Conversing at a cocktail party: Linking individual abilities to neural coding

Barbara G. Shinn-Cunningham

Imagine yourself at a trendy restaurant on a busy Friday night. Boisterous conversations ebb and flow, glasses clink, chairs scrape, and all of these sounds reflect off the floor, walls, and tables, adding to the cacophony. In order to converse with your witty dinner companion, you have to be able to tune out other sounds, including the oenophile opining about the hint of apricot is his chardonnay and the businessman arguing with the maitre d' over his reservation time. No existing machine algorithms are able to accomplish what you do in these settings: segregating the speech you care about from the mixture of sounds reaching your ears, and analyzing its content to extract meaning.

In order to accomplish this kind of *selective attention*, your brain relies on the detailed structure of natural sound, grouping together sound elements that turn on and off simultaneously, share a common fundamental frequency or pitch, come from the same location, or have other spectro-temporal features that suggest they were generated by the same source. Since different sources are independent of each other, they typically do not share spectro-temporal features; this enables your brain to segregate the sources from each other and then analyze each of the different perceived objects in the auditory scene, one by one. Unfortunately, you may often find that your ability to selectively attend is diminished in a restaurant with hard, reflective walls: the reflected sound energy smears out the spectro-temporal structure in sound, weakening the features that support perceptual segregation and interfering with selective auditory attention.

Figure 1 illustrates these concepts through visual analogy. When sources are too similar to each other (Fig. 1a), the brain has difficulty separating out individual words in the scene and instead tends to analyze the entire scene at once as one mass of overlapping letters; as a result, extracting the meaning of any given word is difficult and time consuming. However, if independent sources are different in some attribute (such as their pitch, somewhat analogous to visual color; see Fig. 1b), the brain perceives each word as a separate object, and can more quickly focus on and analyze each word. When the scene's structure is smeared out (Fig. 1c, analogous to the effects of reverberant energy), selective attention is challenging, as different objects are less distinct.

Anecdotally, some listeners with normal hearing thresholds seem to have more difficulty with selective attention than others; moreover, listeners in early middle age often complain that it is harder for them to converse in restaurants and other noisy settings than when they were younger. Inspired by the work of Nina Kraus and her research group at Northwestern University, Dr. Dorea Ruggles, PhD candidate Hari Bharadwaj, and I wondered whether these individual differences in normal-hearing listeners relate to the fidelity with which early sensory portions of the auditory pathway encode spectro-temporal structure. Specifically, we noted that typical hearing screenings assess hearing sensitivity by simply asking listeners to *detect* the presence of pure sinusoids at different frequencies. To carry the visual analogy further, imagine a vision test in which, rather than describing what letter you see on a chart, all you are asked is whether or not there is some kind of letter present! We realized that a physiological measure of how well spectro-temporal details in audible (supra-threshold) sound are *encoded* in the auditory pathway might better reflect the ability to extract the meaning of real-world sounds. Indeed, Prof. Kraus and her colleagues have found a clear relationship between brainstem encoding (measured by analyzing the voltage on the scalp of the listener) and many factors, including musical training, familiarity with a tonal language, and even reading proficiency.

We recruited a large number of listeners, ranging in age from young adult to middle aged, and tested their ability to understand one talker in the presence of two competing talkers in both a simulated anechoic space (with no reflected energy) and in a simulated room with ordinary walls. We also measured how well the brainstem of each listener encoded a period-

Fig. 1. In a loud setting like a coffee bar, lots of competing sounds add up acoustically to create the auditory scene. a) If the sources are too similar in their acoustic attributes, it is difficult to segregate the sources and analyze them, as visualized here by words of identical color. b) When the sources are different in pitch or other attributes, each word is perceived as a distinct object, and is easy to analyze. c) When the scene is blurry (such as from reverberant energy), objects are less distinct and harder to analyze.

continued on page 8

Conversing at a Cocktail Party

continued from page 7

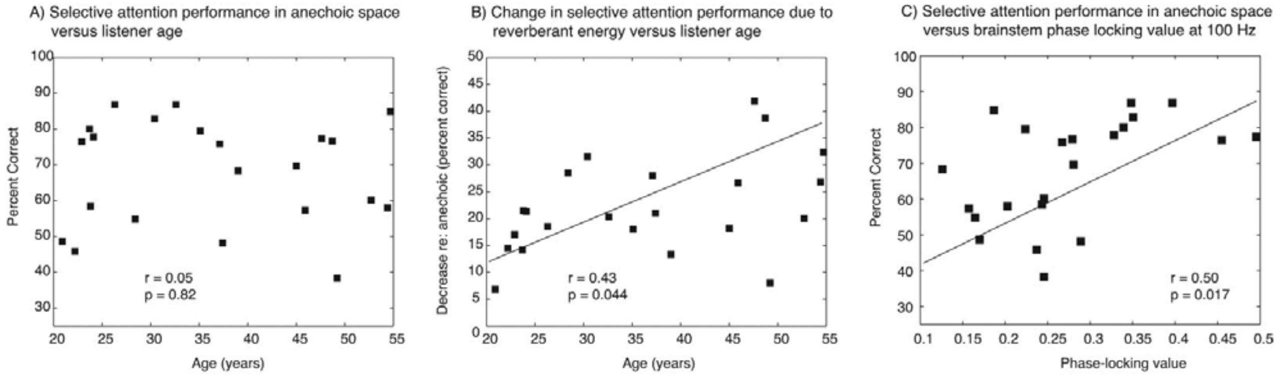

Fig. 2. a) Selective attention performance in anechoic space varies greatly from one listener to another, but is unrelated to age. b) For all listeners, selective attention performance decreases when reverberant energy is added to the scene; importantly, the effect of reverberation increases with age. c) Across all listeners, selective attention performance is correlated with the strength of the auditory brainstem encoding of the fundamental frequency of an input periodic sound (a synthesized syllable "dah" with a pitch of 100 Hz).

ically repeating sound.

We found that the ability of our normal-hearing listeners to selectively attend to a listener in a room varied dramatically from listener to listener, ranging from 33% correct up to nearly 90% correct in anechoic space. While we thought we might see that early aging hurt selective attention ability, performance was unrelated to age (see Fig. 2a). As expected, we found that for every individual listener, adding reverberation hurt performance. Importantly, even though age didn't predict how well listeners could direct selective attention, we found the negative effects of reverberation increased with listener age (see Fig. 2b). In other words, middle-aged listeners, as a group, are no worse at understanding speech in the presence of competing speech than young adults; however, on an individual basis, reverberant energy interferes more with performance the older a listener is. We also found that the strength with which the brainstem encodes the fundamental frequency of a periodic input sound is related to performance (Fig. 2c). The greater the fidelity of the brainstem in encoding the spectro-temporal structure of input sound, the better an individual listener is deploying selective auditory attention.

These kinds of studies can help us to understand the reasons for large individual differences in how well a person can function in ordinary social settings. For instance, spectro-temporal structure, which is critical for segregating and selecting a sound source from an auditory scene, is not represented equally well in the sensory pathway of all "normal hearing" listeners, which in turn explains differences in how well listeners can understand speech in a complex acoustic scene. In addition, we now can say that middle-aged listeners most likely are having greater problems communicating in everyday settings than they did when they were younger: reverberation in everyday settings truly impacts an older listener more than it does a young adult. By teasing apart different factors that affect real-world communication, we may ultimately identify distinct mechanisms important to everyday function, and find new methods for aiding listeners with different forms of perceptual difficulties.

Barbara Cunningham is Director of the Center for Computational Neuroscience and Neural Technology and a Professor of Biomedical Engineering at Boston University, 677 Beacon St., Boston, MA 02215. This article is based, in part, on paper 2aAA6 at the 162nd meeting of ASA in San Diego.