Topographic organization is essential for pitch perception

Shihab A. Shamma*

Department of Electrical Engineering and Institute for Systems Research, University of Maryland, College Park, MD 20742

Pitch is the perceptual attribute we associate most with melodies in music, patterns of bird songs, and the distinctions between speakers' voices. It plays a key role in the organization, segregation, and identification of sound sources in cluttered auditory scenes because it is derived from acoustic cues that closely reflect the material and geometric properties of resonating objects. Pitch has been the subject of intensive psychoacoustic studies for well over a century. In recent decades, physiological investigations in humans and animals have attempted to locate and understand the biological substrate underlying pitch perception at various levels of the auditory nervous system. However, despite all efforts, a deep understanding of the mechanisms that give rise to the pitch percept remains elusive. This uncertainty has generated passionate debates between the proponents of two very different theories of pitch, one based on the *place* or *location* of neural activation patterns, and the other on their *temporal* modulations. This state of affairs is now likely to change dramatically in favor of the place theories with the publication of results of intricately designed psychoacoustic experiments by Oxenham *et al.* (1) reported in this issue of PNAS.

There is universal agreement that pitch is basically a correlate of the periodicity of a sound waveform. The simplest such waveform is the pure tone (Fig. 1*A*); it is composed of a single sinusoid that gives rise to a pitch percept correlated with the frequency (or periodicity) of the sinusoid. A richer percept results when a complex periodic waveform is composed of several tones that are harmonically related, i.e., are integer multiples of a common fundamental frequency (Fig. 1*B*). Such a waveform evokes a salient and unified sense of pitch that we typically associate with the fundamental frequency regardless of the relative amplitudes of the harmonics, their dynamics, and even whether the fundamental component itself is physically present or missing. For example, the perceived pitch of A_3 is the same (440 Hz) whether produced by a piano, violin, or a soprano because all emanating waveforms share the same fundamental periodicity. The distinctive timbre of each sound is primarily related to the relative amplitudes and dynamics of the harmonic components (Fig. 1*C*).
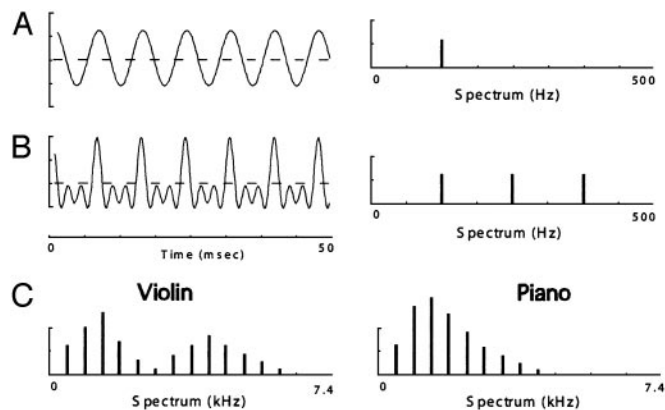

Fig. 1. Time waveforms and/or spectra. (*A*) Single (pure) tone at a frequency of 125 Hz. (*B*) Three harmonically related tones with a fundamental frequency of 125 Hz (125, 250, and 375 Hz). (*C*) Schematic spectra of A_3 played on a violin and a piano. Both spectra have the same fundamental frequency (440 Hz) and hence are perceived at the same pitch. The amplitudes of the harmonic components are quite different between the two instruments, giving rise to their distinctive timbres.

The universal agreement on the nature of pitch vanishes at the first synapse of the auditory system when one delves into the question of how the auditory system computes from sound periodicity (or, equivalently, harmonicity) the pitch. The reason is that the cochlea of the inner ear acts much like a frequency analyzer, directing the harmonic components to excite separate, relatively confined, and frequency-ordered places

> ## Pitch can be derived by pattern recognition on the auditory nerve.

along its length, hence creating the tonotopic or frequency axis of the auditory system. Because the resulting neural activation pattern induced on the auditory nerve is uniquely related to fundamental frequency of the sound, it is readily conceivable that pitch can be derived by a pattern identification stage, for instance, one that finds the ideal harmonic pattern that best matches that of the incoming sound (Fig. 2*A*). This scheme would work even if the fundamental and/or some of the higher harmonics were missing. In fact, numerous other implications of this hypothesis have been thoroughly tested and confirmed over many decades, and conse-

quently it is now well known as the spectral or "*place*" *theory* of pitch (2–4).

This would have been the sole theory of pitch were it not for the unique ability of the auditory system to encode information about the frequency content of a signal by an alternative parallel means. Specifically, a tone not only excites the auditory nerve at a particular place but also induces a neural response that is modulated temporally at a rate that equals the frequency of the tone. These modulations occur up to fairly high rates in mammals, somewhat coinciding with the full pitch range (4,000 Hz). Consequently, a harmonic pattern is conveyed to the central auditory system both by its pattern of activation in place and by pattern of modulations in time (Fig. 2*B*). With the emergence of the "*temporal*" *theory* of pitch during the last decade, a hot debate has ensued on whether the central auditory system exploits place or temporal cues to derive pitch, and what biologically plausible mechanisms could perform these computations. This debate has largely been fueled by the remarkable lack of clear biological evidence in support of either theory.

Against the backdrop of this controversy come the ingenious experiments by Oxenham *et al.* (1) to offer a simple unambiguous test of the two theories.
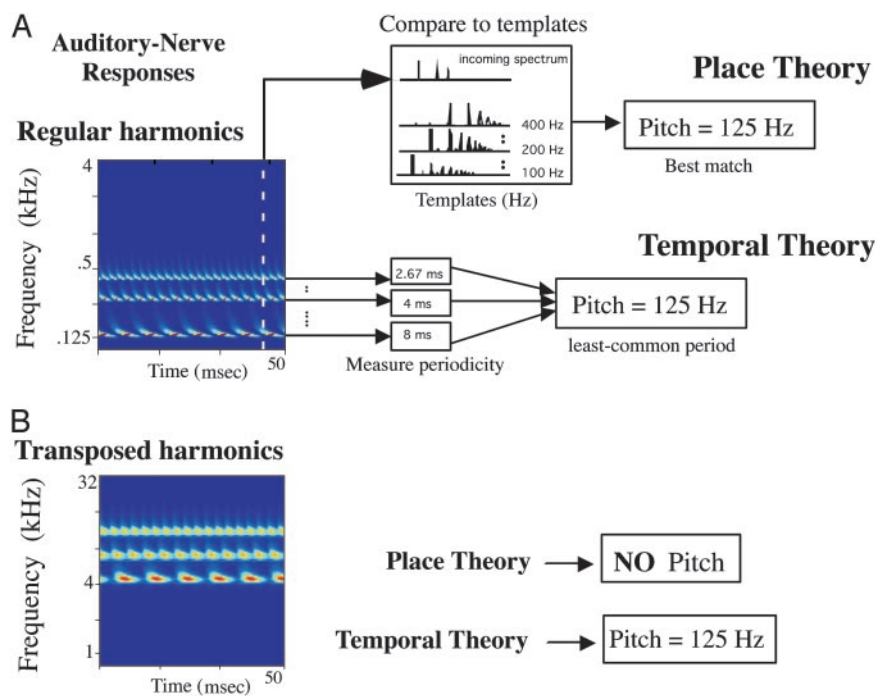
---

*E-mail: sas@isr.umd.edu.

Fig. 2. Schematic of the place theory and the temporal theory of pitch. (*A*) Pitch theories. The pattern of activation evoked by a three-harmonic stimulus (125, 250, and 375 Hz) on the auditory nerve. The place theory assumes that the neural response pattern across the tonotopic axis (white dashed line) is compared with the patterns from a bank of harmonic templates stored in the central auditory system. Pitch is then assigned to the fundamental of the best-matching template. The temporal theory postulates that first the response periodicity is measured in parallel in all auditory-nerve channels. The pitch is then evaluated by pooling the measurements and selecting the fundamental period common to all channels (or the least common period). (*B*) Predictions of the place and temporal theories of the pitch of the transposed stimuli. Because the spectral pattern of these stimuli does not resemble that of any harmonic pattern, the place theory predicts no particular clear pitch percept. The temporal theory predicts an unaltered pitch percept because the temporal modulations remain well represented in the transposed tones.

The design of the experiments overcomes a difficult conceptual hurdle: How can one manipulate the closely intertwined spectral and temporal cues independently of each other, so as to monitor their effects on the perceived pitch? The logic of the experiments is based on the fact that the two theories ascribe starkly different assessments of the importance of the two types of cues. In the extreme formulation of the place theory, the temporally modulated responses are unnecessary and are not used at all in evaluating the pitch. By contrast, in the temporal theory, the location of the modulated responses along the tonotopic axis is immaterial to the derived pitch. Therefore, if one could arbitrarily displace the temporally modulated responses from their "normal" tonotopic locations without changing the modulation rates in any way, one could test directly whether the perceived pitch is seriously disrupted, as the place theory would predict, or whether it remains unaltered, as the temporal theory asserts.

To accomplish this manipulation and, for instance, move a 125-Hz modulation from its regular place in the cochlea to a faraway place, e.g., to the location normally reserved for a 4-kHz tone (Fig. 2*B*), Oxenham *et al.* used a clever technique proposed originally by van de Par and Kohlrausch (5) and used recently by Les Bernstein (6) to study the role of temporal signals in binaural sound localization. The technique is to impose the 125-Hz modulation on a high-frequency carrier tone (4-kHz), much as speech is carried by AM radio signals. When the ear receives this modulated high-frequency carrier, it directs it to the 4-kHz place, where it becomes subsequently demodulated at the cochlea/nerve interface, resulting in low-frequency modulated responses on the nerve. Bernstein (6) and Oxenham *et al.* (1) confirm this scenario and the integrity of the low-frequency modulations by demonstrating that they in fact are induced at their new places with sufficient fidelity to serve accurate binaural localization (Experiment 1).

Oxenham *et al.* (1) then proceed to the heart of this study to demonstrate that transposing temporally modulated waveforms to arbitrary tonotopic locations has a severe impact on the perceived pitch. Subjects were largely unable to compare the pitch of the transposed stimuli to those of regular harmonics (Experiment 2), nor were they able to assign them a clear pitch value by matching it to the frequency of a pure tone (Experiment 3). These remarkably clean results strongly argue in favor of a fundamental role for tonotopic place in the computation of pitch. Furthermore, if temporal cues are involved, they must be somehow intrinsically linked to the correct place.

This elegantly simple and intellectually compelling study goes a long way toward clarifying the constraints that a valid theory of pitch must satisfy. Its results also are consistent with conclusions of a series of other recent studies that have highlighted the shortcomings of the temporal theory in explaining the pitch of resolved versus unresolved harmonics (7, 8) and the perception of regularity in periodic and random click trains (9). The results of the Oxenham *et al.* (1) experiments are sure to intensify the search for the biological substrate of pitch, and to provide further impetus for a clearer understanding of the role of place and time in auditory perception.

1. Oxenham, A. J., Bernstein, J. G. W. & Penagos, H. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 1421–1425.
2. Goldstein, J. (1973) *J. Acoust. Soc. Am.* **54,** 1496–1516.
3. Terhardt, E. (1979) *Hear. Res.* **1,** 155–182.
4. Wightman, F. (1973) *J. Acoust. Soc. Am.* **54,** 397–406.
5. van de Par, S. & Kohlrausch, A. (1997) *J. Acoust. Soc. Am.* **101,** 1671–1680.
6. Bernstein, L. (2001) *J. Neurosci. Res.* **66,** 1035–1046.
7. Bernstein, J. & Oxenham, A. (2003) *J. Acoust. Soc. Am.* **113,** 3323–3334.
8. Carlyon, R. (1998) *J. Acoust. Soc. Am.* **104,** 1118–1121.
9. Kaernbach, C. & Demany, L. (1998) *J. Acoust. Soc. Am.* **104,** 2298–2306.