NeuroImage xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Previous exposure to intact speech increases intelligibility of its digitally degraded counterpart as a function of stimulus complexity

Maria Hakonen [a,b,*], Patrick J.C. May [c], Jussi Alho [a], Paavo Alku [d], Emma Jokinen [d], Iiro P. Jääskeläinen [a], Hannu Tiitinen [a,b]

[a] Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering (NBE), School of Science, Aalto University, PO Box 12200, FI-00076 AALTO, Finland
[b] BioMag Laboratory, PO Box 340, FI-00029 HUS, Helsinki University Central Hospital, Finland
[c] Special Laboratory Non-Invasive Brain Imaging, Leibniz Institute for Neurobiology, Brenneckestraße 6, D-39118 Magdeburg, Germany
[d] Department of Signal Processing and Acoustics, School of Electrical Engineering, Aalto University, PO Box 13000, FI-00076 AALTO, Finland

A R T I C L E I N F O

Article history:
Received 27 June 2015
Accepted 10 October 2015
Available online xxxx

Keywords:
Speech
Comprehension
Intelligibility
Acoustic distortion
Magnetoencephalography
Auditory neuroscience

A B S T R A C T

Recent studies have shown that acoustically distorted sentences can be perceived as either unintelligible or intelligible depending on whether one has previously been exposed to the undistorted, intelligible versions of the sentences. This allows studying processes specifically related to speech intelligibility since any change between the responses to the distorted stimuli before and after the presentation of their undistorted counterparts cannot be attributed to acoustic variability but, rather, to the successful mapping of sensory information onto memory representations. To estimate how the complexity of the message is reflected in speech comprehension, we applied this rapid change in perception to behavioral and magnetoencephalography (MEG) experiments using vowels, words and sentences. In the experiments, stimuli were initially presented to the subject in a distorted form, after which undistorted versions of the stimuli were presented. Finally, the original distorted stimuli were presented once more. The resulting increase in intelligibility observed for the second presentation of the distorted stimuli depended on the complexity of the stimulus: vowels remained unintelligible (behaviorally measured intelligibility 27%) whereas the intelligibility of the words increased from 19% to 45% and that of the sentences from 31% to 65%. This increase in the intelligibility of the degraded stimuli was reflected as an enhancement of activity in the auditory cortex and surrounding areas at early latencies of 130–160 ms. In the same regions, increasing stimulus complexity attenuated mean currents at latencies of 130–160 ms whereas at latencies of 200–270 ms the mean currents increased. These modulations in cortical activity may reflect feedback from top-down mechanisms enhancing the extraction of information from speech. The behavioral results suggest that memory-driven expectancies can have a significant effect on speech comprehension, especially in acoustically adverse conditions where the bottom-up information is decreased.

Introduction

Despite increasing efforts in the study of the neural basis of speech comprehension, the processes related to speech intelligibility, which is reflected as correctly identified speech content and arises out of the successful matching of bottom-up acoustic information to top-down memory representations, have remained largely unknown. One reason for this is that studies on speech intelligibility have typically either manipulated the acoustic structure of the speech signal or masked the speech stimulus using varying levels and types of noise. However, both the processing of acoustic features of the stimulus and cognitive operations related to the recognition of the content of speech sounds are reflected in brain responses, and it is therefore difficult to distinguish their overlapping contributions from one another.

Only a limited number of studies have examined the brain mechanisms related to speech comprehension by manipulating stimulus intelligibility without changing the acoustic structure of the stimulus. Our recent magnetoencephalography (MEG) study (Tiitinen et al., 2012) introduced an experimental paradigm where the same set of speech stimuli was presented to the subject in a distorted, undistorted, and again in a distorted form. The intervening exposure to the undistorted versions of the sentences increased the intelligibility of the distorted sentences considerably (i.e. the recognition rate increased from 30% to 80%), and this was reflected as stronger activation to the intelligible sentences in the auditory cortex and surrounding areas. A similar approach to control acoustic variability was used by Giraud et al. (2004) who measured functional magnetic resonance imaging (fMRI) responses to a set of vocoded sentences before and after the subject was trained to perceive these sentences correctly in a learning phase

* Corresponding author at: Department of Neuroscience and Biomedical Engineering (NBE), School of Science, Aalto University, PO Box 12200, FI-00076 AALTO, Finland.
E-mail address: maria.hakonen@aalto.fi (M. Hakonen).

http://dx.doi.org/10.1016/j.neuroimage.2015.10.029
1053-8119/© 2015 Published by Elsevier Inc.

where normal speech and vocoded speech were paired. Since the left inferior frontal gyrus (Broca's area) responded more strongly to noise-vocoded speech after training, the activation in this area was concluded to reflect speech intelligibility. Hannemann et al. (2007) described an electroencephalography (EEG) experiment where the subject first listened to unintelligible, digitally degraded words, after which half of the words were presented in undistorted, intelligible form and, finally, all degraded words were presented again. Those items which had been heard in the non-degraded form in the exposure sequence were more likely to be perceived as intelligible in the consecutive test sequence. Correct identification of the words was associated with an increase in induced gamma-band activity at left temporal electrode sites at around 350 ms. Taken together, these studies suggest that top-down cognitive processes, observable in both behavioral and brain measures, enhance speech comprehension and clearly warrant further exploration.

Studies using fMRI have shown how the processing of intelligible speech takes place in multiple cortical areas: activity spreads from the primary auditory cortex at Heschl's gyrus to the areas of the temporal cortex anterior, posterior and inferior to the primary auditory cortex (Davis and Johnsrude, 2003; Friederici et al., 2010; Leff et al., 2008; Möttönen et al., 2006; Okada et al., 2010), as well as to prefrontal, premotor/motor and posterior inferotemporal regions (Leff et al., 2008; Davis and Johnsrude, 2003; Obleser et al., 2008, for a review, see Peelle et al., 2010). Recent studies have reported that the patterns of intelligibility-related brain activity under unfavorable listening conditions are not identical to those under favorable listening conditions (Davis and Johnsrude, 2007; Giraud et al., 2004; Hervais-Adelman et al., 2012; Shahin et al., 2009; Wild et al., 2012), promoting the hypothesis for the existence of a separate, possibly attention-related, neural mechanism subserving comprehension of degraded speech (Hervais-Adelman et al., 2012). However, the role of, for example, motor areas (Lotto et al., 2009; Scott et al., 2009) and the auditory cortex in speech intelligibility remain controversial (Giraud et al., 2004; Peelle et al., 2010).

In MEG and EEG measurements, auditory stimuli elicit a series of transient responses, the most prominent of which is the auditory N1 response, measured electrically, and its magnetic counterpart, the N1m (for reviews, see Näätänen and Picton, 1987; May and Tiitinen, 2010). In the case of long-duration stimuli (>300 ms), the transient responses are followed by a sustained response that persists for the duration of the sound. The N1m response, generated in the auditory cortex and peaking approximately 100 ms after stimulus onset, is sensitive to the acoustic characteristics of speech sounds, such as the fundamental frequency (Mäkelä et al., 2002), intonation (Mäkelä et al., 2004), periodicity (Tiitinen et al., 2005; Yrttiaho et al., 2009) and phonological features (Obleser et al., 2004). The N1m has also been associated with the process of segregating speech signals from noise contributions (Miettinen et al., 2010, 2011, 2012). Most studies addressing sustained brain activity have used simplified stimuli, such as click trains (Galambos et al., 1981; Gutschalk et al., 2002; Hari et al., 1989), noise signals (Keceli et al., 2012), tones (Huotilainen et al., 1995; Okamoto et al., 2011), or vowels (Eulitz et al., 1995). However, the use of short-duration simplified stimuli may result in an incomplete picture of auditory analysis in the human brain. It is probable that the human brain is optimized for processing complex natural stimuli, such as connected speech (i.e. words and sentences). Therefore, studies geared strictly toward time-locked transient brain responses to brief stimuli lacking in information content should be complemented by investigations focusing on the sustained activity elicited by connected speech. This could potentially reveal how information is integrated over extended time spans, and how complex acoustic streams of sound are translated into meaningful utterances in the human brain.

The objective of the current MEG study was to examine the cortical mechanisms underlying speech comprehension under varying levels of speech intelligibility (i.e. using acoustically distorted and undistorted stimuli) and complexity (i.e. using vowel sounds, words, and sentences). The experimental paradigm introduced in our previous study (Tiitinen et al., 2012) was applied in the current study, with the subject first presented with distorted stimuli, then with undistorted versions of the same set of stimuli, and finally, with the distorted stimuli again. Acoustically identical distorted stimuli were expected to be perceived as either unintelligible or intelligible, depending on whether the subject had previously been exposed to the undistorted (intact) versions of the stimuli. Our hypothesis was that both this behaviorally observable intelligibility effect and variations in stimulus complexity should be accompanied by changes in both the dynamics and spread of brain activity from the auditory cortex to adjacent cortical areas. By exposing the subjects to the undistorted stimuli in the intermediate phase of the experiment, the current experimental setup allows manipulation of the intelligibility of the distorted stimuli without introducing any acoustic changes to these stimuli. Thus, any difference in brain activity elicited by the first and the second presentations of the distorted stimuli cannot be attributed to changes in the acoustic structure but, rather, to the processes directly involved with speech intelligibility. The overall goal of this study was, therefore, to provide further insight into how the top-down cognitive operations triggered by prior information are able to turn even severely distorted acoustic signals into meaningful cognitive entities by enhancing the extraction of relevant acoustic features.

Methods

Subjects

Behavioral and MEG measurements were carried out for two separate groups of sixteen healthy volunteers, aged 19–33 years (average age 22.4 years, SD 3.7 years; 8 male and 8 female; 15 right-handed) in the behavioral measurements and 20–26 years (average age 22.7 years, SD 1.6 years; 8 male and 8 female; 15 right-handed) in the MEG measurements. The use of different sets of subjects was necessary to avoid possible carry-over effects, whereby the presentation of the intact stimuli in the first experiment would renders the distorted stimuli intelligible in the second experiment, already at their first presentation. All volunteers had normal hearing and provided written informed consent. The experiments were approved by the Ethical Committee of Helsinki University Central Hospital.

Stimulus material

Vowels, words, and sentences were constructed using the Bitlips TTS synthesizer (http://www.bitlips.fi/). The sentence set consisted of 192 Finnish sentences, comprising 3 to 7 words (sentence duration: 1.7–4.6 s; mean 3.1 s; SD 0.6 s). Each sentence started with the vowel /a/, /e/, /i/ or /u/. The word set was created by separating the first word of each sentence. Thus, the words (0.31–1.40 s in duration, mean 0.65 s; SD 0.18) in the word set were acoustically identical to the initial words of the sentences. The vowel set included 200-ms instances of all eight vowels of the Finnish language (/a/, /e/, /i/, /o/, /u/, /y/, /ä/, /ö/). The stimuli were recorded at a sampling rate of 44.1 kHz with an amplitude resolution of 16 bits.

In addition to the above undistorted (16-bit) stimuli, the experiment utilized their distorted (1-bit) counterparts. The distorted versions of the stimuli were produced by first resampling the undistorted stimuli at 4.41 kHz using Matlab resample routine. Second, the resampled signals were compressed digitally through reduction of the amplitude resolution (bit rate) of the signals with the 1-bit uniform scalar quantification (USQ) method (see Liikkanen et al., 2007; Gray, 1990). USQ approximates each sample of the speech signal waveform to the nearest permitted level, the number of these depending on the number of bits used in the quantization. For example, using 16-bit USQ, there are a total of approx. 65 000 quantization levels which allows precise

Please cite this article as: Hakonen, M., et al., Previous exposure to intact speech increases intelligibility of its digitally degraded counterpart as a function of stimulus complexity, NeuroImage (2015), http://dx.doi.org/10.1016/j.neuroimage.2015.10.029