Neuropsychologia 95 (2017) 173–181

Contents lists available at ScienceDirect

Neuropsychologia

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

Effects of acoustic periodicity and intelligibility on the neural oscillations in response to speech

Kurt Steinmetzger*, Stuart Rosen

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

ARTICLE INFO

Keywords:
EEG
Speech
Delta
Theta
Intelligibility
Periodicity

ABSTRACT

Although several studies have investigated neural oscillations in response to acoustically degraded speech, it is still a matter of debate which neural frequencies reflect speech intelligibility. Part of the problem is that effects of acoustics and intelligibility have so far not been considered independently. In the current electroencephalography (EEG) study the amount of acoustic periodicity (i.e. the amount of time the stimulus sentences were voiced) was manipulated, while using the listeners' spoken responses to control for differences in intelligibility. Firstly, the total EEG power changes in response to completely aperiodic (noise-vocoded) speech and speech with a natural mix of periodicity and aperiodicity were almost identical, while an increase in theta power (5–6.3 Hz) and a trend for less beta power (11–18 Hz) were observed in response to completely periodic speech. These two effects are taken to indicate an information processing conflict caused by the unnatural acoustic properties of the stimuli, and that the subjects may have internally rehearsed the sentences as a result of this. Secondly, we separately investigated effects of intelligibility by sorting the trials in the periodic condition according to the listeners' spoken responses. The comparison of intelligible and largely unintelligible trials revealed that the total EEG power in the delta band (1.7–2.7 Hz) was markedly increased during the second half of the intelligible trials, which suggests that delta oscillations are an indicator of successful speech understanding.

1. Introduction

In order to shed light on the underlying neural mechanisms and cognitive processes involved when attempting to understand spoken speech, a growing number of magneto- and electroencephalographic (M/EEG) studies focus on the time-frequency properties of the neural signals rather than traditional waveform analyses (for reviews see Giraud and Poeppel, 2012; Weisz and Obleser, 2014). The current experiment adds to existing knowledge by investigating effects of acoustics and intelligibility separately, two factors that usually vary together when speech signals are acoustically manipulated. Specifically, we manipulated the amount of acoustic periodicity, while controlling for differences in intelligibility, and vice versa. In the context of speech, periodicity denotes that a sound is produced by the periodic vibrations of the vocal folds, resulting in voiced speech with a pitch corresponding to the vibration rate. Unvoiced speech sounds, in contrast, emanate from constrictions in the vocal tract and have aperiodic fluctuations in energy, leading to a noisy sound quality and the absence of a pitch.

A popular speech processing technique that has been used in the neurosciences (e.g. Davis and Johnsrude, 2003; Obleser and Weisz, 2012; Peelle et al., 2013; Scott et al., 2000) as well as in psychoacoustic studies concerned with the simulation of cochlear implants (e.g. Qin and Oxenham, 2003; Schoof et al., 2013; Shannon et al., 1995), is noise-vocoding (henceforth referred to as the *aperiodic* condition). By filtering the unprocessed input speech into a specified number of frequency bands, it allows the spectral resolution of the synthesised output speech to be varied in a controlled manner, a feature that is closely related to speech intelligibility. At the same time, using noise as source results in a loss of the natural mix of voiced and voicelessness, and consequently also any voice pitch information, making it resemble whispered speech.

Nevertheless, our previous behavioural work (Steinmetzger and Rosen, 2015) has shown that preserving periodicity information in a vocoder (henceforth the *mixed* condition) does not lead to improved intelligibility rates. This suggests that periodicity information, despite its salience, is a redundant cue, at least in non-tonal languages and quiet listening conditions. The first question the current study addresses, is thus whether EEG time-frequency responses are similarly unaffected by the absence of periodicity.

To enable a more comprehensive investigation of the effects of periodicity, we included a third processing condition in which the same

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

http://dx.doi.org/10.1016/j.neuropsychologia.2016.12.003
Received 14 October 2015; Received in revised form 7 September 2016; Accepted 5 December 2016
Available online 07 December 2016
0028-3932/

speech materials were synthesised with a completely periodic source (henceforth the *periodic* condition). Acoustically this condition is in fact closer to natural speech (which is voiced about 50% of the time – Dellwo et al., 2007; Fourcin, 2010), than aperiodic noise-vocoded speech. However, because natural speech does not contain periodic sounds with much energy in the frequency region above 4 kHz, it sounds very unnatural. Additionally, periodicity is such a salient cue that it obscures weaker cues such as intensity differences, thereby making the information transmitted contradictory. For unvoiced fricatives like /s/ and /ʃ/, for example, aperiodic high-frequency energy is missing and replaced by periodic energy, which makes it difficult to identify these sounds. Consequently, periodic speech has substantially lower intelligibility rates than the other two conditions (Ardoint et al., 2014; Steinmetzger and Rosen, 2015).

In order to control for this expected difference in intelligibility, the single trials were sorted according to the spoken responses of the participants. This approach was also chosen to enable a direct comparison of intelligible and unintelligible trials in the periodic condition. Consequently, the current study also provides the opportunity to investigate how the EEG time-frequency responses are affected by the intelligibility of the speech materials after controlling for systematic acoustic differences. This approach is akin to studies generating a pop-out effect by presenting the same stimulus materials twice, first without any additional information and then again after providing a written transcript (Sohoglu et al., 2012) or the unprocessed recording (Millman et al., 2015), but avoids any predictive top-down processing.

In addition to the acoustically similar unintelligible trials, the current study included completely unintelligible spectrally rotated speech as a second control condition (henceforth the *rotated* condition). Rotated speech has a similar spectro-temporal complexity as unrotated speech and has been used in several of the studies mentioned above (e.g. Becker et al., 2013; Peelle et al., 2013; Scott et al., 2000). Yet, apart from not being a precise acoustic match, the obvious meaninglessness casts doubt on whether it is indeed an adequate control condition. The design of the current experiment thus also serves to directly compare these two control condition types.

Importantly, in contrast to the event-related potentials (ERPs), the EEG time-frequency analyses in the current paper were not assumed to be affected by the perception of a voice pitch. Direct cortical recordings and MEG experiments have shown that the presence of a pitch coincides with increased high gamma power (> 80 Hz; e.g. Griffiths et al., 2010; Sedley et al., 2012). However, due to potential muscle artefacts and the low signal strength when recorded with cortical EEG, we did not include these frequencies in the current analysis. Moreover, functional magnetic resonance imaging (fMRI) signals, which fluctuate at rates of less than 0.5 Hz (He and Raichle, 2009), have been shown to be larger for signals with a pitch (e.g. Norman-Haignere et al., 2013; Patterson et al., 2002). Yet, frequencies below 0.5 Hz similarly lie outside the possible frequency range of EEG time-frequency analyses, because they would require excessively long baseline and stimulus windows.

Based on the results of Strauß et al. (2014a), who have recently reported increased theta activity (here 3–7 Hz) in a left-lateralised fronto-temporal network for so-called ambiguous pseudo-words in an auditory word recognition task, we expected increased theta power in the periodic condition. The pseudo-words used by Strauß et al. (2014a) were characterised by a wrong core vowel, making them resemble the periodic condition in the current study to some extent. Theta oscillations have also been associated with the storage of sequentially presented verbal information in working memory and the phonological loop (Roux and Uhlhaas, 2014). Based on this idea, Strauß et al. (2014a) suggested that subjects may have internally rehearsed the unusual pseudo-words in order to classify them as words or non-words. More generally, this effect was taken to indicate an information processing conflict (Botvinick et al., 2001; Botvinick et al., 2004), although studies eliciting response conflicts in non-speech tasks, for example by using the Stroop paradigm, have usually reported mid-frontal theta power increases (Cohen and Donner, 2013; Hanslmayr et al., 2008).

A recent theoretical approach has linked increased alpha power (~7–13 Hz) to the selective inhibition of brain areas that are not currently task relevant (Jensen and Mazaheri, 2010). Applied to speech perception, it has been proposed that alpha oscillations might be actively enhanced in order to cope with a demanding task, particularly listening to speech in the presence of background noise (Obleser et al., 2012; Strauß et al., 2014b; Wöstmann et al., 2015). For words presented in quiet listening conditions, on the other hand, alpha activity was found to be increasingly suppressed with higher intelligibility levels (Becker et al., 2013; Obleser and Weisz, 2012). However, in these studies the intelligibility of the mostly noise-vocoded stimuli varied along with their acoustic properties (i.e. the number of frequency bands in the vocoder) and hence also the subjective listening effort, which is similarly thought to depend on the degree of acoustic degradation (Obleser and Weisz, 2012; Wöstmann et al., 2015). Sorting the trials in the periodic condition according to the spoken behavioural responses provided the opportunity to test whether there is indeed a direct relation between alpha suppression and speech intelligibility.

2. Material and 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 HL at frequencies between 125 and 8000 Hz. All subjects gave written consent and the study was approved by the UCL research 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 characterised by similar phonetic content and difficulty across the lists, as well as an overall low semantic predictability. Every sentence contains five keywords (nouns, verbs, or adjectives; e.g. <u>Say</u> it <u>slowly</u> but <u>make</u> it <u>ring</u> <u>clear</u>.).

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 0.1–8 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 white noise carrier. In the mixed condition, the envelope of each band was also multiplied with a broadband white noise, but only in time windows where the original speech was unvoiced. Sections that were voiced in the original