Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Research report

The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene

Johanna M. Rimmele [a,b,*], Elana Zion Golumbic [c,d], Erich Schröger [e] and David Poeppel [a,f]

[a] Department of Psychology and Center for Neural Science, New York University, New York, NY, USA
[b] Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[c] Gonda Center for Brain Research, Bar Ilan University, Israel
[d] Department of Psychiatry, Columbia University, New York, NY, USA
[e] Institute of Psychology, University of Leipzig, Leipzig, Germany
[f] Max-Planck Institute for Empirical Aesthetics, Frankfurt, Germany

ARTICLE INFO

Article history:
Received 7 August 2014
Reviewed 10 September 2014
Revised 14 October 2014
Accepted 10 December 2014
Published online 7 January 2015

Keywords:
Auditory cortex
Fine structure
Linguistic
Predictions
Speech envelope

ABSTRACT

Attending to one speaker in multi-speaker situations is challenging. One neural mechanism proposed to underlie the ability to attend to a particular speaker is phase-locking of low-frequency activity in auditory cortex to speech's temporal envelope ("speech-tracking"), which is more precise for attended speech. However, it is not known what brings about this attentional effect, and specifically if it reflects enhanced processing of the fine structure of attended speech. To investigate this question we compared attentional effects on speech-tracking of natural versus vocoded speech which preserves the temporal envelope but removes the fine structure of speech. Pairs of natural and vocoded speech stimuli were presented concurrently and participants attended to one stimulus and performed a detection task while ignoring the other stimulus. We recorded magnetoencephalography (MEG) and compared attentional effects on the speech-tracking response in auditory cortex. Speech-tracking of natural, but not vocoded, speech was enhanced by attention, whereas neural tracking of ignored speech was similar for natural and vocoded speech. These findings suggest that the more precise speech-tracking of attended natural speech is related to processing its fine structure, possibly reflecting the application of higher-order linguistic processes. In contrast, when speech is unattended its fine structure is not processed to the same degree and thus elicits less precise speech-tracking more similar to vocoded speech.

Abbreviation: ECoG, electrocorticography; EEG, electroencephalography; ERF, event-related field; FAR, false alarm rate; fMRI, functional magnetic resonance imaging; HR, hit rate; ICA, independent component analysis; ITPC, intertrial phase coherence; ITPowC, intertrial power coherence; MEG, magnetoencephalography; RMS, root mean square.

* *Corresponding author*. Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Martinistraße 52, D – 20246 Hamburg, Germany.
E-mail addresses: j.rimmele@uke.de (J.M. Rimmele), elana.zion-golumbic@biu.ac.il (E. Zion Golumbic), schroger@rz.uni-leipzig.de (E. Schröger), david.poeppel@nyu.edu (D. Poeppel).
http://dx.doi.org/10.1016/j.cortex.2014.12.014

1. Introduction

Listening to speech in multi-speaker situations is challenging, particularly for populations such as older adults or hearing impaired listeners (for review, see Rimmele, Sussman, & Poeppel, in press). These situations require the segregation of speech streams originating from different speakers and the selection of one of these streams for further processing. The neural mechanisms through which attentional selection is achieved and that facilitate the processing of attended speech over competing stimuli are not fully understood. One major question concerns the degree to which the ability to establish a robust representation of speech in auditory cortex (e.g., as required to attend to a particular speaker) is driven by the acoustic properties of the stimulus including both the speech envelope and its fine structure.

Mechanistically speaking, phase-locking of low-frequency neural activity in auditory cortex ("speech-tracking response") has been proposed to indicate a robust "*object-level*" representation of speech (Luo & Poeppel, 2007). The speech-tracking response has been related to both the temporal envelope of the stimulus, which carries information regarding fluctuations in stimulus energy over time, as well as to its fine structure which contains the more detailed spectro-temporal information of speech (Ding & Simon, 2014). Crucially, speech-tracking has been shown to be more robust for attended-compared to unattended speech presented simultaneously (Ding & Simon, 2012a, 2012b; Horton, D'Zmura, & Srinivasan, 2013; Horton, Srinivasan, & D'Zmura, 2014; Kerlin, Shahin, & Miller, 2010; Zion Golumbic, Ding et al., 2013), suggesting that this mechanism is influenced by attentional selection. The goal of the current study was to clarify the role of two levels of speech acoustics – the temporal envelope and fine structure – in speech-tracking by investigating how they interact with attention in multi-speaker listening situations. Specifically we asked, whether speech's fine structure is utilized by selective attention to enhance speech-tracking of natural speech.

1.1. *The speech-tracking response*

Phase-locking of neural activity in auditory cortex to the temporal envelope of speech is observed primarily in the theta frequency range (3–7 Hz), corresponding to the syllabic time scale in speech. It is well established that the low-frequency fluctuations in the speech envelope, which carry temporal information about syllable onsets/offsets as well as prosodic cues, are crucial for speech intelligibility (Doelling, Arnal, Ghitza, & Poeppel, 2014; Ghitza, Giraud, & Poeppel, 2013; Giraud & Poeppel, 2012; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Zion Golumbic, Poeppel, & Schroeder, 2012). Ghitza and Greenberg (2009) showed that the intelligibility of time compressed speech (with a very low intelligibility <50% words correct), increased dramatically when a theta-range "syllabic rate" was artificially induced by adding periods of silence. In light of these results, it has been proposed that the theta-band 'speech-tracking response' in auditory cortex plays a role in segmenting the speech stream into smaller linguistically-meaningful units (Giraud et al., 2007; Giraud & Poeppel, 2012; Luo & Poeppel, 2007; Zion Golumbic et al., 2012).

Significant phase-locking to the sound envelope is also observed for unintelligible, time-inverted, or noise-vocoded speech as well as non-speech sounds (Ding, Chatterjee, & Simon, 2014; Howard & Poeppel, 2012; Hämäläinen, Rupp, Soltész, Szücs, & Goswami, 2012; Lalor, Power, Reilly, & Foxe, 2009; Millman, Prendergast, Hymers, & Green, 2013; Peelle, Gross, & Davis, 2013; Steinschneider, Nourski, & Fishman, 2013; Wang, Zhu, & Bastiaansen, 2012). Nonetheless, speech-tracking is more robust for natural compared to noise-vocoded speech, in which the fine structure information is removed but the low-frequency temporal fluctuations contained in the speech envelope are preserved (Ding et al., 2014; Howard & Poeppel, 2010; Luo & Poeppel, 2007; Peelle et al., 2013; Wild, Davis, & Johnsrude, 2012). This observation has been interpreted by some as reflecting the application of higher order linguistic processing to natural compared to vocoded speech, since vocoded speech is less intelligible than natural speech (Peelle, Johnsrude, & Davis, 2010; Wild, Davis et al., 2012). Others suggest that these effects may be due to the difference in acoustical information in natural compared to vocoded speech, and that the increased speech-tracking response for natural speech reflects its richer acoustic features (Ding et al., 2014; reviewed in Ding & Simon, 2014). It is difficult to distinguish between these alternatives, as differences in speech acoustics and speech intelligibility are inherently confounded when directly comparing natural versus vocoded speech. Nonetheless, these two interpretations make different predictions regarding the consistency of this phenomenon. Under the acoustic-processing perspective, the advantage for speech-tracking of natural over vocoded speech should remain robust under different cognitive manipulations, since it is primarily due to the acoustic structure of the stimuli. In contrast, under the linguistic-processing perspective, the differences in speech-tracking of natural and vocoded speech may be affected by task demands and the degree of linguistic processing applied. Following this rationale, in the current study we tested how the speech-tracking response of natural and vocoded speech was affected by selective attention as a means for investigating the interaction between the acoustic richness of a stimulus and top-down processing demands. As reviewed below, higher order top-down processes, such as linguistic processing or selective attention, influence processing in auditory cortex therefore reconciling their interaction with bottom-up acoustic processing is critical for understanding the neural architecture and hierarchy underlying speech processing.

1.2. *Effects of linguistic processing on sensory responses*

There is much evidence that sensory processing of speech in auditory cortex can be modulated by higher order processing, such as syntactic or semantic analysis (Kalikow, Stevens, & Elliott, 1977; Miller & Isard, 1963; Peelle et al., 2013; Peelle, 2013), speaker familiarity (Johnsrude et al., 2013) or linguistic expectations set up by visual cues (Jacoby, Allan, Collins, & Larwill, 1988; Sohoglu, Peelle, Carlyon, & Davis, 2012; Zekveld, Kramer, Kessens, Vlaming, & Houtgast, 2008; for review: Peelle et al., 2010). Sohoglu and colleagues (using EEG and MEG) showed that a visual cue, which provides prior knowledge of the speech content, increases the perceived speech clarity in a similar manner as altering the physical parameters of the