



Published in final edited form as:

Trends Cogn Sci. 2012 July ; 16(7): 363–364. doi:10.1016/j.tics.2012.05.004.

Attention modulates ‘speech-tracking’ at a Cocktail Party

Elana Zion-Golumbic and Charles E. Schroeder

Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, 19062, USA, Department of Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY 10032, USA

Abstract

Recent findings by Mesgarani and Chang demonstrate that signals in auditory cortex can reconstruct the spectrotemporal patterns of attended speech tokens better than those of ignored ones. These results help extend the study of attention into the domain of natural speech, posing numerous questions and challenges for future research.

The remarkable human ability to attend to one conversation among many in a noisy social scene, otherwise known as the ‘Cocktail Party Problem,’ has been studied and debated for over 50 years [1]. Selective attention has a critical role in this essential cognitive capacity; however, the physiological mechanisms by which the brain implements attentional selection remain unclear [2]. Classic findings of studies using simple tonal stimuli indicate that the neural response in auditory cortex is reduced for ignored stimuli compared to attended stimuli [3, 4], indicating top-down influence of attention on the sensory processing of stimuli, depending on their task-relevance. A new study by Mesgarani and Chang [5] help to extend these classic findings to the domain of human speech selection in the Cocktail Party Paradigm. They demonstrate that fluctuations in neuroelectric signal power in the 70–150 Hz (‘high gamma’) range over auditory cortex preferentially ‘track’ attended relative to ignored speech tokens. This is important because the high gamma signal is widely believed to reflect variations in the massed firing of ensembles of neurons (i.e., multiunit activity) that underlie the recording electrodes, and thus provides relatively direct access to neuronal processing of speech, which is not often possible in humans.

Mesgarani and Chang recorded intracranial EEG from electrodes implanted on the pial surface of the brain in patients undergoing clinical evaluation for epilepsy surgery. This technique has sufficient specificity and signal-to-noise ratio to obtain a reliable high-gamma range response. Subjects listened to short speech samples from a standardized corpus, presented either alone or simultaneously (‘cocktail party’). The stimuli had a rigid structure, containing a ‘call sign’ (Ringo/Tiger) and a color-number combination; e.g. “ready *Ringo* go to *Blue-5* now”. When speech samples were presented individually, subjects were required to report the color-number combination. When presented simultaneously, subjects were cued with one of the call signs and were initially required to monitor both speakers to identify which uttered the call sign and then rapidly focus on that speaker so as to glean the relevant color-number combination. This clever design allows analytic comparison of the responses to a particular token when it is designated as attended versus ignored, as well as correlating this neural selectivity with behavioral performance. It also allows examination of brain dynamics associated with the process of initial allocation of attention and its time course.

Mesgarani and Chang show that it is possible to reconstruct both the attended and ignored stimuli from the time course of high-gamma power fluctuations, indicating that both stimuli are represented to some degree in the neural response to the stimulus-mixture. However, in correct trials the attended token was more reliably reconstructed than the ignored token, indicating it was preferentially represented. Error trials in contrast, carried no advantage for the attended stimulus, and in fact, there was a tendency to preferentially represent the other

(supposedly ignored) stimulus, suggesting attention was allocated to the ‘wrong’ speaker in those trials. The demonstrable relationship between performance on this challenging task and neuronal selectivity provides a potentially powerful tool for studying the neural correlates of when selective attention succeeds and when it fails, in both normal and clinical populations.

The key finding of attentional modulation of the high-gamma response to speech complements a previous study, not discussed by Mesgarani and Chang, which showed similar attentional-modulated neuronal selectivity in low-frequency phase-locking to speech [6]. Thus far, speech-tracking responses of low-frequency phase and high-gamma power have mostly been studied separately [7]. The similarity in the way these two speech-tracking responses are modulated by attention, together with findings that high-gamma power is coupled to low-frequency phase, heightens interest in the question of the relationship between these two signal domains. Low-frequency phase indexes the slower synaptic current fluxes that cause and gate firing changes reflected in MUA/high-gamma power. Indeed, several recent papers suggest that the combination of high-gamma power/MUA and low-frequency phase is essential for efficient encoding and perceptual selection of natural auditory stimuli [8–10], however this is yet to be fully explored in the context of speech processing. The Mesgarani and Chang paper serves as a proof of concept that MUA correlates of selective speech encoding can be recorded in humans, and as such paves the way for future research that will deepen our mechanistic understanding of speech processing and selective attention.

Mesgarani and Chang’s findings underscore several intriguing questions about the ‘Cocktail Party’ problem and invite future research on the power and range of auditory selective attention. First, the use of language stimuli (albeit drastically simplified) was an important step, and it will now be important to determine how well these findings generalize to the use of the more complex forms that occur in natural conversations. Second, it is clear that language processing often benefits from use of coincident visual contextual cues and extends well beyond the auditory cortical regions identified in this study. In particular, future studies will need to determine how speech tracking in the auditory cortices relates to activity in other speech-related regions as well as regions involved in attentional control. Finally, while the findings of Mesgarani and Chang show strong preferential neural representation of the attended speech stream, just as is widely demonstrated for simple stimuli, the ignored stream is still represented to some degree.

This relates to long-standing questions regarding the ‘attentional bottleneck.’ To what degree is the ignored stimulus processed? What resources are allocated to process it? How does that influence processing of the attended stimulus? When/where, if ever, does the representation become exclusive only for the attended speaker? While the Mesgarani and Chang paper, and several others recently published (reviewed by Giard & Poeppel, 2012), show that mechanistic investigation of selective attention extends to complex natural stimuli, the view beyond the bottleneck remains obscure.

References

1. Cherry EC. Some experiments on the recognition of speech, with one and two ears. *J Acoust Soc Am.* 1953; 25:975–979.
2. McDermott JH. The cocktail party problem. *Curr Biol.* 2009; 19:R1024–R1027. [PubMed: 19948136]
3. Woldorff MG, et al. Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proceedings of the National Academy of Sciences.* 1993; 90:8722–8726.

4. Hillyard S, et al. Electrical signs of selective attention in the human brain. *Science*. 1973; 182:177–180. [PubMed: 4730062]
5. Mesgarani, N.; Chang, EF. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*. 2012. advance online publication, <http://dx.doi.org/10.1038/nature11020>
6. Ding N, Simon JZ. Neural Coding of Continuous Speech in Auditory Cortex during Monaural and Dichotic Listening. *J Neurophysiol*. 2012; 107:78–89. [PubMed: 21975452]
7. Zion Golumbic, EM., et al. Temporal Context in Speech Processing and Attentional Stream Selection: A Behavioral and Neural perspective. *Brain Lang*. 2012. In press, <http://dx.doi.org/10.1016/j.bandl.2011.1012.1010>
8. Giraud A-L, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci*. 2012; 15:511–517. [PubMed: 22426255]
9. Kayser C, et al. Spike-Phase Coding Boosts and Stabilizes Information Carried by Spatial and Temporal Spike Patterns. *Neuron*. 2009; 61:597–608. [PubMed: 19249279]
10. Schroeder CE, Lakatos P. Low-frequency neuronal oscillations as instruments of sensory selection. *Trends Neurosci*. 2009; 32:9–18. [PubMed: 19012975]