

# Imaging the Listening Brain

Adrian KC Lee

Postal:

Department of Speech and  
Hearing Sciences  
Institute for Learning & Brain Sciences  
(I-LABS)  
University of Washington  
Portage Bay Building, Room 206  
Box 357988  
Seattle, Washington 98195-7988  
USA

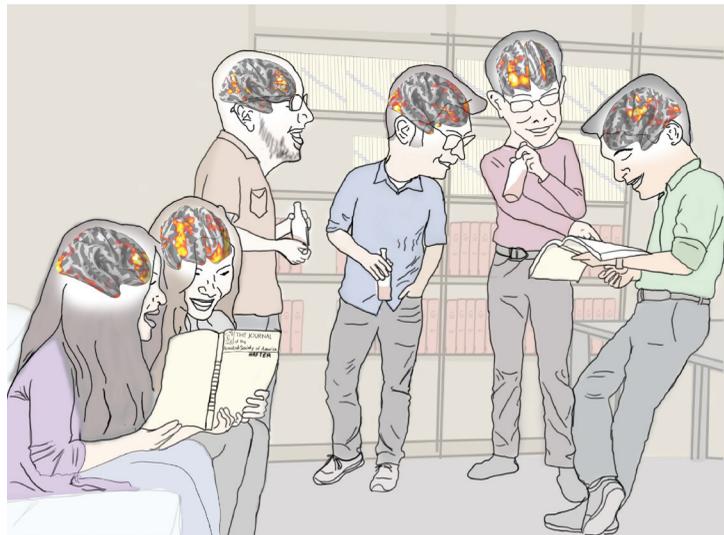
Email:

akclee@uw.edu

*Different neuroimaging tools can help us discover how our brains selectively listen to one voice in a crowded environment.*

## Introduction

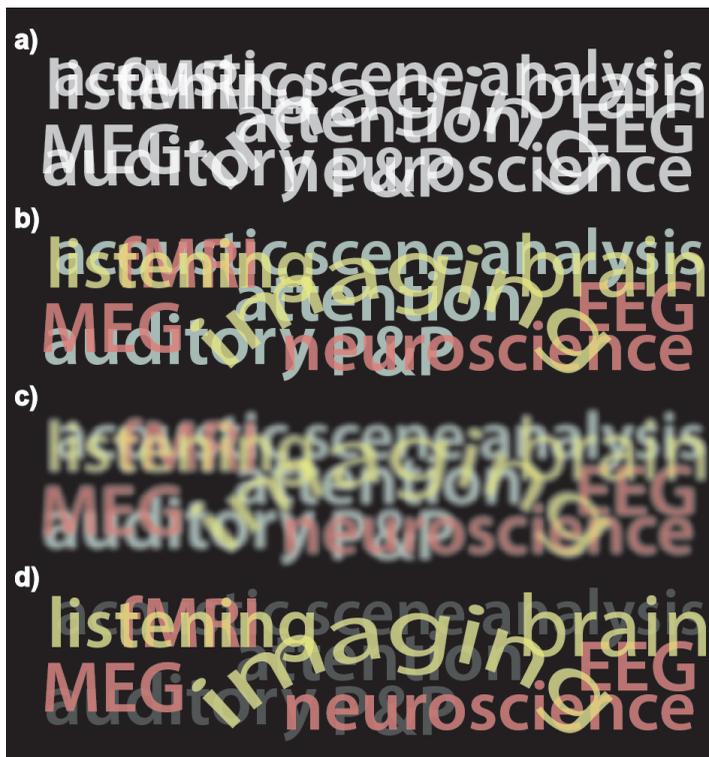
At every Acoustical Society of America (ASA) meeting, we mingle in one big hall at the conference venue and socialize. Some like to catch up with old friends; others continue in deep discussions about all things acoustic. Most listeners without hearing loss can “tune in” to a voice of interest and “tune out” everything else in a crowded environment; more than six decades ago, this behavioral challenge was first coined in *The Journal of the Acoustical Society of America* (Cherry, 1953) as the “cocktail party problem.” Many listeners, especially the younger trainees, can achieve this feat effortlessly (**Figure 1**). However, picking out a single acoustical source from a mixture available to the listener’s ears is not trivial; even state-of-the-art machine-learning algorithms can not yet reliably achieve this task (but see the latest deep-learning approach taking us closer to this lofty computational goal; Wang, 2017).



**Figure 1.** A casual gathering of the Lee lab discussing different research articles published in *The Journal of the Acoustical Society of America*. It can be hard to listen to one speaker with so many simultaneous conversations occurring in this cocktail party environment.

Why is it so difficult to solve the cocktail party problem (Middlebrooks et al., 2017)? It is impossible to work out what (and how many) sound sources have been added to create a given acoustic mixture because this is a mathematically ill-posed problem.

Fortunately, there are auditory features that can provide clues on how to separate sounds in a mixture and recover their veridical sources. This general problem is known as auditory scene analysis (Bregman, 1990). To illustrate this, consider **Figure 2** as a visual analogy. Different acoustical sources arriving at the ears contain spectrotemporal elements that overlap one another, making it very hard to understand each source separately (**Figure 2a**). Acoustical features associated with each source (such as pitch, loudness, and location cues, depicted in different colors in **Figure 2b**) can help disambiguate separate sound objects. If the spectrotemporal resolution is compromised, perhaps due to hearing loss, or the acoustical fea-



**Figure 2.** Visual analogy of parsing an auditory scene. **a:** Different sources overlapping each other, making them hard to separate. **b:** Acoustical features, here depicted in different colors, help separating auditory objects. **c:** Degraded signals can make scene analysis more difficult. **d:** Attention can be maintained or switched across two groups of texts.

tures are degraded such as in a highly reverberant concert hall, auditory scene analysis is even more difficult (Figure 2c). Finally, communication in social settings often involves selectively maintaining attention to a particular auditory object (e.g., your colleague giving constructive criticism on the paper you just presented) and switching attention to other sounds in the environment (e.g., the waiter just passed by asking whether you care for a canapé). This flexibility of attending one sound source and then switching to another at the next instant (e.g., switching attention between two groups of texts in Figure 2d as a visual analogy) is the hallmark of active listening.

How does the brain accomplish this active listening task? Here, I describe our current understanding of how the brain analyzes an auditory scene and forms auditory objects as well as how neural activity in the brain is modulated when we selectively attend to one sound in a mixture. But first, I briefly describe the neuroimaging techniques we use to study the listening brain.

## Imaging Techniques

### Functional Magnetic Resonance Imaging

Functional magnetic resonance imaging (fMRI) is a widely used noninvasive neuroimaging technique. When part of the brain works harder, its oxygen consumption increases; fMRI measures this blood oxygenation level-dependent (BOLD) signal. Even though the BOLD signal is only a proxy for neural activity, there is good correspondence of local field potential reflecting the underlying neural activity and the BOLD response (Logothetis, 2008). This neuroimaging technique has excellent spatial resolution but poor temporal resolution (Figure 3).

Acoustically, fMRI as a neuroimaging technique presents a unique challenge, especially in studies involving auditory attention. Noise in the fMRI environment can reach 110-dB sound pressure level (SPL), approximately as loud as a chainsaw. Hearing protection is used to mitigate noise exposure, but this scanner noise is an unavoidable component of the auditory scene that a subject hears during an fMRI study and can directly affect auditory experimentations (Cusack, 2005). An alternative strategy is to take brain images not continuously but sparsely in time and present an auditory stimulus during the silent periods between imaging acquisitions (Hall et al., 1999). However, this approach lowers the signal-to-noise ratio of the image, and the corresponding sporadic scanner noise can potentially be more distracting than having scanner noise present continuously. These factors illustrate the challenges experimenters face when designing psychoacoustic experiments in an fMRI environment.

	fMRI	Source space		Sensor space			
		M/EEG	MEG	EEG	M/EEG	ECoG	Approx. resolution (mm) (ms)
Spatial	20	10	1	1	N/A	1	20 10 1
Temporal	1	1	1	1	1	1	1000
Coverage	whole brain	whole brain	whole brain	whole brain	whole brain	partial	

**Figure 3.** Trade-offs across neuroimaging modalities in terms of spatial resolution, temporal resolution, and brain coverage. Each neuroimaging technique has its own strengths and weaknesses. Functional magnetic resonance imaging (fMRI) has millimeter resolution but temporal resolution is worse than with other techniques. Magnetoencephalography (MEG) and electroencephalography (EEG; M/EEG when combined) have millisecond temporal resolution but poorer spatial resolution. Electrooculography (ECoG) has both great spatial and temporal resolution but can only record part of the brain, whereas the other techniques can cover the whole brain.

### ***Magnetoencephalography and Electroencephalography***

Magnetoencephalography (MEG) and electroencephalography (EEG or M/EEG when combined) record the magnetic fields outside the head and potentials on the scalp, respectively. These signals reflect the synchronous activation of thousands or millions of neurons (Hämäläinen et al., 1993) compared with fMRI, M/EEG directly measures the electromagnetic activity of the brain. Both technologies can detect activity on the millisecond timescale; this fine temporal resolution is of particular interest when studying auditory processing given the importance of temporal information in audition. However, their spatial resolution is not as fine as fMRI. MEG and EEG have different sensitivity profiles, and when they are used simultaneously, they can provide complementary information about the underlying cortical activities (Sharon et al., 2007).

### ***Magnetoencephalography and Electroencephalography Neural Source Analysis***

Interpreting M/EEG data is a challenge. One can start by analyzing the M/EEG topographical patterns directly (extrapolated from the measurements at each sensor location). However, we face yet another ill-posed problem similar to the cocktail party problem. It is physically impossible to completely and uniquely determine which brain areas were active given the measured M/EEG data. Still, there are two general approaches to relate M/EEG data to the neural sources: (1) source localization and (2) inverse imaging.

The source localization approach assumes that only a small handful of brain regions are active and uses equivalent current dipoles to represent that activity when modeling the M/EEG data. This approach is often favored when the experimenter wants to make inferences about how activity at specific regions of interest (e.g., the bilateral auditory cortex) differs across experimental conditions (e.g., actively attending to a sound vs. passively listening while watching a silent movie). However, a disadvantage of this approach is that other neural activities originating outside the modeled dipoles can change the estimated activity at the modeled locations, leading to erroneous interpretations of results (e.g., failing to account for premotor activities in an auditory attention task can influence our interpretation of the attentional modulated activities modeled in the auditory cortex; Bharadwaj et al., 2014).

Inverse imaging is an alternative approach that also models M/EEG data, but unlike source localization, it estimates activity across the entire brain. Using other information to constrain the ill-posed problem (e.g., by constraining the location of neural activity to the cortex by using the individual's MRI structural scans) allows the experimenter to map M/EEG sensor data to neural sources directly interpretable on the cortex. However, a disadvantage of this approach is the idiosyncrasies associated with the specific inverse model chosen; for example, the often-used minimum-norm model produces estimates of brain activity that are potentially over-smoothed in space.

### ***Electrocorticography***

Electrocorticography (ECoG), sometimes also called intracranial EEG (iEEG), measures brain signals using electrodes that are implanted either above or below the dura, the tough outermost membrane enveloping the brain. Before surgical intervention, some epilepsy patients are implanted with ECoG electrodes to better localize the part of the brain responsible for the seizure onsets. These patients sometimes volunteer for different sensory and cognitive tasks while they are in an extended hospital stay, which provides an opportunity for experimenters to record brain activity from a relatively large patch of cortex with unmatched temporal and spatial resolution. For ethical reasons, the sites of recording are dictated by clinical needs, and thus, cortical coverage varies across patients.

### ***Different Imaging Techniques to Answer Different Questions***

As outlined above, each recording technique has its own strengths and weaknesses (Figure 3), so experimenters choose different imaging techniques to answer different questions. For example, to carefully examine the subdivisions of the human auditory cortical areas, fMRI is often preferred due to its superior spatial resolution. Conversely, to track how the cortical dynamics change with ongoing sound, MEG and EEG are preferred due to their high temporal resolution. ECoG has both fine spatial and temporal resolution, but the cortical coverage is limited and out of the hands of the experimenter. This limits the opportunity to address an array of system neuroscience questions such as how one area of the brain is functionally correlated with another region to coordinate active listening. Overall, the experimental question must be matched with an appropriate recording modality.

## Analyzing the Auditory Scene in the Brain

How does the brain group acoustic energy across frequencies and process them as a unifying sound object through time, such as the grouping harmonics of a voice? One theory, called the temporal coherence model (Shamma et al., 2011), posits that the brain solves this scene analysis problem by collating information across neural populations encoding various sound features, such as pitch and spatial location, that are temporally correlated with each other. Intuitively, this theory makes sense: components from a sound source tend to be coherent with each other. For example, consider a baby crying and being soothed by his/her father's voice. Specifically, the harmonics of the father's voice will be coherently modulated by the same vocal tract. Coincidentally, the spatial cues associated with these harmonics will also be coherently modulated because they all came from the same location. However, these harmonics will be uncorrelated with the crying baby because the spectral energy came from two different independent sources.

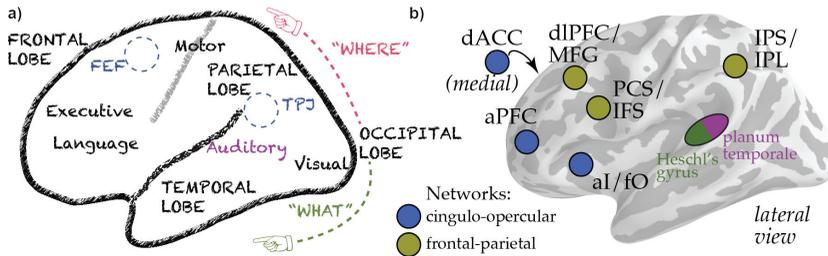
In the laboratory, experimenters use a random figure-ground stimulus to study this hypothesis (Teki et al., 2011). As a visual analogy, psychologists refer to how you group words on this page as "figure" and the white space as "ground." In this auditory figure-ground experiment, the number of tones coherently modulated in frequency is systematically changed and also presented with other random tones. As the number of coherently modulated tones increase, the subjects are more likely to detect this moving figure amid the random ground. A recent study recorded EEGs while the subjects were exposed to these figure-ground stimuli (O'Sullivan et al., 2015). When subjects were watching a silent movie and are passively exposed to these auditory stimuli, these investigators found that the brain tracks this temporal coherence with an onset as early as about 115 ms. This is remarkably fast and suggests that the brain begins to group sounds early in the auditory processing hierarchy. Furthermore, this processing can exist without explicit attention. When subjects are actively engaged in listening to these auditory stimuli (viz., press a button when they hear a specific trained pattern), the neural representation of this temporal coherence is even more pronounced. This suggests that active listening and selective attention can further enhance this temporal coherence processing. The use of source-modeling techniques to analyze these temporal coherence responses has shown that the neural sources originate from the bilateral temporal areas in both passive and active conditions, where-

as previous fMRI research has shown activation bilaterally in the parietal region in response to varying coherence levels (Teki et al., 2011).

## Attention Modulation in the Auditory Cortex

Many studies have asked whether auditory-related brain regions show evidence of "tuning in" to a sound of interest in selective attention tasks. Back in the 1970s, a seminal EEG study (Hillyard et al., 1973) tested whether the attentional state of the listener would modulate the N100 response, a negative deflection in the electric potential of auditory sensory areas that occurs around 100 ms after the onset of an auditory stimulus. Listeners were presented with two streams of tone pips, one ear with a lower frequency relative to that of the other ear, and they were tasked to pick out when a rare "oddball" tone was introduced into the attended stream. Using only one electrode recorded on top of the head at the vertex, the researchers found that the strength of the N100 response varied depending on which stream listeners were instructed to attend. Due to the properties of electromagnetism, every EEG component has an MEG counterpart. The magnetic counterpart of the N100 response, called the M100, was also found to be modulated by attention. Furthermore, using the source localization technique, this attention-modulated M100 component was localized to the auditory cortex (Woldorff et al., 1993). Furthermore, a modulatory effect was found in components preceding the M100, suggesting that the auditory response might be modulated even before the initial stages of sound processing in the cortex, most likely originating from other cortical attentional control centers, a topic that we discuss later. Subsequent fMRI studies using an array of behavioral paradigms also showed attentional modulation of BOLD activity in the auditory sensory areas and that this attention modulation is frequency specific. Like a radio, it seems that the human auditory cortex can tune into a preferred frequency channel on demand (Da Costa et al., 2013).

The aforementioned studies focused on the simple laboratory situation of selectively attending to one of two auditory streams of tones. In recent years, more advanced signal-processing and neural-modeling techniques allow experimenters to examine listening situations that are akin to our original "cocktail party problem." In one MEG experiment (Ding and Simon, 2012), listeners were presented with two lengthy speech streams from different talkers (either the same or different gender) mixed into a single acoustic channel. This



**Figure 4.** Beginner's guide to the listening brain. Shown is a lateral view of the left hemisphere of the cerebral cortex, with the medial portion hidden behind; the front of the head is to the left. **a:** Different lobes in the human cerebral cortex with approximate functional areas and the “what”/“where” pathways. **b:** Different regions of the cingulo-opercular and frontoparietal systems (**dark gray**, valleys/sulci; **light gray**, crests/gyri). FEF, frontal eye field; TPJ, temporoparietal junction; aPFC, anterior prefrontal cortex; al/fO, anterior insular/frontal operculum; dACC, dorsal anterior cingulate cortex; diPFC/MFG, dorsolateral prefrontal cortex/middle frontal gyrus; IPS/IPL, intraparietal sulcus/inferior parietal lobule; PCS/IFS, precentral sulcus/inferior frontal sulcus.

acoustic mixture was then presented identically to each ear without any spatial cues available. Listeners were asked to attend to only one of the two speakers.

Similar to previous findings based on tone pips, the neural representation of the attended speech stream was stronger than the ignored stream. Specifically, the neural response is more phase-locked to the envelope of the attended speech stream in the presence of a competing speech stream. Although neural sources were also detected 50 ms after stimulus onset, the difference in attended versus unattended neural signals primarily arose 100 ms poststimulus. The earlier sources did not differ between the attended and the unattended streams, in agreement with other previous EEGs. Localization analysis of the earlier sources suggests that the neural source for these earlier components originates from Heschl's gyrus (**Figure 4b**), the structure containing the human primary auditory cortex, whereas the neural sources for the later component originate from the planum temporale, the cortical area just posterior to Heschl's gyrus. A possible interpretation is that the entire auditory scene is processed by the primary auditory cortex and it is only weakly sensitive to selective attention, whereas the higher order auditory areas in the planum temporale receive the processed neural signals, with the speech streams already segregated. At this level of processing, perhaps more neural sources are devoted to processing the attended than the unattended speech stream, leading to a stronger difference in neural signal (Simon, 2017).

### Speech Reconstruction Based on Neural Signals

ECoG studies have also found stronger representations of the attended speaker in a multitalker environment (Mesgarani and Chang, 2012). Leveraging the high spatial and temporal resolution available in ECoG, experimenters used these fine-grained neural signals to decode the neural spectrogram, such as reconstructing the attended speech at high temporal and spectral fidelity from the original acoustics of the speech signals. Furthermore, the success of these speech reconstructions correlates with the listener's behavior; reconstruction was successful only in trials when the subjects correctly reported the target words (and not during error trials) and the reconstructed neural spectrogram better reflects the portion of the speech stream when the subjects only attended to one speaker compared

with the earlier period of the task when subjects had to listen for the target call sign from both speakers.

The stimulus-reconstruction method employed in both M/EEG and ECoG studies differs in the frequency content of the neural signals used to recreate the attended speech signals. In ECoG studies, the low-frequency fluctuation (<8 Hz) of the high-frequency gamma signals (>70 Hz) is often used to analyze the attended and unattended speech streams. In M/EEG studies, both attended and unattended speech representations can be seen at the low-frequency representations (<8 Hz), but the signal-to-noise ratio in the high-frequency gamma range is too low to detect meaningful changes. Besides the difference in usable frequency content separating these two recording techniques, there is another, perhaps more important, differentiation between ECoG and M/EEG approaches: the former is invasive and the latter is noninvasive. Despite recording further away from the neural sources in M/EEG, the stimulus-reconstruction method can be effective enough to classify which of two speakers the listener was attending to in single trials (with speech segments of about 60 s; O'Sullivan et al., 2014). This technological development is an important advance toward designing a hearing aid that can follow the user's attention to selectively amplify the sound of interest, the holy grail of a futuristic hearing aid design (Lee et al., 2013a).

### Modulation Beyond the Auditory Cortex

The previous section discussed how the auditory sensory areas seem to faithfully follow the attended speech signal when

maintaining attention on one source in a multitalker environment. However, are there other areas of the brain recruited to help us listen for one talker or to coordinate switching of attention from one talker to another? Does it depend on what features we are attending to, such as listening for the speaker on my left side or listening for the speaker with the highest pitch?

In the visual attention literature, there is a prevalent dual-path theory for spatial and nonspatial feature processing. The dorsal stream (or the “where” pathway) is involved with processing spatial information, whereas the ventral stream (or the “what” pathway) is involved with object identification and recognition (Desimone and Duncan, 1995; **Figure 4a**). A similar dual pathway has also been proposed for auditory information processing (Rauschecker and Scott, 2009), with areas of the primary auditory cortex and surrounding areas organized spatially to feed into this dual-pathway configuration (Rauschecker, 1998). Even before a sound begins, we can direct our attention volitionally to “listen out” for a specific feature of sound. In a recent M/EEG study using inverse imaging constrained by MRI information, our laboratory asked subjects to report back one of two spoken digits (Lee et al., 2013b). These two spoken digits were monotonized (one with a higher pitch and the other with a lower pitch) and convolved with two different head-related transfer functions, one simulating the source coming from the left and the other from the right. We cued the subjects using a visual arrow (viz., left, right, up, or down), indicating which feature they should prepare to attend to even before the onset of the stimulus. We found that the neural response in the left frontal eye fields, part of the dorsal “where” pathway, was enhanced in the preparation and during a spatial attention task when subjects were prompted to pay attention to a location by a left or right arrow. When, instead, subjects were prompted to attend to a higher or lower pitch by an up or down arrow, the left posterior superior temporal sulcus, an area previously implicated for pitch categorization and possibly in the auditory “what” pathway, was greater in preparation for a pitch attention task. These findings are in line with other neuroimaging studies (e.g., fMRI, EEG; Diaconescu et al., 2011), showing that the deployment of auditory spatial attention engages similar cortical regions associated with visual attention, but selectively attending to nonspatial features invokes cortical areas that are specific to the attended sensory modality. These observations generally agree with the notion that the frontoparietal attention system comprised of the precentral sulcus/inferior frontal sul-

cus, dorsolateral prefrontal cortex/middle frontal gyrus, and intraparietal sulcus/inferior parietal lobule regions (**Figure 4b**) is engaged when we want to listen out for a talker in a crowded environment (Eckert et al., 2016).

To carry on social conversations, we also need to switch attention from one talker to another. Converging evidence suggests that the frontoparietal attention system also participates in attention switching in both the visual and auditory modalities. When listeners switched attention between a male and female stream, elevated activation was found in the posterior parietal cortex (Shomstein and Yantis, 2006). A similar region in this cortical area was also more active when listeners had to switch attention from one ear to the other. Combined with other converging evidence found in the visual attention literature, this suggests that the posterior parietal cortex participates in both spatial and nonspatial attention in auditory and visual modalities (Serences et al., 2004). However, one recent M/EEG study suggests that switching based on different features might recruit different behavioral strategies, leading to recruitment of different cortical areas (Larson and Lee, 2014). In this study, listeners were cued initially to attend to one of two talkers. These two talkers were either separated in auditory space (to the left or right) or monotonized to have different pitches. Subjects were cued to listen to one of the two talkers in the beginning of the trial and were asked to either maintain or switch attention during an inserted silent gap in this two-speaker mixture. The result shows that the right temporoparietal junction (TPJ; **Figure 4a**) was more activated when listeners had to switch attention based on the available spatial cues only and that the activity in right TPJ during the attention-switching gap period correlates significantly with how well the listeners perform this task. However, when listeners were told to switch attention and the only cue available was the pitch difference between talkers, the left inferior parietal supramarginal area was more engaged throughout the trial in the maintain trials compared with the switch trials. This suggests that the listeners might use different behavioral strategies to accomplish this nonspatial switching task, thereby recruiting a different cortical area with a different activation time course.

What if we make our listening condition even more challenging than switching attention between speakers, such as reading the conference abstract while attending to the speaker? In a recent meta-analysis study of 10 speech-recognition experiments using fMRI, it was found that the cortical regions in the cingulo-opercular executive-control system consistently showed elevated activity during the

challenging listening condition compared with a controlled baseline listening condition (Eckert et al., 2016). This cingulo-opercular system (**Figure 4b**) is composed of the dorsal anterior cingulate, anterior prefrontal, and anterior insula/frontal operculum regions and generally shows increased neural activity when a perceptual or cognitive task becomes more difficult while they are still engaged in the task. Specifically, this cingulo-opercular executive-control system is thought to reflect task monitoring over time and increased activity here has also been associated with increased effort (Burke et al., 2013).

Age is another interesting factor that influences how this executive-control system is called on to help us understand speech in complex environments. Recruitment in the cingulo-opercular system in older adults performing a language task (Harris et al., 2009; Erb and Obleser, 2013), with or without hearing loss compared with their younger counterparts, is hypothesized to reflect the compensatory mechanism for age-related declines in sensory perception and cognition. However, the increased activity of this cingulo-opercular system can come at a cost; the amount of listening effort is so taxing for the listeners that at some point, they would rather give up than attempt to understand speech in these challenging settings.

### Conclusion and Outstanding Questions

Six decades since Cherry (1953) first outlined the “cocktail party problem,” we are slowly uncovering the neurobiological mechanism that enables us to understand speech in a daily multitalker environment. However, there are still many outstanding questions in our field. For example, looking at the talker’s face helps us understand speech better, especially in extremely noisy settings. Although there has been great work in both humans and primates exploring how auditory and visual processing are combined in the cortex, exactly how visual information can aid us in segregating sounds has not been systematically investigated. Additionally, although we know that listeners with hearing loss struggle to communicate in multitalker environments, the ways in which hearing loss impacts object formation in an auditory scene (see **Figure 2c**) as well as how the attention and executive-control networks are recruited to compensate in these conditions are not well understood. Obviously, much more work needs to be done to answer these vexing questions. But next time you are at the ASA social gathering, take a moment to marvel at all the neurobiological machinery underlying your ability to “tune in” to a single conversation in such a crowded auditory environment.

### Acknowledgments

I thank the National Institute on Deafness and Other Communication Disorders (NIDCD), National Institutes of Health (NIH), for the initial support combining my post-doctoral neuroimaging work with my graduate study on auditory attention through the Pathway to Independence Award Mechanism (K99/R00-DC-010196) as well as the continual support through the Research Grant Program (R01-DC-013260). I thank Tiffany Waddington and Dr. Mark Wronkiewicz for their artistic contributions to **Figure 1**.

### Biosketch



**Adrian KC Lee** obtained his BEng (electrical) from The University of New South Wales, Sydney, Australia, in 2002 and his ScD from the Harvard-MIT Division of Health Sciences and Technology in 2007. He is an

associate professor in the Department of Speech & Hearing Sciences and the Institute for Learning & Brain Sciences, University of Washington, Seattle. His research focuses on mapping the cortical dynamics of auditory attention and developing new neuroengineering approaches to improve hearing aid technology. He is an associate editor and coordinating editor of *The Journal of the Acoustical Society of America* in physiological and psychological acoustics.

### References

- Bharadwaj, H. M., Lee, A. K. C., and Shinn-Cunningham, B. G. (2014). Measuring auditory selective attention using frequency tagging. *Frontiers in Integrative Neuroscience* 8, Article 6. doi:10.3389/fnint.2014.00006.
- Bregman, A. (1990). *Auditory Scene Analysis*. MIT Press, Cambridge, MA.
- Burke, C. J., Brünger, C., Kahnt, T., Park, S. Q., and Tobler, P. N. (2013). Neural integration of risk and effort costs by the frontal pole: Only upon request. *The Journal of Neuroscience* 33(4), 1706-1713.
- Cherry, E. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America* 25(5), 975-979. doi:10.1121/1.1907229.
- Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience* 17(4), 641-651.
- Da Costa, S., van der Zwaag, W., Miller, L. M., Clarke, S., and Saenz, M. (2013). Tuning in to sound: Frequency-selective attentional filter in human primary auditory cortex. *The Journal of Neuroscience* 33(5), 1858-1863.
- Desimone, R., and Duncan, J. (1995). Neural mechanisms of selective visual-attention. *Annual Review of Neuroscience* 18, 193-222.

- Diaconescu, A. O., Alain, C., and McIntosh, A. R. (2011). Modality-dependent “what” and “where” preparatory processes in auditory and visual systems. *Journal of Cognitive Neuroscience* 23(7), 1-15.
- Ding, N., and Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America* 109(29), 11854-11859.
- Eckert, M. A., Teubner-Rhodes, S., and Vaden, K. I. (2016). Is listening in noise worth it? The neurobiology of speech recognition in challenging listening conditions. *Ear and Hearing* 37, Suppl. 1, 101S-110S.
- Erb, J., and Obleser, J. (2013). Upregulation of cognitive control networks in older adults’ speech comprehension. *Frontiers in Systems Neuroscience* 7, 116.
- Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., Gurney, E. M., and Bowtell, R. W. (1999). “Sparse” temporal sampling in auditory fMRI. *Human Brain Mapping* 7(3), 213-223.
- Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., and Lounasmaa, O. (1993). Magnetoencephalography—Theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics* 65(2), 413-497.
- Harris, K. C., Dubno, J. R., Keren, N. I., Ahlstrom, J. B., and Eckert, M. A. (2009). Speech recognition in younger and older adults: A dependency on low-level auditory cortex. *The Journal of Neuroscience* 29(19), 6078-6087.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., and Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science* 182(108), 177-180.
- Larson, E., and Lee, A. K. C. (2014). Switching auditory attention using spatial and non-spatial features recruits different cortical networks. *NeuroImage* 84, 681-687.
- Lee, A. K. C., Drews, M. K., Maddox, R. K., and Larson, E. (2013a). Brain imaging, neural engineering research, and next-generation hearing aid design. *Audiology Today* Jan/Feb 2013, 41-47.
- Lee, A. K. C., Rajaram, S., Xia, J., Bharadwaj, H., Larson, E., Hämäläinen, M. S., and Shinn-Cunningham, B. G. (2013b). Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience* 6, Article 190. doi:10.3389/fnins.2012.00190.
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature* 453(7197), 869-878.
- Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*, 485, 233-236.
- Middlebrooks, J. C., Simon, J. Z., Popper, A. N., and Fay, R. R. (Eds.). (2017). *The Auditory System at the Cocktail Party*. Springer International Publishing, New York.
- O’Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., and Lalor, E. C. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cerebral Cortex* 25(7), 1697-1706.
- O’Sullivan, J. A., Shamma, S. A., and Lalor, E. C. (2015). Evidence for neural computations of temporal coherence in an auditory scene and their enhancement during active listening. *The Journal of Neuroscience* 35(18), 7256-7263.
- Rauschecker, J. P. (1998). Cortical processing of complex sounds. *Current Opinion in Neurobiology* 8(4), 516-521.
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12(6), 718-724.
- Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., and Yantis, S. (2004). Control of object-based attention in human cortex. *Cerebral Cortex* 14(12), 1346-1357.
- Shamma, S. A., Elhilali, M., and Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences* 34(3), 114-123.
- Sharon, D., Hämäläinen, M. S., Tootell, R. B. H., Halgren, E., and Belliveau, J. W. (2007). The advantage of combining MEG and EEG: Comparison to fMRI in focally stimulated visual cortex. *NeuroImage* 36(4), 1225-1235.
- Shomstein, S., and Yantis, S. (2006). Parietal cortex mediates voluntary control of spatial and nonspatial auditory attention. *The Journal of Neuroscience* 26(2), 435-439.
- Simon, J. Z. (2017). Human auditory neuroscience and the cocktail party problem. In Middlebrooks, J. C., Simon, J. Z., Popper, A. N., and Fay, R. R. (Eds.), *The Auditory System at the Cocktail Party*. Springer International Publishing, New York.
- Teke, S., Chait, M., Kumar, S., von Kriegstein, K., and Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *The Journal of Neuroscience* 31(1), 164-171.
- Wang, D. L. (2017). Deep learning reinvents the hearing aid. *IEEE Spectrum* 54, 32-37.
- Woldorff, M. G., Gallen, C. C., Hampson, S. A., Hillyard, S. A., Pantev, C., Sobel, D., and Bloom, F. E. (1993). Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences of the United States of America* 90(18), 8722-8726.

## ASA School 2018



# Living in the Acoustic Environment

5-6 May 2018  
Chaska, Minnesota

- **Two-day program:** Lectures, demonstrations, and discussions by distinguished acousticians covering interdisciplinary topics in nine technical areas
- **Participants:** Graduate students and early career acousticians in all areas of acoustics
- **Location:** Oak Ridge Hotel and Conference Center
- **Dates:** 5-6 May 2018, immediately preceding the ASA spring meeting in Minneapolis
- **Cost:** \$50 registration fee, which includes hotel, meals, course materials, transportation from Oak Ridge to the ASA meeting location
- **For information:** Application form, preliminary program, and more details will be available in November, 2017 at [www.AcousticalSociety.org](http://www.AcousticalSociety.org).

