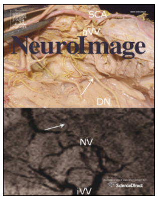
NeuroImage xxx (2015) xxx–xxx

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

ARTICLE INFO

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

Keywords:
Speech
Comprehension
Intelligibility
Acoustic distortion
Magnetoencephalography
Auditory neuroscience

ABSTRACT

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 modeling of the original speech waveform. In the case of 1-bit USQ, the number of levels is reduced to two which results in speech being represented by a series of rectangular pulse forms. Because quantization is a non-linear process, the 1-bit USQ process is capable of producing new frequencies which is seen as degradation of the spectral fine structure of speech by new noisy harmonics. In addition to this, the USQ generates quantization noise which is manifested as flattening of the spectral envelope (Miettinen et al., 2010). In overall, the degraded speech stimuli of the study were flat in terms of the spectral envelope and consisted of the low frequencies of speech (frequencies higher than 2.2 kHz were filtered out in resampling) and new noisy harmonics (generated by 1-bit USQ).

The stimuli were delivered as a mono signal to the subject's ears through Sennheiser HD headphones in the behavioral experiment and through a pair of plastic tubes and ear pieces (Etymotic Research Inc., IL, USA) in the MEG experiment. Sound intensity of the stimuli was set at 70 dB SPL. In the MEG experiment, the intensity was adjusted by measuring it with a sound level meter (Velleman DVM 805) at the tips of the tubes.

Experimental design

The study comprised two experiments. First, a behavioral experiment was designed to test the subject's ability to recognize the auditory stimuli. Second, the effects of acoustic distortion and intelligibility of the auditory stimuli on cortical activity was studied in an MEG experiment (Fig. 1).

The behavioral experiment, carried out in a soundproofed listening booth, consisted of three blocks: one comprising vowels, the second words, and the third sentences. The presentation order of the stimulus blocks was counterbalanced across subjects. During each stimulus block, the subject was first presented with distorted stimuli, then with the undistorted versions of the same set of stimuli, and finally with the distorted stimuli again. The vowel block contained 12 repetitions of eight vowels (/a/, /e/, /i/, /o/, /u/, /y/, /ä/, /ö/) presented in random order (i.e. 96 stimuli in total). At each vowel presentation, the eight alternative answers were presented on a computer screen to the subject whose task was to indicate with a mouse click which vowel had been presented or whether the vowel was unintelligible. Vowels which were correctly identified were classified as intelligible, those incorrectly identified as unintelligible. The set of 192 words and sentences was divided into four subsets (48 words/sentences per subset) and, similarly, the 16 subjects were divided into four subgroups (four subjects per subgroup). Each subgroup of the subjects was presented with one of the word/sentence subsets. The total set of 48 words/sentences presented for each subject was randomized. Following the presentation of each word and sentence, the subject used a keypad to type what he/she had heard. Correctly identified words were classified as identifiable and misidentified words as unintelligible. Intelligibility scores for sentences were computed by scoring the stems and suffixes of inflected words separately after obvious spelling errors had been corrected.

In the MEG experiment, the stimuli were presented in a passive (no task) recording condition during which the subject was under instruction to watch a film without its soundtrack while ignoring the auditory stimuli. Given the novelty of the experimental paradigm, the passive recording condition in MEG was an essential starting point for the investigation since it provides brain events uncontaminated by the effects of attentional engagement (arousal level, selective and/or sustained attention, etc.) as well as of planning and the execution of motor responses. Similarly as the behavioral experiment, the MEG experiment was divided into vowel, word and sentence blocks, in each of which the same set of stimuli was presented in a distorted, undistorted, and again in a distorted form. The presentation order of the stimulus blocks was counterbalanced across subjects. In the vowel block, four vowels (/a/, /e/, /i/, /u/) were repeated 120 times in random order. These vowels were selected because they are maximally displaced from each other in the two-dimensional space spanned by the first and the second formant. To keep the duration of the MEG experiment bearable to the subjects, a subset of 160 sentences was selected from the total set of 192 sentences presented in the behavioral experiment. The corresponding starting words of the sentences comprised the stimuli in the word block. The total set of 160 words/sentences was presented in random order to each subject. The offset-to-onset interstimulus interval (ISI) in all three blocks was 1 s. The duration of the experiment was ~1.75 h.

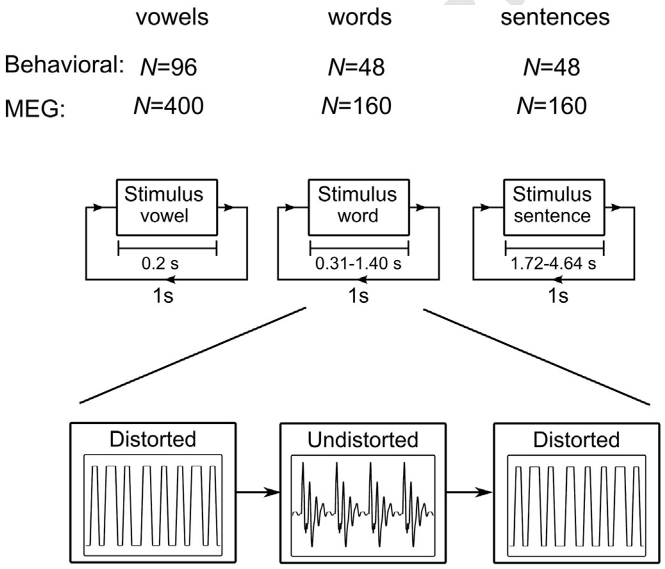


Fig. 1. Experimental design. The study was divided into behavioral and MEG experiments, each of which consisted of three blocks: one comprising vowels, the second words, and the third sentences. In each of the stimulus blocks, the subject was presented with acoustically distorted stimuli, followed by the undistorted versions of the same set of stimuli, after which the distorted stimuli were presented again. N = number of stimuli.

MEG data acquisition

Brain responses were recorded with a 306-channel whole head MEG device (Vectorview 4-D, Neuromag Oy, Finland) in a magnetically shielded room with a three-layer μ-metal and aluminum cover (ETS-Lindgren Euroshield Oy, Eura, Finland). The sampling rate was set at 1.2 kHz. Horizontal and vertical eye movements were measured with electro-oculography (EOG) using two electrode pairs placed above and below the left eye and lateral to the eyes. Before recording, four head-position indicator (HPI) coils were used to determine the position of the subject's head relative to the MEG sensor array. A 3-D digitizer was used to determine the locations of the HPI coils with respect to the three anatomical landmarks (the nasion and the bilateral preauricular points) that define a head-based coordinate system where the x-axis passes through the preauricular points (positive to the right), the y-axis passes through the nasion (positive to the front), and the z-axis unit vector is the vector cross product of the x and y unit vectors. The subject was instructed to remain stationary and to avoid blinking during the measurement.

MEG data preprocessing

The raw MEG data was manually inspected to exclude gradiometer sensors with a low signal-to-noise ratio (SNR). External noise was removed from the raw data using the temporal extension of Signal-Space Separation (tSSS; Taulu and Simola, 2006) as implemented with the MaxFilter software (Elekta-Neuromag). For the transient analysis, the raw data was band-pass filtered at 2–30 Hz with a 4th order Butterworth infinite impulse response (IIR) filter with a length of 10 s.

For the sustained field analysis, the raw data was low-pass filtered at 30 Hz, since high-pass filtering abolishes sustained fields (May and Tiitinen, 2010; Mäkinen, 2006). After filtering, the epochs were computed and corrected with respect to a 100-ms pre-stimulus baseline. For averaging the data, the data epochs were time-locked to stimulus onsets using information recorded on the trigger channel. A 500-ms post-stimulus time window was used for averaging transient responses. For the sustained field analysis, the responses elicited by words and sentences were averaged over a 600-ms post-stimulus time window. To exclude the possible contributions caused by transient offset responses, a subset of 112 words exceeding 560 ms in duration was selected for the sustained field analysis. For the responses to sentences, the analysis was restricted to the sustained fields elicited by the corresponding subset of 112 sentences starting with the words exceeding 560 ms. Epochs with magnetic field gradient amplitudes exceeding 2000 fT/cm were automatically discarded from both the transient and sustained filed analyses. EOG artifacts were removed from the epochs using fast independent component analysis (Gramfort et al., 2013; Hyvärinen and Oja, 2000). Fitting ICA after filtering and epoching allowed more reliable identification of the components that reflect EOG artifacts since the high-frequency noise and drifts as well as contaminated epochs were removed before EOG artifact identification. After removing the EOG artifacts, the epochs were averaged. The preprocessing was performed with the MNE software (Gramfort et al., 2013, 2014).

Gradiometer analysis

For the gradiometer analysis, nine gradiometer pairs centered over the left and right auditory cortices were divided into anterior, medial, and posterior subsets, and an average was calculated over the responses from the three gradiometer pairs within each subset. Response amplitude was determined as the magnitude of the gradiometer pair vector sum. The latencies and amplitudes of the N1m and P2m responses were estimated from the peak values of the anterior, medial and posterior vector sum magnitudes. The peak amplitudes of the N1m and P2m responses were identified as the local maximum within the respective time intervals of 110–170 ms and 180–300 ms. Because the onsets and offsets of the vowels were smoothed with a 10-ms Hann window (ramp length 5 ms) whereas for words and sentences a Hann window of 20 ms (ramp length 10 ms) was used, the difference between the lengths of Hann window ramps (5 ms) was added to the latencies of the transient responses elicited by vowels to make them comparable to the latencies of the transient responses elicited by words and sentences.

Sustained fields were analyzed by dividing the gradiometer pairs into ten location-based subsets: occipital, parietal, and left and right frontal, temporal, sensorimotor and occipitotemporal subsets. For each subset, the magnitudes of the gradiometer-pair vector sums were averaged, and this average was then used to quantify the magnitude of the sustained field as the mean of the response in the 400–560 ms time window.

Current distribution estimates

To estimate the spatial distribution of cortical activity, depth-weighted minimum-norm estimates (MNEs; Hämäläinen and Ilmoniemi, 1994; Lin et al., 2006a,b) were generated with the MNE Software (Gramfort et al., 2014). Moreover, dynamic statistical map (dSPM) estimates (Dale et al., 2000) were generated to provide an indication of the cortical locations where the MNE estimates had the highest SNR. For the MNE and the dSPM estimates, noise-covariance matrices were computed from the 100-ms pre-stimulus baselines of the data. Forward solutions and inverse operators were calculated for each stimulus by employing a single-compartment boundary-element model (BEM) computed using average head and skull surface reconstructions provided by the FreeSurfer software. A loose orientation constraint was used to control the source orientations (Lin et al., 2006a). The MNE and the dSPM estimates were calculated separately for each individual subject, for each type of stimulus (i.e. vowels, words, and sentences), and for each condition (i.e. the first and the second presentations of the distorted stimuli and the presentation of the undistorted stimuli). The dSPM estimates were averaged over a 40-ms time window centered at the peaks of the N1m and P2m responses. The peak latencies of the N1m and P2m were calculated as the time instants when the noise-normalized current estimates reached their maximum values within the respective time windows of 110–170 ms and 190–310 ms. For visualization purposes, dSPM estimates for the N1m and P2m responses were grand-averaged across subjects and conditions. Additionally, grand-averaged dSPM estimates for the sustained field were calculated as mean values in the time range of 400–560 ms.

Region of interest analysis

Regions of interest (ROI) were determined by dividing the dSPM activation areas into subregions on the basis of an anatomical parcellation following the Desikan–Killiany–Tourville atlas (see Fig. 5; Desikan et al., 2006). Activations were analyzed in the following brain areas: the transverse temporal gyrus (TTG), the superior temporal gyrus (STG), the superior temporal sulcus (STS), the supramarginal gyrus (SMG), the insula, the pars opercularis of the inferior frontal gyrus (POp), and the precentral sulcus (PCS). The TTG and STG were combined into the same ROI (TTG + STG). Because of the lack of individual structural MRI data, the ROIs represent only approximations of the corresponding brain areas. The original MNEs without noise normalization were averaged over the source locations to obtain a time course of current strength for each ROI. The mean currents within the ROIs were obtained by averaging the time courses of the currents using a 40-ms time window centered at the peaks of the N1m and P2m responses. The peak latencies were calculated as the time instants when the time courses of the currents exhibited their maximum values within a 110–170-ms time interval for the N1m and a 190–310-ms time interval for the P2m response. For responses to words and sentences, the mean currents were averaged over the time interval of 400–560 ms.

Dipole modeling

The single equivalent current dipole (ECD) was used to estimate the source locations of the N1m and P2m responses. The ECDs were modeled separately in each hemisphere by using a set of 12 gradiometer pairs over each temporal region. A spherical model was used to estimate the conductivity of the head. The ECD analysis was performed with the Elekta Neuromag xFit Source Modeling Software.

Statistical analyses

The data from the behavioral and MEG measurements were analyzed using repeated-measures analysis of variance (ANOVA). Mauchly sphericity tests were run, and Greenhouse–Geisser-corrected p values Q6 and epsilons (ε) were reported when the assumption of sphericity was violated. For behavioral data, a 3 × 3 ANOVA was carried out, with the within-subject factors of complexity (vowel/word/sentence) and condition (degraded & non-intelligible, i.e. 1st distorted; non-degraded & intelligible, i.e. undistorted; degraded & intelligible, i.e. 2nd distorted). In the gradiometer analysis, the transient responses for vowels, words and sentences were analyzed in separate 2 × 3 ANOVAs (hemisphere × condition). Separate ANOVAs were used because the SNR was very low for the responses to words and sentences especially in the posterior and anterior channels, and therefore, including all the complexity levels in the same ANOVA would have required the rejection of a large number of subjects. To study the interactions between condition and complexity as well as the main effect of complexity, an additional analysis was made for the responses measured from the medial channels by including the transient responses to vowels, words and sentences in the same ANOVA. The mean currents during the transient responses also were investigated in separate 2 × 3 ANOVAs (hemisphere × condition) for vowels, words and sentences. This was because the number and shape of the ROIs determined using dynamic statistic map estimates (dSPM) varied between complexity levels. Similarly as in the gradiometer analysis, the interactions between condition and complexity as well as the main effect of complexity were studied by conducting an additional analysis where the responses to vowels, words and sentences were included in the same ANOVA table. This analysis was restricted to a subset of four ROIs (TTG + STG, SMG, STS and insula) that showed the highest SNR for sentences on the basis of the dSPM estimates. These ROIs were selected since they overlapped with the brain regions that showed the highest SNR also for vowels and words. Sustained fields elicited by words and sentences were included in a corresponding 2 × 2 × 3 ANOVA (complexity × hemisphere × condition). The effects of intelligibility and speech degradation on the cortical activity measures were analyzed by post-hoc (Newman–Keuls) comparisons of the conditions. If the responses to the first and the second presentations of the distorted stimuli were unequal, the response was assumed to reflect intelligibility. If the responses to the first and the second presentations were equal with each other, but of a different magnitude than the response to the undistorted stimuli, it is likely that these differences reflect sensitivity to the acoustic structure of the stimulus.

Results

The behaviorally measured intelligibility of the stimuli

Intelligibility (defined as the proportion of correct identifications) was 98.6% (±0.5%) for the undistorted vowels, words and sentences. As depicted in Fig. 2, the overall intelligibility was lower for the first presentation of the distorted stimuli than for the second presentation, increasing from 25.7 ± 2.5% to 45.7 ± 2.7% ($F(1,21) = 575.7$, $\varepsilon = 0.7$, $p < 0.001$). This increase in intelligibility by 20 percentage points for acoustically identical stimuli demonstrates how a single presentation of intact speech material can alter the subject's ability to comprehend distorted speech.

The intelligibility of the undistorted stimuli was high for all complexity levels: 97.6% (±0.8%) for sentences, 98.8% (±0.5%) for words, and 99.4% (±0.3%) for vowels. The magnitude by which the presentation of the undistorted stimuli increased the intelligibility of the distorted stimuli depended on the complexity level ($F(4,60) = 27.1$; $p < 0.001$). This increase was 33.8 percentage points for sentences (from 31.0 ± 4.4% to 64.9 ± 4.6%; $p < 0.001$), 25.9 percentage points for words (from 19.1 ± 3.9% to 45.1 ± 3.7%; $p < 0.001$) and, somewhat surprisingly, there was no improvement of vowel intelligibility (from 26.8 ± 3.0% to 27.1 ± 2.5%, $p =$ n.s.). Thus, it appears that the complexity of the stimuli had a considerable effect on how well the subject was able to comprehend the stimulus, with intelligibility increasing with stimulus complexity.

Transient responses

Amplitudes

As depicted in Fig. 3, all complexity levels elicited prominent N1m and P2m responses in both hemispheres. Fig. 4 shows the peak amplitudes, peak latencies, and mean currents for these transient responses. In anterior areas, the N1m amplitude increased from 23.6 ± 3.0 fT/cm as elicited by the undistorted vowels to 30.5 ± 4.8 fT/cm and 33.9 ± 5.3 fT/cm as elicited by the first and the second presentations of the distorted vowels, respectively ($F(1,16) = 11.2$, $\varepsilon = 0.6$, $p < 0.01$). This effect of speech distortion was hemispherically asymmetric ($F(1,18) = 8.3$, $\varepsilon = 0.7$, $p < 0.01$), and post-hoc tests revealed that the N1m amplitude was stronger for the distorted vowels only in the right hemisphere (1st distorted: 36.0 ± 5.4 fT/cm, 2nd distorted: 40.4 ± 6.3 fT/cm, undistorted: 23.7 ± 2.9 fT/cm, $p < 0.001$). In medial and posterior areas, the distorted and the undistorted vowels elicited N1m amplitudes of equal magnitude. The magnitude of the N1m amplitude elicited by words showed hemispheric asymmetry ($F(2,24) = 3.1$, $p < 0.07$) in the anterior channels, where the right-hemispheric amplitude was stronger for the second presentation of the distorted words (22.2 ± 3.1 fT/cm) than for their first presentation (17.8 ± 3.4 fT/cm, $p < 0.05$) as well as for the undistorted words (13.7 ± 1.8 fT/cm, $p < 0.001$). No significant differences were found between the magnitudes of the left-hemispheric N1m responses to words. Hemispheric asymmetry was also found in the N1m amplitudes in the posterior channels ($F(2,22) = 6.7$, $p < 0.01$): only the right-hemispheric N1m response increased from 10.1 ± 2.0 fT/cm for the undistorted words to 17.3 ± 2.7 fT/cm for the second presentation of the distorted words ($p < 0.05$). The amplitudes of the N1m responses elicited by sentences were of equal magnitude in all cases. In the medial channels, vowels (34.8 ± 3.6 fT/cm) elicited 10.6 fT/cm stronger response than words (24.2 ± 2.2 fT/cm, $p < 0.001$) and 13.6 fT/cm stronger response than sentences (21.2 ± 1.8 fT/cm, $p < 0.001$; see Fig. 4). The P2m amplitudes were insensitive to intelligibility, speech degradation, and complexity.

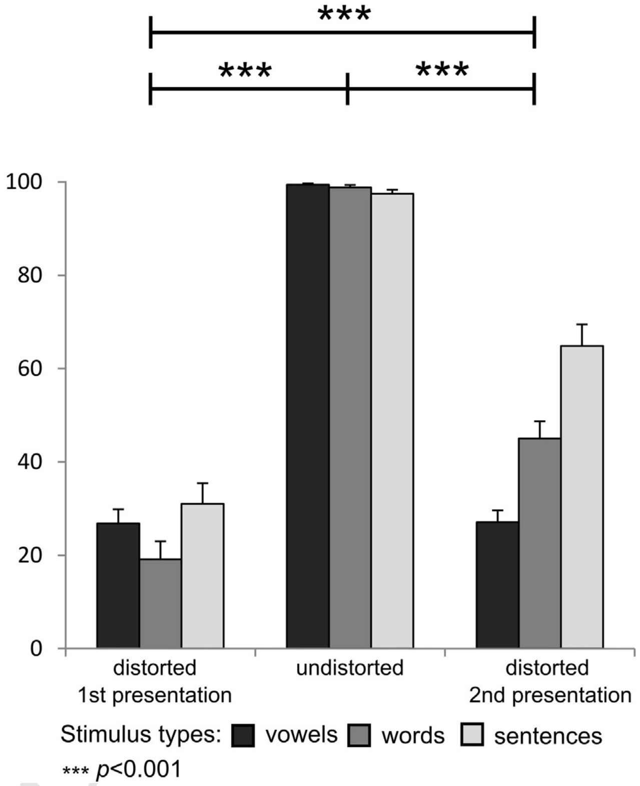


Fig. 2. Behavioral results. At their first presentation, distorted stimuli were difficult to understand (mean subjective intelligibility rating = 25.7%). After an intervening presentation of the same stimuli in an undistorted form (98.6%), the intelligibility of the words and sentences increased considerably (45.7%). Error bars indicate SEM. Significance stars indicate differences between intelligibilities of speech stimuli in their three presentations (averages over complexity levels).

Latencies

The distorted vowels, words, and sentences resulted in N1m and P2m responses which had a longer peak latency than the responses elicited by the corresponding undistorted versions of the stimuli. In anterior channels, this delay was 7.2 ms for vowels, 9.1 ms for words and 12.0 ms for sentences (see Table 1 for details). For vowels, the delay was hemispherically asymmetric ($F(1,19) = 4.5$, $\varepsilon = 0.7$, $p < 0.05$): in the

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

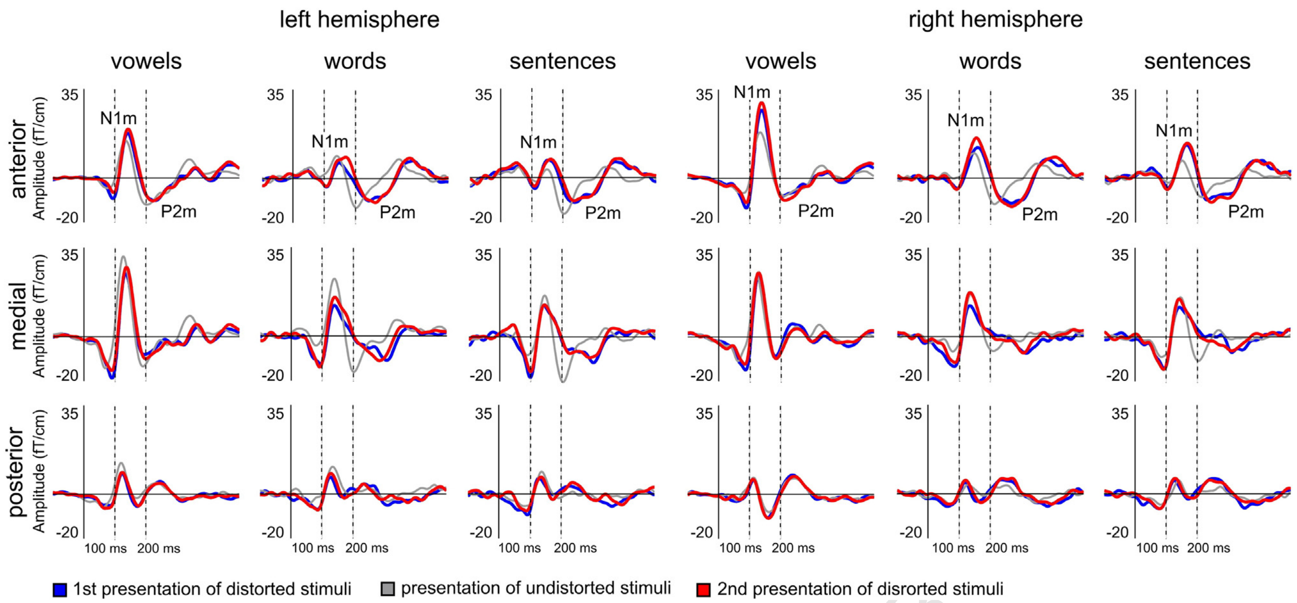


Fig. 3. The transient evoked fields elicited by the distorted and undistorted presentations of vowels, words and sentences. The N1m response was more prominent in anterior and medial than in posterior areas. The P2m response was strongest in the anterior and weakest in the posterior areas. In anterior and medial areas, the N1m amplitude was stronger for vowels than for words and sentences, whereas the P2m amplitude was stronger for sentences than for vowels. Stimulus degradation lead to an increased amplitude of the right-hemispheric N1m and to a decreased amplitude of the left-hemispheric N1m. Stimulus degradation delayed the N1m and P2m latencies in anterior and medial channels. These effects were stronger for responses elicited by words and sentences than for those elicited by vowels.

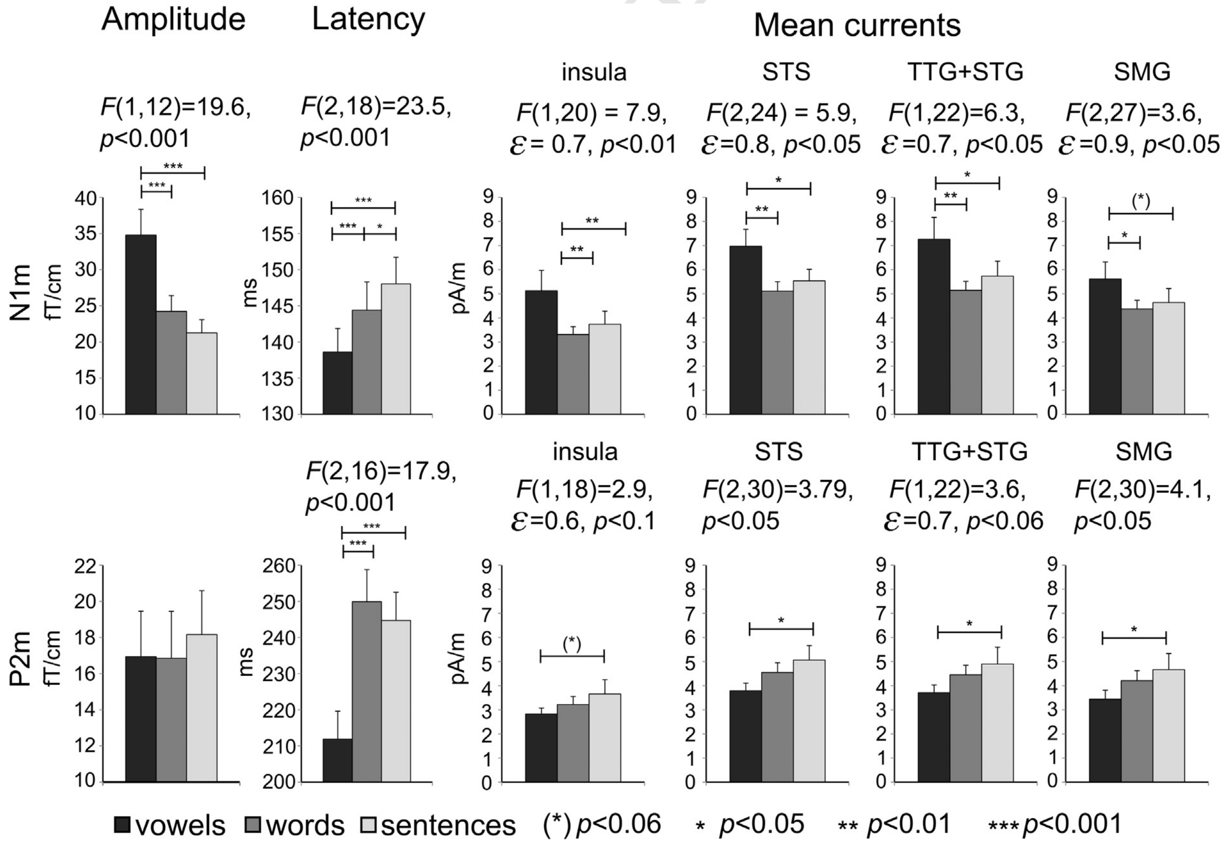


Fig. 4. The effect of speech complexity on the amplitudes, latencies, and mean currents of the N1m and P2m responses. The N1m amplitude was stronger, and its latency was shorter for vowels than for words and sentences. Also, the N1m latency was longer for words than for sentences. During the P2m time range, the mean currents following vowel presentation were stronger than those following sentence presentation. In comparison to words and sentences, vowels elicited a P2m response with a shorter latency. The amplitudes and latencies were analyzed from medial channels. *F* and *p* values describe the main effect of complexity. The mean currents were computed from the ROIs determined separately for the N1m and P2m responses on the basis of the dSPM activation areas for sentences shown in Fig. 5. Error bars indicate SEM.

Table 1
The N1m and P2m latencies in three different conditions. Distortion of the stimuli delayed the responses bilaterally. The results for vowels, words and sentences were computed using separate ANOVAs with different number of subjects. The responses for the medial channels were additionally analyzed in the same ANOVA. However, no interaction between complexity and degradation was found in this analysis. The main effect of complexity is shown in Fig. 4.

	F	*df1*	*df2*	*p*	1st distorted	Undistorted	2nd distorted
N1m latency							
Anterior							
Vowels	8.3	2	28	<0.01	145.8 ± 2.2	139.8 ± 2.8	148.1 ± 2.2
Words	7.2	2	24	<0.01	149.9 ± 2.9	141.4 ± 2.9	151.1 ± 3.7
Sentences	9.3	2	14	<0.01	157.8 ± 2.5	144.7 ± 3.3	155.7 ± 3.8
Medial							
Vowels	11.5	2	30	<0.001	137.3 ± 2.8	132.7 ± 2.4	140.5 ± 2.3
Words	3.3	2	24	<0.06	146.2 ± 4.1	138.9 ± 2.8	145.9 ± 3.7
Posterior		2					
Vowels	3.6	2	24	<0.05	138.5 ± 2.9	134.9 ± 4.2	141.3 ± 3.0
P2m latency							
Anterior							
Vowels	4.4	2	28	<0.05	230.3 ± 5.6	216.8 ± 5.7	231.0 ± 4.4
Words	21.9	2	24	<0.001	263.9 ± 7.1	220.6 ± 5.8	263.7 ± 7.3
Sentences	13.7	2	20	<0.001	250.6 ± 9.2	213.4 ± 5.4	263.3 ± 10.0
Medial							
Words	16.4	2	26	<0.001	264.8 ± 10.4	217.4 ± 5.4	261.0 ± 9.7
Sentences	5.8	2	20	<0.05	251.5 ± 10.4	220.6 ± 7.4	263.0 ± 12.3
Posterior							
Words	4.2	2	14	<0.05	245.1 ± 11.9	212.9 ± 8.9	248.7 ± 15.7

left hemisphere, the distorted vowels (1st distorted presentation: 146.7 ± 2.7 ms, 2nd distorted presentation: 150.0 ± 2.7 ms) resulted in 10.4 ms longer N1m latencies than the undistorted vowels (138.0 ± 3.7 ms, $p < 0.001$), whereas in the right hemisphere, this increase was only 4.0 ms (undistorted: 141.6 ± 2.5 ms; distorted: 144.9 ± 2.0 ms; 146.2 ± 2.0 ms, $p < 0.07$). In medial channels, speech degradation delayed the N1m responses 6.2 ms for vowels and 7.2 ms for words (see Table 1). In posterior channels, this degradation-related delaying effect was observed only in the case of vowel stimuli where it was 5 ms. The N1m latency for vowels was 5.8 ms shorter (138.6 ± 3.3 ms) than that for words (144.4 ± 3.9 ms, $p < 0.001$) and 9.4 ms shorter than that for sentences (148.0 ± 3.7 ms, $p < 0.001$; see Fig. 4). The N1m latency was 3.6 ms shorter for words than for sentences ($p < 0.05$).

The P2m latencies were delayed even more than the N1m latencies as a result of speech degradation: In the anterior channels, the latencies were prolonged by 13.9 ms for vowels, 43.2 ms for words, and 43.5 ms for sentences. In the medial channels, the corresponding delay was 45.4 ms for words and 36.6 ms for sentences. In the posterior channels, the P2m responses to the distorted words peaked 34.0 ms later than the responses elicited by the undistorted words. The P2m latency was 38.0 ms shorter for vowels (211.9 ± 7.7 ms) than for words (249.9 ± 8.9 ms, $p < 0.001$), and it was 32.8 ms shorter for vowels than for sentences (244.7 ± 7.8 ms, $p < 0.001$). The differences in the peak latency of the transient responses elicited by words and sentences were not significant.

Source locations

The source locations of the N1m and P2m responses were modeled with a single ECD at their respective peak latencies in each hemisphere. The average goodness-of-fit values were 90%. There was no difference between the source locations of the N1m responses elicited by the first and the second presentations of the distorted vowels in the anterior–posterior direction (mean $y = 11.1 \pm 2.3$ mm). In comparison, the source of the N1m elicited by the undistorted vowels ($y = 6.6 \pm 2.1$ mm) was 4.5 mm posterior to the N1m source for the distorted vowels ($F(2,24) = 4.2$, $p < 0.05$). In the case of word stimulation, the source of the P2m response to the undistorted stimuli was shifted 9.9 mm in the inferior direction compared to the source of the response to the distorted counterparts of the stimuli (distorted: mean $z = 59.2 \pm 1.9$ mm, undistorted: $z = 49.3 \pm 3.5$ mm; $F(2,8) = 5.9$, $p < 0.05$). The source location of the P2m elicited by vowels was dependent on condition (i.e. first distorted, undistorted, second distorted) and hemisphere ($F(1,9) = 6.5$, $\varepsilon = 0.6$, $p < 0.05$): In the right hemisphere, the second presentation of the distorted vowels elicited a more medial P2m response ($x = 41.5 \pm 4.3$ mm) than the undistorted vowels ($x = 54.6 \pm 2.1$ mm, $p < 0.001$) and the first presentation of the distorted vowels ($x = 50.4 \pm 1.9$ mm, $p < 0.05$), but this medial shift was not observed in the left hemisphere. In general, the source locations of the N1m responses for vowels and sentences were more anterior in the right (vowels: $y = 15.0 \pm 2.2$ mm; sentences: $y = 5.2 \pm 1.7$ mm) than in the left hemisphere (vowels: $y = 4.1 \pm 2.5$ mm, $F(1,12) = 15.2$, $p < 0.01$; sentences: $y = -2.8 \pm 2.8$ mm, $F(1,8) = 9.6$, $p < 0.05$).

Mean currents

A number of brain regions exhibited sensitivity to the intelligibility of speech during the N1m time range (see Fig. 5). In the left-hemispheric TTG + STG, SMG, STS, insula, and PSC, cortical activity increased from 2.2–4.4 pA/m elicited by the first presentation of the distorted vowels to 3.8–7.9 pA/m elicited by the presentation of the undistorted vowels and to 3.0–6.2 pA/m elicited by the second presentation of the distorted vowels. A comparable intelligibility effect was also found in the right-hemispheric TTG + STG, STS and insula. In these brain areas, the second presentation of the distorted vowels (4.4–7.9 pA/m) yielded on average 1.4 pA/m stronger currents than the first presentation of the same vowels (3.5–6.2 pA/m; see Fig. 5 for statistical results). For sentences, the left-hemispheric insula showed sensitivity to speech intelligibility, with the unintelligible sentences resulting in 1.0 pA/m weaker currents (1st distorted: 3.3 pA/m) than the intelligible ones (2nd distorted: 4.3 pA/m, undistorted: 4.1 pA/m). In the left-hemispheric POp, speech degradation decreased the mean currents from 5.0 pA/m elicited by the presentation of the undistorted vowels to 2.3 pA/m and 3.0 pA/m elicited by the first and the second presentations of the distorted vowels, respectively. For words, a corresponding effect was found in the left-hemispheric STS where brain activity decreased from 6.5 pA/m to 4.9 pA/m (1st distorted) and 4.4 pA/m (2nd distorted) as a result of speech degradation. As illustrated in Fig. 4, the mean currents during the N1m time range were on average 1.5 pA/m stronger for vowels (5.1–7.3 pA/m) than for words (3.3–5.1 pA/m) and sentences (3.7–5.7 pA/m) in the TTG + STG, SMG, STS and insula, whereas the responses to words were of equal magnitude to those elicited by sentences.

During the P2m time range, speech degradation affected cortical activity in the auditory cortex and surrounding areas. In the left hemisphere, the TTG + STG, SMG, STS and insula showed sensitivity to acoustic degradation in that speech degradation decreased the mean currents from 4.1–5.0 pA/m to 2.1–3.3 pA/m for vowels and from 5.0–7.2 pA/m to 3.8–5.3 pA/m for sentences (currents for 1st and 2nd distorted presentations averaged). With word stimuli, sensitivity to speech degradation was found in the left-hemispheric TTG + STG and insula where the mean currents decreased from 4.2–5.9 pA/m to 3.1–4.7 pA/m. In the right hemisphere, cortical activity decreased from 3.8–4.1 pA/m to 2.9–3.3 pA/m as a result of vowel degradation in the STS and SMG. In the case of words and sentences, the right hemisphere showed no sensitivity to speech degradation. As illustrated in Fig. 4, the mean currents increased from 2.8–3.8 pA/m to 3.7–5.1 pA/m (on average 1.1 pA/m) when responses to vowels were contrasted to those elicited by sentences in the TTG + STG, SMG, and STS.

Sustained fields

Amplitudes

The mean amplitude of the sustained field was on average 3.7 fT/cm stronger for the distorted (16.5–18.6 fT/cm) than for the undistorted words (12.9–14.3 fT/cm) in the sensorimotor, occipitotemporal and

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

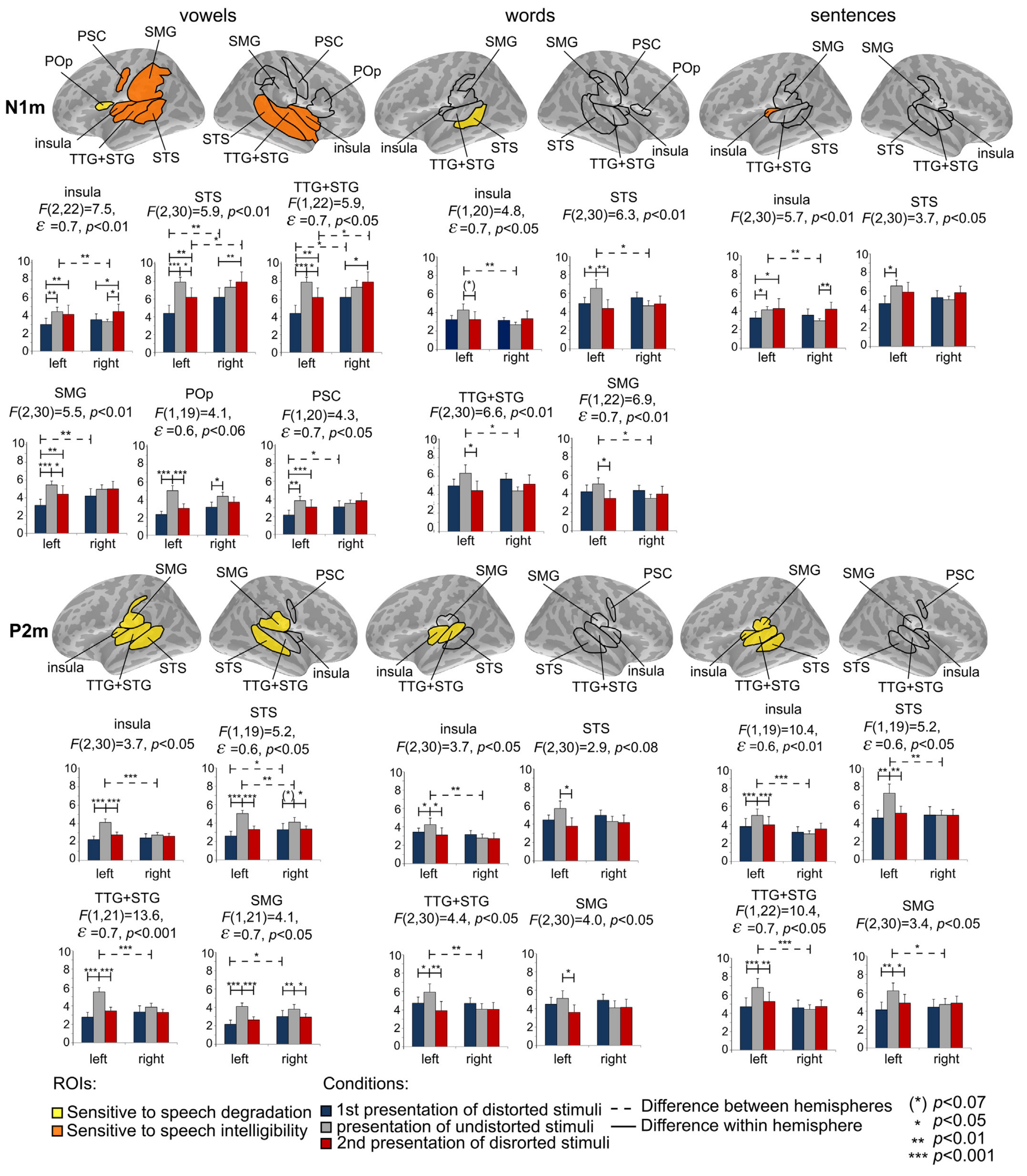


Fig. 5. Regions of interest (ROIs) affected by speech degradation and intelligibility during the N1m and P2m time ranges. The results for vowels, words and sentences were computed using separate ANOVAs with different number of subjects. The interactions between condition and complexity as well as the main effect of complexity were studied by conducting an additional analysis where the responses to vowels, words and sentences were included in the same ANOVA table. This analysis was restricted to a subset of four ROIs (TTG + STG, SMG, STS and insula) that showed the highest SNR for sentences on the basis of the dSPM estimates. However, no interaction between complexity and condition was found in this analysis. See Fig. 4 for the main effects of complexity. *F* and *p* values describe interactions between condition (i.e. 1st distorted, undistorted, 2nd distorted) and hemisphere. Abbreviations for the ROIs: transverse temporal gyrus and superior temporal gyrus (TTG + STG), superior temporal sulcus (STS), supramarginal gyrus (SMG), pars opercularis of the inferior frontal gyrus (POp) and precentral sulcus (PCS). Error bars indicate SEM.

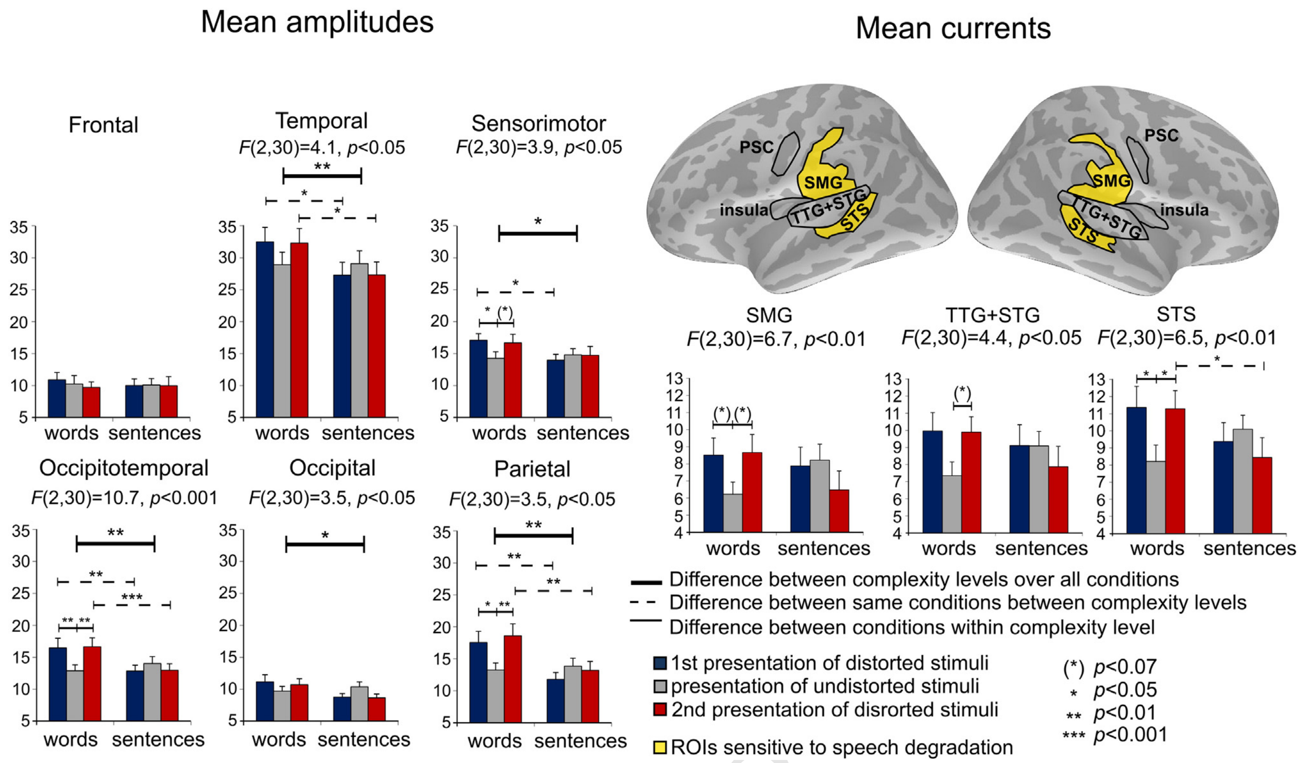


Fig. 6. The effect of stimulus degradation and stimulus complexity on the sustained fields elicited by words and sentences. Speech degradation increased the amplitudes of sustained fields for words in all but frontal and occipital areas. This degradation-related response enhancement was evident also in the mean currents computed in the ROI analysis. *F* and *p* values describe interactions between condition (i.e. 1st distorted, undistorted, 2nd distorted) and complexity. The sustained fields are averages over the time range of 400–560 ms. Error bars indicate SEM.

parietal areas (see Fig. 6). Further, the mean amplitudes were equally strong for the first and the second presentations of the distorted words. Words elicited on average 2.3 fT/cm stronger sustained fields (10.5–31.2 fT/cm) compared to sentences (9.3–27.9 fT/cm) in all but the frontal area. In the temporal, occipitotemporal and parietal areas, this effect was due to the stronger sustained fields for the distorted words (16.5–32.5 fT/cm) when compared to the sustained field elicited by the distorted sentences (11.8–27.3 fT/cm). In the sensorimotor area, the first presentation of the distorted words resulted in a 3.1 fT/cm stronger sustained field (17.1 fT/cm) compared to that elicited by the first presentation of the distorted sentences (14.0 fT/cm).

Mean currents

The mean currents during the sustained field for words were sensitive to speech degradation in the SMG and STS, with the first and the second presentations of the distorted stimuli resulting in 2.7 pA/m stronger responses (8.5–11.4 pA/m) compared to the responses elicited by the undistorted stimuli (6.2–8.2 pA/m). In the TTG + STG, the distorted words elicited 2.5 pA/m stronger mean current than the undistorted words (7.4 pA/m) only at their second presentation (9.9 pA/m). The results are summarized in Fig. 6.

Discussion

We studied how speech comprehension is modified by varying the level of intelligibility (i.e. by using acoustically distorted and undistorted stimuli) and stimulus complexity (i.e. by using isolated vowel sounds, words, and sentences). In both the behavioral and MEG experiments, the subject was first presented with a set of the distorted stimuli, then with the undistorted (intact) version of the same set, and finally the set of distorted stimuli was presented once more. We were particularly interested in comparing the responses to the two instances of the distorted stimulus sets which were acoustically identical. While the intelligibility of the undistorted stimuli was 99%, the overall intelligibility of the distorted speech sounds increased from 26% at the first presentation to 46% at the second. Thus, only a single intervening exposure to an intact (undistorted) version of the stimulus was sufficient to render the originally incomprehensible (distorted) speech sounds comprehensible to the subjects. Interestingly, we also found that intelligibility increased considerably as a function of stimulus complexity, indicating the importance of both word- and sentence-level information to speech comprehension under adverse acoustic conditions. Speech intelligibility was reflected in cortical activity already at latencies of 130–160 ms, suggesting that top-down modulations from higher-order cortical areas take place very rapidly. Speech complexity correlated with cortical activity at latencies of 200–270 ms.

Effects of top-down information and stimulus complexity on behaviorally measured speech intelligibility

The distorted stimuli were difficult to understand at their first presentation (average intelligibility across the three complexity levels: 26%), but after an intervening presentation of the same stimuli in the undistorted, intelligible form (99%), the comprehensibility of the distorted stimuli increased considerably (46%). This result corroborates the findings of our recent attempt on this particular issue (Tiitinen et al., 2012) and that of Hannemann et al. (2007) who found that hearing low-resolution, initially unintelligible nouns in intact form renders them more likely to be perceived as intelligible in a consecutive test sequence. It is unlikely that these perceptual changes resulted from the repeated presentation of the distorted stimuli since there was a gap of several minutes (10 min for vowels, 4 min for words, and 11 min for sentences) between the distorted presentations. Therefore, it is improbable that the subject could have been drawing on any echoic or short-term memory resources in this case. This conclusion is supported by the results obtained in the MEG measurements: stimulus repetition within a time window of seconds usually leads to diminished brain responses, that is, to an opposite pattern to the one observed in this study. As also the contaminating effects of acoustic variation in the stimulation can be ruled out as an explanation for the performance enhancement in these experimental paradigms, it appears that only a single encounter with the intact stimulus significantly facilitates the identification of its distorted counterpart, probably by activating (Diamond and Rozin, 1984; Dorfman, 1994; Forster et al., 1990) and/or modifying (Morton, 1979) corresponding preexisting memory representations or by creating new memory traces (for a review, see Bodner and Masson, 2014). This kind of effect of a stimulus on the response to a later stimulus is referred to as priming (Tulving and Schacter, 1990), and it has typically been examined in word-stem completion (WSC) tests (Tulving et al., 1982). In WSC tests, the subject is first presented with a list of words, and after a delay (typically from several minutes to several hours) he/she is given word stems that have multiple possible completions. Priming occurs when the subject completes the stem more often according to words that had been presented to him/her earlier than according to words that were not presented previously. Priming effects have been reported for different types of stimuli, including pseudowords (Bowers, 1996), familiar and unfamiliar objects (Diamond, 1990), and visual patterns (Ahissar and Hochstein, 2004; Tallon-Baudry et al., 1997). Our results parallel these findings by suggesting that perception relies not only on incoming sensory information but also on facilitative memory-related cognitive processes. Further, this top-down support can produce dramatic and rapid changes to the percept when the eliciting stimulus is under extreme forms of distortion and when not enough bottom-up acoustic cues are available to achieve comprehension.

The distorted vowels were difficult to understand (27%), both before and after the presentation of their undistorted and intelligible counterparts. In contrast, the initially challenging words (19%) became considerably more comprehensible (45%) following the presentation of the undistorted words. This effect was even more pronounced for the distorted sentences whose intelligibility increased from 31% to 65%. Thus, the comprehension of the distorted speech stimuli appears to become easier the more complex the signal is. The increase of intelligibility for 48 sentences in this study was 15 percentage points smaller than the increase achieved with 120 sentences in our previous study (Tiitinen et al., 2012). One possible reason for this is that the sentence set in our previous study was constructed from only seven starting words, three sentence stubs, and four ending words, whereas in the recent study each sentence was unique. Thus, it seems that decreasing the information content of the stimulus set and repeating the same words and sentence stubs increases the priming effect, making the sentences easier to recognize in their distorted form.

The current data merely suggests that the retrieval on bottom-up information from degraded speech is enhanced by memory-driven top-down processes and, at least if low-frequency information is still present in the signal, this effect increases as a function of stimulus complexity. However, it remains unclear which particular aspects of complexity the increase in intelligibility relies on. A possible explanation would be syntactic (Miller and Isard, 1963), semantic (Boothroyd and Nittrouer, 1988; Bradlow and Alexander, 2007; Kalikow et al., 1977; Obleser et al., 2007; Smiljanic and Sladen, 2013; Valentini-Botinhao and Wester, 2014), and lexical influences (McClelland et al., 2006) that have shown to exert contextual influences over speech decoding and intelligibility. Our observations coupled with these previous ones support many of the current psycholinguistic theories, such as the TRACE model (McClelland and Elman, 1986) and the distributed coherent model (Marslen-Wilson, 1987), according to which the mapping of sensory input to stored speech representations occurs in a cascading manner as the speech stream evolves, and the number of possible lexical candidates is gradually reduced until recognition is achieved. This is nicely demonstrated in our experiment: speech intelligibility increases as information accumulates when vowels evolve into words, and words into sentences. Another aspect of speech complexity that may have increased intelligibility is the temporal envelope of speech (i.e. the acoustic power at a given time in a given frequency range; Peelle and Davis, 2012). This contains cues for speech parsing on both the syllabic and phrasal levels (for reviewers, see Golumbic et al., 2012; Peelle and Davis, 2012). These low-frequency fluctuations were still present in the degraded signals of the current study, although distorted by their harmonic frequencies. Replacing the signal waveform with rectangular pulses in the USQ procedure clearly decreases the envelope information but retains distinguishable "bursts" of energy. Since these bursts indicate distinct syllables (Golumbic et al., 2012), they may have supported parsing the signal into syllable-size packages, thereby increasing the intelligibility of words and sentences. Thus, it may be that less top-down facilitation was required to render words and sentences intelligible compared to vowels and, consequently, the intelligibility increased as a function of stimulus complexity after priming.

The novelty of the experimental paradigm used here also points to several issues that would need further clarification. First, the contribution of different aspects of speech complexity on intelligibility could further be evaluated by studying whether a similar increase in intelligibility can be achieved with other forms of distorted speech or with syntactically violated sentences and syntactically correct meaningless sentences. Second, it would be interesting to see whether the human brain can dynamically tune the perceptual system to optimally process degraded speech so that the repetition of distorted stimuli in itself increases their intelligibility even without the intervening presentation of their undistorted counterparts. However, it is unlikely that, in this study, perceptual learning occurring during the first presentation of the degraded stimuli would have resulted in such a dramatic increase in intelligibility at the second presentation of the degraded stimuli. This is supported by the results of Hannemann et al. (2007) who found increase in the identification accuracy and changes in the related brain responses only for distorted stimuli that had also been presented in undistorted form, but not for stimuli presented only in distorted form. An interesting question also is how the number of stimuli and the delay between the distorted and undistorted presentations affect intelligibility and behavioral performances. Assuming that the memory system probed with the current paradigm has a capacity limitation, increasing the number of sentences and/or the intervening interval should at some point lead to decreased performance.

Neural correlates of speech intelligibility

In the current study, a number of cortical areas exhibited activation during the N1m time range that could be associated with speech intelligibility: the mean currents elicited by the distorted stimuli upon their second presentation increased on average 34% as a result of the exposure to their undistorted counterparts. In the case of vowel stimulation, these areas included the transverse temporal gyrus (TTG) and the superior temporal gyrus (STG; posterior STG in the left hemisphere; STG and TTG studied within the same ROI), as well as the superior temporal sulcus (STS; posterior STS in the left hemisphere), the posterior insula, the left-hemispheric anterior supramarginal gyrus (SMG) and the left-hemispheric inferior postcentral sulcus (PSC). A corresponding effect was absent for words, and it was restricted to the posterior insula for sentences. This poverty of effects in the case of words and sentences was probably due to these stimuli eliciting overall much weaker responses, resulting in a decreased signal-to-noise ratio (SNR) compared to the case when vowels were used as stimuli.

The more prominent N1m responses for the distorted signals upon their second presentation suggests that speech processing is guided, or at least modified, by auditory long-term memory representations already at very early processing stages, even in conditions where incoming sounds are processed without attentional demands (given the passive recording condition). Tentatively, this may be explained by a filtering mechanism wherein the receptive fields of neurons are reshaped at hierarchically multiple levels to “filter” certain features from noise (Jääskeläinen et al., 2007). Probably, the enhanced N1m reflects sensitivity to speech sounds, as suggested by studies that have documented a stronger N1m amplitude and/or a shorter N1m latency for speech sounds compared to non-speech sounds (for a review, see Salmelin, 2007). Because an enhancement of the N1(m) has also been reported when, for example, listeners are familiar with speech (Ylinen and Huotilainen, 2007), animal (Kirmse et al., 2009) and musical instrument sounds (Pantev et al., 2001), it may alternatively be that the stronger N1m reflects enhanced processing of behaviorally relevant sounds in general.

The robust intelligibility effects observed in the behavioral experiment lacked prominent correlates in terms of neuronal activity in time ranges after the N1m. One reason for this might be that the resolution of MEG may not be suitable for measuring intelligibility-related brain activity when it spreads from the temporal cortex to other brain regions. Indeed, there is evidence that anterior and frontal brain sources may be too distant to be reliably picked up by MEG sensors, especially, if the subjects lean against the back wall of the measurement helmet (Marinkovic et al., 2004). Given that recent fMRI studies have found strong evidence for cortical activity associated with speech intelligibility (for a review, see Peelle et al., 2010), an obvious future extension of this study would be to use a modified version of the current experimental setup in an fMRI experiment.

Neural correlates of sound degradation

Acoustic degradation increased the right-hemispheric N1m amplitude for vowels by 36% on average and delayed bilaterally the latencies of both the N1m (by 6 ms for vowels, 8 ms for words, and 12 ms for sentences) and the P2m response (by 14 ms for vowels and by 40 ms for words and sentences) for all complexity levels. Evidence for cortical activations being sensitive to acoustic features during the N1m and P2m responses was also found in the ROI analysis. During the N1m time range, the mean currents associated with vowel processing were attenuated 47% by stimulus distortion in the left-hemispheric POp. During the P2m time range, speech degradation decreased left-hemispheric cortical activity, on average, by 41% for vowels and 25% for sentences in the TTG and in posterior parts of the STG (studied within the same ROI), the STS and the insula, as well as in the anterior SMG. For words, speech degradation decreased the mean currents by 25% in the TTG, posterior parts of the STG (studied within the same ROI) and the insula. In the right hemisphere, this attenuation effect was present only for vowels and remained restricted to the STS and the anterior SMG where the mean currents were 20% weaker for the distorted vowels.

The right-hemispheric increase of the amplitude of the N1m to the distorted vowels is in line with the results of previous studies using vowels and employing the same distortion method based on amplitude quantization as used here (Liikkanen et al., 2007; Miettinen et al., 2010, 2012). These studies suggested that the amplitude increase is related to a generation of new prominent frequency components by the nonlinear process utilized in the sound degradation. These additional harmonics possibly activate a larger number of neurons involved in the pitch extraction process. However, in the current study, acoustic degradation of words and sentences was not reflected in the N1m and P2m amplitudes, whereas in our previous study (Tiitinen et al., 2012), the amplitudes increased bilaterally for degraded sentences. A possible reason for this is that the responses were stronger in our previous study (amplitudes: N1m: 43–64 fT/cm vs. 20–24 fT/cm; P2m: 22–34 fT/cm vs. 17–18 fT/cm), and therefore also the SNR has likely been higher. Probably, the shorter offset-to-onset interstimulus interval (ISI; 1 s vs. 4 s) used in this study caused stronger adaptation or tuning of the auditory system to the temporal statistical structure of speech attenuating the responses. These issues are discussed more in the Neural correlates of speech complexity section . Another reason for differences may be that the stimulus sets were not the same and thus not acoustically identical. The frequency range of the distorted stimuli also was different in these two studies because in this study the stimuli were additionally downsampled before the USQ procedure, and therefore frequencies higher than 2.2 kHz were filtered out to avoid aliasing. Thus, the responses to the distorted stimuli may have been weaker because smaller number of neurons have been responding to the distorted stimuli with narrower frequency band. Moreover, our previous and current studies estimated the amplitudes and mean currents using different gradiometers and ROIs which has likely affected on the results.

The cortical activity during the sustained field elicited by words showed sensitivity to speech degradation in that both the first and the second presentations of the distorted stimuli resulted in stronger responses compared to the activity elicited by the intact stimuli. In the gradiometer analysis, this effect was observed in the occipitotemporal, parietal, and sensorimotor areas (on average, 28% stronger mean amplitudes for the distorted words), and in the ROI analysis, it was observed in the SMG and STS (on average, 37% stronger mean currents for the distorted words). This increased activity for the distorted words is in accord with our previous study which found the same effect for sentence-induced activity in the auditory cortex as well as in central inferior parietal and posterior superior temporal areas (Tiitinen et al., 2012).

Neural correlates of speech complexity

The N1m amplitudes were on average 54% stronger for vowels than for words and sentences. This result was confirmed by ROI analyses in the TTG and the posterior STG (studied within the same ROI), as well as in the anterior SMG, the posterior STS and the posterior insula, where words and sentences elicited, on average, 32% weaker currents than vowels. Also, the latency of the N1m seemed to reflect stimulus complexity, with vowels resulting in the shortest latency and sentences in the longest (with an average delay of 9 ms). In contrast, while the P2m amplitude was insensitive to speech complexity, the mean current during the P2m time range was, on average, 25% stronger for sentences than for vowels in all of the four studied ROIs covering the TTG, the posterior STG, the inferior SMG and the posterior STS. Also, vowels resulted in a peak latency of the P2m which was some 35 ms earlier than that associated with words and sentences. Further, the mean amplitude of the sustained field was on average 15% stronger when elicited by words than when elicited by sentences in temporal, sensorimotor, occipitotemporal, occipital, and parietal areas (i.e. in all the studied areas but the frontal area).

The modulations of the N1m and P2m associated with stimulus complexity tentatively suggest that continuous speech is processed differently than isolated vowels in the auditory cortex during the first few hundred milliseconds. Specifically, our results point to a possible context effect, whereby the response to a vowel sound (be it isolated or the initial sound of a word or sentence) seems to be modulated by the complexity and/or the sound duration of the preceding stimulus material. A related finding has been made in a recent EEG study by Lanting et al. (2013) who found the N1–P2 response to a probe tone to be reduced as a function of the duration of the preceding adapter tone. However, it is unclear to what extent this kind of adaptation due to stimulus repetition was occurring in the current experiment where the various stimuli in the word and sentence sets, while certainly overlapping in spectral content, were nonetheless not identical to one another. Moreover, in the study by Lanting et al., the amplitude of the P2 response to pure tones was shown to decrease with the duration of the adapter whereas in our study, the mean currents during the P2m time range were, interestingly, stronger for vowel presentation than for sentence presentation. An alternative explanation could be that the modulations of the N1m and P2m reflect inherent tuning of the auditory system to the temporal statistical structure of speech. Indeed, there is evidence of a systematic relationship between the phase of neural signals and the phase of the temporal envelope of speech that might arise from a tendency of neural systems to utilize rhythmic regularities of speech to form predictions about upcoming events (for reviews, see Golumbic et al., 2012; Peelle and Davis, 2012). The divergent characteristics of the N1m and the P2m responses associated with speech complexity may reflect their different neural origins and functional significances, a conclusion which is supported by recent EEG studies (Crowley and Colrain, 2004; Lanting et al., 2013; Ross and Tremblay, 2009; Tremblay et al., 2014). The differences between the transient responses elicited by vowels and to those elicited by words and sentences may also to some extent be related to the acoustic structure of the stimuli. The words in the word set were acoustically identical with the initial words of the sentences in the sentence set. In contrast, the acoustic structure of the isolated vowels slightly deviated from the acoustic structure of the initial vowels of the words and the sentences. First, the isolated vowels were 200 ms in duration whereas the duration of the initial vowels varied (i.e., due to the presence of both single and double vowels). Second, the onsets and offsets of the vowels were smoothed with a shorter (ramp length 5 ms) Hann window than the words and sentences (ramp length 10 ms). However, it is unlikely that this difference in the window length can alone explain the variations of brain activity as a function of stimulus complexity. For example, sentences that were acoustically identical with words also resulted in delayed N1m responses and diminished sustained fields compared to words.

Conclusions

The present findings suggest that the human ability to understand speech even under acoustically compromised conditions relies on memory-related top-down processes correlating the degraded auditory information with long-term memory traces for speech. Already a single exposure to intelligible, undistorted speech stimuli was shown to be sufficient to render their initially unintelligible, acoustically distorted counterparts intelligible, thus reflecting rapid neuroplasticity. Our results demonstrate that this increase in intelligibility depends on the complexity of the speech stimulus: the more complex the stimulus, the easier it is to recognize. This result is in line with the current models of speech perception suggesting that smaller linguistic units are encoded as part of a longer temporal unit, and intelligibility is achieved by integrating information over time. At the neural level, speech intelligibility was reflected by increased brain activity in the auditory cortex and surrounding areas for distorted speech stimuli after the exposure to their intact counterparts at latencies of 130–160 ms. Thus, top-down information would seem to modify the processing of speech signals already at very early cortical stages of speech comprehension.

Q8

Uncited reference

Pantev et al., 1996

Acknowledgments

This research was supported by the Academy of Finland (grant No. 257811).

Conflict of interest

The authors declare no competing financial interests.

References

Ahissar, M., Hochstein, S., 2004. The reverse hierarchy theory of visual perceptual learning. Trends Cogn. Sci. 8, 457–464. http://dx.doi.org/10.1016/j.tics.2004.08.011.

Bodner, G.E., Masson, M.E.J., 2014. Memory recruitment: a backward idea about masked priming. Psychology of Learning and Motivation, 1st ed. Elsevier Inc. http://dx.doi.org/10.1016/B978-0-12-800283-4.00005-8.

Boothroyd, A., Nittrouer, S., 1988. Mathematical treatment of context effects in phoneme and word recognition. J. Acoust. Soc. Am. 84, 101–114. http://dx.doi.org/10.1121/1.396976.

Bowers, J.S., 1996. Different perceptual codes support priming for words and pseudowords: was Morton right all along? J. Exp. Psychol. Learn. Mem. Cogn. 22, 1336–1353. http://dx.doi.org/10.1037/0278-7393.22.6.1336.

Bradlow, A.R., Alexander, J.A., 2007. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. J. Acoust. Soc. Am. 121, 2339-49. http://dx.doi.org/10.1121/1.2642103.

Crowley, K.E., Colrain, I.M., 2004. A review of the evidence for P2 being an independent component process: age, sleep and modality. Clin. Neurophysiol. 115, 732–744. http://dx.doi.org/10.1016/j.clinph.2003.11.021.

Dale, A.M., Liu, A.K., Fischl, B.R., Buckner, R.L., Belliveau, J.W., Lewine, J.D., Halgren, E., 2000. Dynamic statistical parametric mapping. Neuron 26, 55–67. http://dx.doi.org/10.1016/S0896-6273(00)81138-1.

Davis, M.H., Johnsrude, I.S., 2003. Hierarchical processing in spoken language comprehension. J. Neurosci. 23, 3423–3431 (doi:10.1.1.58.2754).

Davis, M.H., Johnsrude, I.S., 2007. Hearing speech sounds: top-down influences on the interface between audition and speech perception. Hear. Res. 229, 132–147. http://dx.doi.org/10.1016/j.heares.2007.01.014.

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage 31, 968–980. http://dx.doi.org/10.1016/j.neuroimage.2006.01.021.

Diamond, A., 1990. The development and neural bases of higher cognitive functions. Introduction. Ann. N. Y. Acad. Sci. 608 (xiii–lvi).

Diamond, R., Rozin, P., 1984. Activation of existing memories in anterograde amnesia. J. Abnorm. Psychol. 93, 98–105. http://dx.doi.org/10.1037/0021-843X.93.1.98.

Dorfman, J., 1994. Sublexical components in implicit memory for novel words. J. Exp. Psychol. Learn. Mem. Cogn. 20, 1108–1125. http://dx.doi.org/10.1037/0278-7393.20.5.1108.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., Elbert, T., 1995. Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. J. Neurosci. 15, 2748-55.

Forster, K., Booker, J., Schachter, Daniel, Davis, L., C., 1990. Masked repetition priming: lexical activation or novel memory trace? Bull. Psychon. Soc. 28, 341–345. http://dx.doi.org/10.3758/BF03334039.

Friederici, A.D., Kotz, S., Scott, S.K., Obleser, J., 2010. Disentangling syntax and intelligibility in auditory language comprehension. Hum. Brain Mapp. 31, 448–457. http://dx.doi.org/10.1002/hbm.20878.

Galambos, R., Makeig, S., Talmachoff, P.J., 1981. A 40-Hz auditory potential recorded from the human scalp. Proc. Natl. Acad. Sci. U. S. A. 78, 2643–2647.

Giraud, A.L., Kell, C., Thierfelder, C., Sterzer, P., Russ, M.O., Preibisch, C., Kleinschmidt, A., 2004. Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. Cereb. Cortex 14, 247–255. http://dx.doi.org/10.1093/cercor/bhg124.

Golumbic, E.M.Z., Poppel, D., Schroeder, C.E., 2012. Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang. 122, 151–161. http://dx.doi.org/10.1016/j.bandl.2011.12.010.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., Hämäläinen, M., 2013. MEG and EEG data analysis with MNE-Python. Front. Neurosci. 7, 267. http://dx.doi.org/10.3389/fnins.2013.00267.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460. http://dx.doi.org/10.1016/j.neuroimage.2013.10.027.

Gray, R.M., 1990. Quantization noise spectra. IEEE Trans. Inf. Theory 36, 1220–1244. http://dx.doi.org/10.1109/18.59924.

Gutschalk, A., Patterson, R.D., Rupp, A., Uppenkamp, S., Scherg, M., 2002. Sustained magnetic fields reveal separate sites for sound level and temporal regularity in human auditory cortex. Neuroimage 15, 207–216. http://dx.doi.org/10.1006/nimg.2001.0949.

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42. http://dx.doi.org/10.1007/BF02512476.

Hannemann, R., Obleser, J., Eulitz, C., 2007. Top-down knowledge supports the retrieval of lexical information from degraded speech. Brain Res. 1153, 134–143. http://dx.doi.org/10.1016/j.brainres.2007.03.069.

Hari, R., Hämäläinen, M., Joutsiniemi, S.L., 1989. Neuromagnetic steady-state responses to auditory stimuli. J. Acoust. Soc. Am. 86, 1033–1039. http://dx.doi.org/10.1121/1.398093.

Hervais-Adelman, A.G., Carlyon, R.P., Johnsrude, I.S., Davis, M.H., 2012. Brain regions recruited for the effortful comprehension of noise-vocoded words. Lang. Cogn. Process. 27, 1145–1166. http://dx.doi.org/10.1080/01690965.2012.662280.

Huotilainen, M., Tiitinen, H., Lavikainen, J., Ilmoniemi, R.J., Pekkonen, E., Sinkkonen, J., Laine, P., Näätänen, R., 1995. Sustained fields of tones and glides reflect tonotopy of the auditory cortex. Neuroreport 6, 841–844. http://dx.doi.org/10.1097/00001756-199504190-00004.

Hyvärinen, A., Oja, E., 2000. Independent component analysis: algorithms and applications. Neural Netw. 13, 411–430. http://dx.doi.org/10.1016/S0893-6080(00)00026-5.

Jääskeläinen, I.P., Ahveninen, J., Belliveau, J.W., Raij, T., Sams, M., 2007. Short-term plasticity in auditory cognition. Trends Neurosci. 30, 653–661. http://dx.doi.org/10.1016/j.tins.2007.09.003.

Kalikow, D.N., Stevens, K.N., Elliott, L.L., 1977. Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. J. Acoust. Soc. Am. 61, 1337–1351. http://dx.doi.org/10.1121/1.381436.

Keceli, S., Inui, K., Okamoto, H., Otsuru, N., Kakigi, R., 2012. Auditory sustained field responses to periodic noise. BMC Neurosci. 13, 7. http://dx.doi.org/10.1186/1471-2202-13-7.

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

Kirmse, U., Jacobsen, T., Schröger, E., 2009. Familiarity affects environmental sound processing outside the focus of attention: an event-related potential study. Clin. Neurophysiol. 120, 887–896. http://dx.doi.org/10.1016/j.clinph.2009.02.159.

Lanting, C.P., Briley, P.M., Sumner, C.J., Krumbholz, K., 2013. Mechanisms of adaptation in human auditory cortex. J. Neurophysiol. 110, 973–983. http://dx.doi.org/10.1152/jn.00547.2012.

Leff, A.P., Schofield, T.M., Stephan, K.E., Crinion, J.T., Friston, K.J., Price, C.J., 2008. The cortical dynamics of intelligible speech. J. Neurosci. 28, 13209–13215. http://dx.doi.org/10.1523/JNEUROSCI.2903-08.2008.

Liikkanen, L., Tiitinen, H., Alku, P., Leino, S., Yrttiaho, S., May, P.J.C., 2007. The right-hemispheric auditory cortex in humans is sensitive to degraded speech sounds. Neuroreport 18, 601–605. http://dx.doi.org/10.1097/WNR.0b013e3280b07bde.

Lin, F.-H., Belliveau, J.W., Dale, A.M., Hämäläinen, M.S., 2006a. Distributed current estimates using cortical orientation constraints. Hum. Brain Mapp. 27, 1–13. http://dx.doi.org/10.1002/hbm.20155.

Lin, F.-H., Witzel, T., Ahlfors, S.P., Stufflebeam, S.M., Belliveau, J.W., Hämäläinen, M.S., 2006b. Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates. Neuroimage 31, 160–171. http://dx.doi.org/10.1016/j.neuroimage.2005.11.054.

Lotto, A.J., Hickok, G.S., Holt, L.L., 2009. Reflections on mirror neurons and speech perception. Trends Cogn. Sci. 13, 110–114. http://dx.doi.org/10.1016/j.tics.2008.11.008.

Mäkelä, A.M., Alku, P., Mäkinen, V., Valtonen, J., May, P., Tiitinen, H., 2002. Human cortical dynamics determined by speech fundamental frequency. Neuroimage 17, 1300–1305. http://dx.doi.org/10.1006/nimg.2002.1279.

Mäkelä, A.M., Alku, P., Mäkinen, V., Tiitinen, H., 2004. Glides in speech fundamental frequency are reflected in the auditory N1m response. Neuroreport 15, 1205–1208. http://dx.doi.org/10.1097/01.wnr.0000126212.56622.38.

Mäkinen, V., 2006. Analysis of the Structure of Time-frequency Information in Electromagnetic Brain Signals. Helsinki University of Technology.

Marinkovic, K., Cox, B., Reid, K., Halgren, E., 2004. Head position in the MEG helmet affects the sensitivity to anterior sources. Neurol. Clin. Neurophysiol. 30, 1–4. http://dx.doi.org/10.1016/j.biotechadv.2011.08.021.Secreted.

Marslen-Wilson, W.D., 1987. Functional parallelism in spoken word-recognition. Cognition 25, 71–102. http://dx.doi.org/10.1016/0010-0277(87)90005-9.

May, P.J.C., Tiitinen, H., 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Psychophysiology 47, 66–122. http://dx.doi.org/10.1111/j.1469-8986.2009.00856.x.

McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. Cogn. Psychol. 18, 1–86. http://dx.doi.org/10.1016/0010-0285(86)90015-0.

McClelland, J.L., Mirman, D., Holt, L.L., 2006. Are there interactive processes in speech perception? Trends Cogn. Sci. 10, 363–369. http://dx.doi.org/10.1016/j.tics.2006.06.007.

Miettinen, I., Tiitinen, H., Alku, P., May, P.J.C., 2010. Sensitivity of the human auditory cortex to acoustic degradation of speech and non-speech sounds. BMC Neurosci. 11, 24. http://dx.doi.org/10.1186/1471-2202-11-24.

Miettinen, I., Alku, P., Salminen, N., May, P.J.C., Tiitinen, H., 2011. Responsiveness of the human auditory cortex to degraded speech sounds: reduction of amplitude resolution vs. additive noise. Brain Res. 1367, 298–309. http://dx.doi.org/10.1016/j.brainres.2010.10.037.

Miettinen, I., Alku, P., Yrttiaho, S., May, P.J.C., Tiitinen, H., 2012. Cortical processing of degraded speech sounds: effects of distortion type and continuity. Neuroimage 60, 1036–1045. http://dx.doi.org/10.1016/j.neuroimage.2012.01.085.

Miller, G.A., Isard, S., 1963. Some perceptual consequences of linguistic rules. J. Verbal Learn. Verbal Behav. 2, 217–228. http://dx.doi.org/10.1016/S0022-5371(63)80087-0.

Morton, J., 1979. Facilitation in word recognition: experiments causing change in the logogen model. Process. Visible Lang. 13, 259–268. http://dx.doi.org/10.1007/978-1-4684-0994-9.

Möttönen, R., Calvert, G.A., Jääskeläinen, I.P., Matthews, P.M., Thesen, T., Tuomainen, J., Sams, M., 2006. Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus. Neuroimage 30, 563–569. http://dx.doi.org/10.1016/j.neuroimage.2005.10.002.

Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. Psychophysiology 24, 375–425. http://dx.doi.org/10.1111/j.1469-8986.1987.tb00311.x.

Obleser, J., Lahiri, A., Eulitz, C., 2004. Magnetic brain response mirrors extraction of phonological features from spoken vowels. J. Cogn. Neurosci. 16, 31–39. http://dx.doi.org/10.1162/089892904322755539.

Obleser, J., Wise, R.J.S., Alex Dresner, M., Scott, S.K., 2007. Functional integration across brain regions improves speech perception under adverse listening conditions. J. Neurosci. 27, 2283–2289. http://dx.doi.org/10.1523/JNEUROSCI.4663-06.2007.

Obleser, J., Eisner, F., Kotz, S.A., 2008. Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. J. Neurosci. 28, 8116–8123. http://dx.doi.org/10.1523/JNEUROSCI.1290-08.2008.

Okada, K., Rong, F., Venezia, J., Matchin, W., Hsieh, I.H., Saberi, K., Serences, J.T., Hickok, G., 2010. Hierarchical organization of human auditory cortex: evidence from acoustic invariance in the response to intelligible speech. Cereb. Cortex 20, 2486–2495. http://dx.doi.org/10.1093/cercor/bhp318.

Okamoto, H., Stracke, H., Bermudez, P., Pantev, C., 2011. Sound processing hierarchy within human auditory cortex. J. Cogn. Neurosci. 23, 1855–1863. http://dx.doi.org/10.1162/jocn.2010.21521.

Pantev, C., Roberts, L.E., Elbert, T., Roβ, B., Wienbruch, C., 1996. Tonotopic organization of the sources of human auditory steady-state responses. Hear. Res. 101, 62–74. http://dx.doi.org/10.1016/S0378-5955(96)00133-5.

Pantev, C., Roberts, L.E., Schulz, M., Engelien, A., Ross, B., 2001. Timbre-specific enhancement of auditory cortical representations in musicians. Neuroreport 12, 169–174. http://dx.doi.org/10.1097/00001756-200101220-00041.

Peelle, J.E., Davis, M.H., 2012. Neural oscillations carry speech rhythm through to comprehension. Front. Psychol. 3, 1–17. http://dx.doi.org/10.3389/fpsyg.2012.00320.

Peelle, J.E., Johnsrude, I.S., Davis, M.H., 2010. Hierarchical processing for speech in human auditory cortex and beyond. Cereb. Cortex 4, 1–3. http://dx.doi.org/10.1093/cercor/bhp318.

Ross, B., Tremblay, K., 2009. Stimulus experience modifies auditory neuromagnetic responses in young and older listeners. Hear. Res. 248, 48–59. http://dx.doi.org/10.1016/j.heares.2008.11.012.

Salmelin, R., 2007. Clinical neurophysiology of language: the MEG approach. Clin. Neurophysiol. 118, 237–254. http://dx.doi.org/10.1016/j.clinph.2006.07.316.

Scott, S.K., Mcgettigan, C., Eisner, F., 2009. A little more conversation, a little less action — candidate roles for motor cortex in speech perception. Nat. Rev. Neurosci. 10, 295–302. http://dx.doi.org/10.1038/nrn2603.A.

Shahin, A.J., Bishop, C.W., Miller, L.M., 2009. Neural mechanisms for illusory filling-in of degraded speech. Neuroimage 44, 1133–1143. http://dx.doi.org/10.1016/j.neuroimage.2008.09.045.

Smiljanic, R., Sladen, D., 2013. Acoustic and semantic enhancements for children with cochlear implants. J. Speech Lang. Hear. Res. 56, 1085-96. http://dx.doi.org/10.1044/1092-4388(2012/12-0097).

Tallon-Baudry, C., Bertrand, O., Delpuech, C., Pernier, J., 1997. Oscillatory gamma-band (30–70 Hz) activity induced by a visual search task in humans. J. Neurosci. 17, 722–734.

Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51, 1759–1768. http://dx.doi.org/10.1088/0031-9155/51/7/008.

Tiitinen, H., Mäkelä, A.M., Mäkinen, V., May, P.J.C., Alku, P., 2005. Disentangling the effects of phonation and articulation: hemispheric asymmetries in the auditory N1m response of the human brain. BMC Neurosci. 6, 62. http://dx.doi.org/10.1186/1471-2202-6-62.

Tiitinen, H., Miettinen, I., Alku, P., May, P.J.C., 2012. Transient and sustained cortical activity elicited by connected speech of varying intelligibility. BMC Neurosci. 13, 157. http://dx.doi.org/10.1186/1471-2202-13-157.

Tremblay, K.L., Ross, B., Inoue, K., McClannahan, K., Collet, G., 2014. Is the auditory evoked P2 response a biomarker of learning? Front. Syst. Neurosci. 8, 28. http://dx.doi.org/10.3389/fnsys.2014.00028.

Tulving, E., Schacter, D.L., 1990. Priming and human memory systems. Science 247, 301–306. http://dx.doi.org/10.1126/science.2296719.

Tulving, E., Schacter, D., Stark, H., 1982. Priming effects in word-fragment completion are independent of recognition memory. J. Exp. Psychol. Learn. Mem. Cogn. http://dx.doi.org/10.1037/0278-7393.8.4.336.

Valentini-Botinhao, C., Wester, M., 2014. Using linguistic predictability and the Lombard Q9 effect to increase the intelligibility of synthetic speech in noise. INTERSPEECH 2014, pp. 2063–2067 (Signapore).

Wild, C.J., Yusuf, A., Wilson, D.E., Peelle, J.E., Davis, M.H., Johnsrude, I.S., 2012. Effortful listening: the processing of degraded speech depends critically on attention. J. Neurosci. 32, 14010–14021. http://dx.doi.org/10.1523/JNEUROSCI.1528-12.2012.

Ylinen, S., Huotilainen, M., 2007. Is there a direct neural correlate for memory-trace formation in audition? Neuroreport 18, 1281–1284. http://dx.doi.org/10.1097/WNR.0b013e32826fb38a.

Yrttiaho, S., Alku, P., May, P.J.C., Tiitinen, H., 2009. Representation of the vocal roughness of aperiodic speech sounds in the auditory cortex. J. Acoust. Soc. Am. 125, 3177–3185. http://dx.doi.org/10.1121/1.3097471.