ELSEVIER

Short Communication

Theta-band phase tracking in the two-talker problem

CrossMark

Dillon A. Hambrook *, Matthew S. Tata

The University of Lethbridge, Lethbridge, AB, Canada

ABSTRACT

It is usually easy to understand speech, but when several people are talking at once it becomes difficult. The brain must select one speech stream and ignore distracting streams. We tested a theory about the neural and computational mechanisms of attentional selection. The theory is that oscillating signals in brain networks phase-lock with amplitude fluctuations in speech. By doing this, brain-wide networks acquire information from the selected speech, but ignore other speech signals on the basis of their non-preferred dynamics. Two predictions were supported: first, attentional selection boosted the power of neuroelectric signals that were phase-locked with attended speech, but not ignored speech. Second, this phase selectivity was associated with better recall of the attended speech.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

The human auditory system has a striking ability to selectively perceive a single sound source out of a complex mixture. This general phenomenon and the associated computational challenges have been termed the "cocktail party problem" (Cherry, 1953). This problem emerges in any acoustic scene with more than one sound source. The perceptual consequence of failing to maintain selection in a complex scene has been called auditory information masking (Kidd, Mason, Richards, Gallun, & Durlach, 2007), or more generally, *distraction* (Ponjavic-Conte, Hambrook, Pavlovic, & Tata, 2013).

The neural mechanisms by which we deal with complex scenes have been under intense investigation in recent years. A promising recent theory, called *selective entrainment* (Schroeder & Lakatos, 2009a; Zion Golumbic, Ding, et al., 2013), proposes that this problem is solved in part by phase matching between neuroelectric oscillations of the brain and low-frequency dynamics of acoustic signals. It is known now that neuroelectric oscillatory activity can "track" spectrotemporal modulations in speech (Ahissar et al., 2001; Hertrich, Dietrich, Trouvain, Moos, & Ackermann, 2012; Luo & Poeppel, 2007). Furthermore, selective attention modulates the selectivity or strength of this tracking process (Ding & Simon, 2012; Kerlin, Shahin, & Miller, 2010; Lakatos et al., 2013; Mesgarani & Chang, 2012; Zion Golumbic, Ding, et al., 2013). By selectively tracking the phase of a single audio source, oscillating

ensembles might preferentially represent the tracked signal and reject signals that are not phase locked.

Evidence for such a theory has begun to emerge: theta-band phase tracking of speech is more pronounced when the speech signal is well comprehended relative to when it is degraded and difficult to understand (Peelle, Gross, & Davis, 2013). Thus phase-tracking is a correlate of successful perception. Furthermore, using intracranial electrocorticography (ECoG), (Zion Golumbic, Ding, et al., 2013) showed that oscillatory signals in auditory cortex track the acoustic envelope of speech in a non-selective manner – both attended and unattended speech signals were similarly tracked. By contrast, Medial Frontal Gyrus (MFG) exhibited selective tracking such that the attended speech was preferentially tracked. Since this region of cortex is also known to engage in auditory working memory tasks (Arnott, Grady, Hevenor, Graham, & Alain, 2005; Crottaz-Herbette, Anagnoson, & Menon, 2004), these data suggest a role for phase tracking in linking sensory and memory regions. Finally, theta-band phase tracking of speech was more pronounced when the speech signal was accompanied by video of the talker's lip movements (Zion Golumbic, Cogan, Schroeder, & Poeppel, 2013) – suggesting that phase-tracking is associated with communication between ensembles of neurons that are anatomically distinct but functionally linked.

Selective attention in a complex scene is well-known to enhance perception and memory encoding (Broadbent, 1952; Treisman, 1964). If phase tracking of speech dynamics is a mechanism for implementing selective attention, then variation in perceptual performance should mirror variation in the strength of speech-locked EEG signals. In the present study we report that selective listening in a free-field "two-talker" situation strengthens

* Corresponding author. Address: Department of Neuroscience, The University of Lethbridge, 4401 University Drive, Lethbridge, AB T1K3M4, Canada. Fax: +1 (403)329 2775.
E-mail address: dillon.hambrook@uleth.ca (D.A. Hambrook).

a theta-band signal that tracks the acoustic envelope of selected speech, relative to ignored speech. Furthermore, by reassigning trials on the basis of correct or erroneous recall of a probe word, we found evidence that selective phase tracking of an attended stream enhances the ability to recall that stream.

Briefly, participants listened to two different, simultaneously presented, 15-s audiobook clips read by different speakers, presented 60° to either side of the acoustic midline while EEG was recorded. Before each block of 15-s trials participants were cued to attend to one of the two speakers. Following each trial participants were presented a probe word from the target clip, the distractor clip, or a clip that was not presented on that trial (catch probe). The participants' task was a two-alternative forced choice task to indicate if the probe word was present or absent in either of the previously played clips. EEG data from each trial were cross-correlated with the first derivatives of the speech envelopes of the target and distractor speech clips played on that trial (Hertrich et al., 2012). The first 1000 ms of the EEG data for each trial was excluded as it contained transient responses due to the sudden onset of sound. This cross-correlation function selectively separated brain activity that was phase-locked to energy transients in either speech stream. We tested the prediction that EEG signals independently phase locked to target and distractor streams would be differentiated when the target was successfully encoded, but not when encoding of the target was compromised by the distracting stream.

2. Methods

2.1. Participants

19 Undergraduates from the University of Lethbridge were recruited and participated for course credit. Participants provided informed written consent. Procedures were in accordance with the Declaration of Helsinki and were approved by the University of Lethbridge Human Subjects Review Committee. Participants were neurologically normal and reported normal hearing. Two participants were excluded for failing to respond on a significant number of trials (three standard deviations outside the mean across all trials). Only EEG data from participants who correctly responded at a rate higher than chance (>50% correct) to the target stream were analyzed, thus 16 participants contributed to the data analysis (12 female; two left-handed; average age: 22.2 years).

2.2. Stimuli and task

All stimuli were presented in free field by an Apple Mac Pro with a firewire audio interface (M-Audio Firewire 410). Participants sat between two near-field studio monitors (Mackie HR624 MK-2) arranged 1 m away and 60° from the front auditory midline. Stimulus presentation was controlled by a program custom coded using Apple Computer's Core Audio framework (Mac OS 10.6).

The stimuli consist of 20 segments from the book *World War Z* by Max Brooks, narrated by 20 different readers (one female). Each segment was 15 s long and normalized to the same average root-mean square (RMS) sound amplitude. Three unique probe words were selected from each of the 20 speech segments and audio clips of the selected words were obtained from an online dictionary.

Each participant completed 20 blocks of five trials each. Blocks were of 98 s duration. Each speech segment was the target on five trials. Within each block the presentations of speech segments were randomized and an individual speech segment did not occur twice within a single block. Prior to each block participants were instructed to attend to either the left or right speaker. The target and distractor streams were presented simultaneously from separate speakers for 15 s, followed by a 1 s silence, followed by a probe word presented from both speakers. Participants were given 3.5 ± 0.25 s following the probe word to respond before the start of the next trial. Probe words were drawn from the target stream, distractor stream, or a stream that was not presented on that trial (probe absent or "catch" trials). Participants performed a two-alternative forced choice task to indicate if the probe word was present or absent in either of the speech clips.

2.3. EEG analysis

EEG was recorded with 128 Ag/Ag-Cl electrodes in an elastic net (Electrical Geodesics Inc., Eugene, OR, USA). Scalp voltages were recorded at a 500 Hz sampling rate and impedances were maintained under 100 kΩ. Data were high-pass filtered at 0.1 Hz to remove DC offset. Data were first analyzed using the BESA software package (Megis Software 5.3, Grafelfing, Germany). Data were visually inspected for bad channels and the signal from a small number of electrodes (10 or less) was replaced with an interpolated signal. Because of the length of the trials, eye movement artifacts occurred in a majority of trials, therefore eye movement artifacts were corrected using the adaptive artifact correction algorithm (Ille, Berg, & Scherg, 2002). Data were interpolated to an 81-channel 10–10 montage and exported from BESA and further analyzed in MATLAB (MATLAB version 7.10.0; The Mathworks Inc., 2010, Natick, MA, USA) using custom scripts and EEGLAB functions (Delorme & Makeig, 2004).

To isolate EEG activity phase-locked to the competing speech samples, the first derivative of the acoustic envelope was calculated. The acoustic envelope of each sample was calculated by taking the absolute value of the Hilbert transform of the sample and low-pass filtering at 25 Hz. The acoustic envelope was then down-sampled to match the sample rate of the EEG data. The first-derivative of the resulting signal was calculated, half-wave rectified, and normalized such that the sum of the signal across the whole epoch equaled 1 (Hertrich et al., 2012). Thus a signal which captures transient energy increases, an aspect of acoustic stimuli to which the auditory system is known to be tuned, was obtained (Fishbach, Nelken, & Yeshurun, 2001; Howard & Poeppel, 2010). This signal was then cross-correlated with each channel of the time-aligned EEG data to arrive at a cross-correlation function which reflects activity that is phase-locked to acoustic transients in either stream.

To determine the frequency content of the observed phase-locked activity wavelet decomposition was performed on the cross-correlation function. Evoked power was calculated as the power in the trial-averaged cross-correlation function, normalized by the mean evoked power across the whole [−200, 800] ms epoch.

3. Results

Repeated measures t-tests were conducted to compare differences in response rates (Fig. 1) when the probe was drawn from the target stream, the distractor stream, or a stream that was not heard on that particular trial (i.e. a "catch" trial). Participants successfully detected the presence of the probe when it was in the target stream ("responded present" vs. "responded absent", $t = 11.16$, $p < 0.0001$), but not when it was in the distractor stream ($t = -0.72$, $p = 0.4846$). Participants also successfully noted the absence of the probe on "catch" trials ("responded present" vs. "responded absent", $t = -6.4$, $p < .0001$). The proportion of correct detections ("responded present") was greater when the probe was present in the target stream relative to the distractor stream ($t = 4.89$, $p = .0003$).
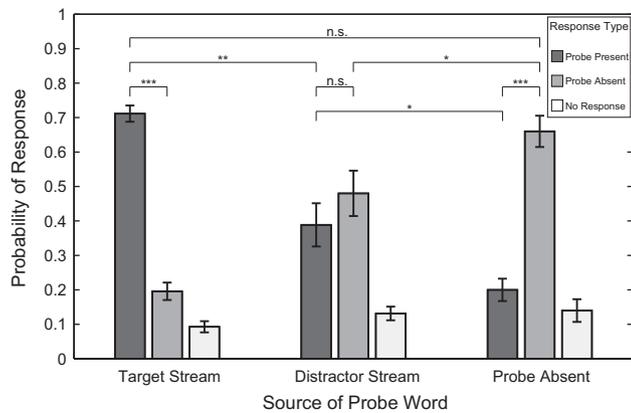
Fig. 1. Probability of response types when the probe word is from the target stream, distractor stream, or an unheard stream. Probability of a correct response is lower when the probe is drawn from the distractor stream, compared to when the probe is from the target stream or from an unheard stream (*$p < 0.05$; **$p < 0.001$; $p < 0.0001$). Error bars indicate standard error of the mean.
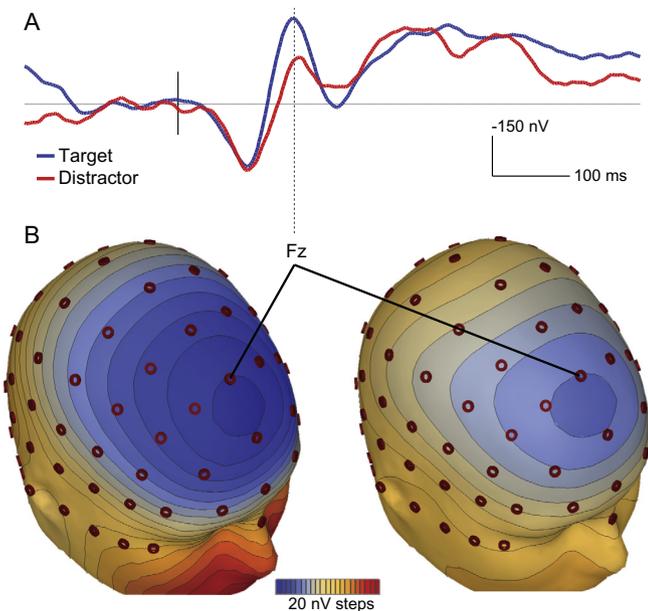

Fig. 2. (A) Cross-correlation of the EEG signal and the first derivative of the speech envelope of target (blue) and distractor (red) speech streams at electrode Fz. (B) Isopotential maps of the cross-correlation function for target (left) and distractor (right) speech streams at 150 ms lag, marked by a dashed line above.

Fig. 2A shows the grand averaged cross-correlation of 14-s EEG epochs with the first derivatives of the acoustic envelopes for the target and distractor speech streams, at electrode Fz. The most robust difference occurred at ∼150 ms lag. The cross-correlation functions for target and distractor speech had similar scalp topographies at this lag (Fig. 2B), suggesting that the difference was due to an increase in power that was phase-locked to the first-derivative of the acoustic envelope, rather than a spatial reconfiguration of cortical generators.

Previous studies suggested that EEG signals were maximally phase-locked to speech in the theta band (∼6 Hz). We used a wavelet time–frequency decomposition to explore the theta-band content of the cross-correlation function for target and distractor speech streams. To assess differences between the theta-band response to the target and distractor speech, a Monte Carlo permutation test was performed on the time–frequency data, averaged

across 4–8 Hz, with a correction to preserve false-discovery rate (FDR) (Benjamini & Hochberg, 1995). There was significantly more theta-band power phase-locked to the target speech than the distractor speech at lags 140–240 ms ($p < 0.0011$, FDR corrected $\alpha = 0.05$) (Fig. 3A). To assess the behavioral importance of this difference in phase-locked theta power, we sorted probe-present trials according to performance: Correct responses were those trials on which participants correctly detected a probe in the target stream and errors were those trials on which the participant missed a probe in the target stream. Phase-locked theta power was averaged across latencies from 140 to 240 ms (Fig. 3B). Differences were assessed using Wilcoxon signed rank tests. Target-locked theta power was significantly greater than distractor-locked theta power on correct trials ($Z = -2.10$, $p = 0.0353$), but not on error trials ($Z = -1.48$, $p = 0.153$). Importantly, there was significantly more power phase-locked to the target stream on correct trials than on error trials ($Z = -1.98$, $p = 0.0494$), but no difference between distractor-locked theta power on correct and error trials ($Z = -1.16$, $p = 0.2676$).

4. Discussion

Our behavioral results show a clear effect of attention on participants' ability to correctly recall words from recently heard speech streams. Participants were significantly more likely to correctly recall words than to miss words when they were presented in the target speech stream. By contrast, participants were not more likely to correctly recall words from the distractor stream (in fact they were slightly more likely to miss words in the distractor stream). However, it is important to note that they were more likely to recall the probe word when it was present in either stream relative to when it was absent. This suggests that some aspects of the distractor speech were at least occasionally encoded.

Our electrophysiological results showed an enhancement of the theta band EEG response that was phase locked to the acoustic dynamics of target compared to distractor speech. This was most evident at latencies between 140 and 240 ms. This result converges with other studies that used different analysis techniques to identify the effect of attention on low frequency, speech-locked activity (Ding & Simon, 2012; Kerlin et al., 2010; Power, Foxe, Forde, Reilly, & Lalor, 2012; Zion Golumbic, Ding, et al., 2013). This result provides confirmation that correlating the acoustic dynamics of competing speech streams with the electrophysiological response to the sound mixture is an effective way to separate the neural response to the individual streams. This is a way to "unmix" the concurrent dynamics of competing speech representations.

Theta-band phase tracking of speech is a neuroelectric correlate of selecting a single speech stream. However, it is likely that other well-known cues such as fundamental frequency and spatial location contribute substantially to successful pre-attentive stream segregation. Our study was not designed to compare the relative contribution of each of these cues to stream segregation.

The cross-correlation method employed in the present study acts as a low-pass filter with a kernel unique to each speech envelope. It therefore retains only the low-frequency activity that is phase-locked to the acoustic dynamics of individual speech segments. It is therefore not surprising that we observed no effects at higher frequencies. Our results suggest that low-frequency amplitude modulations in the acoustic envelope of speech are cues that allow for the entrainment of neuroelectric activity to speech dynamics. However, we do not rule out a role for higher-frequency signals in speech selection and scene analysis.

Phase tracking of speech dynamics is a relatively unexplored phenomenon and, while studies have been conducted to understand the mechanism (Ding & Simon, 2012; Zion Golumbic, Ding,
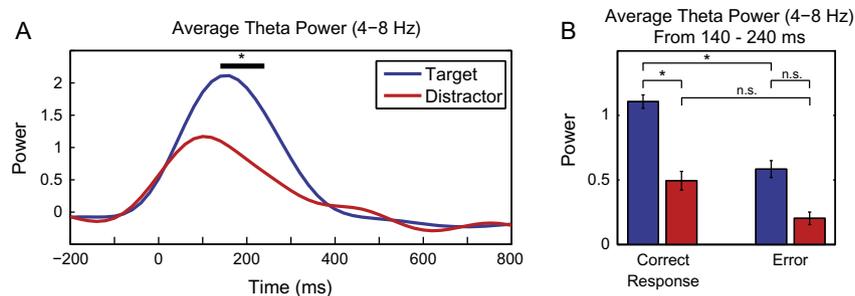
Fig. 3. (A) Grand averaged evoked power in the cross-correlation function for the theta band (4–8 Hz). Black bar indicates time bins in which there was a significant difference between the response to the target and distractor streams ($p < 0.0011$, FDR corrected $\alpha = 0.05$). (B) Grand averaged evoked power for the theta band (4–8 Hz) averaged across latencies from 140 to 240 ms. Trials were split on the basis of the participant's successful identification of the subsequent probe word ($^*p < 0.05$).

et al., 2013) there has been less effort applied to understanding its behavioral correlates. The combination of our behavioral and electrophysiological data illustrates the perceptual importance of phase-tracking. We showed that enhanced phase tracking of the dynamics of speech preceded correct recall of probe words in the selected speech stream. Phase-tracking of the target stream was not significantly enhanced preceding failures of recall. This result aligns with the results of Mesgarani and Chang (2012). That study reconstructed the spectrogram of speech based on local field potentials. They found that on successful trials the target speech stream was preferentially represented whereas on error trials competing streams were equally represented. These results suggest that an increase in phase tracking of target speech is associated with solving the cocktail party problem.

4.1. Selective entrainment

The *selective entrainment* hypothesis, proposed by Schroeder and Lakatos (2009b), Zion Golumbic, Ding et al., (2013) suggests that entrainment of neural oscillations to the temporal dynamics of a single behaviorally relevant stream is a mechanism for attentional selection. What perceptual benefit might phase tracking provide to the listener? The theory is well aligned with the notion that communication between two or more neural ensembles is facilitated when their oscillatory behavior is synchronized – a model known as *communication through coherence* (Fries, 2005). In this view, phase coherence enables a transmitting neuron (or group of neurons) to optimally drive a receiving neuron by aligning their pre-synaptic activity with temporal windows of maximal sensitivity to post-synaptic depolarizations. By modulating phase, such a mechanism might provide a means for some neural assemblies to "ignore" inputs from non-selected cells (Fries, Reynolds, Rorie, & Desimone, 2001).

Taken together, *selective entrainment* and *communication by coherence* provide a possible framework in which to explore perception and distraction in complex scenes. Selective attention has been conceptualized for over a century (James, 1890) as the preferential representation of a single source of information for enhanced perception, at the exclusion of other sources. This implies that attended acoustic signals selectively contribute information to, for example, working-memory, reward-processing, and response-planning mechanisms. At the neuronal level, this implies that networks of cells representing the features of attended sensory input should not only be bound together within sensory cortex (Malsburg & Schneider, 1986; Singer & Gray, 1995), but should also have preferential access to brain-wide non-sensory areas. To be selectively attended, an auditory signal should have exclusive access to brain-wide networks, while representations of distracting signals should be unable to communicate outside of sensory cortex. Selective entrainment might create a computational bias in

the cortex by enabling a phase-predictive process: by entraining brain-wide neural oscillations to the modulation frequency of a single speech stream, neural ensembles across multiple cortical systems can become biased to events in that stream.

This view of the selective entrainment predicts that the EEG signal should exhibit little or no phase tracking of unattended speech. In this way our data were not entirely consistent with the selective entrainment hypothesis. We did find phase-locked activity associated with both streams, with phase-locked theta power being enhanced for the target relative to distractor stream. An important consideration here is that the phase relationship between target and distractor envelopes might be critically important. Purely by chance, the competing speech streams should exhibit periods of transient coherence. Such coherence between target and distractor might make it particularly difficult to maintain selection and would appear as transient phase-locking between the EEG and the distractor stream. A kind of active distraction might result, in which events in the distractor stream could intrude into representations of the target stream. Our study was not designed to differentiate errors of intrusion from simple failures of recall.

The theta-band signal in the EEG that is phase-locked with the target envelope might exhibit a change in power for various reasons: This increase may be due to increased gain in a fixed-latency evoked response triggered by increments in sound energy within the acoustic signal. Alternatively, ongoing theta-band oscillations might be better phase entrained to the attended envelope, without any modulation in the amplitude of those oscillations. Either of these situations, or a combination of the two, would appear as a modulation of phase-locked power. Distinguishing between phase entrainment and additive phase-locked activity as mechanisms in the generation of phase-locked signals in scalp-recorded EEG is highly problematic (Telenczuk, Nikulin, & Curio, 2010). Therefore we can conclude only that the neuroelectric dynamics of the brain exhibit components that are phase-aligned with the envelopes of speech in the auditory scene, that these signals are modulated by attentional selection, and that this modulation is reflective of variation in the subsequent recall of the attended stream.

The precise mechanism by which enhanced phase-locking to a stream enhances recall is uncertain; however, our results suggest it may be due to enhanced perceptual processing of speech. Successful recall of previously heard speech requires that the to-be-remembered speech is accurately perceived, successfully encoded into short-term or long-term storage, and accurately retrieved at the time of recall. An enhancement of any of these processes may result in an improved ability to recall the original speech. Our results show enhanced neuroelectric activity phase-locked to speech that lags that speech by around 200 ms – an enhancement that is associated with improved subsequent memory of that speech. The fact that it is time-locked to the to-be-remembered speech suggests that phase-locking enhances either perceptual

processes, encoding processes, or both; but not retrieval processes which must occur following the probe word. Event related potential (ERP) studies of verbal encoding have identified electrophysiological effects associated with successful encoding that occur 300–1000 ms after presentation of subsequently recalled words (Friedman, 1990; Guo, Voss, & Paller, 2005; Otten, Quayle, Akram, Ditewig, & Rugg, 2006). Studies employing time–frequency analysis of EEG and MEG have identified an increase in theta band synchronization following the presentation of to-be-remembered stimuli, also at latencies of 300–1000 ms post-presentation (Klimesch, Doppelmayr, Schimke, & Ripper, 1997; Osipova et al., 2006). Because our significant findings occur at latencies prior to 300 ms – the earliest latency at which encoding-related effects have been identified in EEG – we believe that increased phase-locking to speech facilitates recall by enhancing the perception of the phase-locked speech prior to encoding to memory. However, we emphasize here that the concept of latency in the cross-correlation function represents a lag between the acoustic signal and the EEG and its connection to latencies in ERP peaks may be complex.

5. Conclusions

In the present study we confirmed two predictions of the selective entrainment hypothesis: that neuroelectric oscillations in the theta-band of the EEG are more phase-locked to attended relative to unattended speech; and that perceptual performance was modulated by the degree of phase-tracking.

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proceedings of the National Academy of Sciences of the United States of America, 98(23), 13367–13372. http://dx.doi.org/10.1073/pnas.201400998.

Arnott, S. R., Grady, C. L., Hevenor, S. J., Graham, S., & Alain, C. (2005). The functional organization of auditory working memory as revealed by fMRI. Journal of Cognitive Neuroscience, 17(5), 819–831. http://dx.doi.org/10.1162/0898929053747612.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate – A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society Series B – Methodological, 57(1), 289–300.

Broadbent, D. E. (1952). Listening to one of two synchronous messages. Journal of Experimental Psychology, 44, 51–55.

Cherry, E. C. (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. http://dx.doi.org/10.1121/1.1907229.

Crottaz-Herbette, S., Anagnoson, R., & Menon, V. (2004). Modality effects in verbal working memory: Differential prefrontal and parietal responses to auditory and visual stimuli. NeuroImage, 21(1), 340–351. http://dx.doi.org/10.1016/j.neuroimage.2003.09.019.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. Journal of Neuroscience Methods, 134(1), 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Ding, N., & Simon, J. Z. (2012). Emergence of neural encoding of auditory objects while listening to competing speakers. Proceedings of the National Academy of Sciences, 109(29), 11854–11859. http://dx.doi.org/10.1073/pnas.1205381109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1205381109.

Fishbach, A., Nelken, I., & Yeshurun, Y. (2001). Auditory edge detection: A neural model for physiological and psychoacoustical responses to amplitude transients. Journal of Neurophysiology, 85(6), 2303–2323. <http://www.ncbi.nlm.nih.gov/pubmed/11387378>.

Friedman, D. (1990). ERP's during continuous recognition memory for words. Biological Psychology, 30, 61–87.

Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. Trends in Cognitive Sciences, 9(10), 474–480. doi:http://dx.doi.org/S1364-6613(05)00242-1, pii:10.1016/j.tics.2005.08.011.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. Science, 291(5508), 1560–1563. http://dx.doi.org/10.1126/science.1055465.

Guo, C., Voss, J. L., & Paller, K. A. (2005). Electrophysiological correlates of forming memories for faces, names, and face-name associations. Brain Research.

Cognitive Brain Research, 22(2), 153–164. http://dx.doi.org/10.1016/j.cogbrainres.2004.08.009.

Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., & Ackermann, H. (2012). Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology, 49(3), 322–334. http://dx.doi.org/10.1111/j.1469-8986.2011.01314.x.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. Journal of Neurophysiology, 104(5), 2500–2511. http://dx.doi.org/10.1152/jn.00251.2010.

Ille, N., Berg, P., & Scherg, M. (2002). Artifact correction of the ongoing EEG using spatial filters based on artifact and brain signal topographies. Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society, 19(2), 113–124. <http://www.ncbi.nlm.nih.gov/pubmed/11997722>.

James, W. (1890). The principles of psychology. New York, NY: Henry Holt and Co.

Kerlin, J. R., Shahin, A. J., & Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 30(2), 620–628. http://dx.doi.org/10.1523/JNEUROSCI.3631-09.2010.

Kidd, G., Mason, C. R., Richards, V. M., Gallun, F. J., & Durlach, N. I. (2007). Informational masking. In W. A. Yost, A. N. Popper, & R. R. Fay (Eds.), Auditory perception of sound sources (pp. 143–189). New York.

Klimesch, W., Doppelmayr, M., Schimke, H., & Ripper, B. (1997). Theta synchronization and alpha desynchronization in a memory task. Psychophysiology, 34(2), 169–176. <http://www.ncbi.nlm.nih.gov/pubmed/9090266>.

Lakatos, P., Musacchia, G., O'Connel, M. N., Falchier, A. Y., Javitt, D. C., & Schroeder, C. E. (2013). The spectrotemporal filter mechanism of auditory selective attention. Neuron, 77(4), 750–761. http://dx.doi.org/10.1016/j.neuron.2012.11.034.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron, 54(6), 1001–1010. http://dx.doi.org/10.1016/j.neuron.2007.06.004.Phase.

Malsburg, C. V., & Schneider, W. (1986). A neural cocktail-party processor. Biological Cybernetics, 54, 29–40. <http://link.springer.com/article/10.1007/BF00337113>.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature, 485(7397), 233–236. http://dx.doi.org/10.1038/nature11020.

Osipova, D., Takashima, A., Oostenveld, R., Fernández, G., Maris, E., & Jensen, O. (2006). Theta and gamma oscillations predict encoding and retrieval of declarative memory. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 26(28), 7523–7531. http://dx.doi.org/10.1523/JNEUROSCI.1948-06.2006.

Otten, L. J., Quayle, A. H., Akram, S., Ditewig, T. A., & Rugg, M. D. (2006). Brain activity before an event predicts later recollection. Nature Neuroscience, 9(4), 489–491. http://dx.doi.org/10.1038/nn1663.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebral Cortex, 23, 1378–1387. http://dx.doi.org/10.1093/cercor/bhs118.

Ponjavic-Conte, K. D., Hambrook, D. A., Pavlovic, S., & Tata, M. S. (2013). Dynamics of distraction: Competition among auditory streams modulates gain and disrupts inter-trial phase coherence in the human electroencephalogram. PLoS ONE, 8(1), e53953. http://dx.doi.org/10.1371/journal.pone.0053953. M.J. Chacron (Ed.).

Power, A. J., Foxe, J. J., Forde, E.-J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. The European Journal of Neuroscience, 35(9), 1497–1503. http://dx.doi.org/10.1111/j.1460-9568.2012.08060.x.

Schroeder, C. E., & Lakatos, P. (2009a). Low-frequency neuronal oscillations as instruments of sensory selection. Trends in Neurosciences, 32(1), 9–18. http://dx.doi.org/10.1016/j.tins.2008.09.012.

Schroeder, C. E., & Lakatos, P. (2009b). Low-frequency neuronal oscillations as instruments of sensory selection. Trends in Neurosciences, 32(1), 9–18. http://dx.doi.org/S0166-2236(08)00250-6, pii:10.1016/j.tins.2008.09.012.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. Annual Review of Neuroscience, 18, 555–586. http://dx.doi.org/10.1146/annurev.ne.18.030195.003011.

Telenczuk, B., Nikulin, V., & Curio, G. (2010). Role of neuronal synchrony in the generation of evoked EEG/MEG responses. Journal of Neurophysiology, 104(6), 3557–3567. http://dx.doi.org/10.1152/jn.00138.2010.

Treisman, A. M. (1964). The effect of irrelevant material on the efficiency of selective listening. American Journal of Psychology, 77, 533–546. <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=14251963>.

Zion Golumbic, E. M., Cogan, G. B., Schroeder, C. E., & Poeppel, D. (2013). Visual input enhances selective speech envelope tracking in auditory cortex at a "cocktail party". The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 33(4), 1417–1426. http://dx.doi.org/10.1523/JNEUROSCI.3675-12.2013.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., Goodman, R. R., Emerson, R., Mehta, A. D., Simon, J. Z., Poeppel, D., & Schroeder, C. E. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "Cocktail Party". Neuron, 77(5), 980–991. http://dx.doi.org/10.1016/j.neuron.2012.12.037.