Acoust. Sci. & Tech. **34**, 6 (2013)

INVITED REVIEW

Revisiting place and temporal theories of pitch

Andrew J. Oxenham*

Department of Psychology, University of Minnesota,
Elliott Hall, N218, 75 East River Parkway, Minneapolis, MN 55455, USA

Abstract: The nature of pitch and its neural coding have been studied for over a century. A popular debate has revolved around the question of whether pitch is coded via "place" cues in the cochlea, or via timing cues in the auditory nerve. In the most recent incarnation of this debate, the role of temporal fine structure has been emphasized in conveying important pitch and speech information, particularly because the lack of temporal fine structure coding in cochlear implants might explain some of the difficulties faced by cochlear implant users in perceiving music and pitch contours in speech. In addition, some studies have postulated that hearing-impaired listeners may have a specific deficit related to processing temporal fine structure. This article reviews some of the recent literature surrounding the debate, and argues that much of the recent evidence suggesting the importance of temporal fine structure processing can also be accounted for using spectral (place) or temporal-envelope cues.

Keywords: Hearing, Pitch, Auditory perception, Temporal fine structure

PACS number: 43.66.Ba, 43.66.Hg [doi:10.1250/ast.34.388]

1. INTRODUCTION

Pitch is a primary auditory sensation. We typically think about pitch in the context of music, where sequences of pitch define melody and simultaneous combinations of pitch define harmony. But pitch also plays a crucial role in speech, where rising and falling pitch contours help define prosody, and improve speech intelligibility [1]. Indeed in several of the world's languages, such as Chinese, pitch contours help define the meaning of words. In addition, in complex acoustic environments, differences in pitch can help listeners to segregate and make sense of competing sound sources [2].

Pitch is the perceptual correlate of the periodicity, or repetition rate, of an acoustic waveform. In general, periodicities between about 30 and 5,000 Hz elicit a sensation of musical pitch [3,4]. Below and above those limits, changes in frequency are perceived but they do not elicit a sufficiently clear sensation of pitch to allow listeners to recognize melodies or make judgments of musical intervals. Interestingly, these psychophysically determined limits correspond quite well with the lowest and highest notes found on modern musical instruments. For instance, the modern grand piano has lowest and highest notes with fundamental frequencies (F0s) of 27.5 Hz and 4,186 Hz, respectively. The mathematically simplest sound is the pure tone, which is generated through sinusoidal motion. According to Fourier's theorem any waveform can be decomposed into constituent sinusoidal waveforms of different frequencies, amplitudes, and phases. The most common form of pitch-evoking sound is a harmonic complex tone, which comprises sinusoids with frequencies at the F0, or waveform repetition rate, as well as integer multiples of the F0, which are known as harmonics. The questions of how these components are represented in the auditory system, and how pitch is extracted from them, have been debated for over 150 years [5,6]. Nevertheless, there are many aspects that remain unknown or controversial, and the study of pitch and its neural underpinnings remains an active topic of research today [7–10].

2. PITCH OF PURE TONES

Pure tones produce a clear pitch, which is often used as the "gold standard" against which the pitches of other stimuli are compared. We are very sensitive to changes in the frequency of pure tones. Just-noticeable differences (JNDs) in the frequency of a pure tone can be as low as 0.2% for well-trained listeners in the mid-frequency range, between about 500 and 2,000 Hz [11]. A semitone, the smallest step in the Western scale system, is a difference of about 6%, or about a factor of 30 greater than the JND in frequency for pure tones. Musicians tend to have lower (better) frequency JNDs than non-musicians, although the

*e-mail: oxenham@umn.edu

difference tends to vanish once non-musicians have had practice of between 4 and 8 hours at the task [12]. This result suggests that most people are able to discriminate very fine differences in frequency with relatively little in the way of specialized training.

There are two "classical" ways in which the frequency of a pure tone might be coded within the peripheral auditory system, using either a place or time code. The first potential code, known as the place code, reflects the mechanical filtering that takes place in the cochlea of the inner ear. The basilar membrane, which runs the length of the fluid-filled cochlea from the base to the apex, vibrates in response to sound. The responses of the basilar membrane are sharply tuned and highly specific: at low to medium sound levels, a certain frequency will cause only a local region of the basilar membrane to vibrate. Because of its structural properties, the apical end of the basilar membrane responds best to low frequencies, whereas the basal end responds best to high frequencies. Thus, every place along the basilar membrane has its own "best frequency" or "characteristic frequency" (CF) — the frequency to which that place responds most strongly. This frequency-to-place mapping is known as tonotopic organization, and it is maintained throughout the auditory pathways up to primary auditory cortex, thereby providing a potential neural code for the pitch of pure tones.

The second potential code, known as the "temporal" code, relies on the fact that action potentials, or spikes, generated in the auditory nerve tend to occur at a certain phase within the period of a sinusoid. This property, known as phase locking, means that the brain could potentially represent the frequency of a pure tone by way of the time intervals between successive spikes, when pooled across the auditory nerve. No data are available from the human auditory nerve, due to the invasive nature of the measurements, but phase locking has been found to extend from very low frequencies up to about 2–4 kHz in other mammals, depending somewhat on the species [13]. Unlike tonotopic organization, phase locking up to high frequencies is not preserved in higher stations of the auditory pathways. At the level of the auditory cortex, the limit of phase locking reduces to at best 100–200 Hz [14]. Therefore, most researchers believe that if timing information is extracted from the auditory nerve then it must be transformed to some form of place or rate-based population code at a relatively early stage of auditory processing.

There is some psychoacoustical evidence for both place and temporal codes. One piece of evidence in favor of a temporal code is that pitch discrimination abilities deteriorate at high frequencies: the JND between two frequencies becomes considerably larger at frequencies above about 4–5 kHz — the same frequency range above which listeners' ability to recognize familiar melodies [4], or notice subtle changes in unfamiliar melodies [15], degrades. This frequency is similar to the one above which phase locking in the auditory nerve is strongly degraded [e.g., 13,16], suggesting that the temporal code is necessary for accurate pitch discrimination and for melody perception. It might even be taken as evidence that the upper pitch limits of musical instruments were determined by the basic physiological limits of the auditory nerve.

Nevertheless, some form of pitch perception remains possible even with very high-frequency pure tones [11,17], where it is unlikely that phase locking information is useful [e.g., 13], suggesting that place information may also play a role. A recent study of pure-tone frequency discrimination found that frequency discrimination thresholds (in terms of percentage change in frequency) worsened up to frequencies of 8 kHz and then remained roughly constant up to the highest frequency tested of 14 kHz [18]. This pattern of results may be explained by assuming that frequency discrimination is based on timing information at low frequencies; the timing information degrades at progressively higher frequencies so that beyond 8 kHz the timing information is poorer than the available place information.

One line of evidence that place information may be important even at lower frequencies comes from a study that used so-called "transposed tones" [19] to present the temporal information that would normally only be available to a low-frequency region in the cochlear to a high-frequency region, thereby dissociating temporal from place cues [20]. These transposed tones are produced by multiplying a half-wave rectified low-frequency tone (the modulator) with a high-frequency tone (the carrier). This procedure results in a high-frequency tone that produces a temporal response in the auditory nerve that is similar (although not identical) to the auditory-nerve response to a low-frequency tone [21]. That study found that pitch discrimination was considerably worse when the low-frequency temporal information was presented to the "wrong" place in the cochlea, even though the same temporal information could be used by the binaural system to discriminate interaural time differences. The results suggested that timing information alone may not be sufficient to produce good pitch perception, and that place information may be necessary.

A difficulty in assessing the importance of timing and place information is the uncertainty surrounding the representations in the auditory nerve. First, as mentioned above, we do not have direct recordings from the human auditory nerve, and so we are uncertain about the limits of phase locking. Second, we do not know how well the higher levels of the auditory system can extract the temporal information from the auditory nerve. Heinz *et al.* [22] used a computational model of the auditory nerve to show that an optimal detector could extract sufficient timing information from auditory nerve firing to exceed human performance even at very high frequencies. On the other hand, it is not clear how realistic it is to assume that higher stages of the auditory system can optimally integrate fine timing information in the auditory nerve; certainly the human binaural system, which must rely on temporal fine structure cues to encode interaural time differences, shows a rapid deterioration in sensitivity above 1,000 Hz, and is not sensitive to temporal fine structure above about 1,500 Hz.

Similar uncertainty surrounds the coding of place cues in the cochlea. There are no direct measurements of tuning or the sharpness of the place representation in the human cochlea. It has generally been assumed that the human cochlea and auditory nerve are similar to those of commonly studied animals, such as the cat, chinchilla, or guinea pig. However, recent studies suggest that human cochlear tuning may be sharper than that in smaller mammals [23,24]. Because there is some disagreement on this topic [25], there is uncertainty regarding the "true" sharpness of tuning and filter slopes, meaning that it is difficult to evaluate place-based models of frequency discrimination in a quantitative manner. In terms of general patterns of performance, however, the fact that relative sharpness of tuning (quality factor, or Q), remains roughly constant [26], or even increases with increasing frequency [23], suggests that a place-based model would not predict the increasing frequency difference limens that have been found with increasing frequency above about 2,000 Hz [11].

In light of this mixed evidence, it may be safest to assume that the auditory system uses both place and timing information from the auditory nerve in order to extract the pitch of pure tones. Indeed some theories of pitch explicitly require both accurate place and timing information [27]. Gaining a better understanding of how the information is extracted remains an important research goal. The question is of particular clinical relevance, as deficits in pitch perception are a common complaint of people with hearing loss and people with cochlear implants. A clearer understanding of how the brain utilizes information from the cochlea will help researchers to improve the way in which auditory prostheses, such as hearing aids and cochlear implants, present sound to their users.

3. PITCH OF COMPLEX TONES

Many sounds we encounter, including voiced speech and most musical notes, are harmonic complex tones. Each harmonic complex tone is comprised of the F0 (corresponding to the repetition rate of the entire waveform) and upper partials, harmonics, or overtones, with frequencies at integer multiples of the F0. The pitch of a harmonic complex tone usually corresponds to the F0, even if the complex tone has no energy at the F0, or the F0 is masked [6,28–30]. This phenomenon has been given various terms, including pitch of the missing fundamental, periodicity pitch, residue pitch, and virtual pitch. The ability of the auditory system to extract the F0 of a sound is important from the perspective of perceptual constancy of objects under different conditions. For instance, a note played on a violin should still have the same pitch whether it is played in a quiet room or in a room where noisy air-conditioning results in the lower harmonics, including the F0, being masked.

The ability to extract the periodicity pitch is clearly an important one, and one that is shared by many different species [31]. However, there is still considerable debate surrounding how this is achieved. Figure 1 shows schematically how a complex tone can be represented first acoustically and then after filtering by the cochlea. The upper row shows the spectral representation of a harmonic complex tone. The next row depicts the filtering that occurs in the cochlea — each point along the basilar membrane can be represented as a bandpass filter that responds to only those frequencies close to its center frequency. The third row shows the average output, or "excitation pattern," produced by the sound. The fourth row shows an excerpt of the time waveform at the output of some of the filters along the array. This is an approximation of the waveform that drives the inner hair cells in the cochlea, which in turn synapse with the auditory nerve fibers to produce the spike trains that the brain must interpret.

Considering the third and fourth rows of Fig. 1, it is possible to see a transition as one moves from the low-numbered harmonics (i.e., the F0, the second harmonic, and so on) on the left to the high-numbered harmonics on the right: The first few harmonics generate distinct peaks in the excitation pattern, because the filters in that frequency region are narrower than the spacing between successive harmonics. Note also that the time waveforms at the outputs of filters centered at the low-numbered harmonics resemble pure tones, because each filter is responding primarily to a single harmonic. At higher harmonic numbers, the bandwidths of the auditory filters become wider than the spacing between successive harmonics, and so individual peaks in the excitation pattern are lost. Similarly, the time waveform at the output of higher-frequency filters no longer resembles a pure tone, but instead reflects the interaction of multiple harmonics, producing a complex waveform that repeats at a rate corresponding to the F0.

Harmonics that produce distinct peaks in the excitation pattern and/or produce quasi-sinusoidal vibrations on the basilar membrane are referred to as being "resolved." Phenomenologically, resolved harmonics are those that can

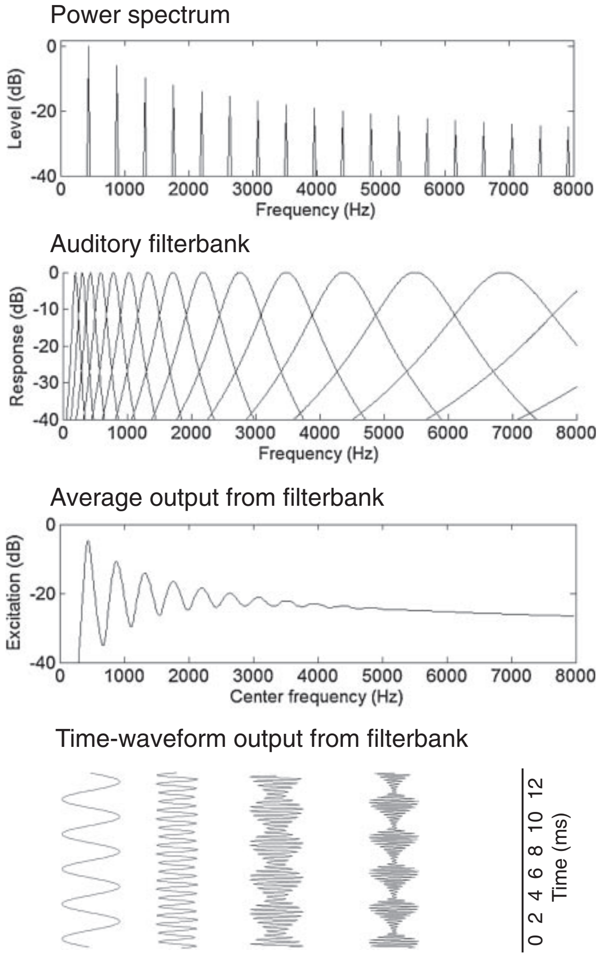

Fig. 1 Representations of a harmonic complex tone with a fundamental frequency (F0) of 440 Hz. The upper panel shows the power spectrum. The second panel shows the auditory filterbank, representing the filtering that occurs in the cochlea. The third panel shows the the time-averaged output of the filterbank, or excitation pattern. The fourth panel shows some sample time waveforms at the output of the filterbank, including filters centered at the F0 and the fourth harmonic, illustrating resolved harmonics, and filters centered at the 8th and 12th harmonic of the complex, illustrating harmonics that are less well resolved and show amplitude modulations at a rate corresponding to the F0.

be "heard out" as separate tones under certain circumstances. Typically, we do not hear the individual harmonics when we listen to a musical tone, but our attention can be drawn to them in various ways, for instance by amplifying them or by switching them on and off while the other harmonics remain continuous [e.g., 32,33]. The ability to resolve or hear out individual low-numbered harmonics as pure tones was already noted by Hermann von Helmholtz in his classic work, *On the Sensations of Tone* [34].

The higher-numbered harmonics, which do not produce individual peaks of excitation and cannot typically be heard out, are often referred to as being "unresolved." The transition between resolved and unresolved harmonics is thought to lie somewhere between the 5th and 10th harmonic, depending on various factors, such as the F0 and the relative amplitudes of the components, as well as on how resolvability is defined [e.g., 32,35–37].

Many theories and models have been proposed to explain how pitch is extracted from complex tones [38]. These can be generally divided into place and time (and place-time) theories, just as with pure tones. With place theories, the auditory system is assumed to extract pitch from the lower-order, resolved harmonics [39–42]. With temporal theories, the time intervals between auditory-nerve spikes, pooled across the auditory nerve, are evaluated using the autocorrelation or an all-interval spike histogram [29,43–46]. Place-time theories have come in different forms, but one version involves coincident timing between neurons with harmonically related CFs that is postulated to lead to a spatial network of coincidence detectors [47]. The available physiological evidence is at least not inconsistent with such proposals [48].

One difficulty with distinguishing between place and temporal (or place-time) models of pitch is that spectral and temporal representations of a signal are mathematically equivalent: any change in the spectral representation is reflected by a change in the temporal representation, and *vice versa*. Discovering what the auditory system does means focusing on the physiological limits imposed by the cochlea and auditory nerve. For instance, the place theory can be tested using known limits of frequency selectivity: if pitch can be heard when only unresolved harmonics are presented (eliminating place information), then place information is not necessary for pitch. Similarly, if all the frequencies within a stimulus are above the upper limits of phase locking, and the temporal envelope information is somehow suppressed, then temporal information is not necessary for pitch perception.

Several studies have demonstrated pitch using either unresolved harmonics [32,35,36,49] or amplitude-modulated noise [50,51], ruling out purely place-based theories of pitch. However, the pitch produced by these stimuli is typically very weak, and may not support very accurate melody perception, or the perception of multiple pitches [52,53].

Low-numbered, resolved harmonics produce a much more robust and salient pitch than do high-numbered, unresolved harmonics. This produces another challenge for temporal models, which typically do not predict a benefit for low-numbered harmonics over high-numbered harmonics [54]. In summary, place models predict performance with unresolved harmonics that is too poor, and temporal models predict performance that is too good. The differences in pitch salience produced by resolved and unresolved harmonics has led to a proposal for two separate pitch mechanisms, one based on the (time or place) information from resolved harmonics, and one based on the temporal-envelope information from unresolved harmonics [55], although there is some question concerning the evidence for this proposal [56,57].

The fact that low-numbered, resolved harmonics are important suggests that place coding may play a role in everyday pitch, or that temporal information from individual harmonics plays a more important role than temporal information from the overall stimulus periodicity. In addition, the study mentioned earlier that used tones with low-frequency temporal information transposed into a high-frequency range [20] studied complex-tone pitch perception by transposing the information from harmonics 3, 4 and 5 of a 100-Hz F0 to high-frequency regions of the cochlea — roughly 4 kHz, 6 kHz, and 10 kHz. If temporal information was sufficient to elicit a periodicity pitch, then listeners should have been able to hear a pitch corresponding to 100 Hz. In fact, none of the listeners reported hearing a low pitch or was able to match the pitch of the transposed tones to that of the missing fundamental. A similar conclusion was reached using bandpass-filtered harmonic complexes, rather than transposed tones [58]. This suggests that, if temporal information is used, it may need to be presented to the “correct” place along the cochlea.

Another line of evidence favoring a role for place coding has come from studying pitch perception using harmonics that are all higher than 5 kHz. An early study found that pitch was not perceived when all the harmonics were above 5 kHz [59], leading to the suggestion that timing information was crucial for periodicity pitch. However, a recent study revisited this conclusion and found that, in fact, listeners were well able to hear pitches between 1 and 2 kHz, even when all the harmonics were filtered to be above 6 kHz, and were sufficiently resolved to ensure that no temporal envelope cues were available [15]. Thus, either temporal information is not necessary for musical pitch, or usable phase locking in the human auditory nerve extends to much higher frequencies than is generally believed [22,60].

Most sounds in our world, including those produced by musical instruments, tend to have more energy at low frequencies than at high; on average spectral amplitude decreases at a rate of about $1/f$, or -6 dB/octave. It therefore makes sense that the auditory system would rely on the lower numbered harmonics to determine pitch, as these are the ones that are most likely to be audible. Also, resolved harmonics — ones that produce a peak in the excitation pattern and elicit a sinusoidal temporal response — are much less susceptible to the effects of room reverberation than are unresolved harmonics. Pitch discrimination thresholds for unresolved harmonics are relatively good (~2%) when all the components have the same starting phase (as in a stream of pulses). However, thresholds are much worse when the phase relationships are scrambled, as they would be in a reverberant hall or church, and listeners' discrimination thresholds can be as poor as 10% — more than a musical semitone [61,62]. In contrast, the response to resolved harmonics is not materially affected by reverberation: changing the starting phase of a single sinusoid does not affect its waveshape — it still remains a sinusoid, with frequency discriminations thresholds of less than 1%.

In summary, the pitch of single harmonic complex tones is determined primarily by the first 5–8 harmonics, which are also those thought to be resolved in the peripheral auditory system. To extract the pitch the auditory system must somehow combine and synthesize information from these harmonics. Exactly how this occurs in the auditory system remains a matter of ongoing research.

4. THE ROLE OF TEMPORAL ENVELOPE AND TEMPORAL FINE STRUCTURE IN PITCH AND SPEECH PERCEPTION

A study by Smith *et al.* [63] combined the temporal fine structure from one sound with the temporal envelope from another sound, and asked listeners what they heard. When the sounds were bandpass-filtered into bands with bandwidths resembling normal auditory filters, speech perception was dominated by the information in the temporal envelope, and the perception of pitch and spatial location was dominated by information in the temporal fine structure. This outcome was in line with earlier speech studies, showing that the temporal envelope was sufficient to convey speech, even with relatively poor spectral information [64], and was consistent with earlier studies showing that pitch is dominated by low-numbered harmonics (as discussed above), and that localization of broadband sounds is dominated by low-frequency interaural time differences in the temporal fine structure [65,66]. In the case of binaural processing, it seems clear that the low-frequency acoustic temporal fine structure is coded temporally in the auditory nerve and brainstem, and that this temporal information is extracted to localize sounds. Phase-locking in the auditory nerve in mammals such as cats and guinea pigs remains strong up to about 1 kHz, and then degrades rapidly beyond that [13]. Structures in the auditory brainstem are specialized to maintain fine time resolution beyond the auditory nerve, and can remain sensitive to minute timing differences (on the order of microseconds) in the inputs arriving from opposite ears [67].

In contrast, despite the tendency to associate temporal fine structure with temporal coding in the auditory system, it is not clear that temporal fine structure is coded temporally for purposes of extracting attributes such as frequency or pitch. Consider, for instance, a single sinusoid, or pure tone. In terms of temporal envelope and fine structure, this stimulus has a flat (unmodulated) temporal envelope, so the information is in the temporal fine structure. However, as discussed above, a pure tone could be coded either by timing information in the auditory nerve, or by place information, based on the position of excitation along the basilar membrane. In fact, one could also consider this information in terms of the temporal envelope: the envelope level is highest at the output of a filter tuned to the frequency of the pure tone, and is progressively lower at the output of filters with CFs progressively further from the frequency of the tone. Thus across-channel *envelope* information can also be used to code the temporal fine structure of pure tones (or any other stimulus).

Over the past decade or so, there has developed a large body of literature on the importance of temporal fine structure for speech perception in noise. It has been argued that, although temporal envelope information is crucial for speech understanding, the information in the temporal fine structure becomes more important in a noise background, and even more important in more complex, fluctuating noise backgrounds. In particular, it has been hypothesized that one reason why hearing-impaired listeners have particular difficulty in complex noise backgrounds is due to a specific deficit in temporal fine structure coding [68]. In a similar vein, most current cochlear implants process only the temporal envelope information from the bandpass-filtered stimulus, and discard the temporal fine structure. This lack of temporal fine structure has been credited with explaining some of the deficits experienced by cochlear-implant users, particularly for speech in fluctuating noise [69,70].

Most of the recent work studying the importance of temporal fine structure has been done using vocoder techniques, where the original temporal fine structure in each frequency subband is replaced either by a tone or by a bandpass noise [71–74]. In other types of study, temporal envelope information is reduced by “flattening” the original temporal envelope, i.e., by keeping the amplitude within each subband constant [68,73,75,76].

All these studies have argued that temporal fine structure information is important for speech perception in noise. In particular, it is claimed that speech masking release — the benefit of introducing amplitude fluctuations in an otherwise steady-state masker — is facilitated by the use of temporal fine structure cues, and that specific deficits in temporal coding associated with hearing loss, or lack of temporal fine structure information in cochlear implants, lead to impaired speech perception in fluctuating noise [68,72]. However, in all these studies, the results can also be explained in terms of spectral cues or temporal envelope cues, as described above for the case of a pure tone. In particular, deficits in temporal fine structure processing in hearing-impaired listeners may reflect poorer spectral resolution (or broader filters), rather than any specific deficit in temporal coding. Thus, it remains unclear whether in fact temporal fine structure deficits really relate to deficits in temporal coding. One study that purported to rule out perceptible spectral cues in the presence of temporal fine structure changes [77] was found to have used stimuli that resulted in audible and spectrally resolved distortion products; when distortion products were masked, the results were no longer consistent with the use of temporal fine structure [78].

Some studies have begun to address the question of temporal fine structure, independent of spectral cues, directly. Two experimental studies have concluded that temporal fine structure does not play an important role in speech masking release. The first study [79] tested the hypothesis by measuring speech masking release in lowpass-filtered and highpass-filtered conditions. The highpass filter cutoff (1,500 Hz) was selected to eliminate any resolved harmonics from speech, where temporal fine structure information might be available. The lack of useable temporal fine structure was confirmed by the finding that pitch discrimination of the highpass stimuli was very poor, and was dependent on the component phases, as would be expected if the judgments were based on temporal envelope cues. The lowpass filter cutoff (1,200 Hz) was selected to produce speech intelligibility scores in steady-state noise that matched the scores found in the highpass-filtered conditions. When the steady-state noise was replaced by a fluctuating noise, or by a single talker, performance improved in both the lowpass-filtered and highpass-filtered conditions by the same amount, suggesting that there was no selective advantage of temporal fine structure in the lowpass-filtered condition.

The second study investigated the intelligibility of whispered speech [80]. Whispered speech is not voiced and so has no periodic temporal fine structure. However, unlike noise-vocoded speech, it retains the same spectral resolution of the spectral envelope of speech, such as the formant frequencies. The prediction was that if periodic temporal fine structure is important for speech masking release, then whispered speech should result in much less speech masking release than normal (voiced) speech. In fact, although whispered speech was less intelligible, the difference in intelligibility between steady-state and fluctuating noise was just as great (and in some cases greater) in whispered speech than in normal speech. Again the results are not consistent with the idea that temporal fine structure is crucial for speech masking release.

A third study used computational modeling to show that the information available in simulated auditory nerve responses to speech and speech-in-noise stimuli were dominated by envelope components of the response and not temporal fine structure [81]. The authors also pointed out how broader filters could lead to the misleading conclusion of poorer temporal fine structure processing, based on the loss of spectral resolution.

In summary, despite the large number of studies that have investigated the role of temporal fine structure, using various types of signal processing, there is very little evidence for its importance in speech masking release, or for the idea that it is coded temporally in the auditory system. In general, it is as difficult to distinguish between place and time codes for speech as it is for pitch. Nevertheless, the potential importance of place information in speech may explain why schemes to improve temporal coding in cochlear implants have not yielded benefits in terms of speech understanding in noise [82,83].

5. SUMMARY

Despite over a century of discussion and dispute concerning the relative importance of place and timing codes in the auditory system for the perception of pitch in music and speech, the question remains somewhat open. The most recent iteration of the debate, involving temporal fine structure and temporal envelope, suffers from the same basic problem that acoustic temporal fine structure can be coded in the auditory system either by a temporal or a place code (or both). So far, despite some claims to the contrary, there remains no conclusive evidence that the temporal coding of temporal fine structure is important for understanding speech in complex fluctuating backgrounds.

ACKNOWLEDGMENTS

This work is supported by a grant from the National Institutes of Health (R01 DC005126). Minoru Tsuzaki and other members of the editorial board provided helpful comments on an earlier version of this paper.

REFERENCES

[1] S. E. Miller, R. S. Schlauch and P. J. Watson, "The effects of fundamental frequency contour manipulations on speech intelligibility in background noise," *J. Acoust. Soc. Am.*, **128**, 435–443 (2010).

[2] C. J. Darwin, "Pitch and auditory grouping," in *Pitch: Neural Coding and Perception*, C. J. Plack, A. J. Oxenham, R. R. Fay and A. N. Popper, Eds. (Springer Verlag, New York, 2005), pp. 278–305.

[3] D. Pressnitzer, R. D. Patterson and K. Krumbholz, "The lower limit of melodic pitch," *J. Acoust. Soc. Am.*, **109**, 2074–2084 (2001).

[4] F. Attneave and R. K. Olson, "Pitch as a medium: A new approach to psychophysical scaling," *Am. J. Psychol.*, **84**, 147–166 (1971).

[5] G. S. Ohm, "Über die Definition des Tones, nebst daran geknüpfter Theorie der Sirene und ähnlicher tonbildender Vorrichtungen [On the definition of tones, including a theory of sirens and similar tone-producing apparatuses]," *Ann. Phys. Chem.*, **59**, 513–565 (1843).

[6] A. Seebeck, "Beobachtungen über einige Bedingungen der Entstehung von Tönen [Observations on some conditions for the formation of tones]," *Ann. Phys. Chem.*, **53**, 417–436 (1841).

[7] T. D. Griffiths and D. A. Hall, "Mapping pitch representation in neural ensembles with fMRI," *J. Neurosci.*, **32**, 13343–13347 (2012).

[8] S. Kumar and M. Schonwiesner, "Mapping human pitch representation in a distributed system using depth-electrode recordings and modeling," *J. Neurosci.*, **32**, 13348–13351 (2012).

[9] A. J. Oxenham, "Pitch perception," *J. Neurosci.*, **32**, 13335–13338 (2012).

[10] X. Wang and K. M. Walker, "Neural mechanisms for the abstraction and use of pitch information in auditory cortex," *J. Neurosci.*, **32**, 13339–13342 (2012).

[11] B. C. J. Moore, "Frequency difference limens for short-duration tones," *J. Acoust. Soc. Am.*, **54**, 610–619 (1973).

[12] C. Micheyl, K. Delhommeau, X. Perrot and A. J. Oxenham, "Influence of musical and psychoacoustical training on pitch discrimination," *Hear. Res.*, **219**, 36–47 (2006).

[13] A. R. Palmer and I. J. Russell, "Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells," *Hear. Res.*, **24**, 1–15 (1986).

[14] M. N. Wallace, R. G. Rutkowski, T. M. Shackleton and A. R. Palmer, "Phase-locked responses to pure tones in guinea pig auditory cortex," *Neuroreport*, **11**, 3989–3993 (2000).

[15] A. J. Oxenham, C. Micheyl, M. V. Keebler, A. Loper and S. Santurette, "Pitch perception beyond the traditional existence region of pitch," *Proc. Natl. Acad. Sci. USA*, **108**, 7629–7634 (2011).

[16] J. E. Rose, J. F. Brugge, D. J. Anderson and J. E. Hind, "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," *J. Neurophysiol.*, **30**, 769–793 (1967).

[17] G. B. Henning, "Frequency discrimination of random amplitude tones," *J. Acoust. Soc. Am.*, **39**, 336–339 (1966).

[18] B. C. J. Moore and S. M. Ernst, "Frequency difference limens at high frequencies: Evidence for a transition from a temporal to a place code," *J. Acoust. Soc. Am.*, **132**, 1542–1547 (2012).

[19] S. van de Par and A. Kohlrausch, "A new approach to comparing binaural masking level differences at low and high frequencies," *J. Acoust. Soc. Am.*, **101**, 1671–1680 (1997).

[20] A. J. Oxenham, J. G. W. Bernstein and H. Penagos, "Correct tonotopic representation is necessary for complex pitch perception," *Proc. Natl. Acad. Sci. USA*, **101**, 1421–1425 (2004).

[21] A. Dreyer and B. Delgutte, "Phase locking of auditory-nerve fibers to the envelopes of high-frequency sounds: Implications for sound localization," *J. Neurophysiol.*, **96**, 2327–2341 (2006).

[22] M. G. Heinz, H. S. Colburn and L. H. Carney, "Evaluating auditory performance limits: I. One-parameter discrimination using a computational model for the auditory nerve," *Neural Comput.*, **13**, 2273–2316 (2001).

[23] C. A. Shera, J. J. Guinan and A. J. Oxenham, "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proc. Natl. Acad. Sci. USA*, **99**, 3318–3323 (2002).

[24] C. A. Shera, J. J. Guinan, Jr. and A. J. Oxenham, "Otoacoustic estimation of cochlear tuning: Validation in the chinchilla,"

J. Assoc. Res. Otolaryngol., **11**, 343–365 (2010).

[25] M. A. Ruggero and A. N. Temchin, "Unexceptional sharpness of frequency tuning in the human cochlea," *Proc. Natl. Acad. Sci. USA*, **102**, 18614–18619 (2005).

[26] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.*, **47**, 103–138 (1990).

[27] G. E. Loeb, M. W. White and M. M. Merzenich, "Spatial cross correlation: A proposed mechanism for acoustic pitch perception," *Biol. Cybern.*, **47**, 149–163 (1983).

[28] J. F. Schouten, "The residue and the mechanism of hearing," *Proc. Kon. Akad. Wetenschap.*, **43**, 991–999 (1940).

[29] J. C. R. Licklider, "A duplex theory of pitch perception," *Experientia*, **7**, 128–133 (1951).

[30] E. de Boer, *On the "Residue" in Hearing* (University of Amsterdam, Amsterdam, 1956).

[31] W. P. Shofner, "Comparative aspects of pitch perception," in *Pitch: Neural Coding and Perception*, C. J. Plack, A. J. Oxenham, R. R. Fay and A. N. Popper, Eds. (Springer Verlag, New York, 2005), pp. 56–98.

[32] J. G. Bernstein and A. J. Oxenham, "Pitch discrimination of diotic and dichotic tone complexes: Harmonic resolvability or harmonic number?" *J. Acoust. Soc. Am.*, **113**, 3323–3334 (2003).

[33] W. M. Hartmann and M. J. Goupell, "Enhancing and unmasking the harmonics of a complex tone," *J. Acoust. Soc. Am.*, **120**, 2142–2157 (2006).

[34] H. L. F. Helmholtz, *On the Sensations of Tone* (Dover, New York, 1885/1954).

[35] A. J. M. Houtsma and J. Smurzynski, "Pitch identification and discrimination for complex tones with many harmonics," *J. Acoust. Soc. Am.*, **87**, 304–310 (1990).

[36] T. M. Shackleton and R. P. Carlyon, "The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination," *J. Acoust. Soc. Am.*, **95**, 3529–3540 (1994).

[37] B. C. J. Moore and H. E. Gockel, "Resolvability of components in complex tones and implications for theories of pitch perception," *Hear. Res.*, **276**, 88–97 (2011).

[38] A. de Cheveigné, "Pitch perception models," in *Pitch: Neural Coding and Perception*, C. J. Plack, A. J. Oxenham, R. R. Fay and A. N. Popper, Eds. (Springer Verlag, New York, 2005), pp. 169–233.

[39] J. L. Goldstein, "An optimum processor theory for the central formation of the pitch of complex tones," *J. Acoust. Soc. Am.*, **54**, 1496–1516 (1973).

[40] E. Terhardt, "Pitch, consonance, and harmony," *J. Acoust. Soc. Am.*, **55**, 1061–1069 (1974).

[41] F. L. Wightman, "The pattern-transformation model of pitch," *J. Acoust. Soc. Am.*, **54**, 407–416 (1973).

[42] M. A. Cohen, S. Grossberg and L. L. Wyse, "A spectral network model of pitch perception," *J. Acoust. Soc. Am.*, **98**, 862–879 (1995).

[43] J. F. Schouten, R. J. Ritsma and B. L. Cardozo, "Pitch of the residue," *J. Acoust. Soc. Am.*, **34**, 1418–1424 (1962).

[44] R. Meddis and M. Hewitt, "Virtual pitch and phase sensitivity studied of a computer model of the auditory periphery. I: Pitch identification," *J. Acoust. Soc. Am.*, **89**, 2866–2882 (1991).

[45] P. A. Cariani and B. Delgutte, "Neural correlates of the pitch of complex tones. I. Pitch and pitch salience," *J. Neurophysiol.*, **76**, 1698–1716 (1996).

[46] R. Meddis and L. O'Mard, "A unitary model of pitch perception," *J. Acoust. Soc. Am.*, **102**, 1811–1820 (1997).

[47] S. Shamma and D. Klein, "The case of the missing pitch templates: How harmonic templates emerge in the early auditory system," *J. Acoust. Soc. Am.*, **107**, 2631–2644 (2000).

[48] L. Cedolin and B. Delgutte, "Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve," *J. Neurosci.*, **30**, 12712–12724 (2010).

[49] C. Kaernbach and C. Bering, "Exploring the temporal mechanism involved in the pitch of unresolved harmonics," *J. Acoust. Soc. Am.*, **110**, 1039–1048 (2001).

[50] E. M. Burns and N. F. Viemeister, "Nonspectral pitch," *J. Acoust. Soc. Am.*, **60**, 863–869 (1976).

[51] E. M. Burns and N. F. Viemeister, "Played again SAM: Further observations on the pitch of amplitude-modulated noise," *J. Acoust. Soc. Am.*, **70**, 1655–1660 (1981).

[52] R. P. Carlyon, "Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker," *J. Acoust. Soc. Am.*, **99**, 517–524 (1996).

[53] H. A. Kreft, D. A. Nelson and A. J. Oxenham, "Modulation frequency discrimination with modulated and unmodulated interference in normal hearing and in cochlear-implant users," *J. Assoc. Res. Otolaryngol.*, **14**, 591–601 (2013).

[54] R. P. Carlyon, "Comments on "A unitary model of pitch perception" [J. Acoust. Soc. Am. 102, 1811–1820 (1997)]," *J. Acoust. Soc. Am.*, **104**, 1118–1121 (1998).

[55] R. P. Carlyon and T. M. Shackleton, "Comparing the fundamental frequencies of resolved and unresolved harmonics: Evidence for two pitch mechanisms?" *J. Acoust. Soc. Am.*, **95**, 3541–3554 (1994).

[56] C. Micheyl and A. J. Oxenham, "Further tests of the "two pitch mechanisms" hypothesis," *J. Acoust. Soc. Am.*, **113**, 2225 (2003).

[57] H. Gockel, R. P. Carlyon and C. J. Plack, "Across-frequency interference effects in fundamental frequency discrimination: Questioning evidence for two pitch mechanisms," *J. Acoust. Soc. Am.*, **116**, 1092–1104 (2004).

[58] J. M. Deeks, H. E. Gockel and R. P. Carlyon, "Further examination of complex pitch perception in the absence of a place-rate match," *J. Acoust. Soc. Am.*, **133**, 377–388 (2013).

[59] R. J. Ritsma, "Existence region of the tonal residue. I," *J. Acoust. Soc. Am.*, **34**, 1224–1229 (1962).

[60] B. C. J. Moore and A. Sęk, "Sensitivity of the human auditory system to temporal fine structure at high frequencies," *J. Acoust. Soc. Am.*, **125**, 3186–3193 (2009).

[61] J. G. Bernstein and A. J. Oxenham, "The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss," *J. Acoust. Soc. Am.*, **120**, 3929–3945 (2006).

[62] J. G. Bernstein and A. J. Oxenham, "The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level," *J. Acoust. Soc. Am.*, **120**, 3916–3928 (2006).

[63] Z. M. Smith, B. Delgutte and A. J. Oxenham, "Chimaeric sounds reveal dichotomies in auditory perception," *Nature*, **416**, 87–90 (2002).

[64] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski and M. Ekelid, "Speech recognition with primarily temporal cues," *Science*, **270**, 303–304 (1995).

[65] F. L. Wightman and D. J. Kistler, "The dominant role of low-frequency interaural time differences in sound localization," *J. Acoust. Soc. Am.*, **91**, 1648–1661 (1992).

[66] E. A. Macpherson and J. C. Middlebrooks, "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.*, **111**, 2219–2236 (2002).

[67] A. Brand, O. Behrend, T. Marquardt, D. McAlpine and B. Grothe, "Precise inhibition is essential for microsecond interaural time difference coding," *Nature*, **417**, 543–547 (2002).

[68] C. Lorenzi, G. Gilbert, H. Carn, S. Garnier and B. C. J. Moore,

"Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Natl. Acad. Sci. USA*, **103**, 18866–18869 (2006).

[69] M. K. Qin and A. J. Oxenham, "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.*, **114**, 446–454 (2003).

[70] G. S. Stickney, P. F. Assmann, J. Chang and F. G. Zeng, "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," *J. Acoust. Soc. Am.*, **122**, 1069–1078 (2007).

[71] K. Hopkins, B. C. J. Moore and M. A. Stone, "Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech," *J. Acoust. Soc. Am.*, **123**, 1140–1153 (2008).

[72] K. Hopkins and B. C. J. Moore, "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," *J. Acoust. Soc. Am.*, **125**, 442–446 (2009).

[73] C. Lorenzi, L. Debruille, S. Garnier, P. Fleuriot and B. C. Moore, "Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal," *J. Acoust. Soc. Am.*, **125**, 27–30 (2009).

[74] M. Ardoint and C. Lorenzi, "Effects of lowpass and highpass filtering on the intelligibility of speech based on temporal fine structure or envelope cues," *Hear. Res.*, **260**, 89–95 (2010).

[75] K. Hopkins, B. C. Moore and M. A. Stone, "The effects of the addition of low-level, low-noise noise on the intelligibility of sentences processed to remove temporal envelope information," *J. Acoust. Soc. Am.*, **128**, 2150–2161 (2010).

[76] S. Sheft, M. Ardoint and C. Lorenzi, "Speech identification based on temporal fine structure cues," *J. Acoust. Soc. Am.*, **124**, 562–575 (2008).

[77] B. C. J. Moore, B. R. Glasberg, H. J. Flanagan and J. Adams, "Frequency discrimination of complex tones; Assessing the role of component resolvability and temporal fine structure," *J. Acoust. Soc. Am.*, **119**, 480–490 (2006).

[78] A. J. Oxenham, C. Micheyl and M. V. Keebler, "Can temporal fine structure represent the fundamental frequency of unresolved harmonics?" *J. Acoust. Soc. Am.*, **125**, 2189–2199 (2009).

[79] A. J. Oxenham and A. M. Simonson, "Masking release for low- and high-pass filtered speech in the presence of noise and single-talker interference," *J. Acoust. Soc. Am.*, **125**, 457–468 (2009).

[80] R. L. Freyman, A. M. Griffin and A. J. Oxenham, "Intelligibility of whispered speech in stationary and modulated noise maskers," *J. Acoust. Soc. Am.*, **132**, 2514–2523 (2012).

[81] J. Swaminathan and M. G. Heinz, "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise," *J. Neurosci.*, **32**, 1747–1756 (2012).

[82] J. T. Rubinstein, B. S. Wilson, C. C. Finley and P. J. Abbas, "Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation," *Hear. Res.*, **127**, 108–118 (1999).

[83] R. Schatzer, A. Krenmayr, D. K. Au, M. Kals and C. Zierhofer, "Temporal fine structure in cochlear implants: Preliminary speech perception results in Cantonese-speaking implant users," *Acta Otolaryngol.*, **130**, 1031–1039 (2010).