2π over this frequency region) (Fig. 4.6). The percept resembles that of a narrowly filtered band of noise, embedded within a wideband noise background. Relatively faint, it nevertheless supports accurate matching (Hartmann, 1993). It becomes stronger if multiple transitions occur at frequencies that follow a harmonic series, forming a ‘Huggins complex tone’ (Bilsen, 1976). Huggins pitch supports melody (Akeroyd *et al.*, 2001), and streaming effects similar to those produced by pure tones (Akeroyd *et al.*, 2005). Huggins pitch necessarily arises from the interaction of neural patterns from the two ears, and this puts a constraint on the locus of pitch extraction (see Section 4.10 below).

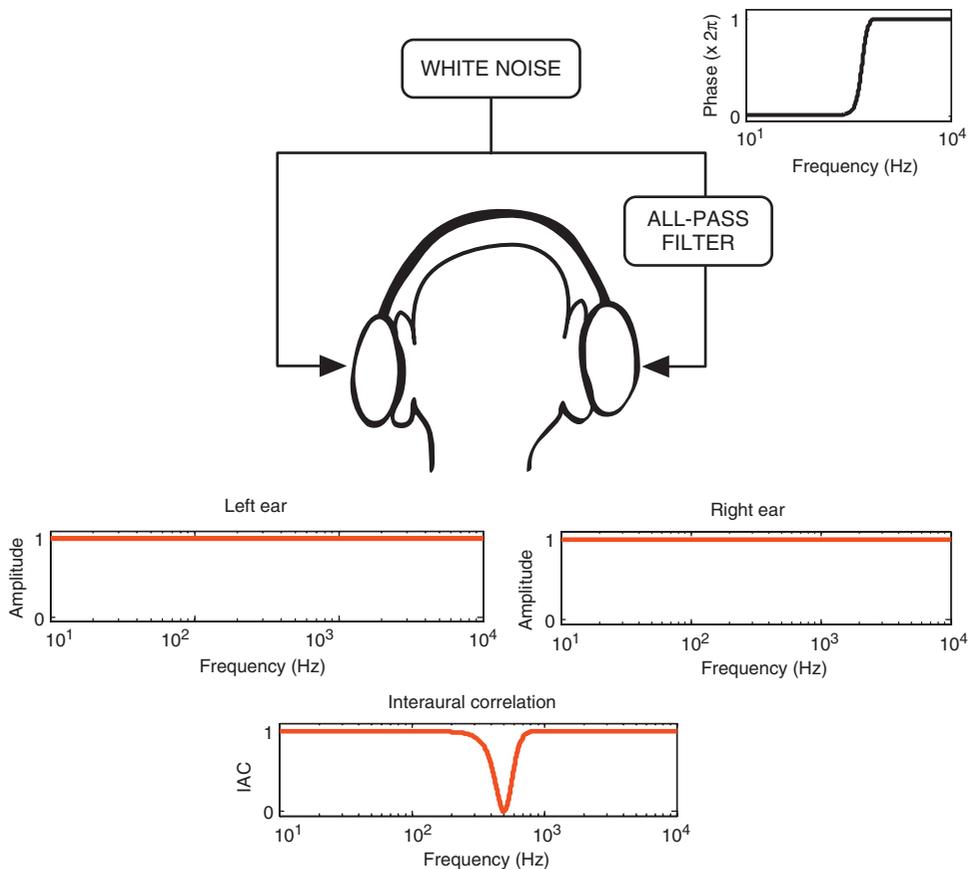


Fig. 4.6 Huggins pitch is obtained by presenting white noise to both ears. The noises at both ears are identical apart from a narrow phase transition created by an all-pass filter in the pathway to one ear. Interaural correlation is high, except at the frequency of the phase transition. The pitch matches this frequency.

Houtsma and Goldstein (1972) presented musically trained subjects with pairs of partials that formed ‘complex tones’ with an F_0 determined by the largest common divisor of their frequencies. They successfully performed an interval recognition task, showing that they could hear a low pitch related to this F_0 , regardless of whether the partials were sent both to the same or to different ears. This is a second example of central formation of pitch from information from both ears. Listening to either ear alone, one hears only a single partial and not the low pitch.

Binaural interactions can create pitch, but they can also weaken it. In a study mentioned earlier (Section 4.7), Bernstein and Oxenham (2003) found that F_0 discrimination thresholds of complex tones increased by an order of magnitude when the rank of the lowest harmonic was increased from the 9th to the 12th (Fig. 4.5). Interestingly, the same occurred when all harmonics went to both ears (diotic condition), or even and odd harmonics to opposite ears (dichotic condition). A complex with even harmonics of F_0 starting from the 12th is identical to a complex with all harmonics of $2 F_0$ starting from the 6th. Listening to that ear alone should therefore give a low threshold. The fact that thresholds were instead high implies that the subjects could not shut off

the contribution of the other ear. Surprisingly, the additional information from the odd harmonics within the other ear was deleterious rather than useful.

With a slightly different paradigm, in which stimuli were sent to the same ear within blocks, rather than to opposite ears on each trial as in their previous study, Bernstein and Oxenham (2008) did find a benefit of binaural presentation, but still not as great as if the ear receiving the odd harmonics could be ignored. Gockel *et al.* (2005) found that a mistuned partial presented contralaterally to the rest of a complex affected the low pitch, but less than when presented to the same ear. Similarly, Gockel *et al.* (1999) found that the interference produced by ‘fringes’ that preceded or followed a complex tone was reduced, but not abolished, by contralateral presentation.

Interestingly, the spatial percept also appears to be affected by interactions with pitch. Huggins pitch is hard to localize (Akeroyd and Summerfield, 2000). In the well-known ‘octave illusion’, Deutsch (1974) found that, when tones of 400 and 800 Hz were presented in alternation such that one ear received the lower when the other received the higher, subjects reported two tones pulsing, the lower tone at one ear and the higher at the other ear. The side that heard the higher tone depended on the subject’s handedness, and was subject to spontaneous reversals. Interactions between pitch and spatial hearing might arise if they shared physiological substrates (see Section 4.10).

4.9 How do we perceive pitch?

How we perceive pitch has been a matter of intense debate for many years (see de Cheveigné, 2005 for a review). It is not yet resolved, and so there is no authoritative explanation of how pitch emerges within the auditory system. The best that we can do is to try to understand the positions in this debate. Two properties need explaining: (1) the *sensitivity* of pitch to small changes in F0, and (2) the relative *invariance* of pitch to large changes in other stimulus parameters. The second property is less often considered, but it is just as important as the first.

According to the *place* hypothesis (Helmholtz, 1877), pitch is determined from the position of maximum excitation along the basilar membrane, within the cochlea. This hypothesis is attractive because it readily accounts for sensitivity: a change in F0 is necessarily accompanied by spectral differences that the excitation pattern should—cochlear frequency resolution permitting—reveal. However, the place hypothesis has a harder time accounting for invariance of pitch across stimuli with different spectra, for example the fact that a pure tone and complex tone can evoke the same pitch.

A pure tone evokes a localized peak of excitation along the basilar membrane, and its frequency could conceivably be discriminated on the basis of this cue, or other cues such as changes along the flanks of the peak of excitation as it shifts with frequency. There are some issues with this idea: excitation patterns measured physiologically are rather broad. They tend to broaden and shift with increasing intensity, without commensurable changes in pitch or discrimination acuity (Chatterjee and Zwillocki, 1997). The parameter dependence of pure-tone pitch discrimination does not fit what is expected of an excitation pattern-based cue (e.g. Moore, 1973, Moore and Sek, 1995, 1998). These and other considerations argue against the place hypothesis, even in the case of pure-tone pitch, except for relatively high frequencies (above about 5 kHz). However, the main problem is that the place hypothesis fails to explain how pure and complex tones might have the same pitch. A complex tone typically evokes multiple peaks, one for each resolved partial (or group of unresolved partials), so the hypothesis needs amending to address this situation. The solution proposed by Ohm (1843) and Helmholtz (1877) was to assume that the pitch of the complex is determined by the peak associated with its *fundamental* partial. That solution floundered on the missing-fundamental phenomenon mentioned previously.

According to the *time* hypothesis, pitch is derived from the periodic pattern of the acoustic waveform, transduced by the cochlea into a pattern of nerve pulses that is processed by the brain. The appeal of this hypothesis is that pitch maps more directly to the stimulus period than to spectral features such as the—possibly missing—fundamental. However, the hypothesis is also incomplete: we still need to explain how the ear reliably extracts one pulse per period, how the pulses are transmitted to the brain, and how the brain counts them. The second point was once contentious because nerve fibers cannot transmit spikes at rates beyond a few hundred spikes per second. This issue was resolved by Wever and Bray (1930) who pointed out that higher rates can be transmitted collectively by groups of fibers. Today it is accepted that periodicities may be coded by the instantaneous probability of spikes within groups of auditory nerve fibers (up to about 5 kHz in cat, Johnson, 1980).

The main difficulty with the time hypothesis is that it is not easy to extract one pulse per period, in a way that is reliable and fully general. If we were dealing only with pure tones, then we could postulate some mechanism that triggers a pulse on a peak, or a zero-crossing of the waveform. However, complex tones often have several peaks and/or zero-crossings per period. Furthermore, the position and number of such cues is highly phase-dependent, which is hard to reconcile with the largely phase-independent nature of pitch (Wightman, 1973*a*). In their simplest form, place and time hypotheses both have insurmountable difficulties in accounting for pitch. Both explain sensitivity to variations in period, but not invariance across stimuli with same period. The debate has now shifted to two newer models, *pattern matching* and *autocorrelation*.

According to the pattern-matching hypothesis, pitch is associated with the harmonic pattern of the partials. The ear is assumed to contain a dictionary of *harmonic templates*, against which the incoming patterns of frequencies are compared. The template that best matches the pattern indicates the pitch. Pattern matching was proposed by de Boer in his thesis (1956) and later promoted by Wightman (1973*b*), Goldstein (1973), and Terhardt (1974); but the seeds of the idea were already in Helmholtz's concept of *unconscious inference*, according to which perception proceeds by matching internal models against incoming sensory evidence. Helmholtz, himself in turn borrowed it from Alhazen who had formulated it in the 11th century (Hatfield, 2002).

Pattern matching allows the place hypothesis to be salvaged by assuming that individual partials (not just the fundamental) give rise to local peaks in excitation along the basilar membrane. For that reason it is sometimes equated with the 'place' hypothesis. However, pattern matching is also compatible with the time hypothesis, if one supposes that frequencies of individual partials are extracted from the temporal patterns that they produce locally at different points along the basilar membrane (supposing that they are resolved). The pattern-matching mechanism ensures invariance across stimuli of differing spectra (but the same F_0), and in particular it solves the 'paradox' of the missing fundamental. For example, a set of partials at 200, 300, and 400 Hz would trigger the same harmonic template as a 100-Hz pure tone, or any other complex of same period.

The pattern-matching hypothesis works if enough partials are resolved so as to constrain the choice of template (Section 4.7). However, pitch is also known to arise for stimuli for which there are no resolved partials. Pattern matching cannot account for such a pitch. This would bring us to discard the hypothesis, if it were not for three conjectures that might save it. The first is that human cochlear frequency resolution might somehow be finer than usually assumed (e.g. Oxenham and Shera, 2003). The second is that stimuli with non-resolved partials might produce distortion products that are resolvable. The third is that pattern matching might coexist with some other mechanism (see below).

The *autocorrelation* hypothesis differs from the time hypothesis in that it does not require spikes to be triggered at a well-defined position within the period. Rather, the periodic neural

pattern is processed by coincidence-detector neurons that calculate the equivalent of an autocorrelation function (Licklider, 1951, 1959; Meddis and Hewitt, 1991*a,b*; Cariani and Delgutte, 1996*a,b*). The spike trains are delayed within the brain by various time lags (using neural delay lines) and combined or correlated with the original. When the lag is equal to the time delay between spikes the correlation is high and outputs of the coincidence detectors tuned to that lag are strong. Spike trains in each frequency channel are processed independently and the results combined into an aggregate pattern (Fig. 4.7). In response to a periodic tone, a ridge appears in the pattern at a lag equal to the period, and this is the cue to pitch. This cue appears for stimuli with *unresolved* partials because the partials beat together at the fundamental (Fig. 4.7, right column). It also appears for stimuli with only *resolved* partials (which produce no beating at the fundamental period in any peripheral channel). This happens because all partials are multiples of the same F_0 , and therefore share a common period multiple equal to $1/F_0$ (Fig. 4.7, middle column).

The autocorrelation process is insensitive to phase, and this addresses the objection against time models mentioned earlier. Phase sensitivity may nevertheless arise from non-linearities in its

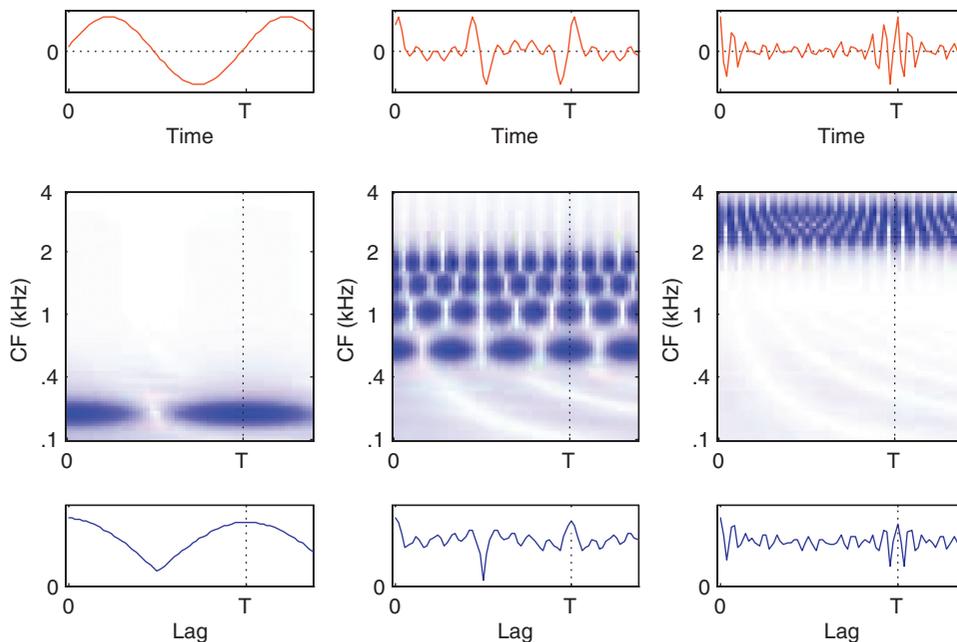


Fig. 4.7 Autocorrelation model of pitch. Top: acoustic waveforms, middle: array of autocorrelation functions (ACFs); bottom: summary autocorrelation functions (SACFs). Stimuli were: 200-Hz pure tone (left column), 200-Hz complex tone with partials 3, 5, 7, 9 (central column), and, 200-Hz complex tone with partials 12, 13, 14, 15, 16 (right column). For the central column, I chose a complex tone with well-spaced partials to illustrate the point that salient within-channel F_0 cues are not necessary to produce a salient F_0 cue in the overall pattern (middle row) or summary (bottom row). The autocorrelation model was modeled with a linear gammatone filterbank followed by half-wave rectification, calculation of a running autocorrelation function, cubic root compression, and summation over filter channels. The cue to pitch is the position of a ridge across channels in the ACF pattern (middle) or a peak in the SACF (bottom).

physiological implementation, see Section 4.10. Coincidence detection is plausible in terms of known physiology, but the hypothesis also requires *neural delays* of up to about 30 ms (to cover the range of musical pitch down to 30 Hz). There is little direct evidence for neural delays that long (Winter, 2005, but see de Cheveigné and Pressnitzer, 2006). Autocorrelation is theoretically related to pattern matching (de Cheveigné, 2005), and indeed it could be proposed that autocorrelation is the way pattern matching is implemented in the brain: the two hypotheses are not mutually exclusive.

An argument sometimes made against autocorrelation is that it works too well: it predicts that pitch should be equally salient for stimuli with resolved and unresolved partials, whereas we saw that such is not the case. This led to the *multiple-mechanism hypothesis*, already mentioned in Section 4.7, according to which pattern matching explains the strong pitch of stimuli with resolved harmonics, and autocorrelation the weaker pitch of stimuli with only unresolved harmonics (or electrical stimulation in cochlear implantees). The alternative to two mechanisms is that a *unitary model* can account for all aspects of pitch (Meddis and Hewitt, 1991*a,b*; Meddis and O'Mard, 1997). This debate is fueled by recent work on the psychophysics of resolved vs. unresolved stimuli (Section 4.7).

In addition to these main theories (place, time, pattern matching, autocorrelation), there are many variants such as the strobed temporal integration model of Patterson *et al.*, (1992) or the cancellation model of de Cheveigné (1998). They address the two main aspects of pitch mentioned at the beginning of this section: (1) sensitivity of pitch to F0; (2) invariance across stimuli with the same F0. Other aspects that also need explaining are: (3) how we perceive pitch in the presence of other sounds; (4) musical properties such as harmony; and (5) the detailed aspects of pitch reported in the psychophysics literature. We are unfortunately rather far from a complete answer to the question 'How do we perceive pitch?'

4.10 The physiological basis of pitch perception

Sound entering the ear is transduced within the cochlea into neural patterns that are processed at several stages within the brainstem, thalamus, and cortex. Each of these stages could be involved in pitch perception, either as processing stage or as a relay of relevant information.

The cochlea is sometimes likened to a 'spectrum analyzer' that transforms the sound waveform into a spectrum coded as a profile of discharge rate across the auditory nerve. However, the auditory system also has access to the temporal patterns (so-called *fine structure*) at the output of each filter, and one could propose instead that the role of the cochlea is to transduce acoustic vibrations into temporal patterns of neural firing. According to this hypothesis, the role of cochlear selectivity would be to improve the quality of transduction and assist scene analysis (de Cheveigné, 2001). These two views (spectrum analyzer vs. frequency-selective transducer) map to the 'place' and 'time' hypotheses. How do they fit with what we now know about cochlear filter properties?

Our knowledge is inferred mainly from psychophysical studies in humans (Patterson, 1976; Glasberg and Moore, 2000) and measurements from the basilar membrane or the auditory nerve in animals (Ruggero, 1992; Robles and Ruggero, 2001; Cedolin and Delgutte, 2005). The results are roughly consistent across species (Ruggero and Temchin, 2005), but there are wide differences in estimates of cochlear filter bandwidth depending on the technique used. For example, bandwidths measured psychophysically (in humans) in forward-masking experiments are narrower than those measured with simultaneous masking by a factor of up to two (Oxenham and Shera, 2003). Likewise, auditory nerve fiber tuning curves measured with pure tones are considerably

narrower than transfer functions estimated by the reverse correlation technique using noise stimuli (Carney and Yin, 1988). This is a problem for our purpose, because the plausibility of different hypotheses depends crucially on the available selectivity.

We can attempt to make sense of these conflicting estimates by recalling that cochlear filtering involves a non-linear active process. In response to a weak *isolated pure tone* at the best frequency of the measurement site, the gain of the cochlear amplifier is large. The gain decreases as the level of the tone increases, implying a concomitant reduction in selectivity (Robles and Ruggero, 2001). The gain also decreases if a second, off-frequency pure tone is added to the on-frequency probe, a phenomenon known as ‘two-tone suppression’ (Ruggero *et al.*, 1992; Jülicher *et al.*, 2001). Thus, pure-tone tuning curves may reflect a sharp selectivity that is available only for isolated pure tones at threshold, and not for more complex stimuli. This would explain the much wider estimates obtained with reverse correlation using wide-band noise stimuli, and also possibly the discrepancy between psychophysical estimates of selectivity from simultaneous and forward masking (Oxenham and Shera, 2003). When speaking of cochlear selectivity as applies to pitch, we must be careful to distinguish between the case of isolated pure tones and that of individual partials of complex tones. Selectivity is less good for the latter.

How does this relate to our different pitch theories? The *place* hypothesis assumes a peak of excitation along the basilar membrane. This is plausible for pure tones at low levels, but at higher levels the peak becomes broader and tends to shift towards a lower frequency place. By contrast, the pitch is rather stable with increasing intensity. For complex tones, it is unlikely that accurate estimates of the frequencies of individual partials, required by the pattern-matching hypothesis, can be derived from peaks in a rate-place representation (but see Cedolin and Delgutte, 2005). They could, however, be derived from *temporal cues*, supposing that each partial is resolved at some point along the basilar membrane so that its periodicity can be measured without interference from its neighbors (Section 4.7, see also Fig. 4.7 middle column). It has also been suggested that the *phase* characteristics of the cochlear filter may contribute to the estimation of the frequencies of individual partials (Shamma and Klein, 2000). Cochlear selectivity is not essential for the *time* and *autocorrelation* hypotheses, but it may facilitate hearing out the pitch of a sound in the presence of competing sounds by improving the signal-to-noise ratio within individual channels (de Cheveigné, 2001). To summarize, cochlear frequency analysis certainly plays an important role in pitch perception, but there is little support for the idea that pitch is derived from peaks in a place-rate representation.

At each point along the basilar membrane, the acoustic stimulus gives rise to vibrations that are transduced into a pattern of firing (‘spikes’) within the auditory nerve. The occurrence of each spike is random, but the *instantaneous probability* of occurrence is not: it follows roughly the half-wave rectified waveform of the mechanical vibration at each locus of transduction. The shape of that vibration reflects the stimulus (spectrum, amplitude), but also the filtering and non-linear properties of the cochlea. For a periodic stimulus, the overall discharge pattern is periodic in terms of instantaneous probability.

For a *pure tone*, basilar membrane vibration is sinusoidal, narrowly localized for a low-amplitude tone, and more spread out for a louder tone. The phase of vibration varies along the basilar membrane, slowly from stapes to just before the locus of maximal sensitivity, and more rapidly thereafter. Thus, there is a phase shift across the population of fibers that respond to a pure tone. The pitch of a pure tone could be derived by measuring the period of the discharge probability, or by locating the point of most rapid phase transition (this is a ‘time’-based version of the place hypothesis) (Shamma and Klein, 2000).

For a *complex tone*, there are three cases of interest. A locus that responds to a partial of low rank may vibrate sinusoidally at that partial’s frequency (Fig. 4.4 bottom left). A locus that

responds to a *combination tone* produced by cochlear distortion may likewise vibrate sinusoidally at the frequency of that combination tone. Other loci, which respond to multiple partials, may vibrate with a complex waveform with an envelope period equal to the fundamental (Fig. 4.4, bottom, middle, and right panels). Thus there are multiple temporal cues within the discharge pattern of auditory nerve fibers that could support the pitch of a complex tone.

Phase-locking of spike trains to the stimulus decreases as stimulus frequency is raised, and is no longer measurable beyond about 5 kHz in cats (Johnson, 1980). The limit is lower in guinea-pigs (Palmer and Russell, 1986), higher in the barn owl (9 kHz; Köppl, 1997), and unknown in humans (Fig. 4.8). A small synchronization index does not necessarily imply that all temporal information

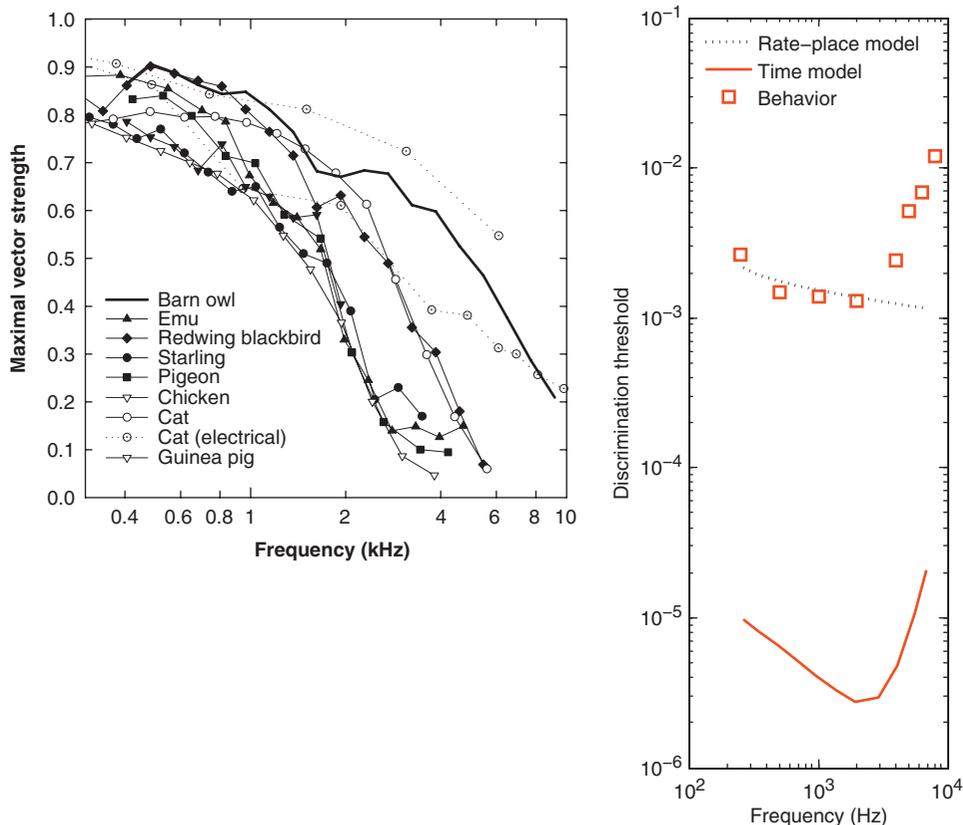


Fig. 4.8 The synchronization index (or vector strength) is used as a measure of quality of the temporal information carried by auditory nerve fiber discharge patterns. It is plotted here as a function of stimulus frequency for several species (reproduced from Köppl, 1997, with permission). A higher value indicates a better representation of a pure tone of that frequency. Data labeled 'electric' are for electric stimulation of the auditory nerve (see Köppl, 1997, for references). Right: human pure-tone frequency discrimination thresholds (symbols) and predictions by a place-rate model (dashed line) and a temporal model fit to cat synchronization data (full line), replotted from Heinz *et al.* (2001). Loss of synchrony at high frequencies produces a degradation of predicted thresholds that parallels that observed behaviorally in humans. However, the predicted thresholds are two orders of magnitude too good: to account for this discrepancy would require the additional assumption of a uniformly inefficient processing of temporal information. The place-rate model does not predict higher thresholds at high frequencies.

is lost: a modeling study based on Johnson's cat data found that residual temporal information might be useful up to 10 kHz (Heinz *et al.*, 2001). As one proceeds to higher relays within the auditory system, synchrony is limited to yet lower frequencies.

Where are temporal cues processed? Obviously this can only occur at a level within the nervous system where they are accurately represented. Auditory nerve fibers terminate within the cochlear nucleus, and from there they are relayed to a series of nuclei within the brainstem and midbrain, that themselves project to the auditory thalamus and cortex. Specializations for time are observed at several levels. For example, within the cochlear nucleus (CN), bushy cells are fed from auditory nerve fibers via large synapses, so that their activity resembles that of their afferents with little loss of temporal accuracy (so-called 'primary-like' response). Other cells (stellate-D and octopus)

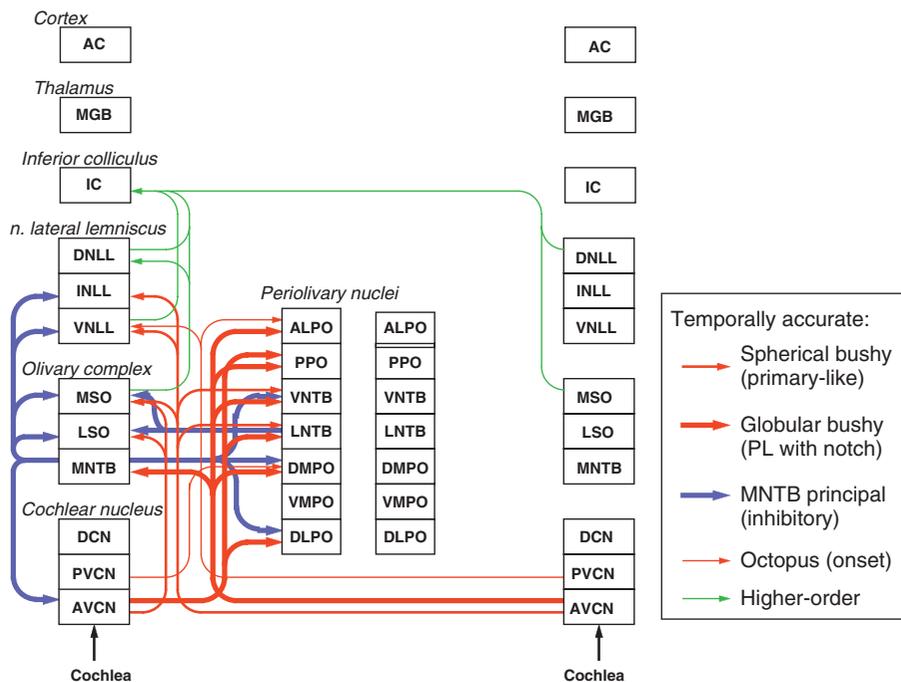


Fig. 4.9 Temporally specialized pathways within the auditory brainstem and midbrain. Spike patterns from the cochlea are relayed by several types of cell within the cochlear nucleus (CN): globular bushy (thick red) and spherical bushy (medium red), with firing patterns similar to auditory nerve fibers. Globular bushy cells feed principal cells of MNTB and LNTB that drive temporally accurate inhibitory pathways (blue). Octopus cells (thin red) have a temporally accurate onset response. Not shown on this schema, certain multipolar (stellate) cells within AVCN have a temporally accurate onset response (onset-C). Only pathways that feed nuclei on one side of the head are shown here (the same pathways exist on both sides). Pitch extraction could, in principle, occur at any level that receives accurate temporal information. Abbreviations: AC, auditory cortex; MGB, medial geniculate body; IC, inferior colliculus; DNLL, INLL, VNLL, dorsal, intermediate, and ventral nuclei of the lateral lemniscus; MSO, LSO, medial and lateral superior olive; MNTB, VNTB, LNTB, medial, ventral, and lateral nuclei of the trapezoid body; ALPO, PPO, DMPO, VMPO, DLPO, anteriolateral, posterior, dorsomedial, ventromedial, and dorsolateral periolivary nuclei; DCN, PVCN, AVCN, dorsal, posterioventral, and anterioventral cochlear nuclei. From Schwartz, (1992); Helfert and Aschoff, (1997); Thompson and Schofield, (2000).

discharge accurately on the onset of a stimulus, or at each period of wideband stimuli such as click trains. These cells project to multiple nuclei within the superior olivary complex (SOC) and lateral lemniscus (LL) (Fig. 4.9).

Spherical bushy cells within the CN project bilaterally to the medial superior olive (MSO) and ipsilaterally to the lateral superior olive (LSO), as well as to other nuclei of the SOC and LL. Globular bushy cells within the CN project to the contralateral medial nucleus of the trapezoid body (MNTB) via secure synapses that ensure spikes are relayed reliably and with low jitter. In turn, MNTB neurons, which are inhibitory, project to ipsilateral LSO and other nuclei within the SOC and to the ventral nucleus of the lateral lemniscus (VNLL), which also receives input from octopus cells of the CN via large synapses. Octopus cell projections to the VNLL are numerous in humans (Adams, 1997). All these nuclei, and others within the brainstem, receive input from temporally specialized CN cells (Thompson and Schofield, 2000). Some of these nuclei, such as MSO, MNTB, and LSO, are thought to subservise binaural processing, but this does not preclude them from other tasks that involve temporally accurate patterns, such as pitch processing. Ascending projections from these nuclei mostly terminate in the inferior colliculus (IC) (Ehret, 1997).

Cells within IC and beyond synchronize to stimulus periodicities of at most a few hundred Hertz (Liu *et al.*, 2005). Therefore it is likely that fast temporal processing is performed at a lower level: *subcollicular* nuclei are potential substrates for the signal processing operations required by pitch.

Tonotopically organized fields exist at all levels up to the thalamus and cortex, reflecting orderly projections from the cochlea, but I argued earlier that pitch is unlikely to emerge from a place–rate representation. Neural activity follows stimulus periodicity up to at most 1000 Hz in the IC, 1200 Hz in the thalamus or 250 Hz in the cortex (Liu *et al.*, 2005; Wallace *et al.*, 2007). However, most neurons have cutoffs well below these limits, and furthermore these ranges cover only part of the range of pitch periodicities. Relaying and processing temporally accurate spike trains entails a cost (in terms of specialized circuitry and metabolism), and it is likely that pitch processing occurs at an early level, possibly as early as the dendritic fields of CN neurons that receive input from the auditory nerve. However, the existence of Huggins pitch suggests a locus beyond the level of binaural interaction.

On the assumption that pattern matching is a ‘high-order’ operation, it is sometimes proposed that pattern matching is performed by secondary auditory cortical fields operating on the output of a tonotopically organized primary cortical field. It is more likely that cortical responses elaborate (and possibly recapitulate) pitch-relevant features (or conjunctions of features) extracted at subcortical levels.

It is frustrating that little direct evidence has been found for any particular locus, or model. Signatures of a pitch extractor, expected from pattern matching and autocorrelation alike, are (1) sensitivity to changes in F0, and (2) invariance across other stimulus dimensions. The first property is readily observed, but alone it is not sufficient to signal pitch. Evidence for a cortical ‘pitch centre’ has been reported based on cortical recordings in animal or brain imaging in humans (Patterson *et al.*, 2002; Bendor and Wang, 2006), but methodological issues complicate the interpretation of experimental results (McAlpine, 2004; Hall and Plack, 2008; Nelken *et al.*, 2008). A recent report describes single unit responses from presurgical recordings from electrodes implanted in the brains of epileptic patients. Tuning to pure tones (embedded within 3-tone random chords) was extremely narrow, limited only by the resolution of stimulus sampling (1/18th of an octave) (Bitterman *et al.*, 2008). The protocol did not test invariance, and thus we cannot exclude that those responses merely reflect a remarkably selective tonotopy, but the sharp frequency resolution is evocative of a sensitivity to pitch.

Physiologists have obviously looked hard for proof of the various pitch theories mentioned earlier, such as Shamma and Klein's (2000) pattern matching model, Licklider's autocorrelation model (Licklider, 1951; Meddis and Hewitt, 1991a), or de Cheveigné's (1998) cancellation model. Direct evidence is still lacking, although most of the ingredients (lateral inhibition, within- and cross-channel coincidence, inhibition, etc.) are ubiquitous. Autocorrelation and cancellation models require delays of up to about 30 ms to accommodate a lower limit of pitch of 30 Hz (Pressnitzer *et al.*, 2001). Evidence for appropriate delays is fragmentary (Behrend *et al.*, 2002; Nayagam *et al.*, 2005), although it has been suggested that delays might arise indirectly from cross-channel interaction (de Cheveigné and Pressnitzer, 2006). Numerous cells have been found to be sensitive to stimulus features relevant for pitch, but they are usually also sensitive to irrelevant features, or otherwise disorderly, making them hard to relate to known models. For example, onset cells within the cochlear nucleus fire accurately to each period of certain stimuli (Winter, 2005), but their phase sensitivity makes them poor candidates for pitch. Winter *et al.* (2001) and Wiegrebe and Meddis (2004) proposed that arrays of periodicity-tuned chopper cells in the cochlear nucleus are 'read out' by coincidence cells in IC. However, the best frequencies of those cells cover a limited range (100–500 Hz) and their properties are level-dependent at low levels (Winter, 2005).

Physiological correlates of pitch processing might escape observation, for example because they are technically hard to measure. It also could be that we *have* observed the correlates of pitch processing, but that they follow a principle that we do not understand. Or it might be the case that pitch is a human trait which is *not* shared by animal models. For example, the selectivity of single neuron responses observed by Bitterman *et al.* (2008) within human auditory cortex is not often reported in animal models.

To summarize, the auditory system is equipped with much neural circuitry to process spectral and temporal features relevant for pitch, but exactly where and how this occurs is still a mystery. The answer may come from progress in recording and imaging techniques in animal models and humans, or it may come from theoretical and modeling efforts to make sense of data that are already available.

4.11 Methods and tools for the study of pitch

Students of pitch are sometimes dismayed, on reading a paper, to find so much devoted to arcane issues of methodology. It is worth understanding what these issues are, if only to filter them out and focus more easily on pitch. It is also important to be able to judge, based on its methodology, whether a study is credible or not.

Introspection once was our only tool to probe the perceptual reality of pitch. Its drawback is that introspection cannot be communicated reliably. Without an external reference to calibrate what one hears, and compare it with others, the researcher is vulnerable to his or her subjectivity, and disagreements are hard to resolve. Psychophysics provides tools to insure that perceptual phenomena are real and general, and codified in a form that can be shared. Careful control of stimuli and design of task ensure that the quantity probed in the experiment is that intended by the experimenter. Objective procedures and signal detection theory (Green and Swets, 1974) allow the performance limits of a subject to be factored from the effects of response bias and criteria. Statistical tests protect us from over-interpreting random quirks in the data, and they can also give some indication about whether patterns observed reflect the idiosyncrasies of the subjects used in the experiment, or instead are of wider validity. These tools help produce results that are reproducible and credible. The downside is that the task (remembering instructions, pushing buttons, etc.) and stimulus repetition may interfere with the

process of perception. Phenomena that do not fit the requirements of these tools may be overlooked.

Psychophysics allows strong inferences to be drawn on mechanisms within the brain, but it is no substitute for direct observation of neural activity. The most detailed observations are made in animal models using invasive techniques, such as the recording of the electrical activity of single or multiple neural units in response to the kind of stimuli that would produce pitch in humans. Other techniques involve measuring local field potentials or optical correlates of neural activity. With a background of knowledge of the brain from a variety of techniques, physiology in animal models gives us the most detailed picture of the neural processes that could be involved in pitch processing (Cariani and Delgutte, 1996*a,b*; Tramo *et al.*, 2005; Winter, 2005). An obvious drawback is that animals may differ from us in some important way in their experience of pitch, or in their processing of pitch-evoking sounds, or both. It is hard to do psychoacoustics in animals (see Chapter 7, this volume) and impossible to get them to describe what they heard, so we cannot be sure whether an observed response reflects pitch.

Brain imaging allows observation of brain activity in the species that interests us most: our own. *Structural* magnetic resonance imaging (MRI) has found correlations between the size of anatomical structures and pitch-related abilities (Gaser and Schlaug, 2003; Schneider *et al.*, 2005). *Functional* magnetic resonance imaging (fMRI) and positron emission tomography (PET) have been used to investigate pitch-related activity within the brain (Griffiths *et al.*, 1998; Patterson *et al.*, 2002; Warren *et al.*, 2003; Griffiths, 2005). Evidence has been found for a 'pitch centre' in lateral or anterolateral Heschl's gyrus (Patterson *et al.*, 2002; Penagos *et al.*, 2004), but this result has recently been questioned (Hall and Plack, 2008). The spatial resolution of fMRI (on the order of 1 cm) allows gross localization of neural activity, but its limited temporal resolution (about 10 s) constrains conclusions concerning the sequence of neural events (Griffiths, 2005).

Electroencephalography (EEG) and magneto-encephalography (MEG) offer better temporal resolution than fMRI, on the order of 1 ms. The onset response to periodic tones (pure or complex) includes a component (N100m) with a latency close to 100 ms that varies systematically with the period (Forss *et al.*, 1993; Stufflebeam *et al.*, 1998; Crottaz-Herbette and Ragot, 2000; Roberts *et al.*, 2000; Lütkenhöner *et al.*, 2001; Seither-Preisler *et al.*, 2006). However, similar onset responses occur for sounds that do not evoke pitch. It is thus imprudent to claim that N100m latency 'codes' pitch, or any other stimulus parameter or qualia. Presenting the pitch-evoking stimulus preceded by a noise-like waveform with similar spectral envelope allows a 'pitch onset response' (POR) to be separated from the generic onset response (Krumbholz *et al.*, 2003; Gutschalk *et al.*, 2004). POR latency and amplitude vary with stimulus period and degree of periodicity, suggesting a pitch-specific response. However, similar responses were observed for transitions between stimuli that differ in regularity but do not evoke a pitch (Chait *et al.*, 2007, 2008). MEG claims relatively good spatial resolution, and has been used to find evidence for a 'tonotopic' or 'periodotopic' organization of human auditory cortex (Romani *et al.*, 1982; Pantev *et al.*, 1988; Langner *et al.*, 1997). Results of different studies are unfortunately contradictory. Localization of sources is highly dependent on a single-dipole model that is not sufficiently accurate to allow such fine-grained conclusions (Lütkenhöner, 2003; Lütkenhöner *et al.*, 2003).

In summary, despite their considerable popularity and authority, non-invasive brain imaging techniques offer only limited insight into pitch perception mechanisms. Issues are limited spatial or temporal resolution, high levels of noise and variability in measured responses, bias towards structures and phenomena that are easy to image, the uncertain relation between the quantities measured (BOLD in fMRI, currents in dendrites of large populations of pyramidal neurons for

MEG and EEG) and relevant activity within the brain, and the need for sophisticated models and statistical procedures to make sense of the data.

Invasive recording techniques potentially provide a more detailed picture of human brain activity (Liégeois-Chauvel *et al.*, 1994; Lachaux *et al.*, 2003; Bitterman, 2008; Schnupp and King, 2008; Schönweiser and Zatorre, 2008). As part of presurgical protocol in epileptic patients, arrays of subdural electrodes are placed on the surface of the cortex (similar to EEG but without the deleterious effect of the high-impedance skull), or depth electrodes are inserted to record both local field potentials and individual neural units within the brain. Proximity to structures of interest, and favorable signal-to-noise ratio, give them an advantage over non-invasive recording techniques, at the expense of severe constraints on the availability of patients and the sampling of brain areas of interest. In common with brain imaging, invasive recording techniques offer only a sparse sampling of the complex activity within the brain.

Models and theories are important tools, indispensable to guide the design of experiments, and stitch their fragmentary results together into a picture that ‘makes sense’ to our understanding. Models of pitch are reviewed by de Cheveigné (2005). Special mention must be made of software models that allow complex experimental data to be compiled and confronted against theoretical hypotheses. A promising new trend is the use of theoretical neuroscience and machine-learning techniques to bridge the gap between experimental data and high-level functions involving cognition and action. Unfortunately it may be difficult for the student of pitch to judge the validity of theories that require mathematical sophistication, or that are embodied by the lines of a computer program. Modeling, to be useful, requires care paid to specification (*what* does the model do?) and pedagogy (*how* does it work?).

Applications, such as automatic speech recognition, music processing, or sensory prostheses, obviously benefit from our understanding of hearing. They also contribute to understanding in an important way: by testing the validity of our hypotheses in the context of ‘real world’ tasks.

As mentioned earlier, the arcane methodological details of many studies on pitch may disorient the reader interested in pitch. Complex-tone stimuli are often presented without their fundamental component. Originally the aim was to probe effects beyond those expected from a classic ‘place’ explanation, an aim that makes sense within the context of the ‘missing fundamental’ debate (Section 4.5). That debate has subsided, but excluding the fundamental component is still useful in physiological studies to distinguish between the selectivity to pure-tone frequency that is ubiquitous in the auditory system, and sensitivity to pitch. Complex tones may also be stripped of additional low-order harmonics (e.g. below the 10th). The aim here is to remove resolved partials so that the stimuli offer only ‘temporal’ cues, as opposed to spectral cues usable by a pattern-matching mechanism. This precaution is relevant in the context of the ‘resolvability’ debate (Section 4.7).

Without additional precautions, these efforts may be compromised by non-linear distortion products that arise within the cochlea. For example, *difference tones* produced by pairs of partials ($f_2 - f_1$, where f_1 and f_2 are partial frequencies) may introduce components at the fundamental and low harmonics (Pressnitzer and Patterson, 2001). Likewise, *cubic difference tones* ($2f_1 - f_2$) may introduce components just below the lowest harmonic of a set of high-frequency partials. To mask these distortion products, that behave like rogue stimulus components, noise is added in the frequency region where they occur. The noise may be white (flat spectrum), or pink (power varies as $1/f$), or adjusted for uniform masking or for equal thresholds in each cochlear channel (threshold-equalized noise, TEN, Moore, 2000). The aim in each case is to mask the distortion products while minimizing any deleterious effect on the stimulus itself.

Stimulus manipulations designed to affect pitch might also affect other perceptual dimensions. To rule out the possibility that a subject uses, for example, a change in loudness instead of pitch

as a cue for a task, the stimulus level may be *roved* between trials. Likewise, a change in F0 may produce physiological responses unrelated to pitch in physiological or brain imaging experiments, and therefore characterizing a genuine pitch effect may require testing over a range of pitch-producing stimuli (Hall and Plack, 2008).

In summary, we have many tools at our disposal to explore pitch. Each has ‘blind spots’ that we must understand, and that we may compensate for to some degree by combining them. Advances in tools and methods may induce major leaps in our understanding, as argued earlier by von Békésy and Rosenblith (1948). In particular, progress in the resolution and signal-to-noise ratio of recording and imaging techniques, coupled with careful use of standard behavioral techniques, might lead to a quantum leap in understanding pitch.

4.12 Conclusions

Pitch is an important quality of sound, the focus of intense inquiry and investigation since antiquity. Pitch is basic to two forms of behavior specific to humans: speech and music. Pitch is usually understood as a one-dimensional percept determined by the period of the stimulus (or its inverse, F0), and insensitive to changes along other stimulus dimensions. However, its complex role within music involves harmonic and melodic effects that go beyond this simple one-dimensional model.

There is still debate as to where, and how, pitch is extracted within the auditory system. Helmholtz’s influential idea that pitch is determined by the locus of maximal vibration within the cochlea is no longer accepted. Rather, it is more likely that pitch is extracted within the auditory nervous system on the basis of temporal patterns transduced from acoustic vibrations within the cochlea. The site of pitch extraction is unknown, but probably relatively peripheral within the lower auditory brainstem or midbrain where temporally accurate neural information is available. Multiple neural substrates appear to be specialized for time, but the shape of many of their responses to sound are complex and difficult to relate to existing models of pitch.

According to the pattern-matching hypothesis, the frequencies of individual partials are estimated and matched against a set of internal harmonic templates. Partial frequencies could be estimated from neural temporal patterns within neural channels within which they are isolated by cochlear filtering, and possibly also from temporal patterns between those channels. As such, pitch perception might be dependent on both the selectivity and the phase properties of a healthy cochlea. According to the autocorrelation hypothesis, the period is instead determined directly from the temporal pattern of nerve activity transduced by the cochlea, by a neural circuit involving an array of delays (or one tunable delay) and coincidence-detecting neurons. According to that hypothesis, cochlear selectivity would not be directly involved in period estimation, but it might be useful to isolate sources within noise and facilitate perception of their pitch. Alternatively, cochlear mechanics might contribute to create the necessary delays, which would then also depend on the healthy condition of the cochlea. Physiological and psychophysical investigations have failed to rule decisively in favor of either hypothesis, and it may be the case that pitch is extracted according to a mechanism that is yet to be discovered.

Investigation tools for pitch include psychophysics, electrophysiology in animal models, brain imaging in humans, and theoretical and engineering approaches to solve similar problems in artificial systems. The study of pitch requires special care paid to methodological questions (e.g. combination tones), that sometimes obscure the pitch-related issues and make the literature hard to read for the newcomer. Once these issues are understood, pitch unfolds itself as a fascinating field where there is still much to be learned.

Acknowledgements

Part of this chapter was written during a scholarship at St John's College, Cambridge, at the invitation of Ian Winter. Thanks to Chris Plack, Maria Chait, Bill Hartmann, Jennifer Linden, Mark Sayles, Marion Cousineau, Shihab Shamma, Laurent Demany, Brian Moore, Andrew Oxenham, and Adrian Fourcin for comments on previous drafts.

References

- Adams, J. C. (1997). Projections from octopus cells of the posteroventral cochlear nucleus to the ventral nucleus of the lateral lemniscus in cat and human. *Auditory Neuroscience* 3:335–50.
- Akeroyd, M. A. and Summerfield, A. Q. (2000). The lateralization of simple dichotic pitches. *Journal of the Acoustical Society of America* 108:316–34.
- Akeroyd, M. A., Moore, B. C. J., and Moore, G. A. (2001). Melody recognition using three types of dichotic-pitch stimulus. *Journal of the Acoustical Society of America* 110:1498–504.
- Akeroyd, M. A., Carlyon, R. P., and Deeks, J. M. (2005). Can dichotic pitches form two streams? *Journal of the Acoustical Society of America* 118:977–81.
- ANSI (1994). *ANSI S1.1-1994. American National Standard Acoustical Terminology*. New York: American National Standards Institute.
- ASA (1960). *Acoustical Terminology SI, 1-1960*. New York: American Standards Association.
- Assmann, P. F. and Paschall, D. D. (1998). Pitches of concurrent vowels. *Journal of the Acoustical Society of America* 103:1150–60.
- Bachem, A. (1948). Chroma fixation at the ends of the musical frequency scale. *Journal of the Acoustical Society of America* 20:704–5.
- Bachem, A. (1950). Tone height and tone chroma as two different pitch qualities. *Acta Psychologica* 7: 80–8.
- Beerends, J. G. and Houtsma, A. J. M. (1989). Pitch identification of simultaneous diotic and dichotic two-tone complexes. *Journal of the Acoustical Society of America* 85:813–19.
- Behrend, O., Brand, A., Kapfer, C., and Grothe, B. (2002). Auditory response properties in the superior paraolivary nucleus of the gerbil. *Journal of Neurophysiology* 87:2915–28.
- Bendor, D. and Wang, X. (2006). Cortical representations of pitch in monkeys and humans. *Current Opinion in Neurobiology* 16:391–9.
- Bernstein, J. G. W. and Oxenham, A. J. (2003). Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number? *Journal of the Acoustical Society of America* 113:3323–34.
- Bernstein, J. G. W. and 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* 117:3816–31.
- Bernstein, J. G. W. and Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: effects of stimulus level. *Journal of the Acoustical Society of America* 120:3912–28.
- Bernstein, J. G. W. and Oxenham, A. J. (2008). Harmonic segregation through mistuning can improve fundamental frequency discrimination. *Journal of the Acoustical Society of America* 124:1653–67.
- Bigand, E. and Tillmann, B. (2005). Effect of context on the perception of pitch structures. In *Pitch—Neural Coding and Perception* (ed. C. J. Plack, A. Oxenham, R. R. Fay, and A. N. Popper), pp. 306–51. New York: Springer.
- Bilsen, F. A. (1976). Pronounced binaural pitch phenomenon. *Journal of the Acoustical Society of America* 59:467–8.
- Bitterman, Y., Mukamel, R., Malach, R., Fried, I., and Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451:197–201.
- Bloothoof, G., Bringmann, E., van Cappellen, M., van Luipen, J. B., and Thomassen, K. P. (1992). Acoustics and perception of overtone singing. *Journal of the Acoustical Society of America* 92:1827–36.
- Bregman, A. S. (1990). *Auditory Scene Analysis*. Cambridge, Mass: MIT Press.

- Burns, E. M. and Viemeister, N. (1976). Nonspectral pitch. *Journal of the Acoustical Society of America* **60**:863–9.
- Burns, E. M. (1981). Circularity in relative pitch judgments for inharmonic complex tones: the Shepard demonstration revisited, again. *Perception and Psychophysics* **30**:467–72.
- Burns, E. M. (1982). Pure-tone pitch anomalies. I. Pitch-intensity effects and diplacusis in normal ears. *Journal of the Acoustical Society of America* **72**:1394–402.
- Burns, E. M. (1983). Pitch of sinusoids and complex tones above 10 kHz. In *Hearing—Physiological Bases and Psychophysics* (ed. R. Klinke and W. M. Hartmann), pp. 327–33. Berlin: Springer-Verlag.
- Cardozo, B. L. (1974). Some notes on frequency discrimination and masking. *Acustica* **31**:330–6.
- Cariani, P. A. and Delgutte, B. (1996a). Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *Journal of Neurophysiology* **76**:1698–716.
- Cariani, P. A. and Delgutte, B. (1996b). Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase-invariance, pitch circularity, rate-pitch and the dominance region for pitch. *Journal of Neurophysiology* **76**:1717–34.
- Carlyon, R. P. (1996a). Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. *Journal of the Acoustical Society of America* **99**:517–24.
- Carlyon, R. P. (1996b). Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics. *Journal of the Acoustical Society of America* **99**:525–33.
- Carlyon, R. P. and Shackleton, T. M. (1994). Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *Journal of the Acoustical Society of America* **95**:3541–54.
- Carney, H. and Yin, T. C. T. (1988). Temporal coding of resonances by low-frequency auditory nerve fibers: single fiber responses and a population model. *Journal of Neurophysiology* **60**:1653–77.
- Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (2002). Auditory phase opponency: a temporal model for masked detection at low frequencies. *Acta Acustica United with Acustica* **88**:334–47.
- Cedolin, L. and Delgutte, B. (2005). Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *Journal of Neurophysiology* **94**:347–62.
- Chait, M., Poeppel, D., de Cheveigné, A., and Simon, J. Z. (2007). Processing asymmetry of transitions between order and disorder in human auditory cortex. *Journal of Neuroscience* **27**:5207–14.
- Chait, M., Poeppel, D., and Simon, J. Z. (2008). Auditory temporal edge detection in human auditory cortex. *Brain Research* **1213**:78–90.
- Cramer, E. M. and Huggins, W. H. (1958). Creation of pitch through binaural interaction. *Journal of the Acoustical Society of America* **30**:413–17.
- Crottaz-Herbette, S. and Rago, R. (2000). Perception of complex sounds: N1 latency codes pitch and topography codes spectra. *Clinical Neurophysiology* **111**:1759–66.
- Chatterjee, M. and Zwislocki, J. J. (1997). Cochlear mechanisms of frequency and intensity coding. I. The place code for pitch. *Hearing Research* **111**:65–75.
- Cramer, E. M. and Huggins, W. H. (1958). Creation of pitch through binaural interaction. *Journal of the Acoustical Society of America* **30**:413–17.
- Culling, J., Summerfield, A. Q., and Marshall, D. H. (1998). Dichotic pitches as illusions of binaural unmasking. I. Huggins pitch and the binaural edge pitch. *Journal of the Acoustical Society of America* **103**:3509–26.
- Culling, J. (1999). The existence region of Huggins pitch. *Hearing Research* **127**:143–8.
- Darrigol, O. (2007). The acoustic origins of harmonic analysis. *Archive for History of Exact Sciences* **61**: 343–424.
- de Boer, E. (1956). On the ‘residue’ in hearing. Unpublished thesis, University of Amsterdam.
- de Boer, E. (1976). On the residue and auditory pitch perception. In *Handbook of Sensory Physiology*, Vol. V–3 (ed. W. D. Keidel and W. D. Neff), pp. 479–583. Berlin: Springer-Verlag.

- de Cheveigné, A. (1998). Cancellation model of pitch perception. *Journal of the Acoustical Society of America* 103:1261–71.
- de Cheveigné, A. (2001). The auditory system as a separation machine. In *Physiological and Psychophysical Bases of Auditory Function* (ed. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven), pp. 453–60. Maastricht, The Netherlands: Shaker Publishing BV.
- de Cheveigné, A. (2005). Pitch perception models. In *Pitch—Neural Coding and Perception* (ed. C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper), pp. 169–233. New York: Springer.
- de Cheveigné, A. and Pressnitzer, D. (2006). The case of the missing delay lines: synthetic delays obtained by cross-channel phase interaction. *Journal of the Acoustical Society of America* 119:3908–18.
- Demany, L. and Semal, C. (2002). Learning to perceive pitch differences. *Journal of the Acoustical Society of America* 111:1377–88.
- Deutsch, D. (1974). An auditory illusion. *Nature* 251:307–9.
- Douglas, K. M. and Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience* 10:915–21.
- Drayna, D., Manichaikul, A., de Lange, M., Snider, H., and Spector, T. (2001). Genetic correlates of musical pitch recognition in humans. *Science* 291:1969–72.
- Ehret, G. (1997). The auditory midbrain, a shunting yard of acoustical information processing. In *The Central Auditory System* (ed. G. Ehret and R. Romand), pp. 259–316. New York: Oxford University Press.
- Forss, N., Mäkelä, J. P., McEvoy, L., and Hari, R. (1993). Temporal integration and oscillatory responses of the auditory cortex revealed by evoked magnetic fields to click trains. *Hearing Research* 68:89–96.
- Gaser, C. and Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *Journal of Neuroscience* 23:9240–5.
- Giangrande, J., Tuller, B., and Kelso, J. (2003). Perceptual dynamics of circular pitch. *Music Perception* 20:241–62.
- Glasberg, B. R. and Moore, B. C. J. (2000). Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *Journal of the Acoustical Society of America* 108:2318–28.
- Gockel, H., Carlyon, R. P., and Micheyl, C. (1999). Context dependence of fundamental-frequency discrimination: lateralized temporal fringes. *Journal of the Acoustical Society of America* 106:3553–63.
- Gockel, H., Carlyon, R. P., and Plack, C. J. (2004). Across-frequency interference in fundamental frequency discrimination: questioning evidence for two pitch mechanisms. *Journal of the Acoustical Society of America* 116:1092–104.
- Gockel, H., Carlyon, R. P., and Plack, C. J. (2005). Dominance region for pitch: effects of duration and dichotic presentation. *Journal of the Acoustical Society of America* 117:1326–36.
- Gockel, H., Moore, B. C. J., Plack, C. J., and Carlyon, R. (2006). Effect of noise on the detectability and fundamental frequency discrimination of complex tones. *Journal of the Acoustical Society of America* 120:957–65.
- Gockel, H., Moore, B. C. J., Carlyon, R. P., and Plack, C. J. (2007). Effect of duration on the frequency discrimination of individual partials in a complex tone and on the discrimination of fundamental frequency. *Journal of the Acoustical Society of America* 121:373–82.
- 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* 54:1496–516.
- Green, D. M. and Swets, J. (1974). *Signal Detection Theory and Psychophysics*. New York: Krieger.
- Griffiths, T. D. (2005). Functional imaging of pitch processing. In *Pitch—Neural Coding and Perception* (ed. C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper), pp. 147–68. New York: Springer.
- Griffiths, T. D., Buchel, C., Frackowiak, R. S., and Patterson, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience* 4:633–7.

- Gutschalk, A., Patterson, R. D., Scherg, M., Uppenkamp, S. and Rupp, A. (2004). Temporal dynamics of pitch in auditory cortex. *NeuroImage* 22:755–66.
- Hall, D. and Plack, C. J. (2008). Pitch processing sites in the human auditory brain. *Cerebral Cortex* 19:576–85.
- Hartmann, W. M. (1978). The effect of amplitude envelope on the pitch of sine wave tones. *Journal of the Acoustical Society of America* 63:1105–13.
- Hartmann, W. M. (1993). On the origin of the enlarged melodic octave. *Journal of the Acoustical Society of America* 93:3400–9.
- Hartmann, W. M. (1997). *Signals, Sound, and Sensation*. Woodbury, NY: AIP Press.
- Hartmann, W. M. and Doty, S. L. (1996). On the pitches of the components of a complex tone. *Journal of the Acoustical Society of America* 99:567–78.
- Hartmann, W. M. and Goupell, M. J. (2006). Enhancing and unmasking the harmonics of a complex tone. *Journal of the Acoustical Society of America* 120:2142–57.
- Hartmann, W. M., Rakerd, B., and Packard, T. N. (1985). On measuring the frequency-difference limen for short tones. *Perception and Psychophysics* 38:199–207.
- Hatfield, G. (2002). Perception as unconscious inference. In *Perception and the Physical World: Psychological and Philosophical Issues in Perception* (ed. D. Heyer and R. Mausfeld), pp. 115–43. New York: John Wiley and Sons.
- Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001). Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve. *Neural Computation* 13:2273–316.
- Helfert, R. H. and Aschoff, A. (1997). Superior olivary complex and nuclei of the lateral lemniscus. In *The Central Auditory System* (ed. G. Ehret and R. Romand), pp. 193–258. New York: Oxford University Press.
- Helmholtz, H. von (1877). *On the Sensations of Tone* (English translation A. J. Ellis, 1st edn, 1885; 2nd edn, 1954). New York: Dover.
- Henning, G. B. and Grosberg, S. L. (1968). Effect of harmonic components on frequency discrimination. *Journal of the Acoustical Society of America* 44:1386–9.
- Houtsma, A. J. M. (1981). Noise-induced shifts in the pitch of pure and complex tones. *Journal of the Acoustical Society of America* 70:1661–8.
- Houtsma, A. J. M. and Goldstein, J. L. (1972). The central origin of the pitch of complex tones. Evidence from musical interval recognition. *Journal of the Acoustical Society of America* 51:520–9.
- Houtsma, A. J. M. and Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America* 87:304–10.
- Hsieh, I. H. and Saberi, K. (2007). Temporal integration in absolute identification of musical pitch. *Hearing Research* 233:108–16.
- Jeffress, L. A. (1948). A place theory of sound localization. *Journal of Comparative and Physiological Psychology* 41:35–9.
- Johnson, D. H. (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *Journal of the Acoustical Society of America* 68:1115–22.
- Jülicher, F., Andor, D., and Duke, T. (2001). Physical basis of two-tone interference in hearing. *Proceedings of the National Academy of Sciences USA* 98:9080–5.
- Kaernbach, C. (1993). Temporal and spectral basis of the features perceived in repeated noise. *Journal of the Acoustical Society of America* 94:91–7.
- Köppl, C. (1997). Phase locking to high frequencies in the auditory nerve and cochlear nucleus magnocellularis of the barn owl *Tyto alba*. *Journal of Neuroscience* 17:3312–21.
- Krumbholz, K., Patterson, R. D., and Pressnitzer, D. (2000). The lower limit of pitch as determined by rate discrimination. *Journal of the Acoustical Society of America* 108:1170–80.
- Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., and Lütkenhöner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschls gyrus. *Cerebral Cortex* 13:765–72.

- Lachaux, J. P., Rudrauf, D., and Kahane, P. (2003). Intracranial EEG and human brain mapping. *Journal of Physiology, Paris* 97:613–28.
- Laguitton, V., Demany, L., Semal, C., and Liégeois-Chauvel, C. (1998). Pitch perception: a difference between right- and left-handed listeners. *Neuropsychologia* 36:201–7.
- Langner, G., Sams, M., Heil, P., and Schultze, H. (1997). Frequency and periodicity are represented in orthogonal maps in the human auditory cortex: evidence from magnetoencephalography. *Journal of Comparative Physiology, A* 181:665–76.
- Levitin, D. J. and Rogers, S. E. (2005). Absolute pitch: perception, coding, and controversies. *Trends in Cognitive Sciences* 9:26–33.
- Licklider, J. C. R. (1951). A duplex theory of pitch perception *Experientia* 7:128–34. [Reproduced in Schubert, E. D. (1979). Psychological acoustics. In *Benchmark Papers in Acoustics*, Vol. 13, pp. 155–60. Stroudsburg, PA: Dowden, Hutchinson & Ross, Inc.]
- Licklider, J. C. R. (1954). Periodicity pitch and place pitch. *Journal of the Acoustical Society of America* 26:945. [Abstract]
- Licklider, J. C. R. (1959). Three auditory theories. In: *Psychology, a Study of a Science* (ed. S. Koch), pp. 41–144. New York: McGraw–Hill.
- Liégeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., and Chauvel, P. (1994). Evoked potentials recorded from the auditory cortex in man: evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology* 92:204–14.
- Liu, L.-F., Palmer, A. R., and Wallace, M. N. (2005). Phase-locked responses to pure tones in the inferior colliculus. *Journal of Neurophysiology* 95:1926–35.
- Lütkenhöner, B. (2003). Single-dipole analyses of the N100m are not suitable for characterizing the cortical representation of pitch. *Audiology and Neuro-otology* 8:222–3.
- Lütkenhöner, B., Lammertmann, C., and Knecht, S. (2001). Latency of auditory evoked field deflection N100m ruled by pitch or spectrum. *Audiology and Neuro-otology* 6:263–78.
- Lütkenhöner, B., Krumbholz, K., and Seither-Preisler, A. (2003). Studies of tonotopy based on wave N100 of the auditory field are problematic. *NeuroImage*, 19:935–49.
- McAlpine, D. (2004). Neural sensitivity to periodicity in the inferior colliculus: evidence for the role of cochlear distortions. *Journal of Neurophysiology* 92:1295–311.
- Mark, H. E. and Rattay, F. (1990). Frequency discrimination of single-, double- and triple-cycle sinusoidal acoustic signals. *Journal of the Acoustical Society of America* 88:560–3.
- Meddis, R. and Hewitt, M. J. (1991a). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: Pitch identification. *Journal of the Acoustical Society of America* 89:2866–82.
- Meddis, R. and Hewitt, M. J. (1991b). Virtual pitch and phase sensitivity of a computer model of the auditory periphery. II: Phase sensitivity. *Journal of the Acoustical Society of America* 89:2883–94.
- Meddis, R. and O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America* 102:1811–20.
- Mersenne, M. (1636). *Harmonie Universelle*. Paris: Cramoisy (reprinted, 1975, Paris: Editions du CNRS).
- Michéyl, C. and Carlyon, R. P. (1998). Effects of temporal fringes on fundamental-frequency discrimination. *Journal of the Acoustical Society of America* 104:3006–18.
- Michéyl, C. and Oxenham, A. J. (2004). Sequential F0 comparisons between resolved and unresolved harmonics: No evidence for translation noise between two pitch mechanisms. *Journal of the Acoustical Society of America* 116:3038–50.
- Michéyl, C. and Oxenham, A. J. (2005). Comparing F0 discrimination in sequential and simultaneous conditions. *Journal of the Acoustical Society of America* 118:41–4.
- Michéyl, C. and Oxenham, A. (2007). Across-frequency pitch discrimination interference between complex tones containing resolved harmonics. *Journal of the Acoustical Society of America* 121:1621–31.
- Michéyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. (2006a). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research* 219:36–47.

- Micheyl, C., Bernstein, J. G., and Oxenham, A. (2006b). Detection and F0 discrimination of harmonic complex tones in the presence of competing tones or noise. *Journal of the Acoustical Society of America* 120:1493–505.
- Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America* 54:610–19.
- Moore, B. C. J. (1981). Relation between pitch shifts and MMF shifts in forward masking. *Journal of the Acoustical Society of America* 69:594–7.
- Moore, B. C. J. (1982, 2003). *An Introduction to the Psychology of Hearing*. London: Academic Press.
- Moore, B. C. J. (2000). A test for the diagnosis of dead regions in the cochlea. *British Journal of Audiology* 34:205–24.
- Moore, B. C. J. and Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *Journal of the Acoustical Society of America* 74:750–3.
- Moore, B. C. J. and Glasberg, B. R. (1991). Effects of signal-to-noise ratio on the frequency discrimination of complex tones with overlapping and nonoverlapping harmonics. *Journal of the Acoustical Society of America* 89:1888. [Abstract]
- Moore, B. C. J. and Ohgushi, K. (1993). Audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America* 93:452–61.
- Moore, B. C. J. and Sek, A. (1995). Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (amplitude modulation versus frequency modulation). *Journal of the Acoustical Society of America* 97:2468–78.
- Moore, B. C. J. and Sek, A. (1998). Discrimination of frequency glides with superimposed random glides in level. *Journal of the Acoustical Society of America* 104:411–21.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1985a). Thresholds for the detection of inharmonicity in complex tones. *Journal of the Acoustical Society of America* 77:1861–7.
- Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (1985b). Relative dominance of individual partials in determining the pitch of complex tones. *Journal of the Acoustical Society of America* 77:1853–60.
- Moore, B. C. J., Glasberg, B. R., and Proctor, G. M. (1992). Accuracy of pitch matching for pure tones with overlapping or nonoverlapping harmonics. *Journal of the Acoustical Society of America* 91:3443–50.
- Moore, B. C. J., Glasberg, B. R., Low, K. E., Cope, T., and Cope, W. (2006). Effects of level and frequency on the audibility of partials in inharmonic complex tones. *Journal of the Acoustical Society of America* 120:934–44.
- Moore, B. C. J., Glasberg, B., Aberkane, I., and Pinker, S. (2007). Dominance region at low frequencies: implications for pitch theories. *Journal of the Acoustical Society of America* 121:3091–2. [Abstract]
- Nayagam, D. A. X., Clarey, J. C., and Paolini, A. G. (2005). Powerful, onset inhibition in the ventral nucleus of the lateral lemniscus. *Journal of Neurophysiology* 94:1651–4.
- Nelken, I., Bizley, J. K., Nodal, F. R., Ahmed, B., King, A. J. and J. W. S. (2008). Responses of auditory cortex to complex stimuli: functional organization revealed using intrinsic optical signals. *Journal of Neurophysiology* 91:1928–41.
- Ohm, G. S. (1843). On the definition of a tone with the associated theory of the siren and similar sound producing devices. *Poggendorfs Annalen der Physik und Chemie* 59:497ff. [Translated and reprinted in Lindsay (1973). *Acoustics: Historical and Philosophical Development*, pp. 242–7. Stroudsburg, PA: Dowden, Hutchinson and Ross].
- Oxenham, A. and Shera, C. A. (2003). Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *Journal of the Association for Research in Otolaryngology* 4:541–54.
- Oxenham, A., Fligor, B. J., Mason, R., and Kidd, G. J. (2003). Informational masking and musical training. *Journal of the Acoustical Society of America* 114:1543–9.
- Palmer, A. and Russell, C. (1986). Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells. *Hearing Research* 24:1–15.

- Pantev, C., Hoke, M., Lehnertz, K., Lutkenhoner, B., Anogianakis, G., and Wittkowski, W. (1988). Tonal organization of the human auditory cortex revealed by transient auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology* 69:160–70.
- Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *Journal of the Acoustical Society of America* 59:640–54.
- Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (1992). Complex sounds and auditory images. In *Auditory Physiology and Perception* (ed. Y. Cazals, K. Horner, and L. Demany), pp. 429–46. Oxford: Pergamon Press.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., and Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron* 36:767–76.
- Penagos, H., Melcher, J. R., and Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience* 24:6810–15.
- Peters, R. W., Moore, B. C. J., and Glasberg, B. R. (1983). Pitch of components of complex tones. *Journal of the Acoustical Society of America* 73:924–9.
- Plack, C. J. and Carlyon, R. P. (1995). Differences in frequency modulation detection and fundamental frequency discrimination between complex tones consisting of resolved and unresolved harmonics. *Journal of the Acoustical Society of America* 98:1355–64.
- Plack, C. J. and Oxenham, A. J. (2005). The psychophysics of pitch. In *Pitch—Neural Coding and Perception* (ed. C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper), pp. 1–6. New York: Springer.
- Plomp, R. (1964). The ear as a frequency analyzer. *Journal of the Acoustical Society of America* 36:1628–36.
- Plomp, R. (1967a). Pitch of complex tones. *Journal of the Acoustical Society of America* 41:1526–33.
- Plomp, R. (1967b). Beats of mistuned consonances. *Journal of the Acoustical Society of America* 42:462–74.
- Pollack, I. (1947). The atonal interval. *Journal of the Acoustical Society of America* 20:146–8.
- Pressnitzer, D. and Patterson, R. D. (2001). Distortion products and the pitch of harmonic complex tones. In *Physiological and Psychophysical Bases of Auditory Function* (ed. D. J. Breebaart, A. J. M. Houtsma, A. Kohlrausch, V. F. Prijs, and R. Schoonhoven), pp. 97–104. Maastricht: Shaker.
- Pressnitzer, D., Patterson, R. D., and Krumbholz, K. (2001). The lower limit of melodic pitch. *Journal of the Acoustical Society of America* 109:2074–84.
- Pressnitzer, D., Winter, I. M., and de Cheveigné, A. (2002). Perceptual pitch shift for sounds with similar waveform autocorrelation. *Acoustic Research Letters Online* 3:1–6.
- Ritsma, R. J. (1962). Existence region of the tonal residue, I. *Journal of the Acoustical Society of America* 34:1224–9.
- Ritsma, R. J. (1963). Existence region of the tonal residue, II. *Journal of the Acoustical Society of America* 35:1241–5.
- Roberts, T. P. L., Ferrari, P., Stufflebeam, S. M., Steven, M., and Poeppel, D. (2000). Latency of the auditory evoked neuromagnetic field components: stimulus dependence and insights toward perception. *Journal of Clinical Neurophysiology* 17:114–29.
- Robinson, K. and Patterson, R. D. (1995). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *Journal of the Acoustical Society of America* 98:1858–65.
- Robles, L. and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiological reviews* 81: 1305–52.
- Romani, G. L., Williamson, S. J., and Kaufman, L. (1982). Tonal organization of the human auditory cortex. *Science* 216:1339–40.
- Ruggero, M. A. (1992). Physiology of the auditory nerve. In *The Mammalian Auditory Pathway: Neurophysiology* (ed. A. N. Popper and R. R. Fay), pp. 34–93. New York: Springer-Verlag.
- Ruggero, M. A. and Temchin, A. N. (2005). Unexceptional sharpness of frequency tuning in the human cochlea. *Proceedings of the National Academy of Sciences USA* 102:18614–19.
- Ruggero, M. A., Robles, L., and Rich, N. C. (1992). Two-tone suppression in the basilar membrane of the cochlea: mechanical basis of auditory-nerve rate suppression. *Journal of Neurophysiology* 68:1087–99.

- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., and Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition* 99:113–29.
- Sauveur, J. (1701). Système général des intervalles du son. *Mémoires de l'Académie Royale des Sciences* 279–300; 347–54. [Transl. by R. B. Lindsay as 'General system of sound intervals and its application to sounds of all systems and all musical instruments'; and reprinted in *Acoustics: Historical and Philosophical Development*, pp. 88–94. Stroudsburg, PA: Dowden, Hutchinson and Ross.]
- Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., Dosch, H. G., Bleek, S., Stippich, C., and Rupp, A. (2005). Structural and functional asymmetry of lateral Heschls gyrus reflects pitch perception preference. *Nature Neuroscience* 8:1241–7.
- Schnupp, J. W. H. and King, A. J. (2008). Auditory neuroscience: neuronal sensitivity in humans. *Current Biology* 18:R382–5.
- Schönwieser, M. and Zatorre, R. J. (2008). Depth electrode recordings show double dissociation between pitch processing in lateral Heschl's gyrus and sound onset processing in medial Heschls gyrus. *Experimental Brain Research* 187:97–105.
- Schouten, J. F. (1938). The perception of subjective tones. *Proceedings of the Koninklijke Akademie van Wetenschappen te Amsterdam* 41:1086–94. [Reprinted in Schubert (1979). *Psychological acoustics*. In *Benchmark Papers in Acoustics*, Vol. 13, pp. 146–54. Stroudsburg, PA: Dowden, Hutchinson & Ross, Inc.]
- Schouten, J. F. (1940). The residue, a new component in subjective sound analysis. *Proceedings of the Koninklijke Akademie van Wetenschappen te Amsterdam* 43:356–65.
- Schwartz, I. R. (1992). The superior olivary complex and lateral lemniscal nuclei. In *The Mammalian Auditory Pathway: Neuroanatomy* (ed. D. B. Webster, A. N. Popper, and R. R. Fay), pp. 117–67. New York: Springer-Verlag.
- Seither-Preisler, A., Patterson, R. D., Krumbholz, K., Seither, S., and Lutkenhoner, B. (2006). Evidence of pitch processing in the N100m component of the auditory evoked field. *Hearing Research* 213:88–98.
- Semal, C. and Demany, L. (1990). The upper limit of musical pitch. *Music Perception* 8:165–76.
- Semal, C. and Demany, L. (2006). Individual differences in the sensitivity to pitch direction. *Journal of the Acoustical Society of America* 120:3907–15.
- Sethares, W. A. (1997). *Tuning, Timbre, Spectrum, Scale*. London: Springer-Verlag.
- Shackleton, T. M. and 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* 95:3529–40.
- Shamma, S. and 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* 107:2631–44.
- Shepard, R. N. (1982). Structural representations of musical pitch. In *The Psychology of Music* (ed. D. Deutsch), pp. 343–90. Orlando, FL: Academic Press.
- Smooenburg, G. F. (1970). Pitch perception of two-frequency stimuli. *Journal of the Acoustical Society of America* 48:924–42.
- Stufflebeam, S. M., Poeppel, D., Rowley, H. A., and Roberts, P. L. (1998). Peri-threshold encoding of stimulus frequency and intensity in the M100 latency. *Neuroreport* 9:91–4.
- Terhardt, E. (1974). Pitch, consonance and harmony. *Journal of the Acoustical Society of America* 55:1061–9.
- Terhardt, E. (1979). Calculating virtual pitch. *Hearing Research* 1:155–82.
- Thompson, A. M. and Schofield, B. R. (2000). Afferent projections of the superior olivary complex. *Microscopy Research and Technique* 51:330–54.
- Tramo, M. J., Cariani, P., Koh, C. K., Makris, N., and Braid, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception: auditory cortex. *Annals of the New York Academy of Science* 1060:148–74.
- Turner, R. S. (1977). The Ohm–Seebeck dispute, Hermann von Helmholtz, and the origins of physiological acoustics. *British Journal for the History of Science* 10:1–24.
- Ueda, K. and Ohgushi, K. (1987). Perceptual components of pitch: spatial representation using a multidimensional scaling technique. *Journal of the Acoustical Society of America* 82: 1193–200.

- von Békésy, G. and Rosenblith, W. A. (1948). The early history of hearing—observations and theories. *Journal of the Acoustical Society of America* 20:727–48.
- Wallace, M. N., Anderson, L. A., and Palmer, A. R. (2007). Phase-locked responses to pure tones in the auditory thalamus. *Journal of Neurophysiology* 98:1941–52.
- Ward, W. D. (1999). Absolute pitch. In *The Psychology of Music* (ed. D. Deutsch), pp. 265–98. Orlando, FL: Academic Press.
- Warren, R. M., Bashford, J. A., and Wrightson, J. M. (1980). Infrapitch echo. *Journal of the Acoustical Society of America* 65:1301–5.
- Warren, J. D., Uppenkamp, S., Patterson, R. D., and Griffiths, T. D. (2003). Separating pitch chroma and pitch height in the human brain. *Proceedings of the National Academy of Sciences USA* 100:10038–42.
- Wever, E. G. and Bray, C. W. (1930). The nature of acoustic response: the relation between sound frequency and frequency of impulses in the auditory nerve. *Journal of Experimental Psychology* 13: 373–87.
- White, L. J. and Plack, C. J. (2003). Factors affecting the duration effect in pitch perception for unresolved complex tones. *Journal of the Acoustical Society of America* 114:3309–16.
- Wiegrefe, L. and Meddis, R. (2004). The representation of periodic sounds in simulated sustained chopper units of the ventral cochlear nucleus. *Journal of the Acoustical Society of America* 115:1207–18.
- Wier, C. C., Jesteadt, W., and Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America* 61:178–84.
- Wightman, F. L. (1973a). Pitch and stimulus fine structure. *Journal of the Acoustical Society of America* 54:397–406.
- Wightman, F. L. (1973b). The pattern-transformation model of pitch. *Journal of the Acoustical Society of America* 54:407–16.
- Winter, I. M. (2005). The neurophysiology of pitch. In *Pitch—Neural Coding and Perception* (ed. C. J. Plack, A. J. Oxenham, R. R. Fay, and A. N. Popper), pp. 99–146. New York: Springer.
- Winter, I. M., Wiegrefe, L., and Patterson, R. D. (2001). The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. *Journal of Physiology* 537: 553–66.
- Yost, W. A. (1996). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America* 100:3329–35.
- Zatorre, R. J. (2003). Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience* 6:692–5.