Brain & Language 164 (2017) 1–8

Contents lists available at ScienceDirect

Brain & Language

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

Effects of acoustic periodicity, intelligibility, and pre-stimulus alpha power on the event-related potentials in response to speech

Kurt Steinmetzger *, Stuart Rosen

Speech, Hearing and Phonetic Sciences, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF, United Kingdom

ARTICLE INFO

Article history:
Received 2 May 2016
Revised 4 August 2016
Accepted 19 September 2016
Available online 28 September 2016

Keywords:
EEG
Speech
Intelligibility
Periodicity
CNV
Pre-stimulus alpha power

ABSTRACT

Magneto- and electroencephalographic (M/EEG) signals in response to acoustically degraded speech have been examined by several recent studies. Unambiguously interpreting the results is complicated by the fact that speech signal manipulations affect acoustics and intelligibility alike. In the current EEG study, the acoustic properties of the stimuli were altered and the trials were sorted according to the correctness of the listeners' spoken responses to separate out these two factors. Firstly, more periodicity (i.e. voicing) rendered the event-related potentials (ERPs) more negative during the first second after sentence onset, indicating a greater cortical sensitivity to auditory input with a pitch. Secondly, we observed a larger contingent negative variation (CNV) during sentence presentation when the subjects could subsequently repeat more words correctly. Additionally, slow alpha power (7–10 Hz) before sentences with the least correctly repeated words was increased, which may indicate that subjects have not been focussed on the upcoming task.

1. Introduction

Acoustically degraded noise-vocoded speech has been used extensively to investigate the neural correlates of speech intelligibility in both magneto- and electroencephalographic (M/EEG) studies (e.g. Becker, Pefkou, Michel, & Hervais-Adelman, 2013; Ding, Chatterjee, & Simon, 2014; Obleser & Weisz, 2012; Peelle, Gross, & Davis, 2013) and imaging work (e.g. Davis & Johnsrude, 2003; Evans et al., 2014; Scott, Blank, Rosen, & Wise, 2000). Noise-vocoding has proven a very useful tool because it allows the parametric reduction of the intelligibility of speech signals by reducing the number of channels in the analysis/synthesis process. However, this signal manipulation alters the acoustic properties of the stimuli as well as their intelligibility, and these two factors have so far not been considered independently.

Furthermore, while the reduction in intelligibility can mainly be attributed to the lowered spectral resolution of the vocoded speech signals, other acoustic properties are affected by the signal processing as well. Most notably, due to the use of a broadband noise as sound source, noise-vocoded speech is completely aperiodic (i.e. unvoiced), making it sound like an intense version of a whisper. In natural speech, on the other hand, voiced and unvoiced segments alternate. Importantly, only voiced speech possesses a pitch. Previous studies that have investigated pitch perception reliably found increased neural responses for stimuli that possess a pitch, when compared to a spectrally matched control condition (e.g. Griffiths et al., 2010; Norman-Haignere, Kanwisher, & McDermott, 2013) or a broadband noise (Chait, Poeppel, & Simon, 2006). In particular, studies analysing MEG signals in the time domain (Chait et al., 2006; Gutschalk, Patterson, Scherg, Uppenkamp, & Rupp, 2004) have shown that following a transient pitch onset response peaking after around 150 ms, a sustained neural response can be observed for several hundred milliseconds. Thus, it appears likely that the neural response elicited by noise-vocoded speech is *per se* attenuated due to the absence of voicing.

In order to address these issues, we have used a vocoding technique that allows the choice between a periodic (voiced) or an aperiodic (unvoiced) source excitation. This technique was used to synthesise speech that is either completely unvoiced (i.e. noise-vocoded, henceforth referred to as the *aperiodic* condition), preserves the natural mix of voiced and voicelessness (henceforth the *mixed* condition; Dudley, 1939), or is completely voiced (henceforth the *periodic* condition). Previous behavioural work (Steinmetzger & Rosen, 2015) has shown that the intelligibility of the aperiodic and mixed conditions is very similar, while the unnatural-sounding fully periodic condition was found to be considerably less intelligible. In order to analyse effects of acoustic periodicity while controlling for differences in intelligibility, the individual trials in the current study were sorted according to

* Corresponding author.
E-mail address: kurt.steinmetzger.12@ucl.ac.uk (K. Steinmetzger).

http://dx.doi.org/10.1016/j.bandl.2016.09.008
0093-934X/

the listeners' spoken responses (i.e. the number of correctly repeated key words) obtained after every sentence, and only fully intelligible trials were considered. In summary, the first hypothesis was that speech with more periodicity would lead to more negative event-related potentials (ERPs), reflecting the increased neural sensitivity to auditory input that possess a pitch. This effect was expected to be observed during an early time window following sentence onset, including the auditory evoked potentials (AEPs) and the acoustic change complex (ACC; Pratt, 2011).

Sorting the individual trials according to the behavioural responses was also intended to enable the separate analysis of more or less intelligible trials in the periodic condition. This second analysis additionally included spectrally rotated speech, a completely unintelligible non-speech analogue that has been used in a number of the previously mentioned studies (Becker et al., 2013; Peelle et al., 2013; Scott et al., 2000), as a baseline condition (henceforth the *rotated* condition). In contrast to several recent M/EEG studies that have investigated the perception of noise-vocoded (Becker et al., 2013; Obleser & Weisz, 2012; Obleser, Wöstmann, Hellbernd, Wilsch, & Maess, 2012) and unprocessed speech (e.g. Kerlin, Shahin, & Miller, 2010; Müller & Weisz, 2012; Wilsch, Henry, Herrmann, Maess, & Obleser, 2015) by analysing neural activity in the frequency domain, the current study focusses on time-domain responses. Few studies to date have investigated ERPs in response to degraded speech (for exceptions see Becker et al., 2013; Obleser & Kotz, 2011; Wöstmann, Schröger, & Obleser, 2015) and it is hence not well understood how they are affected by both the acoustic characteristics and the intelligibility of the speech signals, particularly over the course of whole sentences.

Based on the notion that slow cortical potentials reflect the degree of cortical excitability (Birbaumer, Elbert, Canavan, & Rockstroh, 1990; He & Raichle, 2009), it was hypothesised that ERP amplitudes over the course of the sentences would be larger in response to more intelligible speech. More specifically, slow negative potentials with an anterior scalp distribution have been associated with both working memory load (e.g. Guimond et al., 2011; Lefebvre et al., 2013) and increased attention (e.g. Teder-Sälejärvi, Münte, Sperlich, & Hillyard, 1999; Woods, Alho, & Algazi, 1994) in auditory tasks. A typical slow negative potential is the contingent negative variation (CNV), which emerges in between a warning stimulus and a task-relevant second stimulus, and is larger when subjects expect and prepare to respond to the latter stimulus (McCallum & Walter, 1968; Tecce & Scheff, 1969). Importantly, the second stimulus may also be a response to the first stimulus (Birbaumer et al., 1990; Kononowicz & Penney, 2016), and hence the design of the current experiment, in which subjects are supposed to verbally repeat the stimulus sentence, fits into the CNV framework too.

In order to further investigate differences between intelligible and unintelligible trials, we additionally analysed the amount of alpha power in the silent baseline interval preceding the stimulus sentences. Decreased alpha power in the pre-stimulus window has been shown to be a predictor of successful target identification in studies using low-level visual and somatosensory stimuli (e.g. Hanslmayr et al., 2007; Romei, Gross, & Thut, 2010; Schubert, Haufe, Blankenburg, Villringer, & Curio, 2009; Van Dijk, Schoffelen, Oostenveld, & Jensen, 2008). Strauß, Henry, Scharinger, and Obleser (2015) have recently also reported alpha phase differences before correctly and incorrectly perceived words in a lexical decision task, but no study to date has reported alpha power differences in the baseline window using speech materials presented auditorily. As reviewed by Klimesch (1999, see also Klimesch, Doppelmayr, Russegger, Pachinger, & Schwaiger, 1998), slower alpha frequencies (~7–10 Hz) in particular have been associated with alertness and expectancy, and may thus serve as a measure of the attentional state in the period before sentence onset. We thus additionally hypothesised to observe enhanced slow alpha power, indicating that subjects have not been fully focussed on the upcoming task, before sentences that would turn out to be unintelligible to them.

2. Methods

2.1. Participants

Eighteen normal-hearing right-handed subjects (8 females, mean age = 21.6 years, SD = 2.3 years) took part in the study. All participants were native speakers of British English and had audiometric thresholds of less than 20 dB Hearing Level at octave frequencies from 125 to 8000 Hz. All subjects gave written consent and the study was approved by the UCL ethics committee.

2.2. Stimuli

The stimulus materials used in this experiment were recordings of the IEEE sentences (Rothauser et al., 1969) spoken by an adult male Southern British English talker with a mean F0 of 121.5 Hz that were cut at zero-crossings right before sentence onset and normalised to a common root-mean-square (RMS) level. The IEEE sentence corpus consists of 72 lists with 10 sentences each and is characterized by similar phonetic content and difficulty across lists, as well as an overall low semantic predictability (e.g. *The birch canoe slid on the smooth planks.*). The individual lists are thus supposed to be equally intelligible. Every sentence contains five key words.

All stimulus materials were processed prior to the experiment using a channel vocoder implemented in MATLAB (Mathworks, Natick, MA). For all three vocoding conditions (aperiodic, mixed, and periodic) the original recordings of the IEEE sentences were first band-pass filtered into eight bands using zero phase-shift sixth-order Butterworth filters. The filter spacing was based on equal basilar membrane distance (Greenwood, 1990) across a frequency range of 0.1 to 8 kHz (upper filter cut-offs in Hz: 242, 460, 794, 1307, 2094, 3302, 5155, 8000; filter centre frequencies in Hz: 163, 339, 609, 1023, 1658, 2633, 4130, 6426). The output of each filter was full-wave rectified and low-pass filtered at 30 Hz (zero phase-shift fourth-order Butterworth) to extract the amplitude envelope. The low cut-off value was chosen in order to ensure that no temporal periodicity cues were present in the aperiodic condition.

In order to synthesise aperiodic speech, the envelope of each individual band was multiplied with a broadband noise carrier. In the mixed condition, the envelope of each band was also multiplied with a broadband noise source, but only in time windows where the original speech was unvoiced. Sections that were voiced in the original recordings were synthesised by multiplying the envelopes with a pulse train following the natural F0 contour. The individual pulses had a duration of one sample point, i.e. about 23 μs at a sampling rate of 44.1 kHz. The F0 contours of the original sentences were extracted using ProsodyPro version 4.3 (Xu, 2013) implemented in PRAAT (Boersma & Weenink, 2013), with the F0 extraction sampling rate set to 100 Hz. The resulting F0 contours were corrected manually where necessary and then used to determine the distance between the individual pulses of the pulse train sources. Based on the original intermittent F0 contours, we also produced artificial continuous F0 contours by interpolation through unvoiced sections and periods of silence. These continuous F0 contours were used to produce the pulse train sources for the periodic condition.