

A Novel View of Hearing in Reverberation

Terry Takahashi^{1,*}

¹Institute of Neuroscience, University of Oregon, Eugene, OR 97403, USA

*Correspondence: terry@uoregon.edu

DOI 10.1016/j.neuron.2009.04.004

In nature, sounds of interest arrive at the eardrums accompanied by echoes that reflect off of surfaces. This superposition can distort the cues by which we localize the source of a sound. Yet, we seem to have no difficulty turning precisely toward the source even in the presence of moderately intense echoes. The article by Devore and colleagues in this issue of *Neuron* suggests that the auditory system can perform this feat by being more responsive to the early portion of a sound which includes the earliest portions when the echoes have yet to arrive.

In our day-to-day environment, we are immersed in sounds that arrive from different locations as well as their echoes that arrive after they reflect off of nearby surfaces. Yet, we navigate this cluttered acoustic environment effortlessly, turning our heads, for instance, toward the person speaking to us in the midst of traffic noise or in an echo-rich building. To appreciate the complexity of this process, consider the analogy offered by Albert Bregman in his pioneering book *Auditory Scene Analysis*: imagine going to a lakeside and digging two troughs from the water to the shore, and by watching the waves in these troughs, being able to tell where the boats are, where the rocks are that reflected the wakes of all the boats, and what the screw-rotation frequencies of the motor-boats are. The pattern of the waves in the troughs, i.e., in the ear canals, is what the auditory system has to work with to extract the signal from the acoustical clutter (Cherry, 1953; Bregman, 1990). Remarkably, we do this automatically and even take this process for granted—that is, until our auditory system ages or we lose hearing in one ear.

In this issue of *Neuron*, Sasha Devore, Antje Ihlefeld, Kenneth Hancock, Barbara Shinn-Cunningham, and Bertrand Delgutte report on neural mechanisms that might allow us to locate a sound source when the sounds from that source arrive at the eardrums accompanied by numerous echoes. They measured the ability of neurons in the cat inferior colliculus (IC) that are highly sensitive to the location of sounds along the horizon, to signal the position of the sound source in the presence of a series of echoes that reflect off of surfaces of a virtual, $11 \times 13 \times 3$ m

room. The echoes were simulated using the binaural room impulse response, a mathematical description of the room's reflective characteristics and its dimensions. By filtering the target sound with this impulse response, it is possible to simulate the sounds arriving at the cats' ears directly from the sound-source and after reflecting off of the room's surfaces.

Note that even in such a simple room, with just a floor, ceiling, and four walls, a sound can bounce off of more than one surface so that primary, secondary, and higher-order echoes add to, and mix with, the waveforms arriving at the eardrums directly from the target. This superposition of multiple individual echoes describes the condition called reverberation. A limited amount of reverberation can make the acoustics of a room “warm” and pleasing; too much of it can make a room sound “muddy” or “boomy.” An anechoic (echoless) environment can sound “dry” and plain by contrast.

It is obviously important to locate the source of a sound and not the reflective surfaces (unless you are a bat navigating by active sonar). Reverberation makes this difficult because the superposition of the target sound and echoes causes the phase and amplitude conveyed by each frequency channel in the auditory system to assume values that are vector averages of the values corresponding to the positions of the source and surfaces (reviewed in Blauert, 1997; Figure 3.8 on page 215). The idea that sine waves, the building blocks of any sound, can be treated as vectors is explained in the book *Signals, Sounds, and Sensation* (Hartmann, 1998; page 17). In the auditory system, the interaural difference in the phase angles of the

resultants can be computed for each frequency band to extract the interaural time difference (ITD), the cue that corresponds to the horizontal position of a source and to which the cat's IC neurons were very sensitive. When there are no echoes, all frequency channels signal the same ITD, and a measure of this coherence, called the interaural correlation (see Figure 3A in Devore et al., 2009), is maximal, as is the precision with which a listener localizes the source. The situation changes in the presence of echoes. Because the sounds' travel times to each ear from the source and from the six reflective surfaces of the virtual room are different, and because these travel times constitute different proportions of each spectral band's period, the phase angles of the resultants and the ITDs computed from them will differ across frequency. In other words, in the presence of echoes, different frequency bands will point to different locations along the horizon. Interaural correlation would decline, and so too would the ability to localize the target sound (Rakerd and Hartmann, 1985).

Figures 3A and 3B in the report from Devore and colleagues illustrate both the problem just explained and the authors' insight into a solution. Figure 3A depicts the interaural correlation plotted on a color scale against ITD (vertically) and time (horizontally) for the anechoic condition. There is a crisp, reddish (high correlation), horizontal streak just to one side of 0 ms of ITD indicating a spatially focused source to one side of the midline. (A source at the midline would have an ITD of 0 ms.) Contrast this view with that of Figure 3B, which plots the same quantities for the highly reverberant condition. The streak

looks diffused or blurred, with the ITD changing over time starting 25 to 50 ms after stimulus onset. This is the result of the echoes that begin to “pile on” soon after the onset of the target sound, and were neurons to compute ITD by averaging across the full duration of the stimulus (400 ms), precision would suffer. Indeed, under reverberant conditions, the degree to which the spike rates of IC cells were modulated by a source’s horizontal position was found to be limited compared to their responses under anechoic conditions, i.e., the acuity with which the IC cells could signal the source’s position diminished.

The authors realized, however, that if ITD was computed from only the initial segment of the response when the echoes had yet to arrive, i.e., between sound onset and 50 ms, the neurons should be able to signal the ITD with an acuity rivaling that obtained under anechoic conditions. They confirmed this hypothesis by replotting the rate-versus-horizontal-position curves using only the responses to the first 50 ms of the stimulus and comparing it with the responses evoked by the remainder of the sound when echoes were present. The authors pointed out that most neurons in the IC respond vigorously at stimulus onset and then taper off as sounds progress, thus emphasizing the response to the early portion of a signal.

In order to judge the applicability of their neural mechanism to perception, the authors first confirmed that human sound localization is compromised in reverberant conditions. They then asked what sound localization performance would look like if the human auditory system, like that of the cat, relied on the early, less-contaminated portion of the binaural signal. To this end, Devore and colleagues

fed their neurophysiological data into the hemispheric difference model (van Bergeijk, 1962; McAlpine et al., 2001; Hancock, 2007), which has been proposed to be the mechanism by which mammals determine the location of sounds by computing the interhemispheric difference in the activities of a population of IC-like neurons. The model’s ability to signal the source’s location in reverberant conditions declines as the model incorporates more of the responses elicited by the echo-contaminated portions of the signal. Importantly, human performance was found to be considerably better than the performance of the model when the model averaged data over the entire 400 ms stimulus, but worse than the performance of a model using only the initial 50 ms. In other words, the human auditory system may, in fact, be relying heavily, although not exclusively, on the initial, less-contaminated segment of the signal.

Of course, as with any study, this one does not explain everything. For example, one might expect that the more heavily a neuron’s response is dominated by the onset of a sound, the more resistant it would be to reverberation. This turns out not to be the case, which implies additional processes, as the authors point out. One might also quibble with predicting human performance from cat neurons, but absent psychoacoustical data from the cat, one has to turn to a species, humans, from which psychoacoustical data can be readily obtained. Finally, the hemispheric difference model has its proponents and detractors, and future studies should investigate the performance of other models, such as that proposed by Jeffress (1948). All told, however, one cannot ignore the fact that the authors, working at the neural level in an animal model, have uncovered a simple solution

to an important perceptual problem that is plausible for the human auditory system.

Daydreaming in a typical lecture hall, one cannot help but imagine the myriad copies of the lecturer’s voice impinging upon the eardrums after reflecting off of multiple surfaces, and to be bewildered by the complexity of the sound localization cues that must result from their superposition. The intellectual contribution of Sasha Devore and her colleagues is not only that they proposed a concrete new mechanism, but that they may have simplified the discourse. William of Occam would surely have been pleased.

REFERENCES

- Blauert, J. (1997). *Spatial Hearing: The Psychophysics of Human Sound Localization* (Cambridge: MIT Press).
- Bregman, A.S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound* (Cambridge: MIT Press).
- Cherry, E.C. (1953). *J. Acoust. Soc. Am.* 26, 554–559.
- Devore, S., Ihlefeld, A., Hancock, K., Shinn-Cunningham, B., and Delgutte, B. (2009). *Neuron* 62, this issue, 123–134.
- Hancock, K.E. (2007). A physiologically-based population rate code for interaural time differences predicts bandwidth-dependent lateralization. In *Hearing: From Sensory Processing to Perception*, J. Verhey, ed. (Berlin: Springer Verlag), pp. 389–398.
- Hartmann, W.M. (1998). *Signals, Sounds, and Sensation* (New York: Springer-Verlag).
- Jeffress, L.A. (1948). *J. Comp. Physiol. Psychol.* 41, 35–39.
- McAlpine, D., Jiang, D., and Palmer, A.R. (2001). *Nat. Neurosci.* 4, 396–401.
- Rakerd, B., and Hartmann, W.M. (1985). *J. Acoust. Soc. Am.* 78, 524–533.
- van Bergeijk, W.A. (1962). *J. Acoust. Soc. Am.* 34, 1431–1437.