J Neurophysiol 99: 2755–2756, 2008.
doi:10.1152/jn.90270.2008.

Editorial Focus

Surprising Twist on Auditory Representation. Focus on: "What's That Sound? Auditory Area CLM Encodes Stimulus Surprise, Not Intensity or Intensity Changes"

Timothy Q. Gentner
Department of Psychology, University of California, La Jolla, California

The concept of sensory representation is central to psychological theories of perception and cognition and to neurobiological models of sensory coding. Understanding exactly what a neural response "represents," however, turns out to be a difficult question to answer fully—particularly for neurons many synapses away from the point of sensory transduction. A recent paper by Gill and colleagues (2007) takes a large step forward in understanding stimulus representation in the auditory system. To appreciate the contribution this paper makes, it's helpful to have a little background.

To a first approximation, sensory systems and the receptive fields they comprise are organized hierarchically. The concept of a "receptive field" was first applied to the visual system (Hartline 1938) to describe the sensitivity of optic nerve fibers to light shined on spatially restricted portions of the retina. While neurons in primary visual cortex inherit the spatial receptive fields of the retinal projections, they respond to oriented bars or edges rather than spots of light. At higher levels in the visual pathway, cells respond best to complex visual patterns such as faces or other objects (Suzuki et al. 2006; Tanaka 2003). A similar representational hierarchy exists in the ascending auditory system, where the frequency tuned responses of primary fibers in the eighth nerve give rise to increasingly complex "objects" at more central levels (Gentner and Margoliash 2003). Phenomenological descriptions notwithstanding, the responses of neurons beyond primary auditory cortex have proven very difficult to model. Currently the best receptive field models typically account for ∼25–30%, or less, of a neuron's response to a stimulus (e.g., Sen et al. 2001). To improve the receptive field models for high-level sensory neurons, one might devise a better function to map the stimulus on to the spike train—most models rely on linear regression (cf. Sharpee et al. 2006). Alternatively, as Gill and colleagues did, one might devise a better representation of the input stimulus (e.g., Rust et al. 2006). Their results suggest that we need to change the way we've been thinking about time in the auditory system.

Contemporary models of high-level auditory neural responses are called "spectro-temporal receptive fields" (or STRFs for short) because they are defined in terms of the average dynamic power spectrum of the stimulus that precedes each spike (Aertsen and Johannesma 1981). The original STRF conception, and more recent variations (e.g., Kowalski et al. 1996; Theunissen et al. 2000), treat the temporal and spectral dimensions of the stimulus similarly. That is, a pattern in the stimulus that plays out across time is given no greater (or lesser) weight than a pattern that plays out across frequency bands. Gill et al. devised a novel representation for complex natural signals that takes into account the unfolding temporal structure of the stimulus. Instead of considering the magnitude of the dynamic power spectrum (i.e., the power at each frequency and each point in time) as the input to a neuron, they converted the sound spectrogram to a large set of conditional probabilities, where the value at each frequency at any given time is a function of the likelihood that the observed power at that frequency is preceded by a particular pattern of power at neighboring frequencies.

To help think about this, imagine that you are sitting at a stop light. The light is red, the light is red, the light is red, and then suddenly it snaps to green. If one samples the state of the traffic light at regular time intervals, there are three possible states for each pair of sequential intervals: the light is red in both intervals, the light is green in both intervals, or the light is red and then green. At any given time, the chance of the red-then-green condition is very low. Gill et al. used an analogous representation of conditional probabilities across multiple frequency bands of an acoustic stimulus based on the power in each spectral band instead of the color of the traffic light. This way of representing the stimulus, which they call the "surprise" representation after similar ideas for visual scenes (Itti and Baldi 2005), radically altered the input to the receptive field model by according events with the *lowest* conditional probabilities the *highest* representational weight. They then test the efficacy of models generated from this representation in predicting the responses of neurons at multiple points in the ascending auditory system of a songbird, comparing their predictions to those for classical STRFs and STRFs derived from spectro-temporal derivatives. Comparisons to the spectral derivative (taken with respect to time) enable the authors to rule out the possibility that the magnitude of a spectro-temporal change, the auditory equivalent of a high-contrast visual edge, drives the neurons' responses.

The STRFs generated from the surprise representation yield significantly better receptive field models in both field L, the avian homologue to layer 4 of primary auditory cortex in mammals, and in caudo-lateral mesopallium (CLM), a secondary forebrain-auditory region, and by no small amount. In field L, the surprise-STRF bettered the classic STRF by an average of 24%, and in CLM, the surprise-STRF trumped the "old-fashioned" models by a whopping 67% on average. The surprise-STRFs outperformed the derivative-STRFs by significant margins as well. These are not marginal quantitative gains. These are categorical improvements that show the firing rates of high-level sensory neurons depend more on the probability of natural stimulus features than on stimulus intensity or

Address for reprint requests and other correspondence: Dept. of Psychology, University of California San Diego, La Jolla, CA 92093 (E-mail: tgentner @ucsd.edu).

intensity changes. Interestingly, at nucleus mesencephalious lateralis, pars dorsalis (MLD), a region homologous to the inferior colliculus in mammals, the surprise-STRFs are no different from the classic STRFs, suggesting that the temporal expectancies are themselves the product of computations added to the system in high-level regions.

The functional interpretations and details of this remarkable discovery may take some time to work out of course. For instance, it may be possible to introduce measures that more accurately capture realistic expectancies across multiple stimulus dimensions (e.g., Dubnov 2006). Likewise the source for the posterior probabilities that constrain expectancy need to be explored. But these are the kinds of questions that naturally emerge from any important discovery. Significant progress in understanding receptive fields has always paved the way for broader level understanding of sensory systems, and the clear demonstration by Gill and colleagues that expectations and natural statistics form a key part of the auditory neural code promises to follow in that tradition.

REFERENCES

Aertsen AM, PI Johannesma. The spectro-temporal receptive field. A functional characteristic of auditory neurons. *Biol Cybern* 42: 133–143, 1981.

Dubnov S. Spectral anticipations. *Comput Music J* 30: 63–83, 2006.

Gentner TQ, D Margoliash. Neuronal populations and single cells representing learned auditory objects. *Nature* 424: 669–674, 2003.

Gill P, Woolley SMN, Fremouw T, Theunissen FE. What's that sound? Auditory area clm encodes stimulus surprise, not intensity or intensity changes. *J Neurophysiol* doi: 10.1152/jn.01270.2007.

Hartline HK. The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *Am J Physiol* 121: 400–415, 1938.

Itti, L, P Baldi. A principled approach to detecting surprising events in video. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2005, vol. 1, p 631–637.

Kowalski N, Depireux DA, Shamma SA. Analysis of dynamic spectra in ferret primary auditory cortex. I. Characteristics of single-unit responses to moving ripple spectra. *J Neurophysiol* 76: 3503–3523, 1996.

Rust NC, Mante V, Simoncelli E, Movshon JA. How MT cells analyze the motion of visual patterns. *Nat Neurosci* 9: 1421–1431, 2006.

Sen K, Theunissen FE, Doupe AJ. Feature analysis of natural sounds in the songbird auditory forebrain. *J Neurophysiol* 86: 1445–1458, 2001.

Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD. Adaptive filtering enhances information transmission in visual cortex. *Nature* 439: 936–942, 2006.

Suzuki W, K. Matsumoto, K, Tanaka. Neuronal responses to object images in the macaque inferotemporal cortex at different stimulus discrimination levels. *J Neurosci* 26: 10524–10535, 2006.

Tanaka K. Columns for complex visual object features in the inferotemporal cortex: clustering of cells with similar but slightly different stimulus selectivities. *Cereb Cortex* 13: 90–99, 2003.

Theunissen FE, Sen K, Doupe AJ. Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. *J Neurosci* 20: 2315–2331, 2000.