

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Understanding How Neural Circuits Measure Pitch

Daniel Bendor

Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Review of Cedolin and Delgutte

Pitch perception plays an important role in audition, especially in the context of music and speech. The brain could theoretically compute pitch from either the spatial or temporal pattern of excitation in the auditory periphery. However, computational pitch models relying on only a single information source from the auditory periphery (spatial or temporal) have been unable to account for the entire set of psychophysical phenomenon associated with pitch perception. One attractive alternate model uses a synergistic approach, relying on the spatial pattern of temporal information (Shamma, 1985). Recently, Cedolin and Delgutte (2010) provided the first physiological evidence that spatiotemporal information from auditory nerve responses can be used to extract pitch.

We hear pitch when a sound is harmonic (spectrally periodic) or has a regularly repeating acoustic structure (temporally periodic). The frequency of this periodicity (fundamental frequency) determines what pitch we perceive, with a higher fundamental frequency generating a higher perceived pitch. Harmonic sounds like those generated by a musical instrument or the human vocal tract produce a strong pitch sensation, while white noise (which is aperiodic) does not generate a pitch percept.

No explicit representation for pitch exists in the cochlea for harmonic sounds; however, if the frequency of each harmonic component can be measured, then later stages of the auditory system can theoretically use this information to calculate the fundamental frequency. The frequencies of the harmonics can be detected by the spatial pattern of excitation along the cochlea's basilar membrane. This spatial pattern is then encoded by the firing rates of auditory nerve fibers, which are determined by the proximity of a fiber's best frequency to a harmonic. Although this works well at low sound levels, at higher sound levels the responses of auditory nerve fibers saturate, minimizing the difference in firing rate between fibers with best frequencies close to and far away from a given harmonic. Because we can still hear a sound's pitch at these sound levels, something other than an auditory nerve fiber's firing rate must be used to detect harmonics.

One way to detect harmonics of louder sounds is to synergistically use temporal information to enhance the contrast in the spatial pattern of excitation. This is possible because the cochlear traveling wave generated by the sound does not have a constant speed, but instead slows down (causing a phase transition) at locations along the basilar membrane with the highest resonance. In other words, the point of maximal excitation along the basilar membrane for a spectral peak will also have the largest phase transition in the cochlear traveling wave. This would translate into a larger delay in the responses of

two neighboring auditory nerve fibers when their best frequency is closer to a harmonic's frequency. Thus harmonics could be detected by measuring where the largest differences are in the responses of neighboring auditory nerve fibers, which theoretically could be computed using lateral inhibition at a later stage in the auditory system.

To test the spatiotemporal pitch model, Cedolin and Delgutte (2010) recorded single-unit responses to harmonic complex tones from the auditory nerve of anesthetized cats. Given the difficulty in finely sampling neighboring auditory nerve fibers with single-unit recordings, Cedolin and Delgutte (2010) instead used the property of scaling invariance, where local spatiotemporal response patterns can be inferred from the response of a single auditory nerve fiber. The principle of scaling invariance states that a small shift in cochlear location will have a similar response (magnitude and phase) for a proportional shift in frequency. Therefore by measuring responses to multiple harmonic complex tones (with varying fundamental frequencies) within a single auditory nerve fiber, the authors could infer what the spatial and temporal pattern of activity across a population of auditory nerve fibers would look like for a single harmonic complex tone.

To measure the spatiotemporal representation's ability to encode a complex tone's harmonics, Cedolin and Delgutte (2010) calculated the mean absolute spatial derivative (MASD), which is the mean difference in activity between two neighboring auditory nerve fibers. With a delay

Received Nov. 19, 2010; revised Dec. 9, 2010; accepted Dec. 14, 2010.

Support for this work was provided by a postdoctoral fellowship from Merck and the Helen Hay Whitney Foundation.

Correspondence should be addressed to Daniel Bendor, Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, 43 Vassar Street, Building 46, Room 5233, Cambridge, MA 02139. E-mail: dbendor@mit.edu.

DOI:10.1523/JNEUROSCI.6077-10.2011

Copyright © 2011 the authors 0270-6474/11/313141-02\$15.00/0

between the two fibers, one fiber's response will slightly precede the other fiber's response, causing a difference in activity at that time point, increasing the MASD. When there is no delay, the temporal patterns of the two fibers are similar, minimizing the difference in activity and decreasing the MASD. If the MASD is largest for auditory nerve fibers with best frequencies near the frequencies of the harmonics, the MASD can represent pitch. The MASD representation can be directly compared to the traditional spatial representation of pitch (rate–place profile), in which peaks in the spatial distribution of firing rates correspond to the frequencies of the harmonics.

The authors first compared the MASD and rate–place profiles of harmonic complex tones at different sound levels. Rate–place profiles failed to detect harmonics at louder sound levels because of the response saturation of auditory nerve fibers. On the other hand, the spatiotemporal representation (based on the MASD profile) was still able to detect harmonics at these higher sound levels where rate–place profiles could not extract pitch. Because of the upper limit of phase locking (~3–5 kHz), the spatiotemporal representation's ability to extract pitch was limited by frequency (to fundamental frequencies <1100 Hz). This is not a shortcoming of the spatiotemporal model, because our ability to perceive pitch is also frequency limited. In contrast, a rate–place representation only depends on the spectral resolvability of harmonics, which does not degrade with frequency.

Humans are insensitive to phase differences between harmonics when they are spectrally resolved on the basilar membrane. Because phase differences affect the relative timing of auditory nerve responses, these differences could compromise the spatiotemporal representation of pitch. Using three different phase relationships between harmonics (cosine phase, alternating phase, and Schroeder phase), the authors found no statistically significant difference in the MASD profile, showing an agreement between the spatiotemporal representation of pitch and human psychophysical results.

To extract pitch from a spatiotemporal representation, it is thought that the brain may use a harmonic template. Using responses pooled over the entire population

of auditory nerve fibers, the authors next measured how accurately a harmonic template could estimate the fundamental frequency. A spatiotemporal representation of pitch was accurate (SD < 2%) for fundamental frequencies between 350 Hz and 1100 Hz. The lower limit of this range is similar to what has been shown for a rate–place-based model (Cedolin and Delgutte, 2005). Although this range covers only the upper portion of the range of fundamental frequencies associated with pitch, lower fundamental frequencies could still be measured using a secondary temporally based mechanism relying on unresolved harmonics (Carlyon and Shackleton, 1994).

Delgutte and colleagues have previously tested both the temporal (Cariani and Delgutte, 1996a,b) and rate–place (Cedolin and Delgutte, 2005) representation of pitch. This report by Cedolin and Delgutte (2010) confirms the viability of a third class of pitch models, relying on spatiotemporal information. Whether one or more of these representations of pitch are actually used by the auditory system cannot be determined by the available data from the auditory nerve. Recording from the next stage in the auditory system, the cochlear nucleus, may help identify which pitch mechanism(s) are used. Identification of lateral inhibition or cross-frequency coincidence detection could further support the use of spatiotemporal information in computing pitch.

Both the rate–place and spatiotemporal models require a harmonic template at some later stage of the auditory system to compute pitch. Recordings from auditory cortex have identified neurons tuned to multiple harmonically related frequencies that could act as a harmonic template (Kadia and Wang, 2003). Pitch-selective neurons in the low-frequency region of primate auditory cortex (Bendor and Wang, 2005) may provide another starting point for identifying the pitch mechanism(s) used by the auditory system. These neurons are tuned to a sound's fundamental frequency, and are sensitive to acoustic parameters important for pitch perception such as temporal regularity and repetition rate. By identifying the cortical and subcortical inputs to these neurons, pathways involved in processing pitch might be identified. Pitch may be computed using a single mechanism (Meddis and O'Mard, 1997), or multiple

mechanisms may be used (Carlyon and Shackleton, 1994).

Our understanding of pitch perception is based largely on human psychophysical data. Although missing fundamental pitch perception has been demonstrated in cats, whether cats hear pitch in the same way as humans do is not well understood. Because the cochlear size in cats is smaller than in humans, spectral resolvability is worse. Properties of pitch perception (and potentially the involved mechanisms) are different when harmonics are resolved or unresolved. This means that for a given acoustic stimulus, the harmonics could be unresolved for a cat and resolved for a human, making human psychophysical data incompatible in predicting what the cat may perceive. Understanding species-specific differences in pitch perception (Walker et al., 2009) will be necessary to interpret potential neural mechanism(s) responsible for measuring a sound's pitch.

References

- Bendor D, Wang X (2005) The neuronal representation of pitch in primate auditory cortex. *Nature* 436:1161–1165.
- Cariani PA, Delgutte B (1996a) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J Neurophysiol* 76:1698–1716.
- Cariani PA, 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. *J Neurophysiol* 76:1717–1734.
- Carlyon RP, Shackleton TM (1994) Comparing the fundamental frequencies of resolved and unresolved harmonics: evidence for two pitch mechanisms? *J Acoust Soc Am* 95:3541–3554.
- Cedolin L, Delgutte B (2005) Pitch of complex tones: rate-place and interspike interval representations in the auditory nerve. *J Neurophysiol* 94:347–362.
- Cedolin L, Delgutte B (2010) Spatiotemporal representation of the pitch of harmonic complex tones in the auditory nerve. *J Neurosci* 30:12712–12724.
- Kadia SC, Wang X (2003) Spectral integration in A1 of awake primates: neurons with single- and multi-peaked tuning characteristics. *J Neurophysiol* 89:1603–1622.
- Meddis R, O'Mard L (1997) A unitary model of pitch perception. *J Acoust Soc Am* 102:1811–1820.
- Shamma SA (1985) Speech processing in the auditory system II: lateral inhibition and the central processing of speech evoked activity in the auditory nerve. *J Acoust Soc Am* 78:1622–1632.
- Walker KM, Schnupp JW, Hart-Schnupp SM, King AJ, Bizley JK (2009) Pitch discrimination by ferrets for simple and complex sounds. *J Acoust Soc Am* 126:1321–1335.