Brain & Language 140 (2015) 49–54

Contents lists available at ScienceDirect

Brain & Language

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

Short Communication

Cortical characterization of the perception of intelligible and unintelligible speech measured via high-density electroencephalography

Rene L. Utianski [a,b,*], John N. Caviness [b], Julie M. Liss [a]

[a] Department of Speech and Hearing Science, Arizona State University, United States
[b] Department of Neurology, Mayo Clinic, AZ, United States

ARTICLE INFO

Article history:
Accepted 22 October 2014
Available online 13 December 2014

Keywords:
EEG
Speech
Perception
Disordered
Intelligibility
Listening strategy

ABSTRACT

High-density electroencephalography was used to evaluate cortical activity during speech comprehension via a sentence verification task. Twenty-four participants assigned true or false to sentences produced with 3 noise-vocoded channel levels (1-unintelligible, 6-decipherable, 16-intelligible), during simultaneous EEG recording. Participant data were sorted into higher- (HP) and lower-performing (LP) groups. The identification of a late-event related potential for LP listeners in the intelligible condition and in all listeners when challenged with a 6-Ch signal supports the notion that this induced potential may be related to either processing degraded speech, or degraded processing of intelligible speech. Different cortical locations are identified as neural generators responsible for this activity; HP listeners are engaging motor aspects of their language system, utilizing an acoustic–phonetic based strategy to help resolve the sentence, while LP listeners do not. This study presents evidence for neurophysiological indices associated with more or less successful speech comprehension performance across listening conditions.

1. Introduction

Understanding speech in typical communication is generally effortless, but listening in acoustically adverse situations exposes a wide range of performance variability among healthy listeners (Borrie, McAuliffe, Liss, O'Beirne, & Anderson, 2012; Choe, Liss, Azuma, & Mathy, 2012). It may be hypothesized that this variability is a reflection of the ways in which listeners recover words (or larger units of speech) from the impoverished acoustic signal to extract meaning. A useful model for investigating this variability is the dual stream model of speech perception (Hickok & Poeppel, 2000; Hickok & Poeppel, 2004; Hickok & Poeppel, 2007). Briefly, the dorsal stream contains the articulatory/motor networks of the frontal lobe and the ventral stream contains the conceptual/semantic networks of the temporal lobe. Anatomically, the dorsal stream spans an area of the posterior Sylvian fissure, projecting toward the frontal regions; the ventral stream projects toward the posterior medial temporal gyrus. It is proposed that these two active, bilateral pathways converge for understanding speech. Previous fMRI work has shown dorsal stream (frontal) activation for processing pseudo words and ventral stream (temporal) activation for complex sentences (Saur et al., 2008). While such research demonstrates differential activation for processing isolated units of speech, it is unknown whether or how this extends to processing degraded speech for meaning. If the behavioral performance data map meaningfully to cortical activation patterns within the dual streams, this may lead to an explanation of performance variability. For example, disproportionate activation of the dorsal stream may be indicative of attending to fine structure of the degraded signal, which may have benefit or cost for recovering the spoken message. High-density electroencephalography (EEG) provides both temporal and spatial resolution to identify cortical activation patterns that may distinguish more and less successful listeners.

We used a sentence verification task (true/false) for noise vocoded speech at three levels of intelligibility. The degradation was intended to induce performance variability among healthy listeners to reveal poorer and better groups. The sentence verification task required both deciphering the speech and rendering a true/false decision, thereby tapping potential differences in recruitment of the dual streams. Simultaneous EEG acquisition allowed for examination of the associated cortical activation patterns, where it was expected that listeners who take advantage of both acoustic–phonetic and semantic information would show activation in both the dorsal and ventral streams, with better performance in the sentence verification task.

* Corresponding author at: Department of Neurology, Mayo Clinic, AZ, United States.
E-mail address: utianski.rene@mayo.edu (R.L. Utianski).

http://dx.doi.org/10.1016/j.bandl.2014.10.008
0093-934X/

2. Methods

2.1. Participants

Participants were 24 undergraduate and graduate students (21 female) from Arizona State University. Ages ranged from 20–48 (mean = 25). Listeners self-reported English as native language; right-handedness; and a negative history for speech, language or hearing disorders. Hearing thresholds were within normal limits (detection of 125–4000 Hz < 15 dB in each ear), per pure tone screening conducted before the experiment. Listeners received $20 for participation.

2.2. Speech stimuli

A healthy 46 year-old female recorded stimuli in a sound-attenuating booth via head mount microphone (Plantronics DSP-100). Her acoustic speech characteristics were typical of age and sex peers. She read from a computer screen 240 sentences, half of which were "true" (e.g. zebras have stripes) and half "false" (e.g. donkeys have wings). Speech was acquired via TF32 (Milenkovic, 2004; 16-bit, 44 kHz) and saved for editing (using Audacity: Free Audio Editor and Recorder). All sentences were three words, with the last word determining the veracity of the sentence. Average sentence duration was 1959 ms (SD = 213 ms). Sound files were RMS normalized prior to vocoding. For the experiment, a third of the sentences were vocoded using a 1-channel (unintelligible), 6-channel (moderately intelligible), or 16-channel (intelligible) noise vocoder (PRAAT Boersma & Weenink, 2008). Thus, there were 40 sentences at each combination of intelligibility level and veracity. Noise vocoding was chosen to degrade the speech signal because it minimizes the availability of phonetic information, while preserving other properties of the speech signal (e.g. rhythmic structure); but without adding to the speech signal (e.g. speech in noise).

2.3. Task

Data were collected, analyzed, and interpreted at ASU, in accordance with approved IRB protocols. Participants sat in a hard-backed chair at a comfortable distance from the monitor. STIM2 (STIM2) delivered sentences through inter-aural headphones (90 dB) and the STIM audio box system ensured proper synchronization of audio delivery to the participant and EEG recording via scan (Scan v. 4.5. Compumedics Neuroscan).

On-screen visual prompts guided the participants through the experiment. Following each sentence presentation, participants pressed true or false keyboard buttons. Pre-training ensured task understanding and sufficient muscle relaxation. Each participant then responded true/false for six blocks of 40 sentences (240), whose order of presentation was randomized within blocks, and order of blocks (1–6) was partially counterbalanced amongst participants. Breaks were permitted.

2.4. Electroencephalographic recording

Neuroscan Acquire (v4.5, Compumedics Neuroscan, Charlotte, USA) (Scan v. 4.5. Compumedics Neuroscan), with 128-channel QuickCap, was used for EEG recordings. Accurate electrode positioning was assured through measurement and positioning of Cz, Fz, and Pz, according to 10–20 system. Recordings were acquired with a 1000 Hz-sampling rate and low-pass filtered below 200 Hz. After recording, a 60 Hz notch filter minimized effects of electrical artifact. Impedance of all electrodes was well below 5 k Ω. Continuous recordings were examined for physiologic and non-physiologic artifact. Artifact reduction through linear derivation minimized the presence of blinks. All recordings were of high quality and used in subsequent epoching.

The continuous file for each individual was epoched for each condition – the 40 sentences of each 1-, 6-, and 16-channel intelligibility levels. Epochs were created for 300 ms prior to the onset of the sentence to 1500 ms following onset. Each individual epoch was examined for artifact and those with high levels were removed. On average, 80% of recorded epochs were utilized to create an average file for each condition for each participant. The selected files were concatenated to create a grand average for all listeners, for each condition. Ten electrodes were omitted from the average due to high artifact. Only averaged files were subsequently used to facilitate strong signal-to-noise ratios.

2.5. Data analysis

2.5.1. Behavioral accuracy

STIM2 (Compumedics Neuroscan, Charlotte, USA) presented sentences and recorded participant response. Microsoft Excel was used for subsequent analyses. All data were reviewed and incorrect and no response items were considered together. Therefore, only correct responses, provided within the appropriate time interval were considered as correct. As our previous work has demonstrated a wide range of listener performance for 6-channel condition, we anticipated and found this in the current behavioral data. We selected an 80% accuracy level as the threshold for successful performance based on pilot data to create groups of high performing (HP; $n = 9$) and low performing (LP; $n = 15$). Pilot data explored accuracy of responses to a systematic range of noise-vocoded speech, and provided an expectation for average performance on each level of noise-vocoded speech. These data indicate exceptionally accurate listeners with highly intelligibly noise-vocoded speech (e.g. 16-channel), performed at or above 80% accuracy on the 6-channel speech task. Grouping accomplished sufficient signal–noise ratios for interpretation. The HP and LP group designations were used in subsequent EEG analyses, where analyses described below were completed for all listeners and repeated for both groups. It is of note that participants in each group were of equivalent age (see Table 1), with HP mean of 27 years-old and LP 25 years-old. Thus differences between the groups are unlikely due to age effects but rather listening differences.

2.5.2. Event-related potentials

Utilizing CURRY7 Multi-Modality Imaging Suite (Compumedics Neuroscan, Charlotte, USA) (CURRY 7: Multimodal Imaging Suite), average files were examined for transient activity, indicated by peaks in mean global field power (MGFP), representative of the power of activity, across all electrodes (Murray, Brunet, & Michel, 2008). Independent components analysis (ICA) assessed the transient in the time interval of interest (Lee, Girolami, & Sejnowski, 1999). Current density reconstruction (CDR), via sLORETA, identified component of interest (Pascual-Marqui, 2002). This localizes the source of the component, or the underlying neural generator responsible for the transient activity. sLORETA computes minimum norm least squares (MNLS) current density amplitudes (dipole moments) and divides them by their error bars (and squares the result), taking into account the amplitude of activity; therefore, the F-values provided by sLORETA can be interpreted as magnitudes of activity. CURRY assigns a Brodmann's area to the solution to the MNLS problem, which was subjectively validated by the authors.

2.5.3. Frequency analysis

Frequency spectra for average files were examined via SCAN (v4.5, Compumedics Neuroscan, Charlotte, USA). The time-domain