



This is an electronic reprint of the original article. This reprint may differ from the original in pagination and typographic detail.

Alexandrou, Anna Maria; Saarinen, Timo; Mäkelä, Sasu; Kujala, Jan; Salmelin, Riitta The right hemisphere is highlighted in connected natural speech production and perception

Published in: NeuroImage

DOI: 10.1016/j.neuroimage.2017.03.006

Published: 15/05/2017

Document Version Publisher's PDF, also known as Version of record

Published under the following license: CC BY-NC-ND

Please cite the original version:

Alexandrou, A. M., Saarinen, T., Mäkelä, S., Kujala, J., & Salmelin, R. (2017). The right hemisphere is highlighted in connected natural speech production and perception. *NeuroImage*, *152*, 628-638. https://doi.org/10.1016/j.neuroimage.2017.03.006

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

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



^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

A R T I C L E I N F O

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 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 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

1053-8119/ © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).





CrossMark

Martikainen et al., 2005). The present study aims to provide a datadriven 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 of individuals using naturalistic experimental tasks. group 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 (Grosiean 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 speechrelated 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 neurallygrounded 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 \pm SD 5.6 \pm 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-



Fig. 1. Summary of the experimental paradigm. *A*, Flowcharts demonstrating the structure of the speech production (left) and the speech perception (right) experiments and the stimuli used in each experiment. See Methods for a detailed description. *B*, Examples of waveforms of slow-rate, normal-rate and fast-rate speech generated during the speech production experiment (top) and of speech and noise stimuli used in the speech perception experiment (bottom). Normalized amplitude (in arbitrary units; y-axis) is plotted against time (in seconds; x-axis). Each plot displays a 5-s chunk of data taken from a 40-s auditory file.

imate the timing of the speech conditions.

Ten participants were randomly chosen to additionally perform normal-rate articulated speech. This task consisted of articulating responses to thematic questions (different from those presented in the speech production condition) without vocalization. This articulation-only task was applied to evaluate whether an observed cortical effect was due to auditory feedback during natural speech production. The temporal structure of the task blocks for this condition was the same as for the speech production blocks.

Speech perception experiment

All twenty participants listened to the speech of an unfamiliar male speaker at normal, slow and fast speaking rates. The speaker responded to the same thematic questions as those presented to the participants in the speech production experiment (example waveform in Fig. 1B, bottom left). The speaker carried out speaking rate modulations following the same training procedure as the participants, resulting in mean syllable production frequencies of 4.6 Hz for normal rate, 2.6 Hz for slow rate (55% of normal) and 6.8 Hz for fast rate (157% of normal). Additionally, participants listened to their own responses to the thematic questions recorded during the speech production experiment at normal, slow and fast speaking rates.

As a control condition, participants listened to unintelligible normal-rate rhythmic noise stimuli (example waveform in Fig. 1B, bottom right). These consisted of amplitude-modulated white noise with the same amplitude envelope and spectral content as the normalrate speech stimuli of the unknown speaker. The temporal structure of speech perception blocks was the same as that of speech production blocks and the temporal structure of noise perception blocks was the same as that of /pa/ repetition blocks.

All speech and noise stimuli were normalized to an average intensity of 70 dB (SPL) using Praat software (Institute of Phonetic Sciences, University of Amsterdam). Stimuli were presented binaurally through plastic tubes and intracanal earpieces at a comfortable listening level, adjusted individually. Each 40-s auditory stimulus was presented only once to avoid learning effects. To ensure that the participants attended to the stimuli, one of the six stimuli in each condition included a 2-s auditory segment which was repeated 4 times (in a similar manner to a broken record). The repetitive segment occurred at a random time point during the 40-s stimulus. A repetitive segment was indicated with an index finger lift using an infrared response panel. To further verify that the participants attended to the speech stimuli of the unknown speaker, they were asked to fill in a surprise multi-choice questionnaire at the end of the experiment.

On both experimental days, the conditions were performed in a random order. Participants were instructed to keep their gaze on a fixation point projected on a screen at ~ 1 m from their sitting position. Prior to the first block of each condition a text was presented on the screen informing the participants of the upcoming condition and the speaking rate (for speech production conditions). There were six blocks per experimental condition, thus totalling to 4 min of data for each condition.

Recordings

Audio

During the speech production experiment, audio signals were recorded at 44.1 kHz sampling frequency using an MEG-compatible microphone (Yoga EM-3, Yoga Electronics Co., Taiwan) connected to a portable recording device (FOSTEX FR-2LE Field Memory Recorder, Tokyo, Japan). The speech materials of the unknown speaker were recorded in a soundproof room with a dual diaphragm condenser microphone (B-2 PRO, Behringer) using Cool Edit 2000 (Syntrillium).

Electromyographic (EMG) and Electrooculographic (EOG) signals

EMG signals were recorded during the speech production experiment with reusable circular electrodes (E220X-HS, conductive area diameter of 0.4 cm), low-pass filtered at 330 Hz and sampled at 1500 Hz. Bipolar EMG channels (inter-electrode distance 1.5 cm) were used to record activity, first, from the lower lip muscles (orbicularis oris) and, second, from muscles associated with tongue and jaw movements (primarily from genioglossus and mylohyoid muscles, determined via tactile inspection of the chin and lower jaw area). Electrode resistance was <10 k\Omega for all participants. Blinks and eye movements (saccades) were monitored by recording EOG.

MEG and MRI

MEG signals were recorded with a Neuromag Vectorview wholehead device (Elekta Oy, Helsinki, Finland) in a magnetically shielded room. Data were filtered at 0.03–500 Hz and sampled at 1500 Hz. The position of the head with respect to the MEG sensor array was determined with the help of five head position indicator coils attached to the scalp that were briefly energized before the measurement. The coil locations were determined with respect to anatomical landmarks (nasion and right/left pre-auricular points) using a 3D digitizer (Isotrak 3S1002, Polhemus Navigation Science). Structural magnetic resonance images (MRIs; 3 T Siemens MAGNETOM Skyra; Siemens Medical Systems, Erlangen, Germany) were obtained for each participant. During the analysis phase, the MEG coordinate system was aligned with individual MRIs based on head position coils and anatomical landmarks.

MEG data analysis

External disturbances were first removed from the MEG data using spatiotemporal signal space separation (Taulu and Simola, 2006). Artifacts due to movement of facial articulatory muscles during the speech production experiment were subsequently removed from the notch-filtered (50 Hz) and low-pass filtered (at 150 Hz) MEG data by performing an independent component analysis (ICA) decomposition (FastICA; Hyvärinen, 1999) of all MEG gradiometers and the two EMG channels. Each resulting component was visually evaluated and classified as an artifact based on the inspection of four different criteria (De Vos et al., 2010; Henderson et al., 2013; Hipp et al., 2011; Kujala et al., 2013): the topography of the component (location of the component at the edges of the sensor space), its localization in the source space (the center of the location outside the brain), the correlation of the component's and the two EMG channels' time series and the relative amount of spectral power in the gamma/high-gamma band for the component. In addition to correlation with the EMG channels, other criteria were included in the classification protocol for two main reasons. Firstly, the EMG channels do not necessarily represent all muscle activity present in the data. The EMG sensors were placed to record activity mainly from the lip and tongue muscles. which were presumed to be the primary sources of artifacts. However, these recordings may not be as informative on the activity of muscles in the cheek area or in the back of the head, which may induce less prominent artifacts. Secondly, in several cases the lip and tongue muscle activity were separated into multiple components instead of a single component, thus leading to lower correlation values with the EMG channels for each component. In these cases, the other criteria were necessary for reliable artifact detection (see Supplementary methods for more details).

MEG signals recorded during speech perception conditions were only mildly contaminated by blink artifacts: blink-contaminated time windows (amplitude of either vertical or horizontal electrooculogram > 150 μ V) were excluded from further analysis. In contrast, speech production tasks were all accompanied by frequent blinks (von Cramon and Schuri, 1980), rendering a threshold-based blink rejection procedure unsuitable. Consequently, blink artifacts were removed from MEG signals using a principal component analysis (PCA)-based routine (Wold et al., 1987; Huotilainen et al., 1993).

Cortical distribution of signal power is the way to estimate activation for continuous, non-averaged data. The cortical distribution of signal power was determined in eight frequency bands spanning the range 1–90 Hz (1–4 Hz; 4–7 Hz; 7–9 Hz; 8–13 Hz; 13–20 Hz; 20–30 Hz; 35–45 Hz; 60–90 Hz) (e.g., Gross et al., 2013; Jensen et al., 2012; Palva and Palva, 2007; Salmelin and Hari, 1994) for each condition and each participant, using Dynamic Imaging of Coherent Sources (DICS) (Gross et al., 2001) with a spherical head model. The sensor-level cross-spectral density matrices (CSD) required for the estimation were calculated using Welch's averaged periodogram method (4096-point Hanning windowing; 50% window overlap; 4096-point Fast-Fourier Transform; 0.4 Hz resolution). To enable statistical evaluation of the data at the group level, the cortical-level estimates were obtained in a spatially equivalent search grid across participants. The grid sampled the grey matter surface, excluding the cerebellum

(20482 points; atlas brain; Freesurfer 5.3) (Fischl, 2012). Grid points in the anterior frontal cortex and temporal pole areas, displaying high sensitivity to eye movement artifacts, were excluded from the analysis. This common grid was transformed to each participant's anatomy via a surface-based transformation (Fischl et al., 1999).

Differences in cortical distribution of signal power between conditions were estimated based on the features of linguistic content (speech production vs. /pa/ syllable repetition; listening to speech vs. noise), speaking rate (e.g. production of normal-rate vs. slow-rate speech) and speech ownership (listening to the unknown speaker's normal-rate speech vs. own normal-rate speech). Statistical significance was determined using group-level cluster-based statistics (cluster-based permutation procedure performed on the statistically significant results obtained from a two-tailed t-test; 10000 permutations; cluster threshold p < 0.05; weighted distance algorithm for linking adjacent grid points; 10-mm cut-off threshold). For the linguistic content, where the contrasted conditions were quite different from each other, the statistical significance threshold for the t-test was set to p < 0.01(family-wise error (FWE) corrected). For all other, more modulatory contrasts (e.g., production of normal-rate vs. fast-rate speech), the statistical significance threshold was set to p < 0.05 (FWE corrected). For each round of the permutation testing, the labels of the two conditions being compared were randomized across subjects, and new t-statistics were computed in all grid points. For each permutation, the largest cluster t-value was stored, leading to distribution of 10000 cluster-level t-values. The original t-statistics were then compared to this distribution. Unique and shared parts of cortical activation patterns between speech production and perception were quantified by examining the grid index overlap of all activated cortical regions across frequencies and contrasts. Talairach coordinates and Brodmann area numbers were obtained for each activated region using the Talairach Daemon (Lancaster et al., 2000).

To further examine whether the spatiospectral cortical patterns resulting from the speech production vs. /pa/ syllable repetition contrast were merely due to increased auditory feedback (as one unavoidably listens to one's own voice while speaking), regions of interest (ROIs) were identified from this linguistic content contrast. In each ROI, the mean power was estimated for the normal-rate speech production and articulation-only speech production (a task that simulates natural speech production but occurs without vocalization; a subset of 10 participants) conditions, and their power difference assessed (paired two-tailed t-test; p < 0.01).

Furthermore, to examine the potential contribution of voice identity processing (c.f. Belin et al., 2004; Kriegstein and Giraud, 2004) to the spatiospectral patterns resulting from the perception of the unknown speaker's speech vs. perception of own speech contrast, ROIs were identified from the speech ownership contrast. In each ROI, the difference in power between the two conditions in individual participants was correlated (Pearson's correlation coefficient; p < 0.05) with the difference in the values of several speaker-specific acoustic features estimated using Praat software (fundamental frequency or pitch; 1st formant frequency; spectral flux; central spectroid). This correlation analysis was carried out for both absolute and relative differences in power and acoustic features. The ROI analysis considered both data from all 20 participants, as well as data only from the 11 female participants. This was done to eliminate any possible confound of male participants, whose acoustic features are closer to the unknown speaker.

Results

Attentional control task during speech perception

Attention was monitored during the speech perception experiment through a repetitive segment detection task. For all speech conditions, all participants (20 out of 20) were able to detect the repetitive segment

Table 1

Significant power modulations (Talairach coordinates, Area label, Brodmann area number, frequency band) for all speech production contrasts.

Talairach coordinates	Area label. Brodmann area number	Frequency band (Hz)							
(x, y, z)	, ,	×	~	<u></u>	r3	20	30	45	90
8	Speech complexity: speech > /pa/	~	4	~	Ś	13,	20,	<u></u> 35	60
47 -54 41	Right inferior parietal lobule, 40								
52 -42 21	Right superior temporal gyrus, 13			1					
64 -33 -9	Right middle temporal gyrus, 21								
59 -8 12	Right pre-central gyrus, 43								
47 -52 19	Right superior temporal gyrus, 39								
66 - 25 4	Right superior temporal gyrus, 22			1					
54 -13 42	Right pre-central gyrus, 4								
53 -49 36	Right supramarginal gyrus, 40			1					
51 -58 6	Right middle temporal gyrus, 39								
-58 -51 40	Left inferior parietal lobule, 40								
-58 3 17	Left pre-central gyrus, 6					1			
-56 -45 27	Left supramarginal gryus, 40								
-50 7 -16	Left middle temporal gryus, 21								
-36 1 -10	Left superior temporal gyrus, 38								
-45 4 42	Left middle frontal gyrus, 6								
-64 -29 7	Left superior temporal gyrus, 22								
-38 25 -14	Left inferior frontal gyrus, 47								
	Speaking rate: normal > slow								
57 -24 34	Right post-central gyrus, 2								
65 - 20 3	Right superior temporal gyrus, 22								
6 -35 60	Right paracentral lobule, 4								
44 24 26	Right middle frontal gyrus, 9								
-55 1 32	Left middle frontal gyrus, 6								
-5 -35 59	Left paracentral lobule, 4								
-55 -15 40	Left pre-central gyrus, 4								
-55 1 32	Left pre-central gyrus, 6								
-26 -29 63	Left post-central gyrus, 3								
-58 -24 22	Left post-central gyrus, 40								
-41 -27 23	Left insula, 13								
-41 33 23	Left middle frontal gyrus, 46								
-49 -49 28	Left inferior parietal lobule, 40								
-53 -17 48	Left post-central gyrus, 1								

embedded in one of the stimuli. For the noise condition, the detection success was 65% (14 out of 20 detected the repetitive segment). The answers to the surprise questionnaire presented at the end of the experiment regarding the content of the unknown speaker's speech were 100% correct for all 20 participants.

Cortical signal modulations resulting from variations of speechrelated features

Variations in linguistic content, speaking rate and speech ownership were found to be associated with distinct spatiospectral patterns of cortical signal modulations. All areas and frequency bands that showed significant power modulations for the examined contrasts are listed in Table 1 for speech production and Table 2 for speech perception. Significant cortical signal power modulations observed at lower frequencies (< 30 Hz) coincided with power spectral maxima (Fig. 2A). At higher frequencies (> 35 Hz), the significant effects were detected in the spectral range that does not show salient local maxima (Fig. 2B).

production compared to /pa/ syllable repetition. In the left hemisphere, this effect encompassed frontocentral, superior temporal and temporo-parietal regions (BA 6, 21, 38, 40, 42, 47) (Fig. 3A, top; Table 1). In the right hemisphere, the effect of linguistic content was mainly observed in temporo-parietal and parietal areas (BA 4, 13, 21, 22, 39, 40, 43) (Fig. 3A, top; Table 1). In both hemispheres, modulations in cortical activity were predominantly observed at the high end of the examined frequencies (35-90 Hz) (Table 1). This spatiospectral pattern did not emerge only due to the auditory feedback associated with vocalization: a ROI comparison between speech production and articulation-only speech conditions found no significant differences (p > 0.01 for all comparisons) in parietal, temporo-parietal or frontocentral regions in either hemisphere (1-4 Hz; 13-20 Hz; 35-90 Hz). In speech perception, the contrast of speech versus acoustically matched white noise vielded significant cortical power differences uniquely in the right hemisphere, in temporal and temporo-parietal areas (BA 2, 12, 21, 22) (Fig. 3A, bottom; Table 2). This effect was observed in the 8-30 Hz frequency range (Table 2).

Effect of linguistic content

Bilateral modulations in cortical activity were observed for speech

Effect of speaking rate

In speech production, contrasting normal- to slow-rate speech

Table 2

Significant power modulations (Talairach coordinates, area label, Brodmann area number, frequency band) for all speech perception contrasts.

Talairach coordinates	Area Label, Brodmann area number	Frequency band (Hz)							
(x, y, z)		1-4	4.7	2.9	8-13	3.20	0.30	5.45	06.0
Speech complexity: speech > noise						~	CV.	ŝ	Ø
59 -25 35	Right post-central gyrus, 2								
64 -32 -9	Right middle temporal gyrus, 21								
58 -40 21	Right superior temporal gyrus, 13								
43 -50 18	Right superior temporal gyrus, 22								
	Speaking rate: slow > normal								
-63 -20 1	Left superior temporal gyrus, 22								
	Speaking rate: fast > normal								
-63 -32 7	Left superior temporal gyrus, 22								
Speech ownership: unknown speaker's > own speech									
65 -27 9	Right superior temporal gyrus, 42								
56 -40 17	Right superior temporal gyrus, 13								
61 -9 13	Right transverse temporal gyrus, 42								
62 - 37 6	Right middle temporal gyrus, 21								
52 -56 5	Right middle temporal gyrus, 37								
62 -36 13	Right superior temporal gyrus, 22								
58 -18 19	Right post-central gyrus, 43								
39 10 10	Right insular cortex, 13								

revealed enhanced neural activity: significantly increased spectral power was observed bilaterally in temporo-parietal and frontal regions (BA 2, 9, 46) (Fig. 3B, top; Table 1). Additionally, this effect encompassed ventral frontocentral regions and the insular cortex in the left hemisphere (BA 1, 3, 4, 6, 13, 40), as well as the superior temporal cortex in the right hemisphere (BA 22) (Fig. 3B, top; Table 1). Apart from the left pre-central effects, which were observed at low frequencies (1–4 Hz), the other regions demonstrated modulations at higher frequencies (mainly in the 13–45 Hz frequency range) (Table 1). In contrast, in speech perception, cortical signal power modulations as a result of speaking rate variations were uniquely left-lateralized: power in the left superior temporal cortex (BA 22) was enhanced for slow and fast speaking rates compared to the normal rate (Fig. 3B, bottom). The observed effect was confined to the 35–45 Hz frequency range (Table 2).

Effect of speech ownership

Perceiving the normal-rate utterances of an unknown speaker compared to one's own normal-rate speech resulted in significantly enhanced cortical signal power only in the right hemisphere. The effect of speech ownership extended from inferior and superior middle temporal areas to temporo-parietal and ventral parietal regions (BA 13, 21, 22, 37, 42, 43) (Fig. 3C; Table 2). For all regions, cortical signal power modulations were observed in the 8–30 Hz frequency range (Table 2). This spatiospectral pattern did not appear to be associated with voice identity processing: a ROI analysis revealed that the observed increases in cortical signal power were not significantly correlated with differences in speaker-specific acoustic features between the unknown speaker's speech and one's own speech (p > 0.05 for all correlations).

Speech production vs. speech perception

There were both unique and shared functional modulations in cortical signal power between speech production and perception across the three features (complexity, speaking rate; also ownership in speech perception). Speech complexity and speaking rate effects in natural speech production were detected primarily in parietal, temporal, temporo-parietal and fronto-parietal areas in both hemispheres (Fig. 4, green color). In contrast, natural speech perception, while bilateral as well, demonstrated a right temporal and temporo-parietal emphasis (Fig. 4, blue color). There was a considerable overlap between cortical areas involved in speech production and speech perception (Fig. 4, magenta color). Specifically, natural speech production and perception were found to jointly activate the right temporoparietal junction (rTPJ), right middle temporal gyrus and left superior temporal gyrus.

Discussion

The present results, based on band-limited cortical power modulations in natural speech tasks, suggest a notable involvement of the right hemisphere in speech processing. The right temporo-parietal junction (rTPJ) was highlighted, as it demonstrated modulations of activity in response to variations of all three speech-related features studied here: linguistic content, speaking rate and speech ownership. The involvement of the left superior temporal cortex (ISTG) in natural speech processing was also emphasized, regarding linguistic content and speaking rate. The three different speech-related features were associated with distinct spatiospectral modulation patterns that encompassed cortical regions in both hemispheres and occurred at delimited signaling frequencies. Speech production demonstrated a bilateral activation pattern, with power modulation concentrating to frontocentral, parietal and temporo-parietal areas. While speech perception showed bilateral activation as well, it notably and rather unexpectedly emphasized right temporo-parietal and temporal areas. The involvement of the left perisylvian cortex aligns with the traditional view of speech processing. However, the present findings, based on naturalistic tasks, further revealed salient, consistent modulation of activity in the right hemisphere, especially in the rTPJ.

Production of natural, normal-rate speech requires the involvement



Fig. 2. Individual-level cortical power spectra for selected cortical areas showing significant task effects. The cortical region of interest is plotted in white. Normalized power (in arbitrary units; y axis) is displayed for all participants (n=20) for two conditions (black, gray) as a function of frequency (in Hz; x axis). The frequency band in which significant group-level power modulations are observed is depicted by an asterisk. A, Linguistic content (production): speech vs. /pa/. B, Speaking rate (perception): normal speech vs. fast speech.

of an extensive bilateral cortical system to support the generation of intricate articulatory patterns as well as the production of higher-level, linguistic content with semantic and syntactic complexity and phonological variations. The left-hemisphere component of this observed bilateral pattern was an expected observation. Specifically, the production of normal-rate connected speech predictably engaged the left face sensorimotor area as well as premotor and prefrontal regions, which have been previously linked to higher-level aspects of speech production (e.g. Binder et al., 2009; Blank et al., 2002). Moreover, in contrast to a simple oromotor task, production of meaningful speech required the contribution of the left perisylvian areas, including the lSTG; this was also anticipated based on existing models of language function (e.g. Hickok and Poeppel, 2007; Hickok, 2012).

In addition to speech production, the ISTG was also implicated in



Fig. 3. Spatiospectral maps of significant power modulations in response to experimental modulations in speech-related features. A, Linguistic content: Production (above), normal-rate speech vs. /pa/ syllable. Perception (below), normal-rate speech vs. noise. **B**. Speaking rate: Production (above), normal vs. slow rate. Perception (below), normal vs. slow rate and normal vs. fast rate. C, Speech ownership: Perception, unknown speaker's speech vs. own speech. The number of frequency bands in which a given region demonstrates power modulations is represented by different colors (1; dark orange; 2, light orange; 3, yellow). See Tables 1 and 2 for a detailed listing of area labels and frequency information.

3

encoding the rate of perceived speech stimuli, using gamma-band (35-45 Hz) neural signaling. The role of ISTG in speech comprehension has been emphasized through empirical evidence (e.g. Buchsbaum et al., 2001; Friederici et al., 2003; Scott et al., 2000) as well as in neurophysiological models (Hickok and Poeppel, 2000, 2004) and cortical gamma-band activity has been associated with auditory processing of normal-rate speech (e.g. Giraud et al., 2007; Giraud and Poeppel 2012; Morillon et al., 2012). The present results im-



Fig. 4. Unique and shared cortical areas engaged in speech production and perception. Modulations of band-limited power are visualized for production and perception, pooled across contrasts and frequency bands. Areas sensitive to speech production are plotted in green (above), areas sensitive to speech perception in blue (below), and the overlap between them in magenta (both sets of images).

portantly add to this evidence by highlighting the special nature of normal-rate perceived speech in ISTG. Cortical processing of incoming auditory stimuli is deemed optimal at the habitual syllabic rate of an utterance (~5 Hz) (Drullman et al., 1994; Smith et al., 2002). In contrast, the slow and fast-rate stimuli in our paradigm featured mean syllabic rates on the order of 2.6 Hz and 6.8 Hz, respectively, thus substantially deviating from this optimum. It may thus be argued that the increased gamma-band power in ISTG reflects increased computational demands of encoding and subsequently attempting to comprehend speech whose rate falls outside of the habitual processing range of the auditory system. This interpretation aligns with previous evidence of a link between the amount of neural activity observed in the ISTG and the cognitive effort required for spoken sentence comprehension (Just et al., 1996; Lewis and Bastiaansen, 2015).

In contrast to the expected left-hemisphere effects, the notable right-hemisphere involvement in both natural speech production and perception was an unanticipated finding considering the conventional understanding of a left-lateralized cortical organization of language function. The right hemisphere was involved in generating normal-rate meaningful speech as part of a more extensive bilateral system, but its role was particularly emphasized in natural speech perception: processing linguistic content and speech ownership exclusively engaged the right hemisphere. This right-hemispheric contribution notably included the right middle temporal cortex, previously associated with semantic processing (Donnelly et al., 2011; Laufer et al., 2011), and the rTPJ. In fact, the rTPJ is typically not associated with linguistic processing: existing literature chiefly refers to this area as a heteromodal association hub mediating social cognition and attention which integrates various kinds of input while maintaining reciprocal connections to temporal and frontal cortical areas (Decety and Lamm, 2007; Seghier, 2013). The present experimental paradigm, comprising of variations in speech-related features, enabled characterization of the functional role of the rTPJ in the production and perception of meaningful, connected speech.

The rTPJ seems to be implicated in the active generation of

meaningful, connected speech. An enhancement of rTPJ band-limited power (13-20 Hz; > 35 Hz) was associated with a shift from the less computationally demanding task of /pa/ syllable repetition to natural speech production. The enhanced power in the rTPJ could be due to the greater motor complexity required to produce natural-rate speech compared to /pa/ syllable repetition. However, anatomically, the rTPJ is part of the inferior parietal cortex, which has been associated with intention-driven, goal-directed communicative actions (Carota et al., 2010). The results of the ROI analysis further suggested that the observed power modulations in the rTPJ do not only result from speech-generated auditory feedback, but are presumably associated with the active generation of a linguistic message. The enhanced rTPJ power could thus reflect the intentional. goal-driven nature of natural speech production, an interpretation well in line with existing models suggesting that speech generation is mostly based on internal expectations translated into phonetic actions (e.g. Guenther, 2006). Moreover, the antero-parietal portion of the rTPJ was more involved in the production of normal-rate speech than slow-rate speech. Slow-rate speech has less linguistic content per time unit than normal-rate speech (as manifested by lower syllable and word production frequencies). Hence, it could be suggested that slow-rate speech taxes cognitive resources related to speech generation less than natural-rate speech. This further supports the interpretation that the rTPJ, and the right hemisphere in general, mediate higher-level cognitive processes supporting the production of meaningful, connected speech.

Furthermore, through the linguistic content and speech ownership contrasts, the rTPJ emerged as part of a right-lateralized cortical system during the perception of meaningful, socially relevant speech. These effects highlight the 8-30 Hz frequency range, previously linked to lexical access (Brennan et al., 2014), semantic evaluation of speech (Shahin et al., 2009) and cortical connectivity during naturalistic language comprehension (Saarinen et al., 2015). Numerous reports exist on the role of the rTPJ as a mediator of attentional processes (e.g. Corbetta and Shulman, 2002; Mitchell, 2008), and increased power in the 8-30 Hz range has been extensively associated with attention (e.g., Jensen et al., 2002; Klimesch, 2012). Based on these findings, one might argue that the observed rTPJ effect reflects increased attention to speech stimuli compared to noise stimuli. However, there is evidence to suggest that listening to noise stimuli is challenging for the auditory system whereas speech, an overlearned input type (Smith, 1992), is processed even in the absence of attention (Wild et al., 2012). This notion is corroborated by the present behavioral results indicating a higher detection success of a repetitive segment when it is embedded in speech than noise. Therefore, while various attentional sub-mechanisms may be involved to some extent, it seems unlikely that the higher rTPJ power for processing natural speech than noise stimuli would only reflect enhanced attentional effort as such. Instead, we suggest that the observed rTPJ power modulations could rather reflect cortical mechanisms of social cognition (Lawrence et al., 2006; Saxe and Kanwisher, 2003; Saxe and Baron-Cohen, 2006) and, for the speech ownership contrast, indicate that this area is also involved in distinguishing oneself from others (e.g. Decety and Sommerville, 2003; Farrer and Frith, 2002; Ruby and Decety, 2001). Hence, these observations suggest that the rTPJ, through its role as a cortical associative area, supports the comprehension of communicative, socially relevant spoken language.

The present findings highlight a right-hemispheric contribution in processing linguistic content and a left-hemispheric contribution in processing speaking rate. This pattern of sensitivity could be broadly linked to the suggestion that different aspects of an incoming speech stimulus are processed asymmetrically in the cortex (e.g., Giraud et al., 2007; Morillon et al., 2010). According to this framework, the left hemisphere undertakes syllabic parsing, while the right hemisphere is responsible for more integrative processes. More generally, the fact that increased linguistic content of perceived speech emphasizes the right hemisphere corroborates previous reports suggesting that the right hemisphere is particularly sensitive to the extraction of meaning from natural speech (Horowitz-Kraus et al., 2015; Jung-Beeman, 2005; Koskinen et al., 2013). Indeed, the right hemisphere has been previously suggested to decode abstract, context-dependent linguistic content during the course of a conversation (Federmeier et al., 2008; St George et al., 1999). It is thought to undertake the analysis of general themes and concepts, including the understanding of jokes, metaphors and idioms (e.g. Brownell et al., 1983; Weylman et al., 1989); such colloquial language uses were numerous in our experimental paradigm of naturalistic, free-form language production. Contour and prosody, acoustic features conveying a multitude of extra-linguistic information, such as the affective state of the interlocutor, may also be preferentially processed in the right hemisphere (Bourguignon et al., 2013).

The significance of the present spatiospectral effects for natural speech production and perception across contrasts is twofold. Firstly, in accordance with recent evidence (Hultén et al., 2014; Simonyan and Fuertinger, 2015; Silbert et al., 2014), these findings point out that while the traditional view of a left-hemisphere specialization for language processing remains valid, the right hemisphere also occupies an important role in processing meaningful speech. The present findings partly align with the dual-stream model of speech production and perception (Hickok and Poeppel, 2007): natural speech production demonstrated a bilateral frontoparietal emphasis, whereas perception engaged, expectedly, more temporal regions. Yet, the notable involvement of the right hemisphere in both the generation and comprehension of meaningful speech argues in favor of an emerging, more integrative view of language processing (Federmeier et al., 2008; Poeppel et al., 2012; Rauschecker and Scott, 2009; Scott and Johnsrude, 2003). Hence, in contrast to the well-established view of a conspicuous left-hemispheric specialization for speech processing (e.g. Blank et al., 2002; Peelle, 2012; McGettigan and Scott, 2012), it may be proposed that the production and perception of natural speech, a highly complex and multifaceted entity, requires a balanced engagement of neural resources from both cortical hemispheres.

Secondly, speech production and speech perception show a notable bilateral overlap. Studies examining isolated linguistic forms have previously demonstrated that speech production and speech perception share cortical representations in the left hemisphere (Buchsbaum et al., 2001; Cogan et al., 2014; Indefrey, 2011; Price et al., 1996; Watkins et al., 2003). The present electrophysiological findings reveal that such an overlap also occurs for natural speech and that this joint activation pattern is particularly salient in the right hemisphere. These spatially overlapping right-hemisphere areas seem to process speech meaningfulness in both modalities; however, in speech production, fast gamma-band (35-90 Hz) signaling frequencies are highlighted, whereas in speech perception the emphasis is on a lower frequency range (8–30 Hz). This dissociation in the time domain suggests that, in these areas, active generation of complex linguistic content operates on faster timescales compared to the extraction of meaning which requires integration of information over relatively longer periods of time. Unlike the rTPJ, the lSTG was shown to process distinct speech-related features in speech production and speech perception; indeed, anatomically overlapping areas may also have dissimilar functional roles in different aspects of speech processing (Silbert et al., 2014). Overall, the present findings importantly accentuate partly shared large-scale cortical representations for the production and perception of meaningful connected language.

Conclusions

The transition from experimental paradigms employing isolated language forms to more naturalistic set-ups has brought about a shift in the common conceptions of language processing. While the classical notion of the importance of the left perisylvian cortex in language processing remains well-founded, recent years have brought attention to the notable involvement of the right hemisphere in language function. The present neurophysiological view on natural speech processing obtained by modulating three fundamental speech-related features highlights the right hemisphere. Especially rTPJ emerged as a central cortical region in natural connected speech processing, sensitive to socially relevant linguistic content and playing an important role in both the generation and comprehension of natural, communicative language.

Acknowledgements

This work was financially supported by the Academy of Finland (Grants #255349, #256459 and #283071 to RS, #257576 to JK) and Sigrid Jusélius Foundation. The authors declare no conflicts of interest.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.neuroimage.2017.03. 006.

References

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., Merzenich, M.M., 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc. Natl. Acad. Sci USA 98, 13367–13372. http://dx.doi.org/ 10.1073/pnas.201400998.
- Alexandrou, A.M., Saarinen, T., Kujala, J., Salmelin, R., 2016. A multimodal spectral approach to characterize rhythm in natural speech. J. Acoust. Soc. Am. 139, 215–226. http://dx.doi.org/10.1121/1.4939496.
- Belin, P., Fecteau, P., Bédard, C., 2004. Thinking the voice: neural correlates of voice perception. Trends Cogn. Sci. 8, 129–135. http://dx.doi.org/10.1016/ i.tics.2004.01.008.
- Belin, P., Zatorre, R.J., Lafaille, P., Ahad, P., Pike, B., 2000. Voice-selective areas in human auditory cortex. Nature 403, 309–312. http://dx.doi.org/10.1038/ 35002078.
- Binder, J.R., Desai, R.H., Graves, W.W., Conant, L.L., 2009. Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. Cereb. Cortex 19, 2767–2796. http://dx.doi.org/10.1093/cercor/bhp055.
- Blank, S.C., Scott, S.K., Murphy, K., Warburton, E., Wise, R.J., 2002. Speech production: Wernicke, Broca and beyond. Brain 125, 1829–1838. http://dx.doi.org/10.1093/ brain/awf191.
- Bourguignon, M., De Tiège, X., de Beeck, M.O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., Jousmäki, V., 2013. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum. Brain Mapp. 34, 314–326. http:// dx.doi.org/10.1002/hbm.21442.
- Brennan, J., Lignos, C., Embick, D., Roberts, T.P., 2014. Spectro-temporal correlates of lexical access during auditory lexical decision. Brain Lang. 133, 39–46. http:// dx.doi.org/10.1016/j.bandl.2014.03.006.
- Brownell, H.H., Michel, D., Powelson, J., Gardner, H., 1983. Surprise but not coherence: sensitivity to verbal humor in right-hemisphere patients. Brain Lang. 18, 20–27. http://dx.doi.org/10.1016/0093-934X(83)90002-0.
- Buchsbaum, B.R., Hickok, G., Humphries, C., 2001. Role of left posterior superior temporal gyrus in phonological processing for speech perception and production. Cogn. Sci. 25, 663–678. http://dx.doi.org/10.1016/S0364-0213(01)00048-9.
- Carota, F., Posada, A., Harquel, S., Delpuech, C., Bertrand, O., Sirigu, A., 2010. Neural dynamics of the intention to speak. Cereb. Cortex 20, 1891–1897. http://dx.doi.org/ 10.1093/cercor/bhp255.
- Cogan, G.B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., Pesaran, B., 2014. Sensorymotor transformations for speech occur bilaterally. Nature 507, 94–98. http:// dx.doi.org/10.1038/nature12935.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. Nat. Rev. Neurosci. 3, 201–215. http://dx.doi.org/10.1038/ nrn755.
- Cummins, F., Port, R., 1998. Rhythmic constraints on stress timing in English. J. Phon. 26, 145–171. http://dx.doi.org/10.1006/jpho.1998.0070.
- Davis, B.L., MacNeilage, P.F., 2000. The motor core of speech: A comparison of serial organization patterns in infants and languages. 288, 527–531. http://dx.doi.org/ 10.1126/science.288.5465.527.
- De Vos, M., Ries, S., Vanderperren, K., Vanrumste, B., Alario, F.X., Van Huffel, S., Burle, B., 2010. Removal of muscle artifacts from EEG recordings of spoken language production. Neuroinformatics 8, 135–150. http://dx.doi.org/10.1007/s12021-010-9071-0.
- Decety, J., Sommerville, J.A., 2003. Shared representations between self and other: a social cognitive neuroscience view. Trends Cogn. Sci. 7, 527–533. http://dx.doi.org/ 10.1016/j.tics.2003.10.004.
- Decety, J., Lamm, C., 2007. The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. Neuroscientist 13, 580–593. http://dx.doi.org/10.1177/1073858407304654.
- Donnelly, K.M., Allendorfer, J.B., Szaflarski, J.P., 2011. Right hemispheric participation

in semantic decision improves performance. Brain Res. 1419, 105–116. http://dx.doi.org/10.1016/j.brainres.2011.08.065.

- Dronkers, N.F., Wilkins, D.P., Van Valin, R.D., Redfern, B.B., Jaeger, J.J., 2004. Lesion analysis of the brain areas involved in language comprehension. Cognition 92, 145–177. http://dx.doi.org/10.1016/j.cognition.2003.11.002.
- Drullman, R., Festen, J.M., Plomp, R., 1994. Effect of reducing slow temporal modulations on speech reception. J. Acoust. Soc. Am. 95, 2670–2680. http:// dx.doi.org/10.1121/1.409836.
- Farrer, C., Frith, C.D., 2002. Experiencing oneself vs. another person as being the cause of an action: the neural correlates of the experience of agency. Neuroimage 15, 596–603. http://dx.doi.org/10.1006/nimg.2001.1009.
- Federmeier, K.D., Wlotko, E.W., Meyer, A.M., 2008. What's 'right' in language comprehension: event-related potentials reveal right hemisphere language capabilities. Lang. Linguist Compass 2, 1–17. http://dx.doi.org/10.1111/j.1749-818X 2007 00042 x
- Fischl, B., 2012. FreeSurfer. Neuroimage 62, 774–781. http://dx.doi.org/10.1016/ j.neuroimage.2012.01.021.
- Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M., 1999. High-resolution intersubject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284. http://dx.doi.org/10.1002/(SICI)1097-0193.
- Flavell, J.H., 1968. The Development of Role-taking and Communication Skills in Children. John Wiley & Sons, Inc, New York, 247.
- Flavell, J.H., 2000. Development of children's knowledge about the mental world. Int. J. Behav. Dev. 24, 15–23. http://dx.doi.org/10.1080/016502500383421.
- Friederici, A.D., Ruschemeyer, S., Hahne, A., Fiebach, C.J., 2003. The role of left inferior frontal and superior temporal cortex in sentence comprehension: localizing syntactic and semantic processes. Cereb. Cortex 13, 170–177. http://dx.doi.org/10.1093/ cercor/13.2.17.
- Giraud, A.L., Poeppel, D., 2012. Speech perception from a neurophysiological perspectiveThe Human Auditory Cortex. Springer, New York, 225–260. http:// dx.doi.org/10.3389/fpsyg.2012.00170.
- Giraud, A.L., Kleinschmidt, A., Poeppel, D., Lund, T.E., Frackowiak, R.S., Laufs, H., 2007. Endogenous cortical rhythms determine cerebral specialization for speech perception and production. Neuron 56, 1127–1134. http://dx.doi.org/10.1016/ j.neuron.2007.09.038.
- Grosjean, F., Lane, H., 1976. How the listener integrates the components of speaking rate. J. Exp. Psychol.-Hum. Percept. Perform. 2, 538–543. http://dx.doi.org/ 10.1037/0096-1523.2.4.538.
- Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci USA 98, 694–699. http://dx.doi.org/10.1073/ pnas.98.2.694.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., Garrod, S., 2013. Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol. 12, e1001752. http://dx.doi.org/10.1371/journal.pbio.1001752.
- Guenther, F.H., 2006. Cortical interactions underlying the production of speech sounds. J. Commun. Disord. 39, 350–365. http://dx.doi.org/10.1016/ i.jcomdis.2006.06.013.
- Henderson, J.M., Luke, S.G., Schmidt, J., Richards, J.E., 2013. Co-registration of eye movements and event-related potentials in connected-text paragraph reading. Front. Syst. Neurosci. 28, 1–13. http://dx.doi.org/10.3389/fnsys.2013.00028.
- Hickok, G., 2012. Computational neuroanatomy of speech production. Nat. Rev. Neurosci. 13, 135–145. http://dx.doi.org/10.1038/nrn3158.
- Hickok, G., Poeppel, D., 2000. Towards a functional neuroanatomy of speech perception. Trends Cogn. Sci. 4, 131–138. http://dx.doi.org/10.1016/S1364-6613(00)01463-7.
- Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition 92, 67–99. http://dx.doi.org/10.1016/j.cognition.2003.10.011.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. Nat. Rev. Neurosci. 8, 393–402. http://dx.doi.org/10.1016/j.cognition.2003.10.011.
- Hipp, J.F., Engel, A.K., Siegel, M., 2011. Oscillatory synchronization in large-scale cortical networks predicts perception. Neuron 69, 387–396. http://dx.doi.org/ 10.1016/j.neuron.2010.12.027.
- Horowitz-Kraus, T., Grainger, M., DiFrancesco, M., Vannest, J., Holland, S.K., Consortium, C.A., 2015. Right is not always wrong: DTI and fMRI evidence for the reliance of reading comprehension on language-comprehension networks in the right hemisphere. Brain Imaging Behav. 9, 19–31. http://dx.doi.org/10.1007/s11682-014-9341-9.
- Houde, J.F., Nagarajan, S., Sekihara, K., Merzenich, M.M., 2002. Modulation of the auditory cortex during speech: an MEG study. J. Cogn. Neurosci. 14, 1125–1138. http://dx.doi.org/10.1162/089892902760807140.
- Hultén, A., Karvonen, L., Laine, M., Salmelin, R., 2014. Producing speech with a newly learned morphosyntax and vocabulary: an magnetoencephalography study. J. Cogn. Neurosci. 26, 1721–1735. http://dx.doi.org/10.1162/jocn_a_00558.
- Huotilainen M., Ilmoniemi R., Tiitinen H., Lavikainen J., Alho K., Kajola M., Simola J., Näätänen R., 1993. Eye-blink removal for multichannel MEG measurements. 9th International Conference on Biomagnetism, Vienna, Austria, 14–20.8.1993. pp. 209–210.
- Hyvärinen, A., 1999. Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. 10, 626–634. http://dx.doi.org/10.1109/ 72.761722.
- Indefrey, P., 2011. The spatial and temporal signatures of word production components: a critical update. Front. Psychol. 2, 1–16. http://dx.doi.org/10.3389/ fpsyg.2011.00255.
- Indefrey, P., Levelt, W.J., 2004. The spatial and temporal signatures of word production components. Cognition 92, 101–144. http://dx.doi.org/10.1016/

j.cognition.2002.06.001.

- Jensen, O., Bonnefond, M., VanRullen, R., 2012. An oscillatory mechanism for prioritizing salient unattended stimuli. Trends Cogn. Sci. 16, 200–206. http:// dx.doi.org/10.1016/j.tics.2012.03.002.
- Jensen, O., Gelfand, J., Kounios, J., Lisman, J.E., 2002. Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cereb. Cortex 12, 877–882. http://dx.doi.org/10.1093/cercor/12.8.877.
- Jung-Beeman, M., 2005. Bilateral brain processes for comprehending natural speech. Trends Cogn. Sci. 9, 512–518. http://dx.doi.org/10.1016/j.tics.2005.09.009.
- Just, M.A., Carpenter, P.A., Keller, T.A., Eddy, W.F., Thulborn, K.R., 1996. Brain activation modulated by sentence comprehension. Science 274, 114–116. http:// dx.doi.org/10.1126/science.274.5284.114.
- Klimesch, W., 2012. Alpha-band oscillations, attention, and controlled access to stored information. Trends Cogn. Sci. 12, 606–617. http://dx.doi.org/10.1016/ i tics 2012 10 007
- Kohler, K.J., 2009. Rhythm in speech and language. Phonetica 66, 29–45. http:// dx.doi.org/10.1159/000208929.
- Koskinen, M., Viinikanoja, J., Kurimo, M., Klami, A., Kaski, S., Hari, R., 2013. Identifying fragments of natural speech from the listener's MEG signals. Hum. Brain Mapp. 34, 1477–1489. http://dx.doi.org/10.1002/hbm.22004.
- Kriegstein, K.V., Giraud, A.L., 2004. Distinct functional substrates along the right superior temporal sulcus for the processing of voices. NeuroImage 22, 948–955. http://dx.doi.org/10.1016/j.neuroimage.2004.02.020.
- Kujala, M.V., Törnqvist, H., Somppi, S., Hänninen, L., Krause, C.M., Vainio, O., Kujala, J., 2013. Reactivity of dogs' brain oscillations to visual stimuli measured with noninvasive electroencephalography. PLoS One 5, e61818. http://dx.doi.org/10.1371/ journal.pone.0061818.
- Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, C.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach atlas labels for functional brain mapping. Hum. Brain Mapp. 10, 120–131. http:// dx.doi.org/10.1002/1097-0193.
- Laufer, I., Negishi, M., Lacadie, C.M., Papademetris, X., Constable, R.T., 2011. Dissociation between the activity of the right middle frontal gyrus and the middle temporal gyrus in processing semantic priming. PLoS One 6, e22368. http:// dx.doi.org/10.1371/journal.pone.0022368.
- Lawrence, E., Shaw, P., Giampietro, V., Surguladze, S., Brammer, M., David, A., 2006. The role of 'shared representations' in social perception and empathy: an fMRI study. Neuroimage 29, 1173–1184. http://dx.doi.org/10.1016/ i.neuroimage.2005.09.001.
- Levelt, W.J., Roelofs, A., Meyer, A.S., 1999. A theory of lexical access in speech production. Behav. Brain Sci. 22, 1–38. http://dx.doi.org/10.1017/ S0140525X99001776.
- Levinson, S.C., 2016. Turn-taking in human communication origins and implications for language processing. Trends Cogn. Sci. 20, 6–14. http://dx.doi.org/10.1016/ j.tics.2015.10.010.
- Lewis, A.G., Bastiaansen, M., 2015. A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex 68, 155–168. http://dx.doi.org/10.1016/j.cortex.2015.02.014.
- Martikainen, M.H., Kaneko, K., Hari, R., 2005. Suppressed responses to self-triggered sounds in the human auditory cortex. Cereb. Cortex 15, 299–302. http://dx.doi.org/ 10.1093/cercor/bhh131.
- McGettigan, C., Scott, S.K., 2012. Cortical asymmetries in speech perception: what's wrong, what's right and what's left? Trends Cogn. Sci. 16, 269–276. http:// dx.doi.org/10.1016/j.tics.2012.04.006.
- Mirman, D., Chen, Q., Zhang, Y., Wang, Z., Faseyitan, O.K., Coslett, H.B., Schwartz, M.F., 2015. Neural organization of spoken language revealed by lesion-symptom mapping. Nat. Commun. 6, 1–9. http://dx.doi.org/10.1038/ncomms7762.
- Mitchell, J.P., 2008. Activity in right temporo-parietal junction is not selective for theory-of-mind. Cereb. Cortex 18, 262–271. http://dx.doi.org/10.1093/cercor/bhm051.
 Morillon, B., Liégeois-Chauvel, C., Arnal, L.H., Bénar, C., Giraud, A.L., Mamessier, A.,
- Morillon, B., Liégeois-Chauvel, C., Arnal, L.H., Bénar, C., Giraud, A.L., Mamessier, A., 2012. Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. Front. Psychol. 3, 1–9. http://dx.doi.org/10.3389/ fpsyg.2012.00248.
- Obleser, J., Weisz, J., 2012. Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. Cereb. Cortex 22, 2466–2477. http://dx.doi.org/10.1093/ cercor/bhr325.
- Palva, S., Palva, J.M., 2007. New vistas for α-frequency band oscillations. Trends Neurosci. 30, 150–158. http://dx.doi.org/10.1016/j.tins.2007.02.001.
- Peelle, J.E., 2012. The hemispheric lateralization of speech processing depends on what "speech" is: a hierarchical perspective. Front. Hum. Neurosci. 6, 309. http:// dx.doi.org/10.3389/fnhum.2012.00309.
- 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.
- Poeppel, D., Idsardi, W.J., van Wassenhove, V., 2008. Speech perception at the interface of neurobiology and linguistics. Philos. Trans. R Soc. B 363, 1071–1086. http:// dx.doi.org/10.1098/rstb.2007.2160.
- Poeppel, D., Emmorey, K., Hickok, G., Pylkkänen, L., 2012. Towards a new neurobiology of language. J. Neurosci. 32, 14125–14131. http://dx.doi.org/10.1523/ JNEUROSCI.3244-12.2012.
- Port, F.R., 2003. Meter and speech. J. Phon. 31, 599–611. http://dx.doi.org/10.1016/ j.wocn.2003.08.001.
- Price, C., Wise, R., Warburton, E., Moore, C., Howard, D., Patterson, K., Frackowiak, R., Friston, K., 1996. Hearing and saying. Brain 119, 919–931. http://dx.doi.org/ 10.1093/brain/119.3.919.
- Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman

primates illuminate human speech processing. Nat. Neurosci. 12, 718–724. http://dx.doi.org/10.1038/nn.2331.

- Ruby, P., Decety, J., 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. Nat. Neurosci. 4, 546–550. http://dx.doi.org/ 10.1038/87510.
- Ruspantini, I., Saarinen, T., Belardinelli, P., Jalava, A., Parviainen, T., Kujala, J., Salmelin, R., 2012. Corticomuscular coherence is tuned to the spontaneous rhythmicity of speech at 2–3 Hz. J. Neurosci. 32, 3786–3790. http://dx.doi.org/ 10.1523/jneurosci.3191-11.2012.
- Saarinen, T., Jalava, A., Kujala, J., Stevenson, C., Salmelin, R., 2015. Task-sensitive reconfiguration of corticocortical 6–20 Hz oscillatory coherence in naturalistic human performance. Hum. Brain Mapp. 36, 2455–2469. http://dx.doi.org/10.1002/ hbm.22784.
- Salmelin, R., Hari, R., 1994. Spatiotemporal characteristics of sensorimotor neuromagnetic rhythms related to thumb movement. Neuroscience 60, 537–550. http://dx.doi.org/10.1016/0306-4522(94)90263-1.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people: the role of the temporo-parietal junction in "theory of mind". Neuroimage 19, 1835–1842. http:// dx.doi.org/10.1016/S1053-8119(03)00230-1.
- Saxe, R., Baron-Cohen, S., 2006. Editorial: the neuroscience of theory of mind. Soc. Neurosci. 1, 3–4. http://dx.doi.org/10.1080/17470910601117463.
- Scott, S.K., Johnsrude, I.S., 2003. The neuroanatomical and functional organization of speech perception. Trends Neurosci. 26, 100–107. http://dx.doi.org/10.1016/ S0166-2236(02)00037-1.
- Scott, S.K., Blank, C.C., Rosen, S., Wise, R.J., 2000. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 123, 2400–2406. http:// dx.doi.org/10.1093/brain/123.12.2400.
- Seghier, M.L., 2013. The angular gyrus multiple functions and multiple subdivisions. Neuroscientist 19, 43–61. http://dx.doi.org/10.1177/1073858412440596.
- Shahin, A.J., Picton, T.W., Miller, L.M., 2009. Brain oscillations during semantic evaluation of speech. Brain Cogn. 70, 259–266. http://dx.doi.org/10.1016/ j.bandc.2009.02.008.
- Silbert, L.J., Honey, C.J., Simony, E., Poeppel, D., Hasson, U., 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. Proc. Natl. Acad. Sci USA 111, E4687–E4696. http://dx.doi.org/10.1073/ pnas.1323812111.
- Simonyan, K., Fuertinger, S., 2015. Speech networks at rest and in action: interactions

between functional brain networks controlling speech production. J. Neurophysiol. 113, 2967–2978. http://dx.doi.org/10.1152/jn.00964.2014.

- Smith, A., 1992. The control of orofacial movements in speech. Crit. Rev. Oral. Biol. Med. 3, 233–267. http://dx.doi.org/10.1177/10454411920030030401.
- Smith, A., Goffman, L., Zelasnik, H.N., Ying, G., McGillem, C., 1995. Spatiotemporal stability and patterning of speech movement sequences. Exp. Brain Res. 104, 439–501. http://dx.doi.org/10.1007/BF00231983.
- Smith, Z.M., Delgutte, B., Oxenham, A.J., 2002. Chimaeric sounds reveal dichotomies in auditory perception. Nature 416, 87–90. http://dx.doi.org/10.1038/416087a.
- St George, M., Kutas, M., Martinez, A., Sereno, M., 1999. Semantic integration in reading: engagement of the right hemisphere during discourse processing. Brain 122, 1317–1325. http://dx.doi.org/10.1093/brain/122.7.1317.
- Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51, 1759. http:// dx.doi.org/10.1088/0031-9155/51/0/000.
- Towle, V.L., Yoon, H.-A., Castelle, M., Edgar, J.C., Biassou, N.M., Frim, D.A., Spire, J.-P., Kohrman, M.H., 2008. ECoG gamma activity during a language task: differentiating expressive and receptive speech areas. Brain 131, 2013–2027. http://dx.doi.org/ 10.1093/brain/awn147.
- von Cramon, D., Schuri, U., 1980. Blink frequency and speech motor activity. Neuropsychologia 18, 603-606. http://dx.doi.org/10.1016/0028-3932(80)90164-5.
- Wang, X.-J., 2010. Neurophysiological and computational principles of cortical rhythms in cognition. Phys. Rev. 90, 1195–1268. http://dx.doi.org/10.1152/ physrev.00035.2008.
- Watkins, K.E., Strafella, A.P., Paus, T., 2003. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41, 989–994. http:// dx.doi.org/10.1016/S0028-3932(02)00316-0.
- Weylman, S.T., Brownell, H.H., Roman, M., Gardner, H., 1989. Appreciation of indirect requests by left-and right-brain-damaged patients: the effects of verbal context and conventionality of wording. Brain Lang. 36, 580–591. http://dx.doi.org/10.1016/ 0093-934X(89)90087-4.
- 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.
- Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. 2, 37–52. http://dx.doi.org/10.1016/0169-7439(87)80084-9.