Harvard-MIT Division of Health Sciences and Technology
HST.722J: Brain Mechanisms for Hearing and Speech
Course Instructors: Bertrand Delgutte, M. Christian Brown, Joe C. Adams, Kenneth E. Hancock, Frank H. Guenther, Jennifer R. Melcher, Joseph S. Perkell, David N. Caplan

Brain Attending a Cocktail Party

Introduction

Imagine after our topic presentation, we go to a crowded pub to celebrate. You are trying hard to ignore those eighties tunes blaring out of the loudspeakers while you are chatting with your colleagues. Suddenly, this familiar melody from a Mozart's piano concerto grabs your attention – yes, your cell phone alerts you that your partner just called for the third time…

This classic phenomenon, in which our brain analyzes a scene by perceptually organizing sensory data to form auditory objects (or auditory streams), is often referred to as the "Cocktail Party Effect" (Cherry, 1953). Many cues have been identified that influence perceptual organization, but only little is known about the actual brain mechanisms underlying this phenomenon. In this proposed topic, we look at the latest development in the quest to find the neural basis for auditory stream segregation.

Background – Auditory grouping mechanisms

The ability to form auditory objects is important in the natural environment where sounds arriving at our ears are a resultant of all spectro-temporal components that may have arisen from different auditory events. We constantly analyze an auditory scene by trying to group related components from one source, and segregating out other frequency components that are not in the attended object (Carlyon, 2004). This ability of bringing acoustical events to the attention foreground may increase the chance of survival for a species in the animal kingdom through auditory awareness to the movement of their predators. Humans also rely on auditory grouping mechanisms for daily communication, especially in the presence of noise and other competing sources, since we need to group simultaneous components originating from a single source across frequency, as well as grouping events across time, in order to hear whole words and messages. In his seminal book, Bregman (1990) provides us with working definitions on the terminologies used in the world of auditory scene analysis. He described auditory stream segregation as:

> The general process of auditory scene analysis in which links are formed between parts of the sensory data. These links will affect what is included and excluded from our perceptual descriptions of distinct auditory events.

If these links, which are also commonly referred to as streaming cues, are correct sensory parts across time, we refer to such perceptual grouping as sequential. However, if these sensory data coexist in time, and the formation of multiple auditory objects are as a result of perceptually partitioning of the spectral contents into distinct objects, e.g., harmonics in a vowel, it is referred to as simultaneous grouping. Apparent spatial location, onset / offset synchrony, frequency proximity, and fundamental frequency are but some of the common acoustic cues that we employed in auditory scene analysis.

Experimental paradigm – Buildup of Streaming

How does one systematically investigate the process of stream segregation that occurs in a complex auditory environment? A popular paradigm is to use an ambiguous auditory figure that could either be heard as one stream with a galloping rhythm (commonly labeled as “Horse”) or as two concurrent streams with two different *tempi* (“Morse”) (See Figure 1). The basic stimulus consists of a high tone A alternating with a low tone B, in repeated ABA_ sequences. If the frequency difference (Δf) between the A and B tones is small, then neighboring tones tend to bind together perceptually, resulting in a “Horse” rhythm. Conversely, if Δf is large, the A and B tones are no longer bind to each other, resulting in a “Morse” rhythm. The tone repetition rate also influences the percept: the faster the repetition rate, the more binding there is between the A and B tones (van Noorden, 1975).

At intermediate values of Δf and tone repetition rate, the initial galloping “Horse” percept changes after a few seconds of listening to a “Morse” percept (Anstis and Saida, 1985; Bregman, 1978; Carlyon *et al*, 2001). With this systematic change in auditory percept over time, it is possible to record neural responses at various points during an ongoing sequence of sounds without any change in the evoking stimulus, and compare the neural responses associated with dramatically different percepts.

Figure removed due to copyright considerations.

Figure 1 For the correct parameters, these sequences are ambiguous and can be heard with one or two perceptual organizations with different rhythms: (Left): a characteristic galloping rhythm ("Horse"); (Right): 2 isochronous streams, like Morse code ("Morse"). Colored regions correspond to perceptual streams. (Taken from Fig. 2, Cusack, 2005).

Key Areas of Research

Many investigations into the neural correlates of auditory streaming employ the aforementioned ABA- paradigm, or the variants thereof, but their generic approaches and the specific questions they address can be divided into several distinct classes. One class of approach is to use single-unit recordings in animals to infer perceptual effects of stream segregation. Fisherman *et al* (2001, 2004) showed that multiunit spiking responses to tone sequences in the primary auditory cortex of awake monkeys follow the pattern that one might expect on the basis of published psychophysical data from human subjects. Bee and Klump (2004) performed similar sequential streaming experiments and recorded neural responses in the auditory forebrain of awake starlings. They concluded that while there are preattentive auditory processes, such as frequency selectivity and forward masking, that contribute to the perceptual segregation of sequential acoustic events having different frequencies into separate auditory streams, there may be additional processes that are required to account for all known perceptual effects related to sequential stream segregation. The major confounding factor in interpretations of these conclusions is that the neural response patterns that are putatively associated with the one- and two-stream percepts were always induced using physically different stimuli.

Another class of approach combines human neuromagnetic and behavioral measures. Gutschalk *et al* (2005) used magnetoencephalography (MEG) to examine the neural bases of auditory stream formation. They concluded that there is a tight coupling between auditory cortical activity and streaming perception, suggesting that an explicit representation of auditory streams may be maintained within nonprimary auditory areas. They hypothesized that selective adaptation by feature-specific neurons might be a general neural mechanism subserving perceptual organization. Cusack (2005) combined functional magnetic resonance imaging (fMRI) with psychophysical experiments to test for an effect of perceptual organization across the whole brain.

The other class of approach addresses the specific interaction between auditory streaming and attention. This remains a controversial topic. Due to time constraint, the scope of the proposed discussion will not include literature that explicitly addresses this issue, but some key results are summarized here for completeness. It has been suggested that streaming is a preattentive process based on recording of event-related potentials (ERP) using a mismatch negativity paradigm (MMN) (Sussman *et al*, 1999). It was argued by these authors that since MMN, which can occur in the absence of attention, is elicited only in the infrequent odd-ball presentation within-streams, the two potential streams of sound were segregated without attention being focused on the auditory stimuli. However, Carlyon *et al* (2001) and Cusack *et al* (2003) argue that the buildup of streaming is an attentive process, and previous experiments by van Noorden (1975) suggests that listeners have control over their perception in the ambiguous Δf region. In other words, whether or not streaming is an attentive process is not conclusive, but the techniques used to probe such a high cognitive process are, nonetheless, interesting to follow.

The last class of active research area is computational modeling of stream segregation (e.g., Beauvois and Meddis, 1991; McCabe and Denham, 1997). An interesting neural network model, known as ARTSTREAM (Grossberg, 2004), uses adaptive resonance theory to propose how the brain achieves auditory scene analysis. Even though this neural network structure does not have clear neural correlates, its putative structure may inspire neurophysiologists to find neural units that have similar behaviors. This is perhaps an interesting paper to discuss if the class would like to be exposed to neural network modeling.

Recent Developments and Proposed Papers for Discussion

In this month's publication of *Neuron*, Micheyl and colleague (Micheyl, 2005) reported their comparative experiments on auditory stream segregation of human and awake monkeys. Unlike in previous studies, they used identical stimuli to conduct psychophysical experiments in humans and single-unit extracellular recordings in the primary auditory cortex of awake monkeys. Interestingly, using signal detection theory discussed in previous sessions, the authors can now compare quantitatively the stochastic neural responses with the probabilistic perceptual judgments.

Earlier this year, Cusack (2005) used fMRI to test for an effect of perceptual organization across the whole brain and suggested that the intraparietal sulcus (IPS) might play a role in perceptual organization. This is also exciting in light of the growing evidence in the literature suggesting that the IPS is also important for binding in vision, touch and cross-modally.

In a sequel to Wang's (2005) paper, Bartlett *et al* (2005) extended the investigation on recordings of auditory cortical neurons in awake marmosets to 2-sound sequences. Not only do their results demonstrate that persistent modulations of the responses of an auditory cortical neuron to a given stimulus can be induced by preceding stimuli, they also find that decreases or increases of responses to the succeeding stimuli are dependent on the spectral, temporal, and intensity properties of the preceding stimulus. Such long-lasting modulation by stimulus context in the cortex suggests these response properties may be important for auditory streaming and segregation.

All the proposed papers for discussion highlight some new experimental approaches or latest findings on the neural correlates of auditory stream segregation. This is a very active area of research on many fronts, ranging from neurophysiology to psychoacoustics. However, we are still in the infancy stage in the quest of understanding the brain mechanism behind auditory scene analysis. Next time when you are in a pub drinking a beer, marvel at how the brain accomplishes such an amazing feat!

Papers for Discussion

Bartlett, E., Wang, X. (2005). "Long-Lasting Modulation by Stimulus Context in Primate Auditory Cortex," *Journal of Neurophysiology*, **94**, 83-104.

Cusack, R. (2005). "The Intraparietal Sulcus and Perceptual Organization," *Journal of Cognitive Neuroscience*, **17**, 641-651.

Micheyl, C., Tian, B., Carlyon, B., Rauschecker, J. (2005). "Perceptual Organization of Tone Sequences in the Auditory Cortex of Awake Macaques," *Neuron*, **48**, 139-148.

Background

Carlyon, R. (2004). "How the brain separates sounds," *Trends in Cognitive Sciences*, **8**, 465-471.

Darwin, C.J. (1997). "Auditory grouping," *Trends in Cognitive Sciences*, **1**, 327-333.

Further Reading

Bee, M., Klump, G. (2004). “Primitive Auditory Stream Segregation: A Neurophysiological Study in the Songbird Forebrain,” *Journal of Neurophysiology,* **92**, 1088-1104.

Fishman, Y., Reser, D., Arezzo, J., Steinschneider, M. (2001). “Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey,” *Hearing Research*, **151**, 167-187.

Fishman, Y., Arezzo, J., Steinschneider, M. (2004). “Auditory stream segregation in monkey auditory cortex: effects of frequency separation, presentation rate, and tone duration,” *Journal of Acoustical Society of America*, **116**, 1656-1670.

Grossberg, S., Govindarajan, K., Wyse, L., Cohen, M. (2004). “ARTSTREAM: a neural network model of auditory scene analysis and source segregation,” *Neural Networks*, **17**, 511-536.

Gutschalk, A., Micheyl, C., Melcher, J., Rupp, A., Scherg, M., Oxenham, A. (2005). “Neuromagnetic Correlates of Streaming in Human Auditory Cortex,” *Journal of Neuroscience*, **25**, 5382-5388.

Sussman, E., Bregman, A., Wang, W., Khan, F. (2005). “Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments,” *Cognitive, Affective, & Behavioral Neuroscience,* **5**, 93-110.

Winkler, I., Takegata, R., Sussman, E. (2005). “Event-related brain potentials reveal multiple stages in the perceptual organization of sound,” *Cognitive Brain Research*, **25**, 291-299.

Reference

Beauvois, M., Meddis, R. (1991). “A computer model of auditory stream segregation,” *Quarterly Journal of Experimental Psychology*, **43**, 517-541.

Bregman, A. S. (1978). “Auditory streaming is cumulative,” *Journal of Experimental Psychology: Human Perception and Performance,* **4**, 380-387.

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organization of Sound*, Cambridge, MA: MIT Press.

Carlyon, R. P., Cusack, R., Foxton, J. M., Robertson, I. H., (2001). “Effects of attention and unilateral neglect on auditory stream segregation,” *Journal of Experimental Psychology: Human Perception and Performance*, **27**, 115-127.

Cherry, E. C. (1953). “Some experiments on the recognition of speech, with one and with two ears”, *Journal of the Acoustical Society of America* **25**, 975-979.

Cusack, R., Deeks, J., Aikman, G., Carlyon, R. P. (2004). “Effects of Location, Frequency Region, and Time Course of Selective Attention on Auditory Scene Analysis,” *Journal of Experimental Psychology: Human Perception and Performance*, **30** (4), 643-656.

McCabe, S., Denham, M. (1997). “A model of auditory streaming,” *Journal of Acoustical Society of America*, **101**, 1611-1621.

Van Noorden, L. (1975). “Temporal coherence in the perception of tone sequences,” doctoral thesis, Eindhoven University of Technology, The Netherlands.

Wang, X., Lu, T., Snider, R., Lian, L. (2005). “Sustained firing in auditory cortex evoked by preferred stimuli,” *Nature*, **435**, 341-346.