NeuroImage 152 (2017) 628–638

Contents lists available at ScienceDirect

NeuroImage

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

The right hemisphere is highlighted in connected natural speech production and perception

Anna Maria Alexandrou[a,b,*], Timo Saarinen[a,b], Sasu Mäkelä[a], Jan Kujala[a], Riitta Salmelin[a,b]

[a] Department of Neuroscience and Biomedical Engineering, Aalto University, P.O. Box 12200, FI-00076 AALTO, Espoo, Finland
[b] Aalto NeuroImaging, Aalto University, Espoo, Finland

ARTICLE INFO

Keywords:
Natural speech
Speech production
Speech perception
Right temporo-parietal junction
Magnetoencephalography
Functional neuroimaging

ABSTRACT

Current understanding of the cortical mechanisms of speech perception and production stems mostly from studies that focus on single words or sentences. However, it has been suggested that processing of real-life connected speech may rely on additional cortical mechanisms. In the present study, we examined the neural substrates of natural speech production and perception with magnetoencephalography by modulating three central features related to speech: amount of linguistic content, speaking rate and social relevance. The amount of linguistic content was modulated by contrasting natural speech production and perception to speech-like non-linguistic tasks. Meaningful speech was produced and perceived at three speaking rates: normal, slow and fast. Social relevance was probed by having participants attend to speech produced by themselves and an unknown person. These speech-related features were each associated with distinct spatiospectral modulation patterns that involved cortical regions in both hemispheres. Natural speech processing markedly engaged the right hemisphere in addition to the left. In particular, the right temporo-parietal junction, previously linked to attentional processes and social cognition, was highlighted in the task modulations. The present findings suggest that its functional role extends to active generation and perception of meaningful, socially relevant speech.

Introduction

The current view of cortical speech processing, obtained both through evidence from patients (e.g. Dronkers et al., 2004; Mirman et al., 2015) and healthy individuals (for a review, see e.g. Indefrey and Levelt, 2004) and reflected in neurocognitive models (Hickok and Poeppel, 2007; Hickok, 2012), is primarily based on studies of the production or perception of single words or sentences. However, recent reports suggest that real-life speech production and perception may rely on additional cortical substrates (Hultén et al., 2014; Koskinen et al., 2013; Silbert et al., 2014). For instance, hemodynamic evidence suggests that linguistic aspects of continuous speech production are not predominantly processed in the left hemisphere, which traditionally has been associated with linguistic processing (e.g. Binder et al., 2009; Blank et al., 2002), but instead engage cortical areas bilaterally (Silbert et al., 2014).

These differences in cortical processing appear as a logical consequence of the inherent qualities of natural connected speech which, unlike decontextualized linguistic stimuli, is a continuous stream of meaningful utterances that features a salient temporal structure, also referred to as rhythm (Alexandrou et al., 2016). Speech rhythm is tied to habitual speaking rates which manifest as typical production frequencies of words (2–3 Hz) and syllables (4–5 Hz) (Alexandrou et al., 2016; Levelt et al., 1999; Poeppel et al., 2008). Speech rhythm is also physically quantifiable through spectral analysis of muscular and acoustic signals (Alexandrou et al., 2016). The notable inter-individual constancy of word and syllable production frequencies supports the role of spoken language in communication (Cummins and Port, 1998; Kohler, 2009). The regular occurrence of linguistic units provides a frame onto which a linguistic message featuring semantic, syntactic and phonological information is embedded (Port, 2003). Hence, natural connected speech is not merely a quasi-periodic sequence of isolated words and sentences, but a complex entity which combines linguistic, cognitive and social processes (e.g. Flavell, 2000). As such, real-life speech represents a fundamental human behaviour which is a vital aspect of social interaction across cultures and languages (Flavell, 1968; Levinson, 2016). One's own speech must be differentiated from the socially more relevant utterances of an interlocutor. This is reflected, for instance, in suppression of auditory cortical activity during one's own speech production (Houde et al., 2002;

* Corresponding author at: Department of Neuroscience and Biomedical Engineering, Aalto University, P.O. Box 12200, FI-00076 AALTO, Finland.
E-mail addresses: anna.alexandrou@aalto.fi (A.M. Alexandrou), timo.p.saarinen@aalto.fi (T. Saarinen), sasu.makela@aalto.fi (S. Mäkelä), jan.kujala@aalto.fi (J. Kujala), riitta.salmelin@aalto.fi (R. Salmelin).

http://dx.doi.org/10.1016/j.neuroimage.2017.03.006
Received 12 October 2016; Accepted 3 March 2017
Available online 06 March 2017

Martikainen et al., 2005). The present study aims to provide a data-driven functional characterization of the neural correlates underlying natural speech production and perception based on modulations in three fundamental, complementary speech-related features: the amount of linguistic content present in an utterance, speaking rate and social relevance.

We use magnetoencephalography (MEG) to measure task-dependent changes of neural engagement, quantified as modulation of neurophysiological spatiospectral patterns. Cortical signals recorded with MEG typically contain detectable frequencies up to ~100 Hz. Some frequency ranges have been linked to aspects of speech processing. For instance, the theta-band (4–7 Hz) spontaneous oscillatory neural activity that seems to align with the syllabic rate (~5 Hz) is thought to support cortical processing of spoken language (for a review, see Peelle and Davis (2012)). Comprehension may be compromised when the syllabic rate is artificially increased to exceed the upper frequency limit of theta-band oscillations (Ahissar et al., 2001). Speech processing has further been associated with gamma-band oscillatory (35–45 Hz; e.g. Giraud et al., 2007; Morillon et al., 2012) as well as more arrhythmic broadband activity (> 60 Hz; e.g. Towle et al., 2008). In this study, neural engagement is quantified as task-related local modulations of MEG signal power in different frequency bands, across the cortex. Such modulations of band-limited cortical signal power are observed as a consequence of cognitive activity (Wang, 2010) and have been shown to reflect functional processes underlying speech processing (e.g. Ahissar et al., 2001; Obleser and Weisz, 2012).

Here, speech production and perception were examined in the same group of individuals using naturalistic experimental tasks. Experimental variations were induced to the pattern with which linguistic information is produced and perceived (speaking rate) and, at the normal speaking rate, to the nature of the produced and perceived linguistic information (amount of linguistic content, social relevance). Speech production and speech perception engage the neurocognitive system in markedly different ways: speech production consists of organizing complex motor output to convey a linguistic message, whereas during speech perception a linguistic message is extracted from an incoming acoustic signal. Therefore, the modulations of linguistic content were implemented separately for each speech modality. In speech production, the non-linguistic control task was an easily adoptable oromotor task consisting of syllable repetition (/pa/). Syllable repetition represents a rudimentary form of speech (cf. babbling in infants; Davis and MacNeilage, 2000) and habitually occurs at ~2 Hz (Ruspantini et al., 2012), which coincides with the spontaneous word production frequency during natural speech production (Alexandrou et al., 2016). In speech perception, the non-linguistic control task was attentive listening to amplitude-modulated white noise (e.g., Belin et al., 2000) that matched both the spectral content and amplitude envelope (and thus the acoustic and rhythmic structure) of natural speech. Speaking rate variations consisted of producing and perceiving meaningful speech at three distinct speaking rates: slow, normal and fast. Speaking rate may be modified on demand so that it is substantially slower or faster than the habitual rate (Grosjean and Lane, 1976); these modifications alter the temporal features of an utterance (Smith et al., 1995). Variations in speaking rate were thus carried out to assess how complex linguistic content is encoded and decoded into words and syllables. Finally, manipulation of social relevance focused on speech ownership, examined by contrasting the perception of one's own normal-rate speech to the normal-rate speech of an unknown speaker, thus addressing how socially salient speech is discriminated from less significant auditory input.

The present study aims to contribute to an emerging, more integrative view of cortical speech processing, obtained through more ecologically valid experimental paradigms. The changes in cortical signal power resulting from modulations in the three central speech-related features considered here would, firstly, enable a spectral and spatial functional characterization of the cortical correlates of natural, real-life speech processing, and secondly, the assessment of a potential overlap between natural speech production and perception. Therefore, the data-driven approach employed here provides an insight into a broad spectrum of natural speech processing and, ultimately, the obtained results may also serve as a basis for generating neurally-grounded hypotheses for future studies.

Methods

Participants

Twenty healthy, right-handed, native Finnish-speaking adults (11 females; 9 males; mean age 24.5, range 19–35 years) with normal hearing participated in the study. All participants gave their informed written consent before taking part in the experiment, in agreement with a prior approval of the Aalto University Ethics Committee.

Experimental design

All participants took part in a speech production and a speech perception experiment conducted, in this order, as separate measurement sessions at least one week apart. An overview of the experimental paradigm is presented in Fig. 1A.

Speech production experiment

As the main experimental condition, the participants were instructed to produce spontaneous speech at three distinct speaking rates (natural/normal, slow or fast) (example waveforms in Fig. 1B, top) prompted by questions (in Finnish) derived from the themes own life, preferences, people, culture/traditions, society/politics and general knowledge (see Alexandrou et al. (2016)). The prompts were quite general (e.g., What kind of hobbies do you have or have had during your life? Describe a traditional Christmas holiday. How does the Finnish school system operate?) and considered easy to respond to by all participants and a separate group of 10 individuals. The prompts were only intended to help the participants fluently verbalize their own thoughts. The aim was to successfully modulate speaking rate and to produce fluent, uninterrupted speech at each rate. To avoid repetition and learning effects, each thematic question was presented only once during the experiment. Regarding the slow rate, participants were asked to aim for 50% of their normal speaking rate. For the fast rate, they were instructed to produce fluent, continuous speech at the highest speaking rate possible, however, without rendering the speech unintelligible or committing articulatory errors. After a training phase, during the actual MEG measurement, participants varied their speaking rate without the aid of any external pacing device. Based on transcription analysis, participants were able to produce fluent, coherent speech without notable pauses or excessive use of filler words at all speaking rates. Mean syllable production frequencies were 4.8 Hz for normal rate, 2.1 Hz for slow rate (44% of normal) and 6.3 Hz for fast rate (123% of normal) (for more details, see Alexandrou et al. (2016)). Participants were instructed to aim for small mouth movements in order to reduce contamination of the recorded MEG signals by muscle artifacts.

A single speech production block consisted of a recorded thematic question spoken by a male speaker (duration 3–9 s; mean ± SD 5.6 ± 1.3 s), a 1-s delay before response onset, a 40-s response period, and a 2.5-s rest period between blocks. A signal tone (50-ms, 1-kHz tone) indicated the beginning of a block, and another signal tone (50-ms, 75-Hz tone) signified the beginning and end of the response period. All sounds were presented via panel loudspeakers.

Repeated production of the syllable /pa/, performed at normal rate, served as a control condition (Ruspantini et al., 2012). A /pa/ repetition block consisted of a 40-s repetition period, with a tone signal (50-ms, 75-Hz tone) indicating the beginning and end of the period. Repetition blocks were separated by 10 s of rest, to approx-