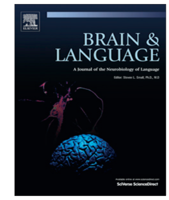
Brain & Language 135 (2014) 52–56

Contents lists available at ScienceDirect

Brain & Language

journal homepage: www.elsevier.com/locate/b&l



Short Communication

Theta-band phase tracking in the two-talker problem

Dillon A. Hambrook *, Matthew S. Tata

The University of Lethbridge, Lethbridge, AB, Canada

ARTICLE INFO

Article history:
Accepted 12 May 2014
Available online 7 June 2014

Keywords:
Electroencephalogram
Phase tracking
Speech
Attention
Cocktail party
Selective entrainment

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.

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).

http://dx.doi.org/10.1016/j.bandl.2014.05.003
0093-934X/

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 ($t = 4.89$, $p = .0003$).