Alexandrou, Anna Maria; Saarinen, Timo; Kujala, Jan; Salmelin, Riitta

Cortical tracking of global and local variations of speech rhythm during connected natural speech perception

Published in:
Journal of Cognitive Neuroscience

DOI:
10.1162/jocn_a_01295

Published: 01/01/2018

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

Please cite the original version:
Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018). Cortical tracking of global and local variations of speech rhythm during connected natural speech perception. *Journal of Cognitive Neuroscience*, *30*(11), 1704-1719. https://doi.org/10.1162/jocn_a_01295

Cortical Tracking of Global and Local Variations of Speech Rhythm during Connected Natural Speech Perception

Anna Maria Alexandrou, Timo Saarinen, Jan Kujala, and Riitta Salmelin

Abstract

■ During natural speech perception, listeners must track the global speaking rate, that is, the overall rate of incoming linguistic information, as well as transient, local speaking rate variations occurring within the global speaking rate. Here, we address the hypothesis that this tracking mechanism is achieved through coupling of cortical signals to the amplitude envelope of the perceived acoustic speech signals. Cortical signals were recorded with magnetoencephalography (MEG) while participants perceived spontaneously produced speech stimuli at three global speaking rates (slow, normal/habitual, and fast). Inherently to spontaneously produced speech, these stimuli also featured local variations in speaking rate. The coupling between cortical and acoustic speech signals was evaluated using audio–MEG coherence. Modulations in audio–MEG coherence spatially differentiated between tracking of global speaking rate, highlighting the temporal cortex bilaterally and the right parietal cortex, and sensitivity to local speaking rate variations, emphasizing the left parietal cortex. Cortical tuning to the temporal structure of natural connected speech thus seems to require the joint contribution of both auditory and parietal regions. These findings suggest that cortical tuning to speech rhythm operates on two functionally distinct levels: one encoding the global rhythmic structure of speech and the other associated with online, rapidly evolving temporal predictions. Thus, it may be proposed that speech perception is shaped by evolutionary tuning, a preference for certain speaking rates, and predictive tuning, associated with cortical tracking of the constantly changing-rate of linguistic information in a speech stream. ■

INTRODUCTION

A large part of human auditory perception consists of listening to natural connected speech in everyday communicative situations. In these situations, speech perception is guided by the temporal regularities in natural connected speech, also referred to as speech rhythm. Speech rhythm is tied to speaking rate, that is, habitual word (2–3 Hz) and syllable (4–5 Hz) production frequencies (Alexandrou, Saarinen, Kujala, & Salmelin, 2016). Especially in natural speech, the speaking rate is remarkably flexible and demonstrates two types of intraindividual variation (Grosjean & Lane, 1976). First, the global speaking rate can be increased or decreased on-demand, as a speaker can voluntarily modulate speech production speed (Alexandrou et al., 2016; Grosjean & Lane, 1976); this induces changes in the global rhythmic structure of an utterance (Smith, Goffman, Zelaznik, Ying, & McGillem, 1995). Second, within a given global speaking rate, local, involuntary variations in speaking rate occur during the course of a single utterance (Miller, Grosjean, & Lomanto, 1984). Local variations in speaking rate underlie the quasi-rhythmic (as opposed to perfectly rhythmic) pattern observed in natural speech (Tilsen & Arvaniti, 2013).

Behavioral work has shown that, during natural speech perception, listeners track both the global rhythmic structure of the input, that is, the overall rate at which the linguistic information is transmitted over time, as well as the local rate of change in linguistic information (Reinisch, 2016; Baese-Berk et al., 2014; Dilley & Pitt, 2010). This behavioral evidence suggests that tracking both these aspects of speaking rate is crucial for successful speech comprehension. Specifically, it has been proposed that tracking of the global speaking rate allows the listener to extract global information in the speech stream, including segmental, prosodic, and nonlinguistic information. These, in turn, can influence perception of local events in speech. On the other hand, tracking the local speaking rate has been suggested to contribute to forming expectations about upcoming events and to help resolve spectrally ambiguous cues in the speech signal. This tracking has been shown to significantly influence the recognition of both function and content words (e.g., Miller, Green, & Schermer, 1984; Summerfield, 1981; Lasky, Weidner, & Johnson, 1976; Liberman, Delattre, Gerstman, & Cooper, 1956). However, the neural correlates of the tracking mechanism during the perception of natural, connected speech remain unknown.

Neuroimaging studies suggest that the cortex tracks incoming acoustic speech signals, which manifests as

Aalto University

Journal of Cognitive Neuroscience 30:11, pp. 1704–1719
doi:10.1162/jocn_a_01295

coupling between cortical signals and the amplitude envelope of the acoustic speech signal (acoustic amplitude envelope; Gross et al., 2013; Peelle, Gross, & Davis, 2013). Nevertheless, existing evidence that cortical signals track the global speaking rate stems from studies focusing on the perception of single sentences with artificially manipulated global speaking rate (Hertrich, Dietrich, & Ackermann, 2013; Ahissar et al., 2001). Moreover, the cortical tracking of local variations in speaking rate in spontaneously produced speech has not been examined as of present (but see Kayser, Ince, Gross, & Kayser, 2015, for a study of phase-locking dynamics when artificially inducing local speaking rate variations in a read-aloud text). These studies have presented evidence that the coupling between brain signals and the acoustic amplitude envelope is modulated as a function of global speaking rate and local variations of speaking rate. However, these studies have employed non-naturalistic linguistic stimuli, and the variations in speaking rate were generated artificially. We suggest that this kind of stimuli, with their quite regular speaking rate and thus salient rhythmic patterning, do not represent instances of real-life language use. Moreover, such stimuli are presented in a highly controlled experimental setting, where the speech segments appear from complete silence and at regular intervals. This is at odds with the considerably quasi-rhythmic rhythm in spontaneously produced speech encountered in everyday life. Finally, artificial increases in speaking rate have been shown to result in linear changes in the duration of linguistic units, whereas in spontaneously produced speech increases in speaking rate are carried out in a nonlinear manner; this, in turn, affects speech comprehension (Janse, 2004; Janse, Nooteboom, & Quené, 2003).

This study aims to provide a cortical-level characterization of the behaviorally observed tracking of speech rhythm during the perception of continuous, spontaneously produced speech, as we encounter it in real-life communicative situations. We assume that this tracking would be neurally implemented through the coupling of cortical signals with acoustic signals, and in line with behavioral evidence, it would be sensitive to both global speaking rate and local variations in speaking rate. Motivated by our previous work focusing on both behavioral and spectral analysis of the acoustic amplitude envelope (Alexandrou et al., 2016), as well as by empirical evidence (Park, Ince, Schyns, Thut, & Gross, 2015; Doelling, Arnal, Ghitza, & Poeppel, 2014; Peelle et al., 2013) and theoretical models (Ghitza, 2011, 2012; Giraud & Poeppel, 2012), we hypothesized that this tracking would occur at frequency ranges corresponding to the word (2–4 Hz) and syllable (4–7 Hz) production frequencies. Although the lower delta band (0.5–2 Hz) would have been potentially of interest (e.g., Molinaro, Lizarazu, Lallier, Bourguignon, & Carreiras, 2016; Vander Ghinst et al., 2016; Bourguignon et al., 2013), we chose to focus on the 2–4 Hz and 4–7 Hz bands for two reasons: First, in the acoustic amplitude, the frequency range of the lower delta band (0.5–1 Hz) corresponds to fluctuations in timescales that are mainly associated with prosodic features in a speech stream (Bourguignon et al., 2013; Poeppel, Idsardi, & van Wassenhove, 2008); these features are presumably not modulated as a function of speaking rate (Ramus, Nespor, & Mehler, 1999). Second, the power spectrum of cortical activity in the lower delta band is characterized by a notable 1/f trend (pink noise; Voss & Clarke, 1975), which would potentially confound subsequent data analyses (e.g., Demanuele, James, & Sonuga-Barke, 2007).

In this study, cortical signals were recorded with magnetoencephalography (MEG) while participants listened to natural connected speech stimuli at three global speaking rates: slow, normal (habitual), and fast. Audio–MEG coherence was used to quantify how neural signals from across the cortex track the acoustic amplitude envelope of the perceived speech stimuli. Reaching beyond single sentences, read-aloud text passages, or rehearsed narratives used in earlier related studies, the present stimuli consisted of unrehearsed, spontaneously produced connected speech, featuring natural variations in the local speaking rate. To our knowledge, there are no reports on the coupling of cortical with acoustic signals during perception of spontaneously produced speech. Spontaneously produced speech differs from the isolated sentences, read-aloud and rehearsed speech stimuli used by previous studies with regard to, for instance, articulation rate (Jacewicz, Fox, O'Neill, & Salmons, 2009), articulatory patterns (Finke & Rogina, 1997), and prosodic features (Nakajima & Allen, 1993). These features influence, in turn, the structure of the acoustic amplitude envelope, which is one main factor that affects the coupling between acoustic and cortical signals. In addition, behavioral evidence suggests that speakers are able to distinguish between spontaneously produced and rehearsed or read-aloud speech stimuli (Chawla & Krauss, 1994), and the type of stimulus—rehearsed or spontaneously produced—shapes subsequent speech comprehension (Brennan & Schober, 2001). Therefore, this study is aimed at elucidating, for the first time, the coupling patterns that emerge during perception of real-life speech and how this coupling is modulated by one of the most important features of running speech: speaking rate. Importantly, the modulations in global speaking rate employed in this study were carried out naturally, without a metronome, and the syllable production frequency for stimuli at all three rates fell within the natural range of syllable production frequencies (see, e.g., Tsao & Weismer, 1997).

We propose that coherence between audio and MEG signals is a neural mechanism that contributes to natural speech comprehension by aiding listeners to track both the global speaking rate and local speaking rate variations. We anticipate a spatial differentiation between the audio–MEG coherence patterns associated with tracking the global speaking rate and the local variations

in speaking rate. Based on previous neuroimaging evidence (Hertrich et al., 2013; Ahissar et al., 2001), we hypothesize that the tracking of global speaking rate engages the auditory cortex bilaterally. Regarding local speaking rate variations, behavioral evidence suggests that such tracking is inherently predictive in nature (Brown, Dilley, & Tanenhaus, 2012; Koreman, 2006). We thus hypothesize that local speaking rate variations would highlight the auditory regions, which have been suggested to be sensitive to temporal regularities in speech stimuli (ten Oever et al., 2017), as well as additional regions, predominantly located in the parietal lobe, that have been suggested to contribute in shaping and updating internal predictive models (Bekinschtein et al., 2009; Andersen & Buneo, 2002).

METHODS

Participants

The participants were 20 healthy, right-handed, native Finnish-speaking adults (11 women, 9 men; mean age = 24.5, range = 19–35 years) with normal hearing. Sample size was determined based on the previous observation that, for more experimentally controlled continuous tasks, a sample size of 10 can be sufficient to examine coherent cortical coupling (Saarinen, Jalava, Kujala, Stevenson, & Salmelin, 2015). In this study, we opted to double that sample size, estimating that it would be sufficient to account for the possibly reduced effect size in more naturalistic tasks. 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.

Stimuli

The participants listened to six 40-sec segments (4 min) of connected speech stimuli spontaneously produced by an unfamiliar to them, untrained male speaker at normal, slow, and fast speaking rates. Speech production was prompted by questions (in Finnish) derived from the following thematic categories: 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. What kind of foods do you like?). The speaker responded to 18 unique thematic questions, six at each speaking rate. The function of the thematic questions was to help the speaker verbalize his own thoughts; he was not required to provide a specific response to each question. Instead, the primary aim was to help the speaker produce fluent, uninterrupted speech at each speaking rate. For the slow-rate, the speaker was asked to reduce his normal speaking rate by 50% by preferably increasing articulation time rather than the length of pauses. For the fast-rate, he was instructed to produce fluent, continuous speech at the highest speaking rate possible while preserving speech intelligibility and minimizing the occurrence of articulatory errors. During speech production, the speaker varied his speaking rate without the aid of any external pacing device. The speech stimuli were recorded in a soundproof room. Raw acoustic signals were collected with a dual diaphragm condenser microphone (B-2 PRO, Behringer) at a 44.1-kHz sampling frequency using Cool Edit 2000 (Syntrillium; see Figure 1 for excerpts of acoustic speech signals at each speaking rate). All stimuli were normalized to the same average intensity using Praat software (Institute of

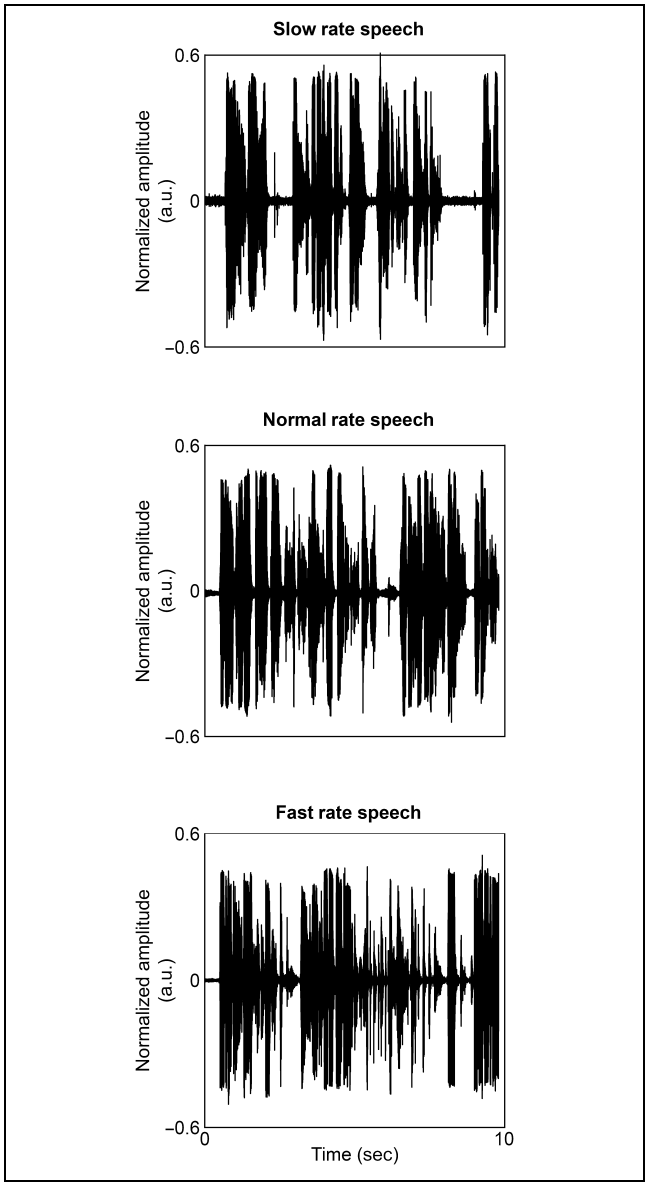

Figure 1. Examples of auditory stimulus waveforms for slow-rate (top), normal-rate (middle), and fast-rate speech (bottom). Normalized amplitude (in arbitrary units; *y* axis) is plotted against time (in seconds; *x* axis). Each plot displays a 10-sec chunk of data taken from a 40-sec auditory stimulus.

Phonetic Sciences, University of Amsterdam). The acoustic signals recorded at each global speaking rate were transmitted to a transcribing company (Tutkimustie Oy, Tampere, Finland) for strict verbatim transcription (i.e., the acoustic signals were transcribed without editions or modifications). The stimuli can be made available upon request.

The mean syllable production frequency at each speaking rate was obtained by averaging the syllable production frequencies across the six responses at each rate. Mean (±*SD*) word production frequencies were 1.8 ± 0.2 Hz for the normal-rate, 1.0 ± 0.1 Hz for the slow-rate (58% of normal), and 2.8 ± 0.2 Hz for the fast-rate (157% of normal). Mean (±*SD*) syllable production frequencies were 4.7 ± 0.5 Hz for the normal-rate, 2.6 ± 0.3 Hz for the slow-rate (55% of normal), and 6.8 ± 0.5 Hz for the fast-rate (148% of normal). Henceforth, the term speaking rate will refer to syllable production frequency. Short pauses in speech were indicated with commas in the transcriptions, allowing a quantitative assessment of the pauses made by the male speaker at each global speaking rate. The mean pause frequency at each global speaking rate was first computed by averaging across the six responses; subsequently, the normalized pause frequency was obtained by dividing the resulting value by the mean syllable production frequency. Mean (±*SD*) normalized pause frequency was 0.07 ± 0.2 for the normal-rate, 0.1 ± 0.2 for the slow-rate, and 0.07 ± 0.1 for the fast-rate. ANOVA for nonparametric data (Friedman test) revealed no significant differences in mean normalized pause frequency across speaking rates, $\chi^2(2) = 4.3$, $p = .12$. A qualitative evaluation of the transcriptions further confirmed that the speaker was able to produce fluent, connected speech with similar pronunciation and without repetitions or excessive use of filler words at all three global speaking rates.

The transcriptions included time stamps every 5 sec, allowing the estimation of syllable production frequencies in 48 separate time segments per global speaking rate. According to these segment-wise production frequencies, the local speaking rate varied as a function of time at each global speaking rate (Figure 2). The segment-wise production frequencies enabled the identification of the time segments of relatively constant speaking rate and changing speaking rate in the spontaneously produced speech stimuli by calculating the first derivative of the local speaking rate. First, the absolute difference in syllable production frequency between consecutive 5-sec segments was computed within each of the six blocks in each global speaking rate. This yielded seven local speaking rate variation values per block. From these seven local speaking rate variation values per block, we identified the two smallest values (indicating the two 5-sec segments in which the speaking rate had remained the most constant relative to the preceding segment; assigned to the constant-rate category) and the two largest values (indicating the two 5-sec segments in which the speaking rate had changed the most relative to the preceding segment; assigned to the changing-rate category). Subsequently, the 5-sec segments assigned to each category based on this behavioral analysis were matched to the corresponding time points in the acoustic and MEG signals. Data from all three global speaking rates were pooled together, yielding a total of thirty-six 5-sec segments of MEG data (180 sec in total) per participant for each category (except for one participant for whom data from one block of fast-rate speech perception were missing due to a technical issue; for this participant, only thirty-four 5-sec segments of MEG data were available in each category).

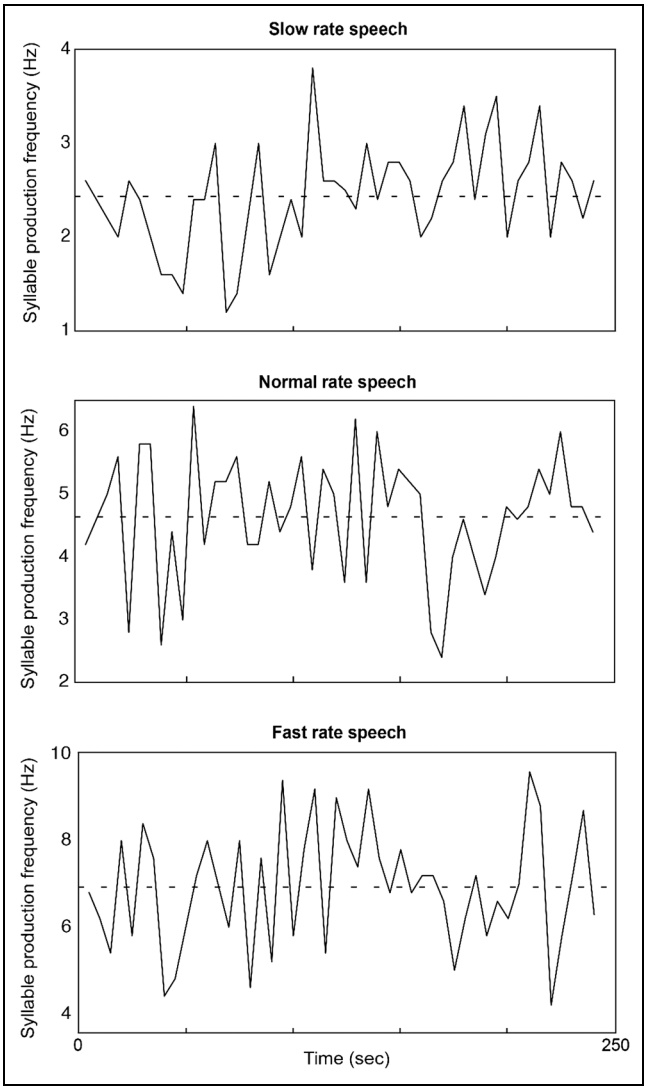

Figure 2. Local variation of speaking rate at different global speaking rates. Syllable production frequencies were estimated for each 5-sec segment (in Hz; *y* axis) across the entire 4 min of data at each speaking rate and are plotted against time (in seconds; *x* axis). The dashed line represents the mean syllable production frequency for the slow (top), normal (middle), and fast (bottom) speaking rate.

The local speaking rate varied between 1.2 and 3.8 Hz for slow-rate speech (mean ± *SD*, 2.5 ± 0.6 Hz), between

2.4 and 6.4 Hz for normal-rate speech (4.6 ± 1.0 Hz), and between 4.0 Hz and 9.4 Hz for fast-rate speech (6.7 ± 1.3 Hz; also see Figure 2). The local variation in speaking rate, from one segment to the next, ranged from 0 to 1.8 Hz (absolute values) for slow-rate speech, from 0 to 3.4 Hz in normal-rate speech, and from 0.4 to 4.6 Hz for fast-rate speech. The mean syllable production frequency did not differ between the constant-rate category (4.8 ± 2.1 Hz) and the changing-rate category (4.6 ± 2.3 Hz; W = 0.001, $p = .46$, Wilcoxon signed-rank test), nor did the distributions of normalized variations differ between the three global speaking rates (normal-rate speech vs. slow-rate speech, $D_{42,\ 42} = 0.10$, $p = .99$; normal-rate speech vs. fast-rate speech, $D_{42,\ 42} = 0.19$, $p = .39$; slow-rate speech vs. fast-rate speech, $D_{42,\ 42} = 0.19$, $p = .39$, two-sample Kolmogorov–Smirnov test).

Experimental Procedure

A single speech perception block consisted of a recorded thematic question spoken by a male voice (duration = 3–9 sec; mean ± *SD,* 5.6 ± 1.3 sec), a 1-sec delay before response onset, a 40-sec speech stimulus (i.e., the unknown male speaker's response to the thematic question), and a 2.5-sec rest period between blocks. A signal tone (50-msec, 1-kHz tone) indicated the beginning of a block, and another signal tone (50-msec, 75-Hz tone) signified the beginning and end of the response. The speech stimuli were grouped into three experimental conditions according to speaking rate (normal, slow, and fast). There were six blocks per experimental condition. Each 40-sec speech stimulus was presented only once to avoid learning effects. The order of the experimental conditions was randomized across participants.

During the experiment, participants were instructed to keep their gaze on a fixation point projected on a screen at ~1 m from their sitting position. Head position was assessed throughout the experiment by observing the participants' head position through a video connection with the MEG room, by examining the measured MEG data for motion-related artifacts in real time and by reminding the participants between conditions, via intercom, to keep their head position constant and avoid, for example, slouching. Stimuli were presented binaurally at an individually adjusted comfortable listening level through plastic tubes and intracanal earpieces. The participants' task was to listen attentively to each stimulus. Task compliance was evaluated by inserting a 2-sec repetitive auditory segment (repeated four times, thus sounding like a broken record) in one of the six stimuli of each experimental condition. The repetitive segment occurred at a random time point during the 40-sec stimulus. The participant was instructed to indicate occurrence of a repetitive segment with an index finger lift, using an optical response panel. To further verify that the participants attended to the speech stimuli, at the end of the experiment they were asked to fill in a surprise multiple-choice questionnaire regarding the content of the stimuli they had heard. After the last block of each condition, participants were asked to rate the overall intelligibility of the six stimuli they had just heard on a scale of 1–10 (1 = *completely unintelligible*, 10 = *completely intelligible*). The effect of global speaking rate on the mean stimulus intelligibility scores (obtained by averaging the individual intelligibility scores across the 20 participants) was tested by ANOVA for nonparametric data (Friedman test).

Recordings

MEG signals were recorded with a 306-sensor (204 gradiometers, 102 magnetometers) Neuromag Vectorview whole-head device (Elekta Oy) in a magnetically shielded room at Aalto NeuroImaging MEG Core. Data were filtered at 0.03–500 Hz and sampled at 1500 Hz. The participants were seated, with the head covered by the MEG helmet. Each participant's head position with respect to the MEG sensor array was determined by attaching five head position indicator coils to the scalp and briefly energizing them before the measurement. The coil locations were determined in reference to anatomical landmarks (nasion and right/left preauricular points) using a 3-D digitizer (Isotrak 3S1002, Polhemus Navigation Science). Blinks and eye movements (saccades) during the MEG measurement were monitored using EOG. Structural MRIs (3T Siemens MAGNETOM Skyra, Siemens Medical Systems) at Aalto NeuroImaging Advanced Magnetic Imaging Center were obtained for each participant after the MEG measurement using a high-resolution T1-weighted 3-D MPRAGE scan (32-channel head coil, 176 slices with 1-mm slice thickness, voxel size = 1 mm × 1 mm × 1 mm, 7° flip angle, 1100-msec inversion time, repetition time/echo time = 2530/3.3 msec). During the analysis process, the MEG coordinate system was aligned with individual MRIs based on head position coils and anatomical landmarks using MRI lab software (Elekta Oy).

Acoustic Signal Analysis

The amplitude envelope of the 4-min-long acoustic signal recorded at each global speaking rate was computed by full-wave rectifying and low-pass filtering (<10 Hz, fourth-order Butterworth filter, forward and backward) the band-pass filtered acoustic signal (80–2500 Hz, fourth-order Butterworth filter, forward and backward). The signal was filtered in this frequency range to emphasize the voiced signal portions that encode the rhythmic features of speech (Alexandrou et al., 2016; Hertrich et al., 2013). The spectrum of the downsampled (by a factor of 10), Tukey-windowed ($r = .2$), and zero-padded envelope was calculated by taking the squared magnitude of the fast Fourier transform using an 8192-point window (for more details, see Alexandrou et al., 2016). The mean rate around which the acoustic amplitude envelope fluctuated as a function of time reflected the global speaking

rate while the quasi-regular pattern of the amplitude envelope fluctuations captured the local variations in speaking rate (5-sec long example; Figure 3, left). The acoustic power spectra estimated across the whole 4-min of data featured salient spectral peaks that approximately aligned with word and syllable production frequencies (Figure 3, right). It is noteworthy that power spectral peaks associated with prosodic frequencies (i.e., <1 Hz) did not demonstrate any shifts as a function of global speaking rate, possibly indicating similar prosodic patterning across stimuli at all three speaking rates.

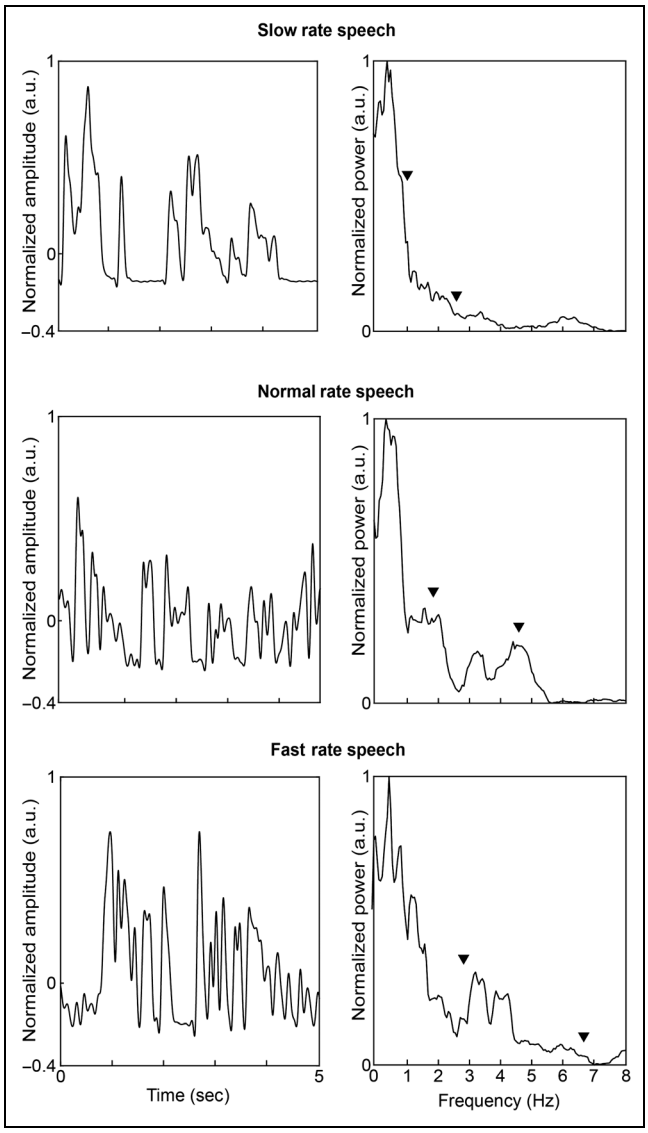

Figure 3. Global speaking rate and the rhythmic structure of acoustic speech signals. Data for slow-rate speech are shown on the top, data for normal-rate speech are shown in the middle, and data for fast-rate speech stimuli are shown at the bottom. Left: The acoustic amplitude envelope (normalized amplitude in arbitrary units; *y* axis) is plotted against time (in seconds; *x* axis). Each plot displays a 5-sec chunk of data taken from a 40-sec auditory stimulus. Right: Acoustic power spectra computed across the 4 min of data at each speaking rate. Normalized power (in arbitrary units; *y* axis) is plotted against frequency (in Hz; *x* axis). The scale of the *x* axis (0–8 Hz) was chosen to cover the range of syllable production frequencies at the different speaking rates. The arrowheads indicate the mean word and syllable production frequencies at each global speaking rate.

MEG Data Analysis

Only gradiometers were included in MEG data analysis as they have a narrow spatial sensitivity pattern and are optimal for recording data from superficial sources; in contrast, magnetometers more readily pick up signals from distant sources, including external artifacts. First, MaxFilter software (Elekta Oy) was used to remove external disturbances from the MEG data with spatiotemporal signal space separation (Taulu & Simola, 2006). Subsequently, blink artifacts were removed from MEG signals using a PCA-based routine (Uusitalo & Ilmoniemi, 1997) implemented in Graph software (Elekta Oy).

The cortical areas showing coherent activity with the amplitude envelope of the acoustic signals were estimated with dynamic imaging of coherent sources (DICS; Gross et al., 2001) using a spherical head model. In DICS, spatial filtering is employed to estimate oscillatory power and coherence in the brain based on a cross-spectral density (CSD) matrix, which represents the oscillatory components and their linear dependencies. DICS is well suited for performing source modeling of continuous MEG data recorded during complex cognitive tasks (Alexandrou, Saarinen, Mäkelä, Kujala, & Salmelin, 2017; Saarinen et al., 2015; Kujala et al., 2007); in addition, it has proven its efficiency for examining coherence between MEG and acoustic speech signals (Peelle et al., 2013).

DICS analysis and the subsequent computation of audio–MEG coherence were carried out in MATLAB software (The MathWorks, Inc.) using custom-made scripts. The covariance computation was done in the form of CSD matrices, which were computed between all planar gradiometer MEG signals and the amplitude envelope of the acoustic signal using Welch's averaged periodogram method (4096-point Hanning windowing, 50% window overlap, 4096-point fast Fourier transform, 0.4 Hz resolution) separately for 10 frequency bins (starting from 1 Hz up to 10 Hz, 1 Hz spacing, 2 Hz spectral width). CSD matrices were calculated for the MEG data recorded during perception of speech at each of the three global speaking rates, as well as for the MEG data in the constant-rate and changing-rate speech categories. Based on these CSD matrices, the cortical-level DICS-based estimates of audio–MEG coherence were computed separately for each frequency bin in a spatially equivalent search grid across participants. The grid sampled the gray matter surface, excluding the cerebellum (20482 points, atlas brain, Freesurfer 5.3; Fischl, 2012). This common grid was transformed to each participant's anatomy via a surface-based transformation (Fischl, Sereno, Tootell, & Dale, 1999). The DICS estimation used a regularization of

0.01% of the maximum eigenvalue of each frequency and condition-specific CSD matrix.

Subsequent statistical tests were carried out using IBM SPSS statistics (IBM) and MATLAB software (The MathWorks, Inc.). The effect of the global speaking rate on the relationship between cortical signals and the acoustic amplitude envelope of the perceived speech signals was evaluated by examining differences in audio–MEG coherence between the main experimental conditions (normal-rate speech vs. slow-rate speech, normal-rate speech vs. fast-rate speech). The effect of local variations in speaking rate on the relationship between cortical signals and the acoustic amplitude envelope of the perceived speech signals was evaluated by examining differences in audio–MEG coherence between the constant-rate speech and changing-rate speech. For group-level statistics, the audio–MEG coherence maps were averaged across the 2–4 Hz and 4–7 Hz frequency bins. The behaviorally estimated syllable production frequencies, spanning the 2–7 Hz range across the three global speaking rates, prompted us to focus subsequent statistical analysis on delta-band (2–4 Hz) and theta-band (4–7 Hz) cortical activity. Moreover, these frequencies have been extensively linked with cortical tracking of speech stimuli (e.g., Kayser et al., 2015; Peelle et al., 2013; Luo & Poeppel, 2007).

Statistical significance was determined using group-level cluster-based statistics controlling for multiple comparisons (cluster-based permutation procedure performed on the statistically significant results obtained from a Student's two-tailed t test for paired samples, 10,000 permutations, statistical significance threshold $p < .05$, family-wise error corrected, cluster threshold $p < .05$, weighted distance algorithm for linking adjacent grid points, 15-mm cutoff threshold for cluster size; Maris & Oostenveld, 2007). In accordance with the procedure described in Maris and Oostenveld (2007), the t values were summed within spatially contiguous clusters (for adjacent voxels with $p < .05$). For each round of the permutation testing, the labels of the two conditions being compared were randomized across participants, and new t statistics were computed in all grid points. For each permutation, the largest cluster t value was collected, yielding a distribution of 10,000 cluster-level t values. Subsequently, the original t statistics were compared with this distribution. An effect was considered significant if the cluster p value exceeded the 95% threshold of the permuted maximum cluster t scores. For each cortical region demonstrating statistically significant differences in audio–MEG coherence, we obtained the Talairach coordinates and Brodmann's area numbers of the center of the region using the Talairach Daemon (Lancaster et al., 2000).

The spatial and spectral specificity of the observed effects was further explored by plotting the frequency spectra in the 1–10 Hz frequency range for each contrast (normal-rate speech vs. slow-rate speech, normal-rate speech vs. fast-rate speech, constant-rate speech vs. changing-rate speech). Audio–MEG coherence spectra in the contrasted conditions were compared in each frequency bin using Student's two-tailed t test (statistical significance threshold $p < .05$, 19 df, in spatially combined clusters in the case of close-by significant clusters).

To further examine the possibility that the modulation in audio–MEG coherence for the constant-rate speech versus changing-rate speech reflects dynamic tracking of local speaking rate, we computed the across-subject standard deviation (*SD*) of the cosine of the instantaneous phases of cortical activity for the constant-rate and changing-rate speech categories. This was done to describe the origin and nature of the observed modulations in audio–MEG coherence and further explore the hypothesis that tracking of the local variations in speaking rate is predictive in nature. Because this kind of analysis is statistically dependent on the results of the contrast between constant-rate speech and changing-rate speech, it was made for illustrative purposes and not intended to yield novel findings. The number of 5-sec segments considered in this analysis was 34, determined by the minimum number of such segments identified in individual participants. Across-subject *SD* was the measure of choice because, in contrast to audio–MEG coherence, it provides a reliable estimate of the phase alignment of cortical activity across subjects even for MEG data sections as short as 5 sec. The segment-wise mean *SD* values, obtained by averaging *SD* values across the time bins in each 5-sec segment, were computed for cortical activity in 10 frequency bins in the 1–10 Hz frequency range (1 Hz spacing, bin width ±0.5 Hz). We compared *SD* values between constant-rate speech and changing-rate speech in each frequency bin using a Student's two-tailed t test for independent samples (statistical significance threshold $p < .05$, 66 df). The range of phase variability in the constant and changing-rate speech category was computed as the difference between the maximum and the minimum phase variability. The mean difference in phase variability between constant-rate speech and changing-rate speech was quantified by averaging the difference in mean *SD* across the thirty-four 5-sec MEG data segments included in the analysis.

RESULTS

Behavioral Results

Attentional Control Tasks

For all three experimental conditions, all participants (20 of 20) were able to detect the repetitive segments embedded in one of the stimuli. The answers to the surprise questionnaire presented at the end of the experiment regarding the content of the speech stimuli were 100% correct for all 20 participants.

Intelligibility Scores

The mean (±*SD*) intelligibility scores were 9.9 ± 0.5 for the normal-rate speech stimuli, 9.9 ± 0.3 for the slow-rate speech stimuli, and 9.5 ± 0.7 for the fast-rate speech stimuli (a score of 10 representing perfect intelligibility). There were no significant differences in intelligibility scores between different rate speech stimuli, $\chi^2(2) = 2$, $p = .37$ (Friedman test).

Effect of Global Speaking Rate on Audio–MEG Coherence Patterns

Global speaking rate was found to be associated with modulations in audio–MEG coherence in the temporal regions bilaterally, as well as in the right paracentral lobule and the right parietal region (Figure 4A). Specifically, there was significantly stronger audio–MEG coherence for perceiving normal-rate speech than fast-rate speech in the superior temporal areas bilaterally (left: −61, −15, 1; BA 22, and right: 61, −9, 1; BA 22) in the 2–4 Hz frequency range (left: cluster p value = .02, cluster size = 140; right: cluster p value = .01, cluster size = 135; cluster threshold $p < .05$, 10,000 permutations) as well as in the right paracentral lobule (7, −33, 67; BA 6) in the 4–7 Hz frequency range (cluster p value = .03, cluster size = 128, cluster threshold $p < .05$, 10,000 permutations; Figure 4A, left). Furthermore, significantly increased audio–MEG coherence for perception of slow-rate than normal-rate speech was observed in the right parietal region (50, −38, 40; BA 40) in the 4–7 Hz frequency range (Cluster 1: cluster p value = .03, cluster size = 134; Cluster 2: cluster p value = .01, cluster size = 144; cluster threshold $p < .05$, 10,000 permutations; Figure 4A, left).

Effect of Local Variations in Speaking Rate on Audio–MEG Coherence Patterns

Local variations in speaking rate were associated with modulation of audio–MEG coherence in the left hemisphere, where significantly stronger audio–MEG coherence was observed for constant-rate speech than changing-rate speech in the 4–7 Hz frequency range of interest (Figure 4B). Specifically, this effect encompassed the left parietal region, particularly highlighting the left postcentral gyrus (−44, −24, 37; BA 2) as well as the left inferior parietal lobule (−33, −20, 41; BA 3; Cluster 1: cluster p value = .03, cluster size = 132; Cluster 2: cluster p value = .04, cluster size = 131; cluster threshold $p < .05$, 10,000 permutations; Figure 4B, left).

Spatial and Spectral Specificity of the Observed Modulations in Audio–MEG Coherence

Next, we examined the spatial and spectral specificity of the observed modulations in audio–MEG coherence. The audio–MEG coherence spectra were plotted in the 1–10 Hz frequency range for each contrasted condition in each of the cortical regions in which modulations of audio–MEG coherence were observed (Figure 4A and B, right). Paired t tests on the coherence spectra in these ROIs confirmed that the modulations of audio–MEG coherence determined in the 2–4 Hz and 4–7 Hz bands were spatially and spectrally specific. In the left and right superior temporal region, in the right paracentral lobule, as well as in the right and left parietal region, significant differences in audio–MEG coherence (based on the paired t tests; shown as vertical lines in the spectral plots) were observed primarily for the contrast for which the statistically corrected effects had been first discovered (cf. Figure 4A and B, left), and only rather sporadic (i.e., mostly observed in a few isolated frequencies) differences in audio–MEG coherence occurred for the other two contrasts. Significant modulations in audio–MEG coherence were also limited to the frequency ranges in which these effects had been first identified. Specifically, in the left superior temporal region, the paired t tests revealed differences in audio–MEG coherence only for the normal-rate speech versus fast-rate speech contrast (2 Hz: $p = .03$, 3 Hz: $p = .01$, 4 Hz: $p = .02$). In the right superior temporal region, differences were found for the normal-rate speech versus slow-rate speech contrast (1 Hz: $p = .03$, 8 Hz: $p = .003$, 9 Hz: $p = .02$) and for the normal-rate speech versus fast-rate speech contrast (1 Hz: $p = .003$, 2 Hz: $p = 5 \times 10^{-4}$, 3 Hz: $p = .003$, 4 Hz: $p = .03$). In the right parietal lobule, differences were observed for all three contrasts (normal-rate speech vs. slow-rate speech, 7 Hz: $p = .02$; normal-rate speech vs. fast-rate speech, 4 Hz: $p = .01$; constant-rate speech vs. changing-rate speech, 2 Hz: $p = .01$). In the right parietal region, differences were observed only for the normal-rate speech versus slow-rate speech contrast (3 Hz: $p = .01$, 4 Hz: $p = .01$, 5 Hz: $p = .02$, 6 Hz: $p = .01$, 7 Hz: $p = .01$). Finally, in the left parietal region, differences were found only for the constant-rate speech versus changing-rate speech contrast (4 Hz: $p = .02$, 5 Hz: $p = .01$, 6 Hz: $p = .02$; Figure 4A and B, left).

Dynamic Readjustment of Cortical Signal Phase in Response to Local Variations in Speaking Rate

The origin and potential dynamic nature of modulations in audio–MEG coherence in response to local variations in speaking rate was further described by computing the variability of the instantaneous phase of cortical signals originating from the left parietal region (Figure 4B, left) for each 5-sec segment included in the constant-rate and changing-rate speech categories (Figure 5). The mean *SD* was significantly larger ($p = .02$), indicating greater across-subject variability of the instantaneous phase of MEG signals, for perception of changing-rate speech (gray) than constant-rate speech (black) in the left parietal region (Figure 5). This effect was found for cortical activity in

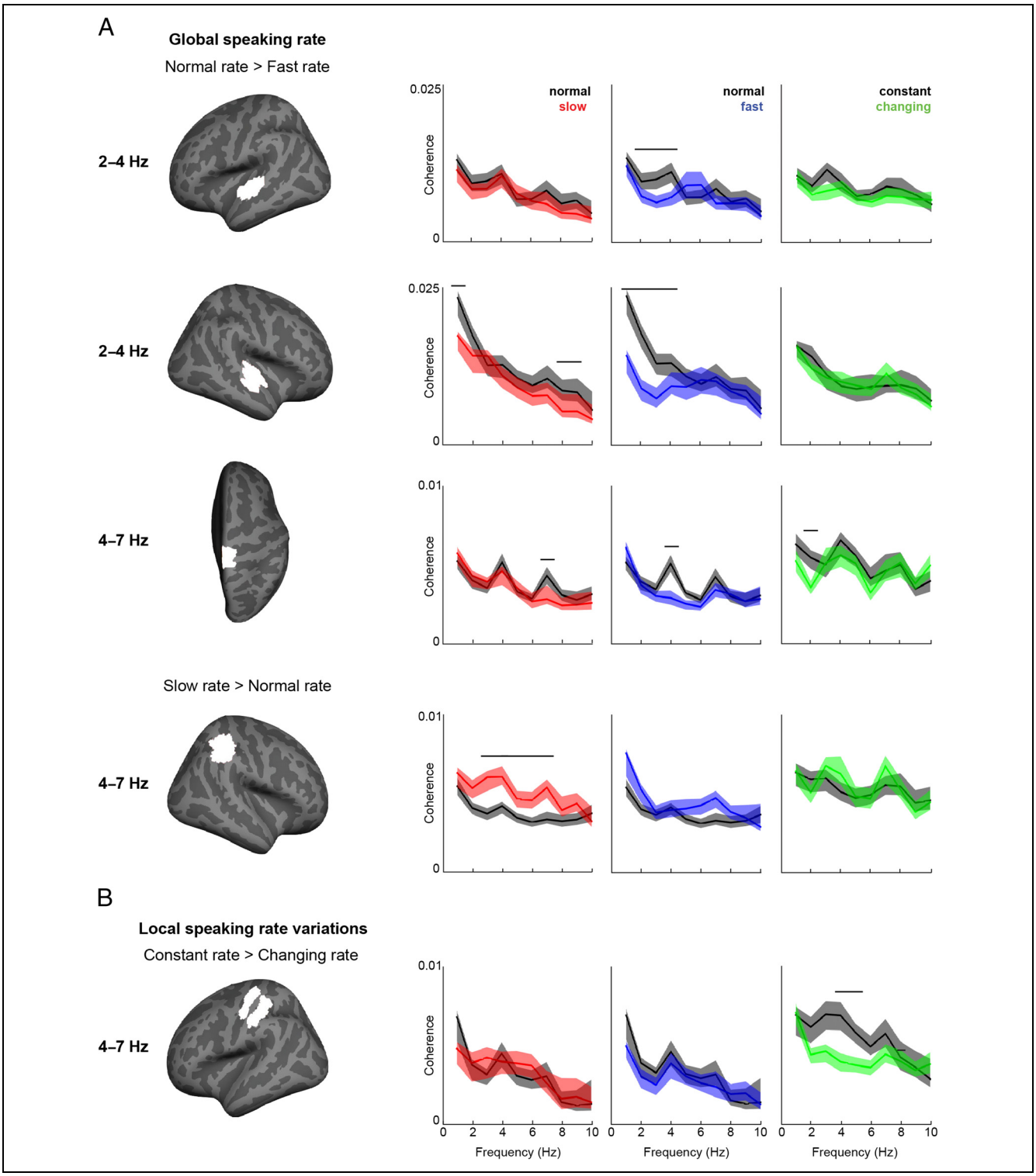

Figure 4. Cortical tracking of global speaking rate and to local variations of speaking rate. Cortical areas (left) demonstrating significant modulations (white color) in audio–MEG coherence at 2–4 Hz or 4–7 Hz as a result of variations in speaking rate and group-level wide-band audio–MEG coherence spectra (right) for these areas. (A) Effect of global speaking rate (normal-rate speech vs. fast-rate speech; normal-rate speech vs. slow-rate speech). (B) Effect of local variations in speaking rate (constant-rate speech vs. changing-rate speech). The frequency range in which significant group-level modulations in audio–MEG coherence (corrected for multiple comparisons) are observed is indicated to the left of each cortical surface image. Audio–MEG coherence (y axis) for each cortical area for which significant effects were observed, plotted as a function of frequency (0–10 Hz; x axis). Data from each contrast are shown in separate columns: on the left, normal-rate speech (black) versus slow-rate speech (red); in the middle, normal-rate speech (black) versus fast-rate speech (blue); on the right, constant-rate (black) versus changing-rate (green). The group-level mean audio–MEG coherence is represented by a solid line. The shaded area around each line demonstrates the *SD* of audio–MEG coherence values. The horizontal lines in the plots represent the frequency bins (width ±0.5 Hz) for which significant effects were observed based on a Student's paired t test with 19 *df* (uncorrected statistics).

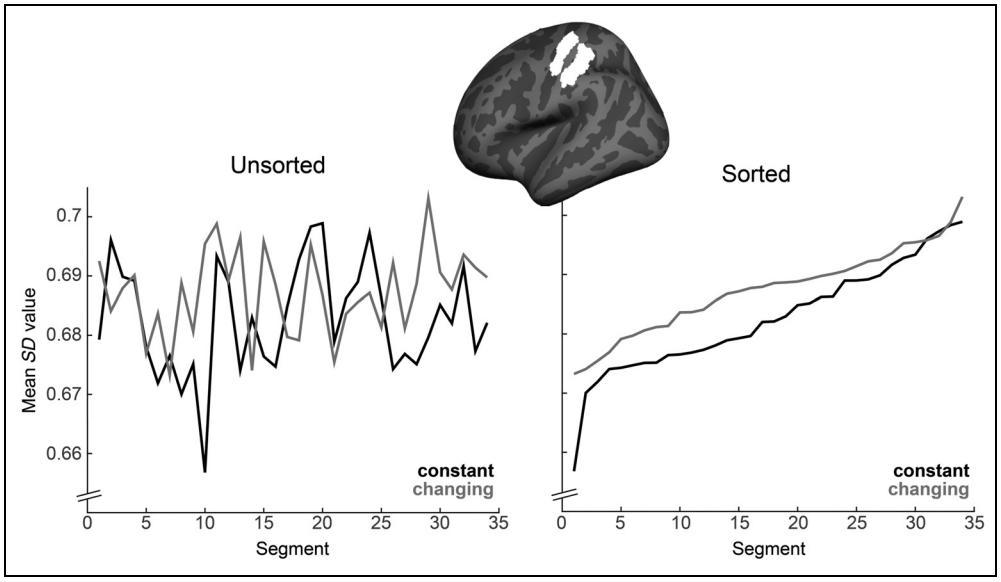

Figure 5. Variability of the instantaneous phase of cortical signals. Results are shown for the 6-Hz frequency bin of the cortical signals generated in the left parietal region (constant-rate > changing-rate speech; see Figure 4B). The mean *SD* values of the cosine of the instantaneous phase (*y* axis) of cortical signals are plotted for each 5-sec segment (*x* axis) for both the constant-rate (black) and changing-rate (gray) speech: unsorted data (left); data in ascending order, sorted by mean *SD* values (right). Mean *SD* was significantly larger for changing-rate speech than constant-rate speech in the 6-Hz frequency bin ($p = .02$).

the 6-Hz frequency bin, thus aligning with the 4–7 Hz frequency range in which modulations in audio–MEG coherence were initially observed (Figure 4B, left). Plotting the mean *SD* for the unsorted 5-sec segments (Figure 5, left) demonstrated that the difference in phase variability between constant-rate and changing-rate speech is quite subtle; this is also illustrated by the mean *SD* range (0.04 for constant-rate speech; 0.03 for changing-rate speech) and by the absolute average difference in mean *SD* between constant-rate speech and changing-rate speech (0.005). Yet, the data sorted in ascending order by the mean *SD* magnitude reveal that the statistically significant difference between constant-rate speech and changing-rate speech reflects a systematically higher phase variability for the changing-rate speech (Figure 5, right), indicating that cortical signal phase is dynamically readjusted in response to local variations in speaking rate.

DISCUSSION

The present results suggest that cortical signals track multiple features of speech rhythm. The global speaking rate (normal-rate speech vs. fast-rate speech, normal-rate speech vs. slow-rate speech) and local variations in speaking rate (constant-rate speech vs. changing-rate speech) were associated with modulations in audio–MEG coherence in the 2–4 Hz (delta) and 4–7 Hz (theta) frequency ranges. These modulations encompassed cortical regions in both hemispheres. Global speaking rate was associated with modulations in audio–MEG coherence in the temporal areas bilaterally, as well as in the right parietal and paracentral regions. Local variations in speaking rate were accompanied by modulations of audio–MEG coherence only in the left parietal region, instead. These spatially and functionally distinct cortical tracking patterns reveal a dual nature of the cortical tracking mechanism of speaking rate.

This study found evidence of two spatially and functionally distinct components of cortical tracking of speech rhythm. The first component is associated with modulations in audio–MEG coherence as a function of global speaking rate. In line with previous studies, there was an emphasis on the 2–4 Hz and 4–7 Hz frequency ranges, and the middle superior temporal regions bilaterally were highlighted (Park et al., 2015; Peelle et al., 2013; Hertrich, Dietrich, Trouvain, Moos, & Ackermann, 2012; Aiken & Picton, 2008; Luo & Poeppel, 2007). We propose that this effect may be conceptualized as evolutionary tuning, which signifies an innate preference in the motor system for certain frequencies of output (~5 Hz). This preference is thought to be reflected in the remarkable constancy of habitual speaking rates across languages (Liberman & Whalen, 2000) and has been suggested to have shaped a similar preference in the auditory system (Assaneo et al., 2016). The preference for the 5-Hz frequency has been proposed to have an evolutionary origin, stemming from primate oromotor vocalizations (Ghazanfar, Morrill, & Kayser, 2013; Morrill, Paukner, Ferrari, & Ghazanfar, 2012), and has been shown to be already present in infants (Telkemeyer et al., 2009). Notably, the global rate of the normal-rate speech stimuli employed here was ~5 Hz, aligning with this proposed innate preference of the motor system and hence with the habitual speaking rates across languages (Alexandrou et al., 2016; Ruspantini et al., 2012; Levelt, Roelofs, & Meyer, 1999). The present findings importantly also point to a preference for the habitual rate of input in the auditory modality: The audio–MEG coherence between auditory cortical activation and the external audio signal was enhanced for normal-rate speech compared with fast-rate speech, an effect thought to facilitate subsequent

processing (Schroeder, Wilson, Radman, Scharfman, & Lakatos, 2010; Lakatos, Karmos, Mehta, Ulbert, & Schroeder, 2008). This interpretation is in line with our earlier analysis of MEG signal power in this same data set, which found decreased power for global speaking rates deviating from the habitual rate (Alexandrou et al., 2017), as well as with the observation that normal-rate speech is processed even in the absence of attention (Wild et al., 2012).

Beyond our original hypothesis, cortical tracking of global speaking rate was additionally associated with modulations of audio–MEG coherence in the right paracentral lobule and right parietal cortex. We propose that this finding is linked to the reliance of comprehension of connected speech on the progressive integration and analysis of information (Bourguignon et al., 2013; Lieberman, 1963) related to semantic context (Federmeier, Wlotko, & Meyer, 2008; St. George, Kutas, Martinez, & Sereno, 1999) and prosody (e.g., Hari & Kujala, 2009; Belin, Fecteau, & Bedard, 2004; Kriegstein & Giraud, 2004). The observed effects in the right paracentral lobule and right parietal area may be seen as evidence of the general organization of integration of linguistic information in postcentral parietal regions (Sepulcre, Sabuncu, Yeo, Liu, & Johnson, 2012). Indeed, the right paracentral lobule has been associated with linking information in memory and gradually building comprehension when listening to a continuous story (Maguire, Frith, & Morris, 1999). Enhanced audio–MEG coherence in the right parietal region for the more slowly unfolding slow-rate speech compared with normal-rate speech may reflect the temporal flexibility of linguistic information processing and integration; this region has been shown to scale its activity to match global speaking rate (Lerner, Honey, Katkov, & Hasson, 2014; Small, Andersen, & Kempler, 1997).

The second component of cortical tracking of speech rhythm was associated with local variations in speaking rate. Although we initially hypothesized contributions of both the auditory cortices and the parietal regions, our empirical results exclusively emphasized the left parietal region. The left parietal region has been linked to predictive timing and the encoding of "when" something happens, a process that also interacts with attention (Arnal & Giraud, 2012; Schwartze, Tavano, Schröger, & Kotz, 2012; Nobre, Correa, & Coull, 2007). The emphasis on the 4–7 Hz frequency range is in line with reports suggesting that low-frequency cortical oscillations play a role in predicting future, behaviorally relevant cues (Calderone, Lakatos, Butler, & Castellanos, 2014; Arnal & Giraud, 2012; Saleh, Reimer, Penn, Ojakangas, & Hatsopoulos, 2010). Indeed, the presently observed modulation in audio–MEG coherence with respect to local variations in speaking rate was found to reflect a dynamic re-adjustment of cortical signal phase.

Based on this spatiospectral pattern, we propose that tracking local speaking rate is a process linked to the inherent predisposition of the brain to continuously seek and extract patterns of temporal regularity from the surrounding, ever-changing sensory environment. This predisposition may be conceptualized as predictive timing and predictive coding, a cortical operation mode that governs perceptual processes (Friston, 2012; Schroeder, Lakatos, Kajikawa, Partan, & Puce, 2008; Bar, 2007; Bonte, Mitterer, Zellagui, Poelmans, & Blomert, 2005) and supports subsequent sensory and cognitive processing (Sohoglu, Peelle, Carlyon, & Davis, 2012). This tracking has been suggested to operate in an anticipatory manner, in which the phase of low-frequency cortical activity is reset before the next relevant sensory event (Arnal & Giraud, 2012). Indeed, based on the present evidence suggesting a dynamic tracking of local variations in speaking rate, it appears that the brain actively makes predictions: the tracking of a quasi-regular stimulus (such as the acoustic amplitude envelope) could be highly predictive and quickly adjustable in nature. In this study, the emphasis on the left parietal region could presumably be associated with a predominant top–down nature of this predictive tuning during spontaneous speech perception (e.g., Andersen & Buneo, 2002; Engel, Fries, & Singer, 2001); future studies could further explore this topic. It has been reported that cortical signals track more accurately temporally regular visual (Cravo, Rohenkohl, Wyart, & Nobre, 2013) and auditory stimuli (ten Oever et al., 2017; Kayser et al., 2015). The presently observed enhanced audio–MEG coherence for the temporally regular constant-rate speech may thus be interpreted as a marker of successful temporal predictions, leading to more efficient subsequent processing (ten Oever et al., 2017; Rohenkohl, Cravo, Wyart, & Nobre, 2012; Schroeder et al., 2010; Lakatos et al., 2008).

Considering the present spatiospectral patterns as a whole, we observe dissociations of both spatial and spectral nature which could potentially reflect differences in cortical processing of natural connected speech. It was found that cortical signals from the right parietal region are sensitive to global speaking rate whereas cortical signals from the left parietal regions track local variations in speaking rate. This functional differentiation might be linked to different cortical processing modes in the right (ad hoc, integrative) and left hemisphere (post hoc, anticipatory) during natural, connected speech perception (Federmeier et al., 2008). Furthermore, this study extends the theoretical framework presented in Giraud and Poeppel (2012), as well as previous evidence gained through experimental paradigms based on the perception of isolated sentences, proposing that cortical signals originating mainly from the auditory regions track speech signals (Peelle & Davis, 2012). Thus, the findings of this study are of special interest: As the first report of entrainment patterns in the context of natural connected speech perception, it extends previous knowledge by suggesting that signals from both auditory and parietal regions track a perceived speech signal. Frontal and parietal regions have been suggested to exert a modulatory influence

on the relationship between cortical signals and acoustical speech signals (Keitel, Ince, Gross, & Kayser, 2017; Kayser et al., 2015; Park et al., 2015; Gross et al., 2013). This study further extends the role of these regions by showing that the bilateral parietal regions do not merely exert a modulatory influence on auditory regions but also directly contribute to tracking an incoming speech signal, aligning with recent findings (Puschmann et al., 2017).

Regarding the spectral aspect of the observed tracking of speaking rate, theoretical and empirical work both highlight the delta and theta frequency bands (e.g., Ghitza, 2012; Peelle & Davis, 2012), which are also observed in this study. Here, the emphasis on this frequency range may be related to exogenous parameters, namely, the frequency content of the stimuli and specifically the timescales of linguistic elements in speech: In the present stimuli, the word and syllable production frequencies spanned the range 2–7 Hz. In line with previous studies, we could potentially interpret the modulations in audio–MEG coherence in the delta band (2–4 Hz) as reflecting the processing of words, whereas modulations in coherence in the theta band (4–7 Hz) could be related to the processing of syllables. Providing support for this view, we found that local variations in speaking rate, which are usually assessed at the syllabic level due to the quick timescale on which they occur (Quené, 2007), were tracked by theta-band cortical activity. An alternative explanation may be related to an intrinsic cortical preference for certain timescales of processing, irrespective of the input. It is especially noteworthy that, in the auditory regions, modulations in audio–MEG coherence were observed in the delta band, whereas in the parietal regions and the right paracentral lobule such modulations were observed in the theta band. This observed dissociation regarding cortical signaling frequencies might provide evidence in favor of the existence of a hierarchy of timescales in the cortex during complex perceptual tasks (cf. Hasson, Yang, Vallines, Heeger, & Rubin, 2008; Kiebel, Daunizeau, & Friston, 2008).

Coherence describes the frequency–domain correlation of two time series. Increased coherence signifies enhanced synchronization between the two signals and, with regard to neural signaling, has been interpreted as more efficient information transfer between two neural populations (Fries, 2005). Here, we computed coherence between cortical signals and the acoustic amplitude envelope. In the context of this study, modulations in coherence signify changes in synchronization and thus in the tracking of a speech signal. Because we experimentally manipulated global speaking rate and also examined the inherently occurring local variations in speaking rate, modulations in coherence were considered as a direct consequence of these two kinds of variations in speaking rate. As reflected in the power spectra of the acoustic amplitude envelope of our stimuli, variations in speaking rate affect the temporal structure of the acoustic amplitude envelope which, from a computational perspective, is the main factor affecting coherence values. However, we also acknowledge that global speaking rate does not only affect the amount of linguistic content per time unit but it may also be accompanied by changes in the syntactic and semantic structure of an utterance (e.g., Cohen Priva, 2017). Variations in global speaking rate may also be associated with altered prosodic features (e.g., Fougeron & Jun, 1998), although this was likely not the case in this study as the low-frequency components of the acoustic amplitude envelope remained essentially constant. Local variations in speaking rate may, in turn, be associated with higher-level linguistic events in speech: For instance, speaking rate has been suggested to change locally at the phrasal level and co-occur with specific lexical events in speech (e.g., Byrd, Kaun, Narayanan, & Saltzman, 2000). Furthermore, it has been shown that coherence is not only dependent on the characteristics of an incoming stimulus but receives top–down modulations as a function of, for instance, expected syntactic structure (Meyer, Henry, Gaston, Schmuck, & Friederici, 2017). Therefore, even though we here base our interpretations primarily on the rate of speech, which was the variable specifically controlled in the present parametric experimental design, we cannot exclude the possibility that discourse-level features might also contribute to the presently observed coherence patterns. Future studies should further examine the origins of the modulation of audio–MEG coherence during perception of spontaneous speech at different rates.

Spontaneous speech, as opposed to single sentences, read-aloud text passages, or rehearsed narratives used in previous related studies, is the kind of speech we are mostly exposed to in real-life social situations. The presently employed experimental paradigm was designed to approximate a real-life listening context. Unlike previous studies examining the effect of global speaking rate on the relationship between cortical and acoustic signals (Hertrich et al., 2013; Ahissar et al., 2001), this study did not involve artificially time-compressed stimuli or impose external pacing on global speaking rate modulations. Because of the noncontrolled nature of the stimuli, the absolute audio–MEG coherence values were lower than those reported in earlier studies; nonetheless, the wide-band audio–MEG coherence spectra demonstrate that the observed effects are salient and occur at clearly delimited frequency ranges. Both theoretical models (Ghitza, 2011) and studies examining artificially time-compressed stimuli (Hertrich et al., 2013; Ahissar et al., 2001) propose that coherence peaks should shift as a function of global speaking rate. In this study, we did not observe such clear-cut effects in the wide-band audio–MEG coherence spectra. This may be due to natural speech showing some overlap in syllable production frequencies between global speaking rates as shown by a previous study from our research group that characterized speech rhythm through acoustic and EMG signals (Alexandrou et al., 2016). The study addressed

spontaneously produced speech at three global speaking rates (normal, slow, fast) in the same 20 individuals considered in this study. However, the fairly stable frequency of coherence peaks across the different global speaking rates may also suggest that, when tracking complex, spontaneously produced speech, the brain employs its own, internal signaling, instead of being passively driven by the frequency content of the input. Finally, stimulus intelligibility did not significantly vary across experimental conditions nor did the attention level of the participants, as assessed through the local attentional task of repetitive segment detection and the surprise multiple-choice questionnaire that probed the level of global stimulus comprehension. In addition, no significant differences were found either in the mean speaking rate between the constant and changing-rate categories or in the distributions of local speaking rate variations among the three global rates. Although it is not possible to completely rule out that attention to aspects of the speech stimulus that are not related to meaning per se may have affected the coherence between speech and cortical signals, we nevertheless suggest that the present findings reliably assess the neural correlates of the dual tracking mechanism of speaking rate during natural speech perception.

The observed audio–MEG coherence patterns offer new insights to the existing framework of cortical tracking of incoming speech signals as a mechanism supporting speech perception. In accordance with an emerging, more integrative view of speech processing (Alexandrou et al., 2017; Poeppel, Emmorey, Hickok, & Pylkkänen, 2012; Federmeier et al., 2008), cortical tracking of natural, connected speech extended beyond the previously reported emphasis on the temporal regions (Peelle et al., 2013; Ahissar et al., 2001) and was found to also engage the parietal regions bilaterally and the right paracentral lobule (see Giordano et al., 2017, for similar findings during perception of continuous, but not spontaneously produced speech). On a broader level, this finding supports the notion that perception is not a passive reaction to incoming stimuli but a highly constructive process that undergoes optimization in both the auditory regions, as well as in the parietal and paracentral regions (Morillon & Schroeder, 2015; Zion-Golumbic et al., 2013; Engel et al., 2001). Indeed, perception of real-life speech entails a continuous perceptual adjustment. First, we encounter a multitude of speakers on an everyday basis, with global speaking rates that demonstrate some individual variation around the habitual speaking rate of ~5 Hz (Alexandrou et al., 2016; Tsao & Weismer, 1997). This natural interindividual variability is analogous to the experimentally generated intraindividual variations in global speaking rate examined here. Tuning in to faster or slower speakers requires an active adaptation from the listener's part (Dupoux & Green, 1997). It could be suggested that tracking the global speaking rate (characteristic of a given speaker) would afford a solid sensory basis onto which the more fine-grained predictions related to local variations in speaking rate are embedded upon. In the auditory cortex, this presumed sensory basis has been thought to be represented by the frequency of neuronal spiking, which would be proportional to the speaking rate. For instance, a speaker demonstrating a high global speaking rate would induce faster neuronal spiking compared with a speaker who demonstrates a lower global speaking rate (Gütig & Sompolinsky, 2009). Second, building up-to-date predictions of future events based on previous input is paramount for adjusting to the dynamically changing, real-life communicative contexts, in which speech is typically heard only once (Schwartze et al., 2012). Thus, in this dynamic sensory mode, the robust evolutionary tuning is proposed to form a hard-wired perceptual basis, whereas predictive tuning continuously updates this basis as a function of expectation, attention, and sensory input.

Conclusion

This study is the first to examine the neural basis of tracking the speaking rate in perception of spontaneously produced natural connected speech. The present results indicate that cortical tracking of incoming speech is a multidimensional phenomenon. The tracking mechanism of speaking rate is dual in nature, manifesting two spatially and functionally distinct components of cortical tuning in speech perception: evolutionary tuning that is associated with global rhythmic structure and predictive tuning that is driven by local changes in speaking rate. These findings complement previous evidence and propose that, during the perception of spontaneously produced natural connected speech, the functional role of cortical tracking of the acoustic amplitude envelope is not merely confined to achieving syllabic segmentation of the input but extends to temporal predictions and expectations.

Acknowledgments

This work was financially supported by the Academy of Finland (Grants 255349, 256459, and 283071 to R. S. and Grant 257576 to J. K.), the Alfred Kordelin Foundation (Grant 160143 to A. A.), the Emil Aaltonen Foundation (Grant 170011 N1 to A. A.), the Finnish Cultural Foundation (Grant 00170944 to T. S.), and the Sigrid Jusélius Foundation (grant to R. S.). MEG and MRI data were recorded at the Aalto NeuroImaging research infrastructure.

Reprint requests should be sent to Anna Maria Alexandrou, Department of Neuroscience and Biomedical Engineering, Aalto-yliopisto Perustieteiden korkeakoulu, Rakentajanaukio 2C, Aalto, 00076, Finland, or via e-mail: anna.alexandrou@aalto.fi.

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. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 13367–13372.

Aiken, S. J., & Picton, T. W. (2008). Human cortical responses to the speech envelope. *Ear and Hearing, 29,* 139–157.
Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2016). A multimodal spectral approach to characterize rhythm in natural speech. *Journal of the Acoustical Society of America, 139,* 215–226.
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.
Andersen, R. A., & Buneo, C. A. (2002). Intentional maps in posterior parietal cortex. *Annual Review of Neuroscience, 25,* 189–220.
Arnal, L. H., & Giraud, A.-L. (2012). Cortical oscillations and sensory predictions. *Trends in Cognitive Sciences, 16,* 390–398.
Assaneo, M. F., Sitt, J., Varoquaux, G., Sigman, M., Cohen, L., & Trevisan, M. A. (2016). Exploring the anatomical encoding of voice with a mathematical model of the vocal system. *Neuroimage, 141,* 31–39.
Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science, 25,* 1546–1553.
Bar, M. (2007). The proactive brain: Using analogies and associations to generate predictions. *Trends in Cognitive Sciences, 11,* 280–289.
Bekinschtein, T. A., Dehaene, S., Rohaut, B., Tadel, F., Cohen, L., & Naccache, L. (2009). Neural signature of the conscious processing of auditory regularities. *Proceedings of the National Academy of Sciences, U.S.A., 106,* 1672–1677.
Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences, 8,* 129–135.
Bonte, M. L., Mitterer, H., Zellagui, N., Poelmans, H., & Blomert, L. (2005). Auditory cortical tuning to statistical regularities in phonology. *Clinical Neurophysiology, 116,* 2765–2774.
Bourguignon, M., De Tiege, X., de Beeck, M. O., Ligot, N., Paquier, P., Van Bogaert, P., et al. (2013). The pace of prosodic phrasing couples the listener's cortex to the reader's voice. *Human Brain Mapping, 34,* 314–326.
Brennan, S. E., & Schober, M. F. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44,* 274–296.
Brown, M., Dilley, L., & Tanenhaus, M. K. (2012). Real-time expectations based on context speech rate can cause words to appear or disappear. *Proceedings of the Annual Meeting of the Cognitive Science Society,* 1374–1379.
Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. *Papers in Laboratory Phonology V,* 70–87.
Calderone, D. J., Lakatos, P., Butler, P. D., & Castellanos, F. X. (2014). Entrainment of neural oscillations as a modifiable substrate of attention. *Trends in Cognitive Sciences, 18,* 300–309.
Chawla, P., & Krauss, R. M. (1994). Gesture and speech in spontaneous and rehearsed narratives. *Journal of Experimental Social Psychology, 30,* 580–601.
Cohen Priva, U. (2017). Not so fast: Fast speech correlates with lower lexical and structural information. *Cognition, 160,* 27–34.
Cravo, A. M., Rohenkohl, G., Wyart, V., & Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience, 33,* 4002–4010.
Demanuele, C., James, C. J., & Sonuga-Barke, E. J. (2007). Distinguishing low frequency oscillations within the 1/f spectral behaviour of electromagnetic brain signals. *Behavioral and Brain Functions, 3,* 62.
Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21,* 1664–1670.
Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage, 85,* 761–768.
Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 914–927.
Engel, A. K., Fries, P., & Singer, W. (2001). Dynamic predictions: Oscillations and synchrony in top–down processing. *Nature Reviews: Neuroscience, 2,* 704–716.
Federmeier, K. D., Wlotko, E. W., & Meyer, A. M. (2008). What's 'right' in language comprehension: Event-related potentials reveal right hemisphere language capabilities. *Language and Linguistics Compass, 2,* 1–17.
Finke, M., & Rogina, I. (1997). Wide context acoustic modeling in read vs. spontaneous speech. In *1997 IEEE International Conference on Acoustics, Speech, and Signal Processing* (*Vol. 3,* pp. 1743–1746). Los Alamitos, CA: IEEE Computer Society Press.
Fischl, B. (2012). FreeSurfer. *Neuroimage, 62,* 774–781.
Fischl, B., Sereno, M. I., Tootell, R. B., & Dale, A. M. (1999). High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human Brain Mapping, 8,* 272–284.
Fougeron, C., & Jun, S.-A. (1998). Rate effects on French intonation: Prosodic organization and phonetic realization. *Journal of Phonetics, 26,* 45–69.
Fries, P. (2005). A mechanism for cognitive dynamics: Neuronal communication through neuronal coherence. *Trends in Cognitive Sciences, 9,* 474–480.
Friston, K. (2012). Prediction, perception and agency. *International Journal of Psychophysiology, 83,* 248–252.
Ghazanfar, A. A., Morrill, R. J., & Kayser, C. (2013). Monkeys are perceptually tuned to facial expressions that exhibit a theta-like speech rhythm. *Proceedings of the National Academy of Sciences, U.S.A., 110,* 1959–1963.
Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology, 2,* 1–12.
Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology, 3,* 238.
Giordano, B. L., Ince, R. A., Gross, J., Schyns, P. G., Panzeri, S., & Kayser, C. (2017). Contributions of local speech encoding and functional connectivity to audio-visual speech perception. *eLife, 6,* e24763.
Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience, 15,* 511–517.
Grosjean, F., & Lane, H. (1976). How the listener integrates the components of speaking rate. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 538–543.
Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., et al. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology, 11,* 1–14.
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. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 694–699.
Gütig, R., & Sompolinsky, H. (2009). Time-warp-invariant neuronal processing. *PLoS Biology, 7,* e1000141.

Hari, R., & Kujala, M. V. (2009). Brain basis of human social interaction: From concepts to brain imaging. *Physiological Reviews, 89,* 453–479.
Hasson, U., Yang, E., Vallines, I., Heeger, D. J., & Rubin, N. (2008). A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience, 28,* 2539–2550.
Hertrich, I., Dietrich, S., & Ackermann, H. (2013). Tracking the speech signal—Time-locked MEG signals during perception of ultra-fast and moderately fast speech in blind and in sighted listeners. *Brain and Language, 124,* 9–21.
Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., & Ackermann, H. (2012). Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. *Psychophysiology, 49,* 322–334.
Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change, 21,* 233–256.
Janse, E. (2004). Word perception in fast speech: Artificially time-compressed vs. naturally produced fast speech. *Speech Communication, 42,* 155–173.
Janse, E., Nooteboom, S., & Quené, H. (2003). Word-level intelligibility of time-compressed speech: Prosodic and segmental factors. *Speech Communication, 41,* 287–301.
Kayser, S. J., Ince, R. A., Gross, J., & Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *Journal of Neuroscience, 35,* 14691–14701.
Keitel, A., Ince, R. A., Gross, J., & Kayser, C. (2017). Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. *Neuroimage, 147,* 32–42.
Kiebel, S. J., Daunizeau, J., & Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Computational Biology, 4,* e1000209.
Koreman, J. (2006). Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America, 119,* 582–596.
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.
Kujala, J., Pammer, K., Cornelissen, P., Roebroeck, A., Formisano, E., & Salmelin, R. (2007). Phase coupling in a cerebro-cerebellar network at 8–13 Hz during reading. *Cerebral Cortex, 17,* 1476–1485.
Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., & Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science, 320,* 110–113.
Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping, 10,* 120–131.
Lasky, E. Z., Weidner, W. E., & Johnson, J. P. (1976). Influence of linguistic complexity, rate of presentation, and interphrase pause time on auditory-verbal comprehension of adult aphasic patients. *Brain and Language, 3,* 386–395.
Lerner, Y., Honey, C. J., Katkov, M., & Hasson, U. (2014). Temporal scaling of neural responses to compressed and dilated natural speech. *Journal of Neurophysiology, 111,* 2433–2444.
Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22,* 1–38.
Liberman, A. M., Delattre, P. C., Gerstman, L. J., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology, 52,* 127–137.
Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4,* 187–196.
Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech, 6,* 172–187.
Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54,* 1001–1010.
Maguire, E. A., Frith, C. D., & Morris, R. (1999). The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. *Brain, 122,* 1839–1850.
Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164,* 177–190.
Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex, 27,* 4293–4302.
Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Attention Perception and Psychophysics, 36,* 329–337.
Miller, J. L., Grosjean, F., & Lomanto, C. (1984). Articulation rate and its variability in spontaneous speech: A reanalysis and some implications. *Phonetica, 41,* 215–225.
Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., & Carreiras, M. (2016). Out-of-synchrony speech entrainment in developmental dyslexia. *Human Brain Mapping, 37,* 2767–2783.
Morillon, B., & Schroeder, C. E. (2015). Neuronal oscillations as a mechanistic substrate of auditory temporal prediction. *Annals of the New York Academy of Sciences, 1337,* 26–31.
Morrill, R. J., Paukner, A., Ferrari, P. F., & Ghazanfar, A. A. (2012). Monkey lipsmacking develops like the human speech rhythm. *Developmental Science, 15,* 557–568.
Nakajima, S., & Allen, J. F. (1993). A study on prosody and discourse structure in cooperative dialogues. *Phonetica, 50,* 197–210.
Nobre, A. C., Correa, A., & Coull, J. T. (2007). The hazards of time. *Current Opinion in Neurobiology, 17,* 465–470.
Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. (2015). Frontal top–down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology, 25,* 1649–1653.
Peelle, J., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology, 3,* 1–17.
Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23,* 1378–1387.
Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience, 32,* 14125–14131.
Poeppel, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 363,* 1071–1086.
Puschmann, S., Steinkamp, S., Gillich, I., Mirkovic, B., Debener, S., & Thiel, C. M. (2017). The right temporoparietal junction supports speech tracking during selective listening: Evidence from concurrent EEG-fMRI. *Journal of Neuroscience, 37,* 11505–11516.
Quené, H. (2007). On the just noticeable difference for tempo in speech. *Journal of Phonetics, 35,* 353–362.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73,* 265–292.
Reinisch, E. (2016). Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics, 37,* 1397–1415.

Rohenkohl, G., Cravo, A. M., Wyart, V., & Nobre, A. C. (2012). Temporal expectation improves the quality of sensory information. *Journal of Neuroscience, 32,* 8424–8428.
Ruspantini, I., Saarinen, T., Belardinelli, P., Jalava, A., Parviainen, T., Kujala, J., et al. (2012). Corticomuscular coherence is tuned to the spontaneous rhythmicity of speech at 2–3 Hz. *Journal of Neuroscience, 32,* 3786–3790.
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. *Human Brain Mapping, 36,* 2455–2469.
Saleh, M., Reimer, J., Penn, R., Ojakangas, C. L., & Hatsopoulos, N. G. (2010). Fast and slow oscillations in human primary motor cortex predict oncoming behaviorally relevant cues. *Neuron, 65,* 461–471.
Schroeder, C. E., Lakatos, P., Kajikawa, Y., Partan, S., & Puce, A. (2008). Neuronal oscillations and visual amplification of speech. *Trends in Cognitive Sciences, 12,* 106–113.
Schroeder, C. E., Wilson, D. A., Radman, T., Scharfman, H., & Lakatos, P. (2010). Dynamics of active sensing and perceptual selection. *Current Opinion in Neurobiology, 20,* 172–176.
Schwartze, M., Tavano, A., Schröger, E., & Kotz, S. A. (2012). Temporal aspects of prediction in audition: Cortical and subcortical neural mechanisms. *International Journal of Psychophysiology, 83,* 200–207.
Sepulcre, J., Sabuncu, M. R., Yeo, T. B., Liu, H., & Johnson, K. A. (2012). Stepwise connectivity of the modal cortex reveals the multimodal organization of the human brain. *Journal of Neuroscience, 32,* 10649–10661.
Small, J. A., Andersen, E. S., & Kempler, D. (1997). Effects of working memory capacity on understanding rate-altered speech. *Aging, Neuropsychology, and Cognition, 4,* 126–139.
Smith, A., Goffman, L., Zelaznik, H. N., Ying, G., & McGillem, C. (1995). Spatiotemporal stability and patterning of speech movement sequences. *Experimental Brain Research, 104,* 493–501.
Sohoglu, E., Peelle, J. E., Carlyon, R. P., & Davis, M. H. (2012). Predictive top–down integration of prior knowledge during speech perception. *Journal of Neuroscience, 32,* 8443–8453.
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.
Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074–1095.
Taulu, S., & Simola, J. (2006). Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. *Physics in Medicine and Biology, 51,* 1759–1768.
Telkemeyer, S., Rossi, S., Koch, S. P., Nierhaus, T., Steinbrink, J., Poeppel, D., et al. (2009). Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience, 29,* 14726–14733.
ten Oever, S., Schroeder, C. E., Poeppel, D., van Atteveldt, N., Mehta, A. D., Mégevand, P., et al. (2017). Low-frequency cortical oscillations entrain to subthreshold rhythmic auditory stimuli. *Journal of Neuroscience, 37,* 4903–4912.
Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *Journal of the Acoustical Society of America, 134,* 628–639.
Tsao, Y.-C., & Weismer, G. (1997). Interspeaker variation in habitual speaking rate: Evidence for a neuromuscular component. *Journal of Speech, Language, and Hearing Research, 40,* 858–866.
Uusitalo, M. A., & Ilmoniemi, R. J. (1997). Signal-space projection method for separating MEG or EEG into components. *Medical & Biological Engineering & Computing, 35,* 135–140.
Vander Ghinst, M., Bourguignon, M., de Beeck, M. O., Wens, V., Marty, B., Hassid, S., et al. (2016). Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. *Journal of Neuroscience, 36,* 1596–1606.
Voss, R. F., & Clarke, J. (1975). "1/f noise" in music and speech. *Nature, 258,* 317–318.
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. *Journal of Neuroscience, 32,* 14010–14021.
Zion-Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron, 77,* 980–991.