NeuroImage 216 (2020) 116788

Contents lists available at ScienceDirect

NeuroImage

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

Neocortical activity tracks the hierarchical linguistic structures of self-produced speech during reading aloud

Mathieu Bourguignon [a,b,c,*], Nicola Molinaro [a,d], Mikel Lizarazu [e], Samu Taulu [f,g], Veikko Jousmäki [h], Marie Lallier [a], Manuel Carreiras [a,d], Xavier De Tiège [b,i]

[a] BCBL, Basque Center on Cognition, Brain and Language, 20009, San Sebastian, Spain
[b] Laboratoire de Cartographie Fonctionnelle du Cerveau, UNI – ULB Neuroscience Institute, Université Libre de Bruxelles (ULB), Brussels, Belgium
[c] Laboratoire Cognition Langage et Développement, UNI – ULB Neuroscience Institute, Université Libre de Bruxelles (ULB), Brussels, Belgium
[d] Ikerbasque, Basque Foundation for Science, Bilbao, Spain
[e] Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'Etudes Cognitives, Ecole Normale Supérieure, EHESS, CNRS, PSL University, 75005, Paris, France
[f] Institute for Learning & Brain Sciences, University of Washington, Seattle, WA, USA
[g] Department of Physics, University of Washington, Seattle, WA, USA
[h] Aalto NeuroImaging, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, PO BOX 15100, FI-00076, AALTO, Espoo, Finland
[i] Magnetoencephalography Unit, Department of Functional Neuroimaging, Service of Nuclear Medicine, CUB – Hôpital Erasme, Brussels, Belgium

ARTICLE INFO

Keywords:
Reading
Speech perception
Speech production
Connected speech
Cortical tracking of speech
Magnetoencephalography

ABSTRACT

How the human brain uses self-generated auditory information during speech production is rather unsettled. Current theories of language production consider a feedback monitoring system that monitors the auditory consequences of speech output and an internal monitoring system, which makes predictions about the auditory consequences of speech before its production. To gain novel insights into underlying neural processes, we investigated the coupling between neuromagnetic activity and the temporal envelope of the heard speech sounds (i.e., cortical tracking of speech) in a group of adults who 1) read a text aloud, 2) listened to a recording of their own speech (i.e., playback), and 3) listened to another speech recording. Reading aloud was here used as a particular form of speech production that shares various processes with natural speech. During reading aloud, the reader's brain tracked the slow temporal fluctuations of the speech output. Specifically, auditory cortices tracked phrases (<1 Hz) but to a lesser extent than during the two speech listening conditions. Also, the tracking of words (2–4 Hz) and syllables (4–8 Hz) occurred at parietal opercula during reading aloud and at auditory cortices during listening. Directionality analyses were then used to get insights into the monitoring systems involved in the processing of self-generated auditory information. Analyses revealed that the cortical tracking of speech at <1 Hz, 2–4 Hz and 4–8 Hz is dominated by speech-to-brain directional coupling during both reading aloud and listening, i.e., the cortical tracking of speech during reading aloud mainly entails auditory feedback processing. Nevertheless, brain-to-speech directional coupling at 4–8 Hz was enhanced during reading aloud compared with listening, likely reflecting the establishment of predictions about the auditory consequences of speech before production. These data bring novel insights into how auditory verbal information is tracked by the human brain during perception and self-generation of connected speech.

1. Introduction

To produce understandable speech, humans rely on self-monitoring of speech output. Such monitoring is based on neural integration of self-generated sensory information, which links speech production to speech perception (for a review, see Hickok, 2012). Still, how this self-produced sensory information is used to control speech remains unclear.

Current theories of language production consider a feedback monitoring system that monitors the sensory consequences of speech output to

* Corresponding author. Laboratoire de Cartographie fonctionnelle du Cerveau, UNI – ULB Neuroscience Institute, Université libre de Bruxelles, 808 Lennik Street, 1070, Brussels, Belgium.
E-mail address: mabourgu@ulb.ac.be (M. Bourguignon).

https://doi.org/10.1016/j.neuroimage.2020.116788
Received 19 September 2019; Received in revised form 19 February 2020; Accepted 20 March 2020
Available online 26 April 2020

correct errors during production (for reviews, see Hickok, 2012 and Houde and Chang, 2015). Evidence about the importance of such a system comes from adaptations of the speaker's speech output to compensate for sensory (i.e., auditory and somatosensory) feedback manipulations (Bauer et al., 2006; Burnett et al., 1998; Guo et al., 2017; Houde, 1998; Liu et al., 2018; Shiller et al., 2009; Tremblay et al., 2003). But such feedback monitoring system cannot account for extremely fast self-corrections of speech observed in humans (Blackmer and Mitton, 1991; Nozari et al., 2011), as they require excessive neural processing time. Hence, most of the current models of language production additionally include an internal speech monitoring system, which makes predictions about motor programs and sensory consequences of speech output before its actual production (Hickok, 2012; Houde and Chang, 2015). Consensus about the neural basis of such an internal system is however lacking (Gauvin et al., 2016). Indeed, some authors consider that the sensory consequences of speech are predicted by sensory networks similar to those involved in monitoring feedback speech (Hickok, 2012; Indefrey, 2011), while others consider that this process recruits distinct neural structures such as, e.g., brain structures involved in conflict monitoring (Hickok, 2012; Nozari et al., 2011).

A potential way to gain insights into the neuronal bases of internal and feedback monitoring systems supporting speech production is to study the coupling between the speaker's voice and his/her own brain activity during connected speech production. Previous magnetoencephalography (MEG) studies focusing on connected speech listening demonstrated speech–sensitive coupling between the slow modulations in the temporal envelope of the speaker's voice and the listeners' (mainly auditory) cortex activity both in quiet and adverse auditory scenes (Alexandrou et al., 2018a; Bourguignon et al., 2013a; Clumeck et al., 2014; Ding et al., 2016; Ding and Simon, 2013; Gross et al., 2013; Molinaro et al., 2016; Peelle et al., 2013a; Vander Ghinst et al., 2019, 2016a; Zion Golumbic et al., 2013). This coupling, henceforth referred to as the cortical tracking of speech, mainly occurs at syllable (4–8 Hz), word (2–4 Hz) and phrasal/sentential (<1 Hz) rates. It is considered to play a pivotal role in parsing connected speech into linguistic units (i.e., syllables, words or phrases/sentences) to promote subsequent speech recognition (Park et al., 2018; Zion Golumbic et al., 2012). Additionally, it might help predict the precise timing of events in the speech stream such as syllables, words and phrases/sentences (Donhauser and Baillet, 2020; Zion Golumbic et al., 2012). Such predictions probably facilitate speech comprehension as well as coordination of turn-taking transitions during verbal conversation (Friston and Frith, 2015; Zion Golumbic et al., 2012). It is then sensible to hypothesize that similar cortical tracking of speech is also at work during connected speech production. Given the crucial role of the cortical tracking of speech in language comprehension (Bourguignon et al., 2013a; Clumeck et al., 2014; Ding et al., 2016; Gross et al., 2013; Molinaro et al., 2016; Peelle et al., 2013a; Vander Ghinst et al., 2016a), such tracking might indeed contribute to self-produced speech monitoring systems. If confirmed, this could bring unprecedented insights into how humans handle self-generated auditory information during language production. Additionally, investigating coupling directionality (i.e., speech → brain vs. brain → speech coupling) during connected speech production could bring critical information about the neural bases of speech production monitoring systems in humans: *feedback* monitoring systems that monitor the sensory consequences of speech output during production should indeed involve speech → brain coupling, while *internal* monitoring systems that generate predictions about motor programs and sensory consequences of speech output before its actual production should involve brain → speech coupling. The demonstration of a difference in coupling directionality between connected speech listening and production might ultimately bring additional clues about the functional roles (i.e., epiphenomenon vs. effective contribution to speech production monitoring systems) of the cortical tracking of speech during speech production.

To address these issues, the present MEG study relied on the comparison of the cortical tracking of speech while subjects listened to recordings of texts read aloud (by a reader or themselves) and while they read themselves a text aloud. This approach was similar to those used in previous studies investigating the cortical tracking of speech during listening to the live (Bourguignon et al., 2013a; Clumeck et al., 2014) or recorded (Clumeck et al., 2014; Destoky et al., 2019; Vander Ghinst et al., 2019, 2016a) voice of somebody reading a text aloud. It was also based on the assumption that some neurocognitive processes are shared by natural speech production and reading aloud (Sulpizio and Kinoshita, 2016). Indeed, although reading aloud differs in several aspects (e.g, rhythmicity, prosody, etc.) from natural speech and could be considered as a non-naturalistic form of speech stimuli (for a detailed review, see Alexandrou et al., 2018b), it is recognized as a form of speech production such as, e.g., spontaneous narrative, narrative recalls, conversation, picture description (see, e.g., Bóna, 2014). The last stages of language production in these different speech situations indeed share various output processes: all include phonological encoding (i.e., assigning a segment to a position in a metrical frame), phonetic encoding (i.e., retrieving the motor plans required for articulation), and articulation (i.e., producing the gestures leading to an acoustic sound) (Kawamoto et al., 2015). Settling on reading aloud as a task also makes it possible to control speech content and linguistic form, which are parts of the main speech features (i.e., speech rhythm, linguistic content, speaker's identity) previously reported to affect brain rhythms (Alexandrou et al., 2017). However, reading aloud decreases the subjects' need to focus on semantic/lexical access, other cognitive processes or speech style, which can potentially affect the cortical tracking of speech and directionality assessments during language production (Bóna, 2014). Still, during reading aloud, read speech provides an external frame of reference to which readers can compare their speech output to, which is not the case during naturalistic speech production (Alexandrou et al., 2018b). Finally, comparing the neural processes at play during listening to somebody reading aloud and during reading aloud allows relying on auditory verbal information that shares common rhythmicity and prosody.

This MEG study assesses coherence and directionality between slow temporal modulations in voice and brain activity to investigate the cortical tracking of speech in subjects who (i) read a text aloud, (ii) listened to a recording of a different text, and (iii) listened to a recording of their own speech while reading aloud (i.e., playback). The study was specifically designed to (i) identify cortical areas that track the slow fluctuations of the speakers' voice during self-produced speech, (ii) determine the causal nature of this tracking in the framework of feedback and internal speech monitoring systems, and (iii) assess tracking and directionality differences between reading aloud and listening. We predicted that (i) the human brain would track the hierarchical linguistic structures (i.e., syllables, words, sentences/phrases) of speech during both reading aloud and listening conditions, (ii) tracking levels, areas and directionality would differ between reading aloud and listening conditions, and (iii) that listening to a recording of a different text vs. playback would modulate the level of internal predictions about the upcoming speech during listening.

2. Methods

2.1. Participants

Eighteen healthy native Spanish speakers without any history of neuropsychiatric disease or language disorders were studied. One participant was excluded from the study due to excessive artifacts in the data. The study therefore reports on 17 participants (range 20–32 years; mean age 23.9 years; 9 females and 8 males). Sixteen participants were right-handed according to Edinburgh handedness inventory (score range 40–100%; mean ± SD, 70.6 ± 19.1%) (Oldfield, 1971). Handedness appraisal was missing from the last participant. Thirteen participants had a university degree, 1 was a master student, and 3 were trained professionals with high school or secondary school degree (degree obtained at age ~18 or ~16 respectively when no grade is repeated). The study

was approved by the BCBL Ethics Committee. Participants were included in the study after written informed consent.

2.2. Experimental paradigm

The experimental stimuli were derived from 2 narrative texts of ~1000 words. The topics of the texts were maximally neutral: the first elaborated on the origin of life and human spirituality, while the second was an attempt to define what is a "discourse". Both texts were read aloud by a male and a female native Spanish speaker and recorded with a high quality microphone. We kept only the first 5 min of these audio recordings. Reading pace was 2.53 ± 0.12 words/s (mean ± SD across the four recordings of the number of words read divided by recording duration).

Participants underwent four experimental conditions (*read*, *listen*, *playback*, and *rest*) lasting ~5 min each while they were sitting in the MEG chair with their head inside the MEG helmet. During the *read* condition, participants continuously read aloud one of the two texts printed on A4 pages for 314 ± 13 s (mean ± SD across participants). During the *listen* condition, they listened to the audio recording of the other text read by the reader of their gender. Texts were assigned to conditions in a counterbalanced manner. During the *playback* condition, participants listened to their own voice recorded (see section 2.3 for data acquisition details) earlier during the *read* condition. Obviously, *playback* condition was performed in all subjects after the *read* condition. This *playback* condition was used (i) to assess the impact of possible sensory prediction about upcoming speech (as subjects had some hints about speech content and production from the prior *read* condition) on the cortical tracking of speech and on tracking directionality, and (ii) to control for potential differences in speech rhythm between *listen* and *read*. For both *listen* and *playback* conditions, sounds were played with VLC running on a MacBook pro and delivered at 60 dB (measured at ear-level in every participant) through a MEG-compatible front-facing flat-panel loudspeaker (Panphonics Oy, Espoo, Finland) placed ~2 m from the participants. During the *rest* condition, participants were asked to fixate the gaze at a point on the wall of the magnetically shielded room (MSR) and try to reduce blinks and saccades to the minimum. The order of the conditions was either *read–listen–rest–playback* or *listen–read–rest–playback*.

2.3. Data acquisition

Neuromagnetic signals were recorded at the Basque Centre on Cognition, Brain and Language (BCBL) with a whole-scalp-covering neuromagnetometer installed in a MSR (Vectorview & Maxshield™; MEGIN Elekta Oy, Helsinki, Finland). The 306-channel MEG sensor layout consisted in 102 sensor triplets, each comprising one magnetometer and two orthogonal planar gradiometers characterized by different patterns of spatial sensitivity to nearby or right beneath cortical sources. The recording pass-band was 0.1–330 Hz and the signals were sampled at 1 kHz. The head position inside the MEG helmet was continuously monitored by feeding current to five head-tracking coils located on the scalp and observing the corresponding coil-induced magnetic field patterns by the MEG sensors. Head position indicator coils, three anatomical fiducials, and at least 150 head-surface points (covering the whole scalp and the nose surface) were localized in a common coordinate system using an electromagnetic tracker (Fastrak, Polhemus, Colchester, VT, USA).

An optical fiber microphone was placed inside the MSR to record participants' voice during the *read* condition. To maximize sound quality, the microphone was taped to the edge of the MEG helmet, ~5 cm away from subjects' mouth. Sound signals were recorded with *Audacity* at a sampling rate of 44.1 kHz. Electrooculograms (EOG) monitored vertical and horizontal eye movements, and electrocardiogram (ECG) recorded heartbeat signals. All these signals were recorded time-locked to MEG signals.

High-resolution 3D cerebral magnetic resonance images (MRI; T1-weighted MPRAGE sequence) were acquired on a 3 T MRI scan (Siemens Medical System, Erlangen, Germany) with a 32-channel head coil.

2.4. Data preprocessing

As reading aloud is typically associated with many sources of high-amplitude artifacts in electrophysiological signals (e.g., head movements, muscle artifacts, eye movements, etc.), special care was taken during data preprocessing to subtract as much as possible these artifacts from raw MEG data.

Fig. 1 (left part) depicts all the preprocessing steps.

Continuous MEG data were first preprocessed off-line using the temporal signal space separation (tSSS) method (with a correlation limit of 0.9 and the segment length of the temporal projection set equal to the recording duration) to subtract external interferences, to correct for head movements, and to dampen movement artifacts induced by reading aloud (Taulu et al., 2005; Taulu and Simola, 2006). To further suppress heartbeat, eye-blink, and eye-movement artifacts, 30 independent components were evaluated from the MEG data low-pass filtered at 25 Hz using FastICA algorithm (dimension reduction, 30; non-linearity, tanh) (Hyvärinen et al., 2001; Vigario et al., 2000). Independent components displaying a correlation exceeding 0.15 with any EOG or ECG signals were subtracted from the full band and full rank MEG data. The mean ± SD of rejected components was 7.2 ± 1.4 (*read*), 5.1 ± 1.8 (*listen*), 4.9 ± 2.0 (*rest*), and 5 ± 2.0 (*playback*). MEG data were then low-pass filtered at 145 Hz and notched at 50 and 100 Hz for landline artifact removal. Finally, when the maximum MEG amplitude exceeded 5 pT (magnetometers) or 1 pT/cm (gradiometers), data within 1 s before and after the excessive amplitude were discarded from further analyses to avoid inclusion of MEG data compromised by any other high-amplitude artifacts (e.g., head movements or excessive muscle artifacts; Muthukumaraswamy, 2013) not removed by tSSS or ICA. Note that we did not attempt to remove muscle artifacts because these are typically not well removed by classical linear regression techniques (for a review, see Muthukumaraswamy, 2013), and most importantly, because their power mainly lies at frequencies above 20 Hz (Muthukumaraswamy, 2013) on which we did not focus. That is, we here focused on brain activity at frequencies below 8 Hz.

To estimate the efficacy of the preprocessing steps described in the paragraph above, MEG data were also minimally preprocessed, with signal space separation (SSS) only. Power spectral densities were then estimated for minimally and fully preprocessed gradiometer data, based on the same time segments (identified based on fully preprocessed data).

Speech temporal envelopes were obtained from all sound recordings as the rectified sound signals low-pass filtered at 50 Hz. Speech temporal envelopes were further resampled at 1000 Hz time-locked to MEG signals.

2.5. Coherence analysis

Fig. 1 (middle part) depicts all the processing steps related to coherence analysis.

To perform coherence analyses, continuous data obtained in all conditions (*listen*, *playback*, *read* and *rest*) were split into 2-s epochs with 1.6-s epoch overlap, leading to a frequency resolution of 0.5 Hz (Bortel and Sovka, 2014). Also, for each participant, only the minimum amount of epochs across all conditions was used for subsequent analyses to enforce similar signal to noise ratio across conditions. These steps led to 703 ± 45 epochs of MEG and voice envelope signals for each participant and condition.

Coherence is an extension of Pearson correlation coefficient to the frequency domain that determines the degree of coupling between two signals, providing a number between 0 (no linear dependency) and 1 (perfect linear dependency) for each frequency (Halliday, 1995).

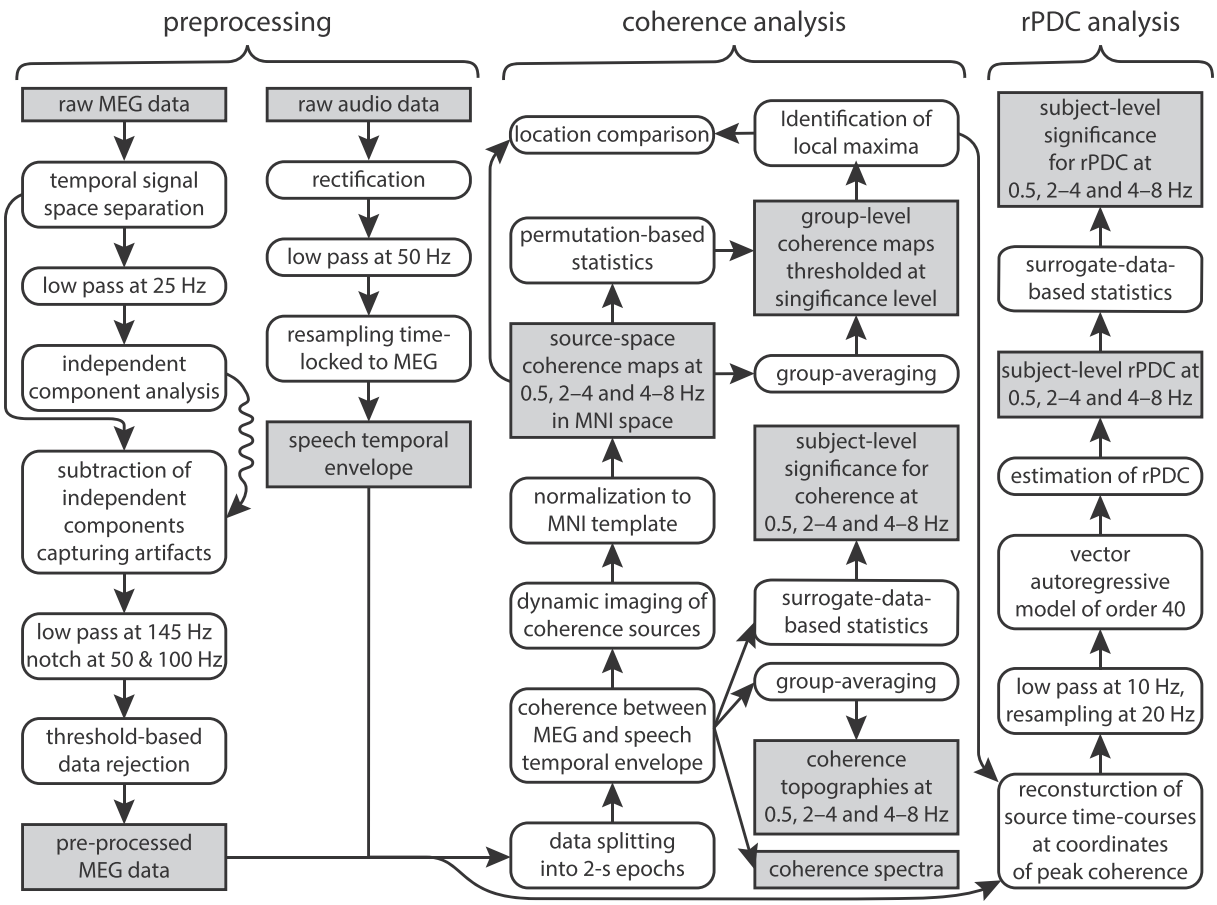

Fig. 1. Main processing steps applied to the data, including preprocessing, coherence analysis and renormalized partial directed coherence (rPDC) analysis. Gray squared corner boxes indicate data input/output. White rounded corner boxes indicate processing steps.

Coherence was previously used to assess the coupling between voice and brain signals at the frequencies corresponding to phrasal/sentential (<1 Hz), word (2–4 Hz) and syllable (4–8 Hz) rates (Bourguignon et al., 2013b; Keitel et al., 2018; Luo and Poeppel, 2007; Molinaro and Lizarazu, 2018; Peelle et al., 2013b; Poeppel, 2003; Vander Ghinst et al., 2016b). Based on these studies, we focused on these predefined frequency ranges. Identifying relevant frequency ranges in a data-driven way might have proven challenging for the *read* condition since we expected remaining artifacts to give rise to significant values of coherence with speech temporal envelope.

Coherence was first estimated at the sensor level. Data from gradiometer pairs were combined in the orientation of maximum coherence as done in Bourguignon et al. (2015). Coherence at phrasal/sentential level was taken at the frequency bin corresponding to 0.5 Hz; coherence at word level was taken as the mean coherence across frequency bins comprised in 2–4 Hz; coherence at syllable level was taken as the mean coherence across frequency bins comprised in 4–8 Hz. These frequency ranges matched well with the occurrence rate of silent periods longer than 100 ms (*listen*, 0.41 ± 0.07 Hz; *read & playback*, 0.45 ± 0.05 Hz), the effective rate of words (*listen*, 3.13 ± 0.13 Hz; *read & playback*, 3.17 ± 0.35 Hz) and the effective rate of syllables (*listen*, 6.86 ± 0.28 Hz; *read & playback*, 6.95 ± 0.76 Hz). As in previous studies (Bourguignon et al., 2020; Vander Ghinst et al., 2019), silent periods were defined as periods when the auditory speech envelope was below a tenth of its mean, and the effective rates were assessed as the number of words or syllables manually extracted from audio recordings divided by the corrected duration of the audio recording. The corrected duration was taken as the total time during which the reader was actually talking, that is the total duration of the audio recording minus the sum of all silent periods lasting at least 100 ms.

Coherence was also evaluated at the source level using a beamformer approach since this method has a high sensitivity to activity coming from locations of interest while attenuating external interferences such as reading-induced head movement, eye movements, or muscle artifacts (Hillebrand et al., 2005). Beamforming also presents the valuable advantage of reconstructing residual muscle artifacts close to the activated muscles, making it easy to identify such artifacts on source-reconstructed images (Muthukumaraswamy, 2013). To do so, individual MRIs were first segmented using Freesurfer software (Martinos Center for Biomedical Imaging, Massachusetts, USA; Reuter et al., 2012). Then, the MEG forward model was computed for three orthogonal tangential current dipoles placed at the nodes of a homogeneous 5 × 5 × 5 mm^3 cubic grid source space covering the entire brain volume, including the cerebellum and brainstem (MNE suite; Martinos Center for Biomedical Imaging, Massachusetts, USA; Gramfort et al., 2014). The 3-D forward model was further reduced to its two first principal components (Hillebrand et al., 2012; Sekihara and Nagarajan, 2008). The last component was discarded because it closely corresponds to the radial orientation with respect to the skull and hence is close to magnetically silent. Finally, coherence maps were produced within the computed source space at 0.5 Hz, 2–4 Hz, and 4–8 Hz using Dynamic Imaging of Coherent Sources (DICS; Gross et al., 2001), and further interpolated onto a 1-mm grid. Both planar gradiometers and magnetometers were used for inverse modeling after dividing each sensor signal by its noise variance. Despite the fact that raw magnetometer signals are considered noisier than planar gradiometers, in the framework of signal space separation, signals from both sensor types are reconstructed from the same inner components, corresponding to the magnetostatic multipole expansion, and have therefore similar levels of residual interference after suppression of signals from external sources (Garcés et al., 2017).

Accordingly, even in the *read* condition where distal artifacts contaminated magnetometers' more than gradiometers' raw signals, it makes little difference to include either or both of the sensor types for source reconstruction. The noise variance was estimated from the preprocessed *rest* MEG data, for each sensor separately. As the analyses described in a further paragraph require extracting the time course of some sources, we used the additional constraint that beamformer weight coefficients are real-valued. This constraint is sensible since one can easily argue that electrical currents in the brain are real¬-valued. Practically, it leads to using the real part of the cross-spectral density matrix in DICS beamformer computation.

To compute group-level coherence maps, a non-linear transformation from individual MRIs to the standard Montreal Neurological Institute (MNI) brain was first computed using the spatial-normalization algorithm implemented in Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London, UK; Ashburner et al., 1997; Ashburner and Friston, 1999) and then applied to individual MRIs and coherence maps. This procedure generated a normalized coherence map in the MNI space with 1-mm cubic voxels for each subject, condition and frequency of interest (i.e., 0.5 Hz, 2–4 Hz, and 4–8 Hz). Group-level maps were obtained by averaging the normalized coherence maps across participants and conditions.

We further identified the coordinates of local maxima in group-level coherence maps. Such local coherence maxima are sets of contiguous voxels displaying higher coherence values than all neighboring voxels. We only report statistically significant local coherence maxima, disregarding the extent of these clusters. Indeed, cluster extent is hardly interpretable in view of the inherent smoothness of MEG source reconstruction (Bourguignon et al., 2018; Hämäläinen and Ilmoniemi, 1994; Wens et al., 2015). We also disregarded some local maxima that, based on their location, were likely to arise out of artifacts.

2.6. Directionality assessment

Fig. 1 (right part) depicts all the processing steps related to directionality assessment.

The directionality of the coupling between the temporal envelope of voice signals and the activity within brain areas displaying a significant local maximum of coherence (see section 2.8), was assessed with renormalized partial directed coherence (rPDC; Schelter et al., 2009, 2006). To this aim, the time-course of brain electrical activity within these brain areas was estimated with the beamformer described in section 2.5, in the orientation maximizing the coherence with speech temporal envelope. Source and voice signals were low-pass filtered at 10 Hz and down-sampled at 20 Hz. Then, for each source separately, a vector autoregressive (VAR) model of order 40 was fitted to the source and the voice data using the ARfit package (Schneider and Neumaier, 2001). The rPDC was then estimated based on the Fourier transform of the VAR model coefficients. This enabled for estimating rPDC at frequencies from 0 to 10 Hz with 0.5 Hz resolution.

2.7. Partial coherence to control for artifacts

In the *read* condition, there was a discrepancy between sensor and source-level results (see Results section). In the sensor space, strong artifacts at the edge of the sensor array obscured the 2–4-Hz and 4–8-Hz cortical tracking of speech. In the source space, artifacts were present but genuine cortical tracking of speech in auditory cortices was clearly visible thanks to the use of the beamformer approach. To verify that this discrepancy pertained to that beamformer did effectively dampen artifacts—and hence strengthen results derived from source-space data—, we estimated the coherence between speech temporal envelope and MEG signals while partialling out the contribution of MEG signals recorded at sensors on the edge of the sensor array.

The following analysis was performed separately at 0.5 Hz, 2–4 Hz and 4–8 Hz. For each gradiometer pair on the edge of the sensor array (23 in total), we estimated the orientation in the 2-d space spanned by both gradiometer signals (Bourguignon et al., 2015) yielding the maximum coherence with speech temporal envelope. Partial coherence was then estimated between speech temporal envelope and all gradiometer signals (again optimizing on the orientation within all pairs) while partialling out edge gradiometer signal in its optimal orientation (Halliday, 1995). This led to as many sensor distribution of partial coherence as there are edge gradiometer pairs. For each sensor, we retained the minimum partial coherence value across all these edge gradiometer pairs.

2.8. Statistical analyses

2.8.1. Reading pace

Participants' word production rate in the *read* condition was compared to that in the texts used in the *listen* condition with a paired *t*-test.

2.8.2. Comparison of power spectral density between conditions

We compared power spectral densities between the *read* and *listen* conditions at the frequencies of interest. For each subject, tested condition and preprocessing scheme (SSS only, and tSSS + ICA), global power spectral density was computed as the mean power spectral density across gradiometer sensors. The global power densities averaged across each frequency range of interest (0.3–0.7 Hz, 2–4, Hz and 4–8 Hz) were log-transformed and compared between conditions with a paired-sample *t*-test. We also compared between conditions the difference between global power for fully- and minimally preprocessed data.

2.8.3. Significance of subject-level coherence in the sensor space

We evaluated the statistical significance of sensor-space coherence values, using surrogate-data-based statistics (Faes et al., 2004). For each participant, condition, and frequency range of interest (i.e., 0.5 Hz, 2–4 Hz, and 4–8 Hz), we extracted the maximum across gradiometer pairs of the mean coherence across the frequency range of interest. This maximum genuine coherence was then compared to a distribution of 1000 surrogate values computed in the same way, but with speech temporal envelope replaced by its Fourier transform surrogate (Faes et al., 2004). Fourier transform surrogate preserves the power spectrum but destroys the phase information by replacing the phase of Fourier coefficients by random numbers in the range $[-\pi; \pi]$ (Faes et al., 2004). Genuine maximum coherence values were deemed significant when they exceeded the 95th percentile of their surrogate distribution.

2.8.4. Significance of group-level coherence in the source space

The statistical significance of—*genuine*—group-level coherence maps was assessed with non-parametric permutation test (Nichols and Holmes, 2002). First, participant- and group-level *null* coherence maps at the frequencies of interest (i.e., 0.5 Hz, 2–4 Hz, and 4–8 Hz) were computed with the MEG signals and the voice signals rotated in time by about 2 min 30 s (i.e., where the first and second halves were swapped, thereby destroying genuine coherence but preserving spectral properties). The exact temporal rotation applied was chosen to match a pause in speech to enforce continuity. Group-level difference maps were obtained by subtracting *f*-transformed *genuine* (*read*, *listen* or *playback*) and *null* group-level coherence maps for each frequency of interest. Under the null hypothesis that coherence maps are the same whatever the experimental condition, the labeling *genuine* or *null* are exchangeable prior to difference map computation (Nichols and Holmes, 2002). To reject this hypothesis and to compute a significance threshold for the correctly labeled difference map, the sample distribution of the maximum of the difference map's absolute value within the entire brain was computed from a subset of 1000 permutations. The threshold at $p < 0.05$ was computed as the 95 percentile of the sample distribution (Nichols and Holmes, 2002). All supra-threshold local coherence maxima were interpreted as indicative of brain regions showing statistically significant coupling with the produced (*read*) or heard (*listen* and *playback*) sounds.

2.8.5. Comparison of source location between conditions

The coordinates of significant local coherence maxima were statistically compared between frequency ranges (0.5 Hz vs. 2–4 Hz, 0.5 Hz vs. 4–8 Hz, and 2–4 Hz vs. 4–8 Hz) and between conditions (*listen* vs. *playback*, *listen* vs. *read*, and *playback* vs. *read*) using the location-comparison approach proposed by Bourguignon et al. (2018). This method uses a bootstrap procedure (Efron, 1979) to estimate the sample distribution of coordinates of the two local coherence maxima under comparison and tests the null hypothesis that the distance between them is zero. Briefly, for each comparison, we generated 1000 group-level maps of the conditions under assessment by random bootstrapping from the subjects, and identified in each map the coordinates of the local maximum closest to the average location of the two maxima in the genuine maps. The resulting sample distribution of coordinate difference was then submitted to a multivariate location test evaluating the probability that this difference is zero (Bourguignon et al., 2018). That test tightly relates to the multivariate T^2 test (Hotelling, 1931) and assumes that the sample distribution of coordinates difference is normal.

For some local maxima, we further tested the—*a posteriori*—hypothesis that their bootstrap coordinate distributions were bimodal rather than unimodal, suggesting that two separate sources would contribute to these single local maxima (for a full description of the methods, see supplementary material subsections 8.1.1 & 8.2.1). Note that this analysis cannot tell apart the two following possibilities: 1) the two sources are present in all participants with inter-individual variability in their relative strength or 2) either of the sources is present in a given participant, so that there is inter-individual variability in the recruited network.

2.8.6. Significance of individual subjects' rPDC values and comparison between coupling directions

We evaluated the number of participants showing statistically significant rPDC at 0.5 Hz, 2–4 Hz, or 4–8 Hz, using surrogate-data-based statistics (Faes et al., 2010). For each participant, selected brain area, and coupling direction, the genuine rPDC value (at 0.5 Hz, or the mean across 2–4 Hz or 4–8 Hz) was compared to a distribution of 1000 surrogate rPDC values derived from causal Fourier transform surrogate data (Faes et al., 2010). Causal Fourier transform surrogate data were generated with the estimated VAR model wherein coupling in the specific causal direction being tested is abolished by setting to 0 the associated coefficients. Because other coefficients in the VAR model were kept unaltered, coupling in the causal direction that was not being tested was preserved. As a consequence, some degree of coherence within the data was preserved; arguably only that ascribable to the preserved coupling direction (Faes et al., 2010). Genuine rPDC values were deemed significant when they exceeded the 95th percentile of their surrogate distribution.

Values of rPDC were compared between speech → brain and brain → speech directions using paired t-tests across participants.

2.8.7. ANOVA assessment of coherence, rPDC, and partial coherence values

Source-level coherence, rPDC and sensor-level partial coherence values were analyzed with 2-way repeated measures ANOVAs. In these assessments, the first factor was the condition (*listen*, *playback*, and *read*). Based on the result that only one—non-artifactual—local maximum was found per hemisphere in all conditions, we also included the hemisphere as a second factor. ANOVAs were run separately for 0.5 Hz, 2–4 Hz and 4–8 Hz coupling, and for speech → brain and brain → speech directions in case of rPDC assessment. This is justified by that coupling values within these two classes had relatively different variances. Analysing data together would have violated the homoscedasticity assumption of the ANOVA. For source-level coherence values, the dependent variable was the maximum coherence across a 10-mm sphere centered on significant local maxima of group-level coherence maps. For sensor-level partial coherence values, the dependent variable was the maximum partial coherence across subsets of gradiometer pairs showing the peaks of coherence. Formally, these subsections comprised the 9 gradiometers of maximum coherence averaged across participants and conditions. There were 2 selections, one for the left and one for the right hemisphere.

2.8.8. Significance of the difference between coupling directions in effects involving rPDC

For each effect revealed by the ANOVA conducted on rPDC values, we evaluated whether this effect was significantly different between coupling directions with a Bootstrap approach. That is, we created a Bootstrap distribution of the difference in F value associated to this effect assessed on rPDC in the brain → speech vs. speech → brain directions. A p-value was computed as the proportion of negative (or positive) F values.

2.9. Data and software availability

Data and analysis scripts are available upon reasonable request to the corresponding author.

3. Results

3.1. Reading pace

In the *read* condition, participants read at a pace (uncorrected) of 2.64 ± 0.28 words/s (mean ± SD) or 5.83 ± 0.62 syllables/s. This pace was not significantly different from the one they heard in the *listen* condition ($t_{16} = 1.26$, $p = 0.23$). Their reading fluency (percentage of words correctly read without hesitation) was 98.3 ± 0.7%.

3.2. Artifact removal in the read condition

Fig. 2 (top) shows that reading-induced artifacts were dampened by the preprocessing steps. When the data were minimally preprocessed (only with signal space separation), the overall MEG power was ~100% higher in *read* than *listen* in the frequency ranges of interest (~0.5 Hz, 118%, $t_{16} = 2.60$, $p = 0.019$; 2–4 Hz, 105%, $t_{16} = 3.01$, $p = 0.0082$; 4–8 Hz, 71%, $t_{16} = 2.42$, $p = 0.028$). For the fully preprocessed data (temporal signal space separation and ICA), overall MEG power was less than 25% higher in *read* than *listen* (~0.5 Hz, 13%, $t_{16} = 2.30$, $p = 0.035$; 2–4 Hz, 23%, $t_{16} = 4.12$, $p = 0.0008$; 4–8 Hz, 5%, $t_{16} = 1.24$, $p = 0.23$). Importantly, the difference in overall power between *read* and *listen* was significantly lower for fully than minimally preprocessed MEG (~0.5 Hz, $t_{16} = 2.39$, $p = 0.030$; 2–4 Hz, $t_{16} = 2.27$, $p = 0.038$; 4–8 Hz, $t_{16} = 2.34$, $p = 0.33$). Overall, this indicates that some artifacts remained in *read* MEG signals, but that their amplitude was low.

Fig. 2 (bottom) shows that reading-induced artifacts affected mainly the sensors on the edge of the sensor array, more so for minimally than fully preprocessed data. After full preprocessing, some sensors at the vertex even showed lower power in *read* than *listen* at ~0.5 Hz and 4–8 Hz.

3.3. Coherence results

3.3.1. Coherence in the sensor space

Fig. 3 illustrates the results of cortical tracking of speech quantified with coherence in the sensor space. The maximum coherence between MEG signals and speech temporal envelope was observed at 0.5 Hz as in previous studies (Bourguignon et al., 2018, 2013a; Clumeck et al., 2014; Molinaro et al., 2016; Vander Ghinst et al., 2019, 2016a). Coherence at 0.5 Hz was significant in all subjects and conditions ($ps < 0.05$). This frequency matches the supra-second phrasal/sentential time-scale. Coherence was also significant in a substantial proportion of the 17 subjects at 2–4 Hz (*listen*, 11; *playback*, 8; *read*, 17) and 4–8 Hz (*listen*, 13; *playback*, 12; *read*, 17). Note that the detection rate of significant coherence in the *read* condition has likely been inflated by the presence of artifacts inherent to speech production. These frequency ranges were further investigated because they match the 250–500 ms word time-scale

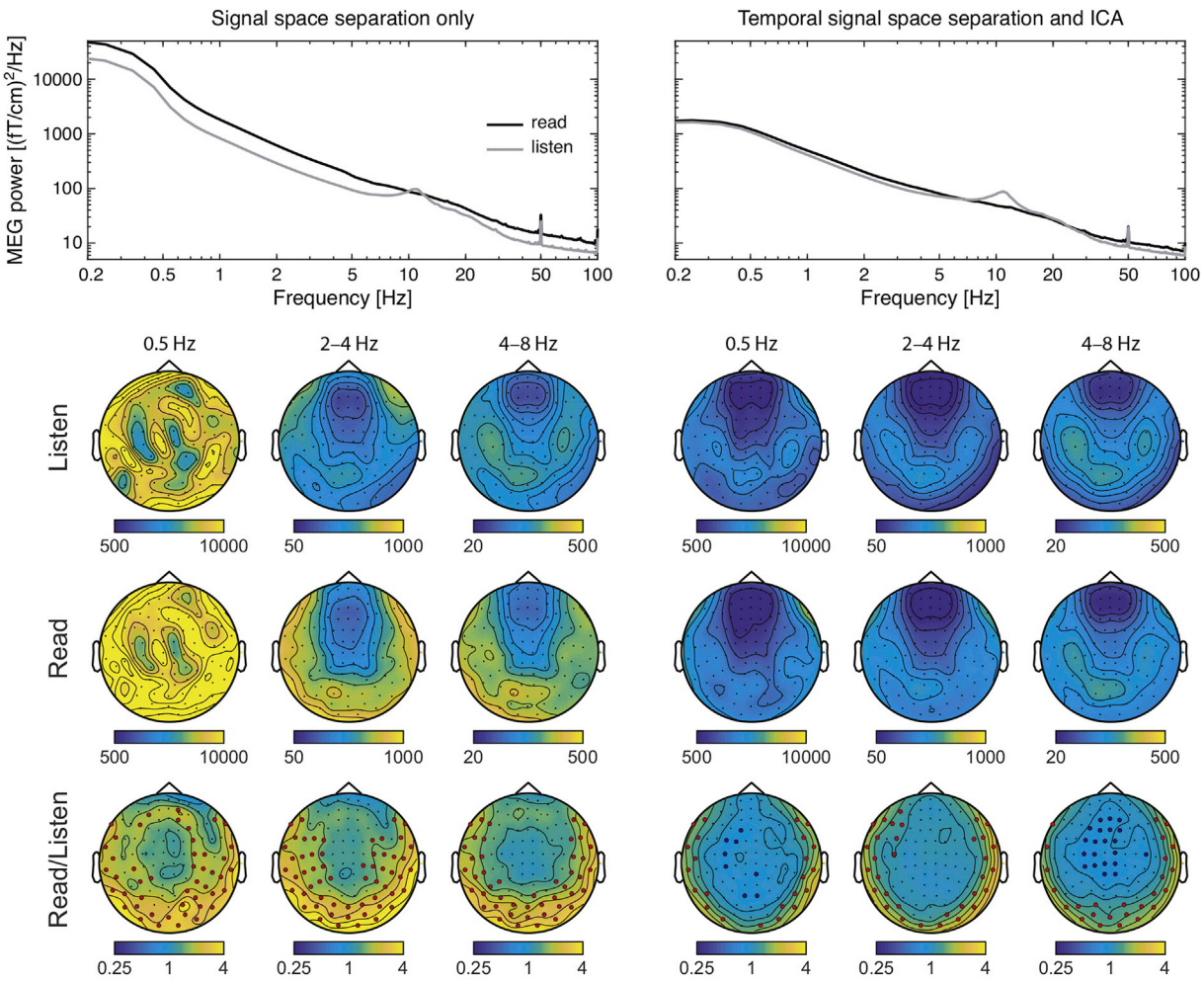

Fig. 2. Adequacy of preprocessing steps to suppress artifacts. *Top* — Power spectral densities of MEG gradiometer signals averaged across sensors and subjects in the *read* (black traces) and *listen* conditions (gray traces). MEG data were either preprocessed with signal space separation only (left) or with temporal signal space separation and independent component analysis (ICA) (right). *Bottom* — Spatial distribution of power spectral densities at frequencies of interest (mean across 0.3–0.7 Hz, 2–4 Hz, and 4–8 Hz) in *listen* and *read*, and ratio thereof. In ratio maps, significant ratios are highlighted with red (*read* > *listen*) or blue (*read* < *listen*) discs.

(2–4 Hz), and the 150–300-ms syllable time-scale (4–8 Hz). In both listening conditions (*listen* & *playback*), coherence topographies were characterized by clusters over bilateral temporal sensors, more posterior at 0.5 Hz than at 2–4 and 4–8 Hz. In the *read* condition, coherence topographies were suggestive of the presence of strong artifacts but also of genuine bilateral activity arising from temporal sensors (more convincingly so at 0.5 Hz than at 2–4 Hz, and 4–8 Hz). At 0.5 Hz, the peak coherence in both hemispheres was one sensor more anterior in the *read* condition than in the listening conditions. However, we did not appraise the statistical significance of this location difference in the sensor space since such comparison was planned (and will be fulfilled) for source space data.

3.3.2. Coherence in the source space

Fig. 4A presents the source-space coherence maps obtained with DICS at 0.5 Hz, 2–4 Hz, and 4–8 Hz separately.

Table 1 presents the MNI coordinates of significant local coherence maxima observed in source-space maps that we treated as non-artifactual based on their location.

In both listening conditions (*listen* & *playback*) significant local coherence maxima localized in bilateral cortex around posterior superior temporal sulcus (pSTS) at 0.5 Hz and at bilateral supratemporal auditory cortex (STAC) at 2–4 Hz, and 4–8 Hz, except for the right-hemisphere source of 2–4-Hz coherence in playback that localized at the pSTS. Still, the location comparison test revealed no statistically significant difference in location between frequencies ($ps > 0.05$; 12 comparisons: 2 conditions × 2 hemispheres × 3 pairs of frequency ranges). This negative result was likely due to limited statistical power since location differences between 0.5-Hz and 4–8-Hz cortical tracking of speech was previously reported to be highly significant in the right hemisphere on a larger sample (Bourguignon et al., 2018). There were also no statistically significant differences in location between the two listening conditions ($ps > 0.5$; 6 comparisons: 3 frequencies × 2 hemispheres).

In the *read* condition, source reconstruction results emphasized the presence of genuine cortical tracking of speech. Some artifacts remained but they did not overshadow coherence local maxima in the auditory/speech regions of each hemisphere and in each frequency range explored (see Fig. 4A and Table 1 for peak coordinates and coherence values). Local maxima of coherence interpreted as artifactual were in the pons (0.5 Hz, MNI [–1 –1 –35]; 2–4 Hz, [1 –4 –38]; 4–8 Hz, [2 –14 –36]), in the right frontal pole (4–8 Hz, [52 34 –9]), and in the right temporal pole (0.5 Hz, [52 17 –26]; 2–4 Hz, [52 19 –29] and [30 13 –41]).

The cortical tracking of speech elicited by the *read* condition appeared to be different from that during listening conditions at 0.5 Hz, 2–4 Hz, and 4–8 Hz. We focus below on the source location comparison between *read* and *listen*, and thereafter, highlight the differences seen when *read* and *playback* were compared.

At 0.5 Hz, right-hemisphere local coherence maxima in *read* and *listen* were distant of only 3 mm, a distance that was not statistically significant ($F_{3,998} = 0.052$, $p = 0.98$). In the left hemisphere, they were distant of 19

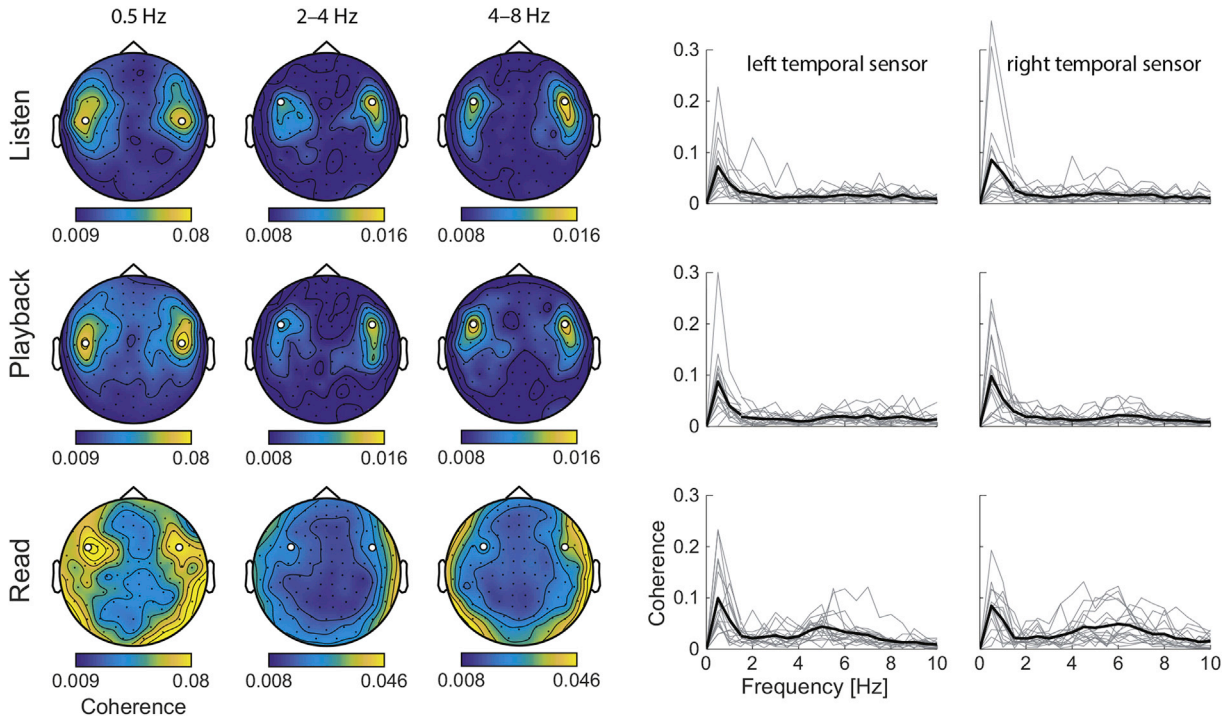

Fig. 3. Coherence at the sensor level. *Left* — Sensor distribution of the coherence at 0.5 Hz, 2–4 Hz, and 4–8 Hz averaged across subjects. White discs highlight the sensors of maximum coherence, or, in the read condition at 2–4 Hz and 4–8 Hz, the sensors suggestive of the presence of genuine speech brain tracking. *Right* — Individual (gray) and group-averaged (black) coherence spectra at the highlighted sensors. Values from 0 to 1.5 Hz are taken from sensors identified in the 0.5 Hz map, and values from 1.5 Hz to 10 Hz from the sensors identified in the 2–4 and 4–8 Hz maps (identical sensors).

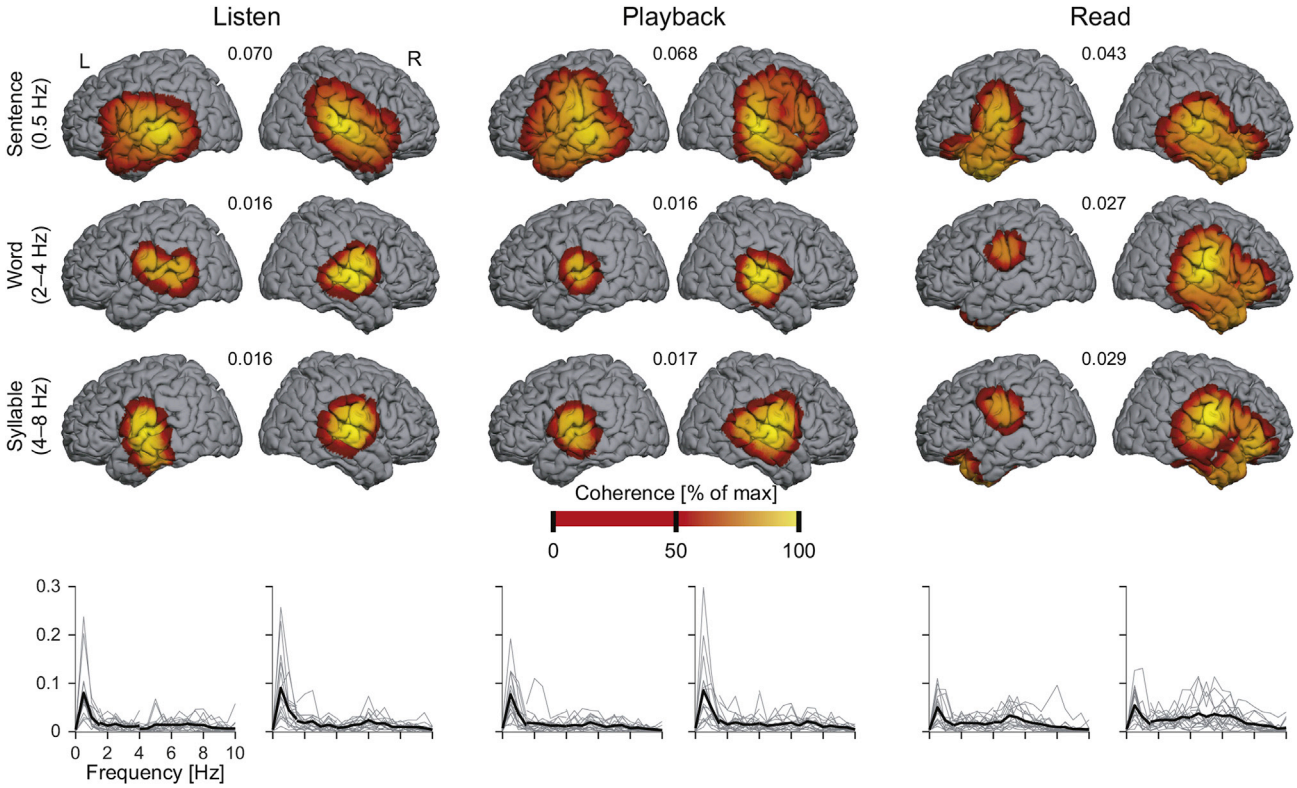

Fig. 4. Coherence in the source space. *Top* — Group-level coherence maps at 0.5 Hz, 2–4 Hz, and 4–8 Hz in the 3 conditions (*listen*, *playback* and *read*) thresholded at statistical significance level. The color scale is tailored to each coherence map: it ranges from 0 to its maximum (indicated in between brain images). *Bottom* — Individual (gray) and group-averaged (black) coherence spectra at the local maxima of coherence.

Table 1
MNI coordinates [mm] and coherence values of maximum speech brain tracking, as well as corresponding sensor-level coherence values controlled for artifacts in sensors at the edge of the sensor array.

	Left hemisphere			Right hemisphere		
	MNI coordinate [mm]	Source coherence	Sensor partial coherence	MNI coordinate [mm]	Source coherence	Sensor partial coherence
Cortical tracking of speech at 0.5 Hz						
listen	[–66 –25 1]	0.068	0.056	[66 –25 7]	0.070	0.060
playback	[–67 –28 –3]	0.063	0.046	[66 –24 3]	0.068	0.046
read	[–62 –10 12]	0.040	0.045	[66 –22 6]	0.043	0.041
Cortical tracking of speech at 2–4 Hz						
listen	[–63 –25 15]	0.0128	0.0106	[67 –11 6]	0.0157	0.0123
playback	[–64 –17 9]	0.0138	0.0104	[68 –23 3]	0.0162	0.0118
read	[–63 –15 30]	0.0178	0.0132	[68 –15 14]	0.0269	0.0169
Cortical tracking of speech at 4–8 Hz						
listen	[–61 –12 7]	0.0159	0.0138	[65 –13 7]	0.0162	0.0133
playback	[–63 –12 9]	0.0153	0.0122	[65 –11 7]	0.0172	0.0135
read	[–62 –13 28]	0.0209	0.0174	[65 –10 18]	0.0286	0.0249

mm, which, surprisingly, was not deemed statistically significant either ($F_{3,998} = 1.41$, $p = 0.24$). Detailed analyses revealed that this lack of significance pertained to that bootstrap coordinates of the local coherence maximum in the *listen* condition peaked at two locations: at the pSTS but also at the STAC (for more details, see supplementary material subsection 8.1.2). There was only one peak in the STAC in the *read* condition (see supplementary material subsection 8.1.3). These results indicate that reading aloud elicits cortical tracking of speech at 0.5 Hz only in STAC while speech listening also recruits the cortex around the pSTS (see supplementary material subsection 8.2.2).

At 2–4 Hz, local coherence maxima in the *read* condition localized in the parietal operculum bilaterally. Although these locations were distant from those in the *listen* condition by 18 mm (left hemisphere) and 9 mm (right hemisphere), the location-comparison test did not deem this difference in location statistically significant (left hemisphere, $F_{3,998} = 2.02$, $p = 0.11$; right hemisphere, $F_{3,998} = 2.51$, $p = 0.058$). Again, the lack of significance for the left-hemisphere assessment pertained to that bootstrap coordinates of the local coherence maximium in the *listen* condition peaked at both the pSTS and the parietal operculum (close to the STAC though) (see supplementary material subsection 8.1.4). There was only one peak in the parietal operculum in the *read* condition (see supplementary material subsection 8.1.5). This pattern of results indicates that reading aloud elicits cortical tracking of speech at 2–4 Hz only in the parietal operculum while speech listening also recruits the cortex around the pSTS (see supplementary material subsection 8.2.3).

At 4–8 Hz, local coherence maxima in the *read* condition localized in bilateral parietal operculum, i.e., more dorsally (above the sylvian fissure) than those in the *listen* condition by 19 mm (left hemisphere) and 11 mm (right hemisphere). The location-comparison test confirmed that this difference in location between *read* and *listen* conditions was statistically significant (left hemisphere, $F_{3,998} = 10.10$, $p < 0.0001$; right hemisphere, $F_{3,998} = 3.49$, $p = 0.015$).

Similar results were obtained for the comparison between *read* and *playback*. Still, the following statistical conclusions were different: the location comparison test did reveal a significant difference in location between *read* and *playback* at 2–4 Hz (left hemisphere, $F_{3,998} = 4.34$, $p = 0.0048$; right hemisphere, $F_{3,998} = 2.70$, $p = 0.044$), and no significant difference at 4–8 Hz in the right hemisphere ($F_{3,998} = 2.47$, $p = 0.061$).

3.3.3. Effect of conditions on the coherence strength

Values of the cortical tracking of speech quantified with coherence at condition-specific dominant sources were compared with repeated measures ANOVA, separately at 0.5 Hz, 2–4 Hz, and 4–8 Hz.

At 0.5 Hz there was a main effect of condition on coherence level ($F_{2,32} = 8.10$, $p = 0.0014$), no significant main effect of hemisphere ($F_{1,16} = 0.20$, $p = 0.66$), and no significant interaction ($F_{2,32} = 1.95$, $p = 0.16$). Post-hoc t-tests revealed that coherence values in *listen* (0.092 ± 0.039, mean ± SD of the mean coherence across hemispheres) and *playback* (0.090 ± 0.046) did not differ significantly ($t_{16} = 0.21$, $p = 0.84$), while values in *read* (0.057 ± 0.022) were significantly lower than those in *listen* ($t_{16} = 3.95$, $p = 0.0012$) and *playback* ($t_{16} = 3.47$, $p = 0.0031$).

At 2–4 Hz there was a main effect of condition on coherence level ($F_{2,32} = 10.2$, $p = 0.0004$), no significant main effect of hemisphere ($F_{1,16} = 4.31$, $p = 0.054$), and no significant interaction ($F_{2,32} = 1.39$, $p = 0.26$). Post-hoc t-tests revealed that coherence values in *listen* (0.0135 ± 0.0049) and *playback* (0.0150 ± 0.077) did not differ significantly ($t_{16} = 0.65$, $p = 0.53$), while values in *read* (0.0218 ± 0.0053) were significantly higher than those in *listen* ($t_{16} = 6.85$, $p < 0.0001$) and *playback* ($t_{16} = 3.16$, $p = 0.0061$).

At 4–8 Hz there was a main effect of condition on coherence level ($F_{2,32} = 16.6$, $p < 0.0001$), no significant main effect of hemisphere ($F_{1,16} = 2.23$, $p = 0.15$), and no significant interaction ($F_{2,32} = 0.06$, $p = 0.94$). Post-hoc t-tests revealed that coherence values in *listen* (0.0183 ± 0.0052) and *playback* (0.0191 ± 0.052) did not differ significantly ($t_{16} = 0.58$, $p = 0.57$), while values in *read* (0.0294 ± 0.0086) were significantly higher than those in *listen* ($t_{16} = 4.28$, $p = 0.0006$) and *playback* ($t_{16} = 4.37$, $p = 0.0005$).

3.4. Directionality results

rPDC was used to separate the relative contributions of signals reacting to speech (i.e., external feedback monitoring system) and signals preceding speech (i.e., internal speech monitoring system) to the cortical tracking of speech. In practice, rPDC was computed between speech temporal envelope and the time-course of each source of significant cortical tracking of speech.

Fig. 5 presents rPDC values in all conditions.

Table 2 details the number of participants displaying significant rPDC in all conditions, directions and frequency ranges of interest.

Paired t-tests revealed that rPDC was systematically higher in the speech → brain direction than in the brain → speech direction ($ps < 0.05$) except at 0.5 Hz in the left hemisphere in the *read* condition ($t_{16} = 1.61$, $p = 0.13$), and at 2–4 Hz in 2 instances ($ps = 0.06$).

The ANOVA assessment of rPDC values was performed with factors condition (*listen*, *playback* and *read*) and hemisphere (left and right) separately at 0.5 Hz, 2–4 Hz, and 4–8 Hz, and for the two coupling directions. There was a significant main effect of condition on speech → brain rPDC at 0.5 Hz ($F_{2,32} = 4.66$, $p = 0.017$) explained by lower values in *read* (10.8 ± 7.2, mean ± SD of the mean rPDC across hemispheres) than in *listen* (16.9 ± 7.9; $t_{16} = 2.70$, $p = 0.016$) and *playback* (17.0 ± 11.9; $t_{16} = 3.45$, $p = 0.0033$), with the two latter that did not differ significantly ($t_{16} = 0.063$, $p = 0.95$). There was also a significant effect of condition on brain → speech rPDC at 4–8 Hz ($F_{2,32} = 8.43$, $p = 0.0011$) explained by higher values in *read* (2.75 ± 0.74) than in *listen* (2.06 ± 0.38; $t_{16} = 2.90$, $p = 0.011$) and *playback* (2.02 ± 0.38; $t_{16} = 3.50$, $p = 0.0030$), with the two latter that did not differ significantly ($t_{16} = 0.30$, p

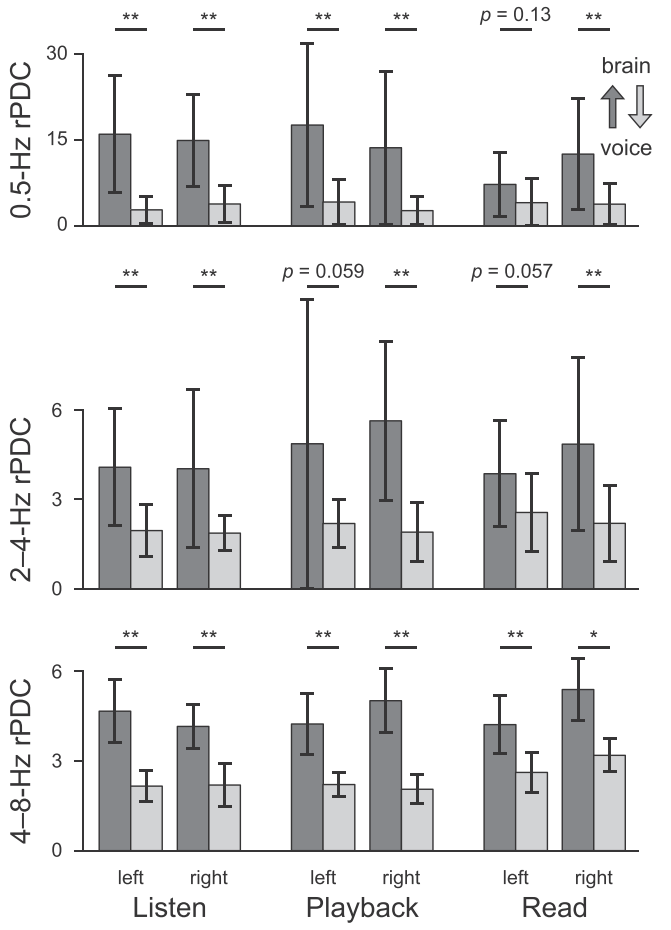

Fig. 5. Directionality assessment with renormalized partial directed coherence (rPDC). Bars display the mean and SD of rPDC values. There is 1 bar per conditions (*listen*, *playback* and *read*), frequency range of interest (0.5 Hz, 2–4 Hz, and 4–8 Hz), hemisphere (left and right), and direction (speech → brain and brain → speech). Significance of the comparison between directions are indicated above each pair of bars (*$p < 0.05$, **$p < 0.01$).

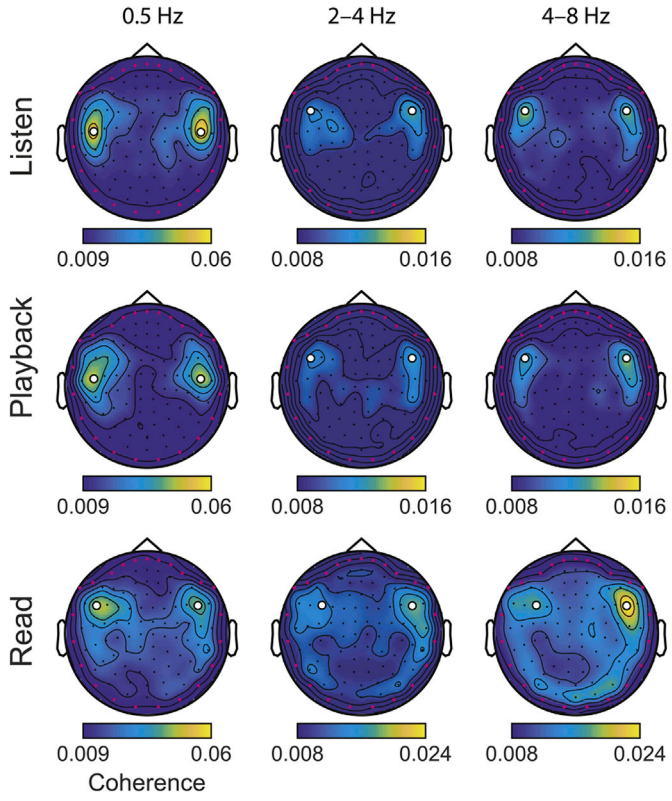

Fig. 6. Speech brain tracking at the sensor level assessed with partial coherence to control for artifacts in edge sensors (highlighted in magenta). Note that topographies at 2–4 Hz and 4–8 Hz are displayed with a different scale for the *read* and listening (*listen* and *playback*) conditions. White discs highlight the same sensors as those in Fig. 1 (sensors of maximum coherence, or, in the *read* condition at 2–4 Hz and 4–8 Hz, the sensors suggestive of the presence of genuine speech brain tracking).

= 0.77). There were no other significant main effects or interactions (*ps* > 0.1). Importantly, the factor condition had a significantly stronger effect on speech → brain than brain → speech rPDC at ~0.5 Hz (p = 0.034; Bootstrap-based assessment), and on brain → speech than speech → brain rPDC at 4–8 Hz (p = 0.028).

As it is unclear how artifacts contributed to these rPDC results at 0.5 Hz and 4–8 Hz, we repeated the rPDC analysis between speech temporal envelope and signals from a sensor that picked up strong artifacts (left hemisphere: MEG153*; right hemisphere: MEG263*). The ANOVA assessment of these rPDC values revealed in all 4 instances (2 coupling directions × 2 frequency ranges) a significant effect of condition (*ps* < 0.05) explained by higher values in *read* than in *listen* and *playback*. Also, the factor condition did not have a significantly stronger effect on speech → brain than brain → speech rPDC at ~0.5 Hz (p = 0.73), and it affected significantly less the brain → speech than speech → brain rPDC at 4–8 Hz (p = 0.002; i.e., the effect opposite to that found at the sources of significant cortical tracking of speech).

Table 2
Number of subjects displaying significant renormalized partial directed coherence (rPDC).

		listen		playback		read	
		left	right	left	right	left	right
0.5 Hz	speech → brain	16	15	14	13	10	12
	brain → speech	4	5	5	3	5	4
2–4 Hz	speech → brain	12	10	8	16	10	10
	brain → speech	2	1	4	2	7	3
4–8 Hz	speech → brain	10	8	9	9	12	9
	brain → speech	0	0	1	1	4	6

3.5. Partial coherence

Fig. 6 illustrates the cortical tracking of speech in sensor space controlled for artifacts in edge sensors using partial coherence. It is noteworthy that in the *read* condition, artifacts were substantially suppressed with partial coherence, while coherence at sensors above bilateral auditory cortices was essentially preserved. Moreover, partial coherence values were quite faithful to the source-space coherence values, as can be seen in group-level values displayed in Table 1 (similarity in source coherence and sensor partial coherence values).

Partial coherence levels were compared with repeated measures ANOVA with factors condition (*listen*, *playback* and *read*) and hemisphere (left and right) separately at 0.5 Hz, 2–4 Hz and 4–8 Hz. At 0.5 Hz, there were no significant effects nor interaction (*ps* > 0.5). At 2–4 Hz and 4–8 Hz there was a main effect of condition (2–4 Hz, $F_{2,32} = 7.75$, $p = 0.0018$; 4–8 Hz, $F_{2,32} = 18.3$, $p < 0.0001$), and no significant main effect of hemisphere nor interaction (*ps* > 0.2). Partial coherence values in *read* (2–4 Hz, 0.0226 ± 0.0057; 4–8 Hz, 0.0292 ± 0.0106; mean ± SD of the mean coherence across hemispheres) were higher than those in *listen*

(2–4 Hz, 0.0158 ± 0.0042, $t_{16} = 5.12$, $p = 0.0001$; 4–8 Hz, 0.0157 ± 0.0049, $t_{16} = 4.38$, $p = 0.0005$) and *playback* (2–4 Hz, 0.0160 ± 0.0085, $t_{16} = 2.86$, $p = 0.011$; 4–8 Hz, 0.0158 ± 0.0046, $t_{16} = 4.41$, $p = 0.0004$), while they did not differ significantly between *listen* and *playback* (2–4 Hz, $t_{16} = 0.0714$, $p = 0.94$; 4–8 Hz, $t_{16} = 0.14$, $p = 0.89$).

4. Discussion

This MEG study investigates how the human brain uses self-generated auditory information during connected speech production in the framework of feedback and internal speech monitoring systems. For that purpose, the cortical tracking of the hierarchically nested linguistic structures of speech was compared between reading aloud and two listening conditions, i.e., listening to someone reading a text or to a recording of the self-generated speech while reading aloud. Main original findings are that (i) the human brain tracks the speech temporal envelope at frequencies matching the occurence rate of the main linguistic structures of speech during both reading aloud and listening conditions, (ii) the auditory cortex tracks sentences/phrases (<1 Hz) with a decreased strength during reading aloud compared with listening conditions, (iii) the parietal operculum tracks syllables and words (2–4 and 4–8 Hz) during reading aloud while this tracking occurs at primary (4–8 Hz) or secondary (2–4 Hz) auditory cortex during listening, (iv) there was no statistically significant difference in tracking directionality between listening to a recording of a different text vs. playback, and (v) the cortical tracking of speech at <1 Hz, 2–4 Hz, and 4–8 Hz is dominated by auditory feedback processing during both reading aloud and listening, with an enhancement of internal speech monitoring at 4–8 Hz during reading aloud compared with listening. Taken together, these data bring unprecedented insights into the neural mechanisms at play for the monitoring of the auditory consequences of self-produced speech while reading aloud.

4.1. Cortical tracking of speech at frequencies < 1 Hz

The cortical tracking of speech at <1-Hz was attenuated during self-produced speech compared with listening to external speech. A control analysis, however, failed to corroborate this finding as it indicated similar rather than lower level of <1-Hz tracking during reading compared with listening. An attenuation would be well in line with the literature on speech-evoked brain responses. Indeed, auditory cortical responses (i.e., N100/M100 evoked response) to self-produced speech are typically attenuated or suppressed compared with those obtained during listening to a playback recording of the same sounds or during silent reading of a text (Curio et al., 2000; Houde et al., 2002; Numminen et al., 1999; Numminen and Curio, 1999). Such attenuation is absent when the auditory feedback is altered (e.g., pitch-shifted or alien speech feedback) (Heinks-Maldonado et al., 2006, 2005).

Our results also indicate that the cortical tracking of speech at <1-Hz while reading aloud is dominated by the speech feedback monitoring system. Indeed, both reading and listening gave rise to similarly low level of <1-Hz brain → speech coupling, which we posit, is the hallmark of reliance on the internal speech monitoring system. Note that the significant brain → speech coupling observed in ~30% of the subjects was most likely spurious, i.e., related to the fact that, in directionality assessment, strong coupling in one direction generates spurious coupling in the other direction (Faes et al., 2010).

Our results also shed light on the neural network involved in monitoring <1-Hz fluctuations in speech temporal envelope. During speech listening, this network seems to include the STAC and cortex around pSTS (possibly owing to inter-individual variability in the recruited area), while it only involves the STAC during reading aloud. This suggests that during self-generated speech, sensory feedback at phrasal/sentential level is mainly processed at early auditory cortices. The fact that we observed similar brain sources at the origin of the cortical tracking of speech <1 Hz also suggests that speech-related artifacts did not substantially influence source space results.

4.2. Cortical tracking of speech at 2–4 and 4–8 Hz

At 2–4 and 4–8 Hz, the cortical tracking of speech was stronger when reading aloud than during passive listening and it peaked in different cortical areas: primary (i.e., STAC for 4–8 Hz) or secondary (i.e., cortex around the pSTS for 2–4 Hz) auditory cortices during listening and parietal operculum during reading aloud. Tracking was mainly driven by the speech → brain contribution during reading aloud similarly to the listening conditions. There was however a significant enhancement in brain → speech coupling at 4–8 Hz during reading compared with listening conditions.

A previous MEG study demonstrated the existence of significant coupling between ventral primary sensorimotor (SM1) cortex (i.e., mouth area) and orbicular oris muscle activities during silent mouthing of a syllable (/pa/) periodically repeated at different frequencies (i.e., 0.8–5 Hz) (Ruspantini et al., 2012). This coupling phenomenon was driven by the mouth movement repetition rate during syllable mouthing and peaked at the individual spontaneous movement rate (i.e., self-paced rate of syllable articulation: ~2–3 Hz). It is therefore probably analogous (for a detailed discussion, see Bourguignon et al., 2019) to the previously described cortico-kinematic coupling (CKC) phenomenon, which is the coupling between the kinematics of finger or toe movements and the activity in the SM1 cortex corresponding to the moved limb (Bourguignon et al., 2012, 2011; Marty et al., 2015a, 2015b; Piitulainen et al., 2015). CKC indeed occurs at movement frequency (and harmonics), which is rather similarly visible in the rectified surface electromyogram and other kinematic-related signals such as acceleration, force and pressure (Piitulainen et al., 2013a,b). Of note, CKC is mainly driven by proprioceptive afferents to SM1 cortex (Bourguignon et al., 2015; Piitulainen et al., 2013a,b). Accordingly, our data suggests that during connected speech production, self-generated proprioceptive and auditory information resulting from word and syllable production are monitored in ventral SM1 cortex. In particular, the multimodal (i.e., somatosensory and auditory) nature of such speech-related sensory monitoring at SM1 cortex is supported by the rather low correlation between rhythmical lip movement and auditory speech temporal envelope during speech production (Bourguignon et al., 2020; Chandrasekaran et al., 2009; Park et al., 2016). The observed frequency-specific auditory feedback monitoring at SM1 cortex is in agreement with the external feedback monitoring system and the sensorimotor transformation theories of speech (Cogan et al., 2014; Hickok, 2012; Houde and Chang, 2015). Critically, the present study suggests that the neocortical areas involved in cortical tracking of speech at 2–4 Hz and 4–8 Hz are different during speech perception and production, which brings novel major insights into the neural bases of speech external feedback monitoring systems. Finally, the fact that the 4–8-Hz brain → speech coupling was significantly enhanced during reading (compared to listening) also suggests that the brain does generate internal sensorimotor representations of upcoming self-produced syllabic sounds, as put forward by the predictive coding theory (Friston, 2010). Importantly, the motor origin of this effect supports the notion that, in this frequency band, the brain computes the time-course of the to-be-produced articulation.

4.3. Methodological considerations

First, there was no difference between *listen* and *playback* conditions in any of the tested aspects of the cortical tracking of speech. This implies that the effects we uncovered (i) were not influenced by priming about upcoming speech content (intrinsic to *playback*) and (ii) not linked to a difference in speech rhythm between *listen* and *read*.

Second, neurophysiological mechanisms involved in overt language production are typically difficult to explore using MEG due to multiple sources of high-amplitude artifacts (e.g., head and jaw movements, muscular activity, etc.) that contaminate brain signals (see, e.g.,

Simmonds et al., 2014). Here, we were minimally concerned with muscle artifacts because their power mainly lies at frequencies above 20 Hz (Muthukumaraswamy, 2013) on which we did not focus. For the removal of other artifacts (at lower frequencies; including those induced by articulatory mouth gestures in *read*), we relied on tSSS, ICA, and threshold-based artifact rejection. Power spectral analyses showed that these preprocessing steps were efficient to dampen artifacts, but not to fully suppress them from sensors at the edge of the sensor array. We then reconstructed brain activity with a minimum variance beamformer, an approach that specifically passes activity coming from locations of interest while cancelling external interferences (Hillebrand et al., 2005). Still, sensor and source maps of cortical tracking of speech in the production condition indicated the presence of remaining movement artifacts characterized by coherence values comparable to those associated with genuine cortical tracking of speech. It is therefore probable that these artifacts were mild and hence not suppressed by tSSS, ICA or beamforming.

Beyond attempting to suppress artifacts, we conducted two control analyses designed to evaluate the impact of remaining artifacts on our results. First, by computing the rPDC between speech temporal envelope and MEG signals at sensors with high amplitude artifacts, we could demonstrate that reading-induced artifacts spuriously inflate rPDC values in both directions. This supports our two main findings since reading (compared with listening) was associated with decreased <1-Hz tracking (rather than increased), and specifically increased 4–8-Hz tracking in the brain → speech direction (rather than in both directions). As further support for the genuineness of these patterns of directionality, the effect of condition on coupling direction at the sensors with high amplitude artifacts was completely absent (<1-Hz) or even opposite to that found at the sources of significant cortical tracking of speech (4–8-Hz). Finally, we used partial coherence analysis in sensor space wherein we subtracted the contribution of MEG signals at sensors on the edge of the sensor array to support our source-level results. This second control analysis corroborated the finding that 2–4 and 4–8 Hz tracking is enhanced during reading compared with listening. However, it suggested similar rather than lower level of <1-Hz tracking during reading compared with listening. Further studies based on artifact free electrophysiological signals (e.g., intracranial recording; Cogan et al., 2014) will be required to confirm source-space results. Also, we cannot exclude that the sources of 2–4 and 4–8 Hz tracking in the reading condition may have been shifted by the artifacts remaining in sensor data. Invasive electrophysiological recordings are therefore warranted to identify the exact cortical network involved in tracking of self-produced speech, and specifically, to determine the relative contribution of STAC and parietal operculum. As a last limitation, this study involved a large number of statistical comparisons and hence, an inflated risk of identifying false positives. Therefore, further replication studies are warranted to confirm our main findings.

Despite these limitations that warrant taking the results of this study with some caution, we demonstrate that the cortical tracking of speech observed at <1 Hz during *listen* and *read* is rather similar in terms of brain areas and tracking level. Furthermore, the results obtained at 4–8 Hz during *read* are in line with those previously reported by Ruspantini et al. (2012) during syllable production. These data therefore suggest the existence of common cortical tracking of speech phenomena during self-generated speech production accompanying reading aloud and perception while listening to somebody reading a text aloud. Still, the generalization of these findings to production and perception of naturalistic speech (e.g., during natural conversation) needs to be done with caution and warrants further investigations. Indeed, as already stressed in the Introduction, reading aloud differs in several aspects (e.g, rhythmicity, prosody, etc.) from naturalistic speech (Alexandrou et al., 2018b). Further studies should also integrate time-locked recordings of speech-related peripheral signals (e.g., surface EMG of some facial muscles, lip movement kinematics through accelerometers or video-taping) that would contribute, e.g., to artifact removal and detection of speech production errors. They should also include a comprehensive behavioural assessment of speech listening and production tasks such as, e.g., task difficulty, or assessment of speech comprehension during listening. Such data would indeed contribute to more advanced electrophysiological data analyses than those done in the present study. Still, they would need to be carefully selected as scores to behavioral assessments such as, e.g., visual analogue scales for speech intelligibility or answers to questions typically tend to plateau in healthy subjects during listening in quiet auditory scenes (Vander Ghinst et al., 2019, 2016b).

Notwithstanding all these methodological considerations, this study represents a first step towards the understanding of the neural bases and functional aspects of the cortical tracking of speech during a form of speech production.

4.4. Conclusions

This study brings insights into how the human brain tracks the slow-temporal features of the auditory feedback during self-generation of speech. That is, while reading aloud, the reader's brain tracks the slow temporal structure of the self-generated speech. Also, the auditory cortex tracks phrases/sentences whereas the parietal operculum tracks words and syllables. Finally, data also suggests that both tracking mainly engage the feedback monitoring system, but with increased involvement of the internal speech monitoring system for syllable tracking at neocortical areas distinct from those recruited during speech perception.

CRediT authorship contribution statement

Mathieu Bourguignon: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Funding acquisition. **Nicola Molinaro:** Conceptualization, Writing - review & editing, Supervision, Funding acquisition. **Mikel Lizarazu:** Methodology, Writing - review & editing. **Samu Taulu:** Methodology, Writing - review & editing. **Veikko Jousmäki:** Writing - review & editing, Supervision. **Marie Lallier:** Writing - review & editing, Funding acquisition. **Manuel Carreiras:** Conceptualization, Writing - review & editing, Supervision, Funding acquisition. **Xavier De Tiège:** Conceptualization, Writing - original draft, Writing - review & editing, Supervision, Funding acquisition.

Acknowledgments

Mathieu Bourguignon has been supported by the program Attract of Innoviris (grant 2015-BB2B-10), by the Spanish Ministry of Economy and Competitiveness (grant PSI2016-77175-P), and by the Marie Skłodowska-Curie Action of the European Commission (grant 743562). Nicola Molinaro has been supported by the Spanish Ministry of Economy and Competitiveness (grant PSI2015-65694-P), the Agencia Estatal de Investigación (AEI), the Fondo Europeo de Desarrollo Regional (FEDER) and by the Basque government (grant PI_2016_1_0014). Mikel Lizarazu has been supported by the Agence Nationale de la Recherche (grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL). Xavier De Tiège is Post-doctorate Clinical Master Specialist at the Fonds de la Recherche Scientifique (F.R.S.-FNRS, Brussels, Belgium).

This research is supported by the Basque Government through the BERC 2018–2021 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation SEV-2015-0490. The MEG project at the CUB Hôpital Erasme is financially supported by the Fonds Erasme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuroimage.2020.116788.

References

Alexandrou, A.M., Saarinen, T., Kujala, J., Salmelin, R., 2018a. Cortical tracking of global and local variations of speech rhythm during connected natural speech perception. J. Cognit. Neurosci. 30, 1704–1719.

Alexandrou, A.M., Saarinen, T., Kujala, J., Salmelin, R., 2018b. Cortical entrainment: what we can learn from studying naturalistic speech perception. Lang. Cognit. Neurosci. 152, 1–13.

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.

Ashburner, J., Friston, K.J., 1999. Nonlinear spatial normalization using basis functions. Hum. Brain Mapp. 7, 254–266.

Ashburner, J., Neelin, P., Collins, D.L., Evans, A., Friston, K., 1997. Incorporating prior knowledge into image registration. Neuroimage 6, 344–352.

Bauer, J.J., Mittal, J., Larson, C.R., Hain, T.C., 2006. Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude. J. Acoust. Soc. Am. 119, 2363–2371.

Blackmer, E.R., Mitton, J.L., 1991. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition 39, 173–194.

Bóna, J., 2014. Temporal characteristics of speech: the effect of age and speech style. J. Acoust. Soc. Am. 136, EL116–E121.

Bortel, R., Sovka, P., 2014. Approximation of the null distribution of the multiple coherence estimated with segment overlapping. Signal Process. 96, 310–314.

Bourguignon, M., Baart, M., Kapnoula, E.C., Molinaro, N., 2020. Lip-reading enables the brain to synthesize auditory features of unknown silent speech. J. Neurosci. 40, 1053–1065.

Bourguignon, M., De Tiège, X., de Beeck, M.O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., Jousmäki, V., 2013a. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum. Brain Mapp. 34, 314–326.

Bourguignon, M., De Tiège, X., de Beeck, M.O., Ligot, N., Paquier, P., Van Bogaert, P., Goldman, S., Hari, R., Jousmäki, V., 2013b. The pace of prosodic phrasing couples the listener's cortex to the reader's voice. Hum. Brain Mapp. 34, 314–326.

Bourguignon, M., De Tiège, X., Op de Beeck, M., Pirotte, B., Van Bogaert, P., Goldman, S., Hari, R., Jousmäki, V., 2011. Functional motor-cortex mapping using corticokinematic coherence. Neuroimage 55, 1475–1479.

Bourguignon, M., Jousmäki, V., Dalal, S.S., Jerbi, K., De Tiège, X., 2019. Coupling between human brain activity and body movements: insights from non-invasive electromagnetic recordings. Neuroimage 203, 116177.

Bourguignon, M., Jousmäki, V., Op de Beeck, M., Van Bogaert, P., Goldman, S., De Tiège, X., 2012. Neuronal network coherent with hand kinematics during fast repetitive hand movements. Neuroimage 59, 1684–1691.

Bourguignon, M., Molinaro, N., Wens, V., 2018. Contrasting functional imaging parametric maps: the mislocation problem and alternative solutions. Neuroimage 169, 200–211.

Bourguignon, M., Piitulainen, H., De Tiège, X., Jousmäki, V., Hari, R., 2015. Corticokinematic coherence mainly reflects movement-induced proprioceptive feedback. Neuroimage 106, 382–390.

Burnett, T.A., Freedland, M.B., Larson, C.R., Hain, T.C., 1998. Voice F0 responses to manipulations in pitch feedback. J. Acoust. Soc. Am. 103, 3153–3161.

Chandrasekaran, C., Trubanova, A., Stillittano, S., Caplier, A., Ghazanfar, A.A., 2009. The natural statistics of audiovisual speech. PLoS Comput. Biol. 5, e1000436.

Clumeck, C., Suarez Garcia, S., Bourguignon, M., Wens, V., Op de Beeck, M., Marty, B., Deconinck, N., Soncarrieu, M.-V., Goldman, S., Jousmäki, V., Van Bogaert, P., De Tiège, X., 2014. Preserved coupling between the reader's voice and the listener's cortical activity in autism spectrum disorders. PLoS One 9, e92329.

Cogan, G.B., Thesen, T., Carlson, C., Doyle, W., Devinsky, O., Pesaran, B., 2014. Sensory–motor transformations for speech occur bilaterally. Nature 507, 94–98.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., Hari, R., 2000. Speaking modifies voice-evoked activity in the human auditory cortex. Hum. Brain Mapp. 9, 183–191.

Destoky, F., Philippe, M., Bertels, J., Verhasselt, M., Coquelet, N., Vander Ghinst, M., Wens, V., De Tiège, X., Bourguignon, M., 2019. Comparing the potential of MEG and EEG to uncover brain tracking of speech temporal envelope. Neuroimage 184, 201–213.

Ding, N., Melloni, L., Zhang, H., Tian, X., Poeppel, D., 2016. Cortical tracking of hierarchical linguistic structures in connected speech. Nat. Neurosci. 19, 158–164.

Ding, N., Simon, J.Z., 2013. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J. Neurosci. 33, 5728–5735.

Donhauser, P.W., Baillet, S., 2020. Two distinct neural timescales for predictive speech processing. Neuron 105, 385–393 e9.

Efron, B., 1979. Bootstrap methods: another look at the jackknife. Ann. Stat. 7, 1–26.

Faes, L., Pinna, G.D., Porta, A., Maestri, R., Nollo, G., 2004. Surrogate data analysis for assessing the significance of the coherence function. IEEE Trans. Biomed. Eng. 51, 1156–1166.

Faes, L., Porta, A., Nollo, G., 2010. Testing frequency-domain causality in multivariate time series. IEEE Trans. Biomed. Eng. 57, 1897–1906.

Friston, K., 2010. The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. 11, 127–138.

Friston, K.J., Frith, C.D., 2015. Active inference, communication and hermeneutics. Cortex 68, 129–143.

Garcés, P., López-Sanz, D., Maestú, F., Pereda, E., 2017. Choice of magnetometers and gradiometers after signal space separation. Sensors 17.

Gauvin, H.S., De Baene, W., Brass, M., Hartsuiker, R.J., 2016. Conflict monitoring in speech processing: an fMRI study of error detection in speech production and perception. Neuroimage 126, 96–105.

Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Parkkonen, L., Hämäläinen, M.S., 2014. MNE software for processing MEG and EEG data. Neuroimage 86, 446–460.

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. 11, e1001752.

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., Salmelin, R., 2001. Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. U. S. A 98, 694–699.

Guo, Z., Wu, X., Li, W., Jones, J.A., Yan, N., Sheft, S., Liu, P., Liu, H., 2017. Top-down modulation of auditory-motor integration during speech production: the role of working memory. J. Neurosci. 37, 10323–10333.

Halliday, D., 1995. A framework for the analysis of mixed time series/point process data—theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog. Biophys. Mol. Biol. 64, 237–278.

Hämäläinen, M.S., Ilmoniemi, R.J., 1994. Interpreting magnetic fields of the brain: minimum norm estimates. Med. Biol. Eng. Comput. 32, 35–42.

Heinks-Maldonado, T.H., Mathalon, D.H., Gray, M., Ford, J.M., 2005. Fine-tuning of auditory cortex during speech production. Psychophysiology 42, 180–190.

Heinks-Maldonado, T.H., Nagarajan, S.S., Houde, J.F., 2006. Magnetoencephalographic evidence for a precise forward model in speech production. Neuroreport 17, 1375–1379.

Hickok, G., 2012. Computational neuroanatomy of speech production. Nat. Rev. Neurosci. 13, 135–145.

Hillebrand, A., Barnes, G.R., Bosboom, J.L., Berendse, H.W., Stam, C.J., 2012. Frequency-dependent functional connectivity within resting-state networks: an atlas-based MEG beamformer solution. Neuroimage 59, 3909–3921.

Hillebrand, A., Singh, K.D., Holliday, I.E., Furlong, P.L., Barnes, G.R., 2005. A new approach to neuroimaging with magnetoencephalography. Hum. Brain Mapp. 25, 199–211.

Hotelling, H., 1931. The generalization of student's ratio. Ann. Math. Stat. 2, 360–378.

Houde, J.F., 1998. Sensorimotor adaptation in speech production. Science 279, 1213–1216.

Houde, J.F., Chang, E.F., 2015. The cortical computations underlying feedback control in vocal production. Curr. Opin. Neurobiol. 33, 174–181.

Houde, J.F., Nagarajan, S.S., Sekihara, K., Merzenich, M.M., 2002. Modulation of the auditory cortex during speech: an MEG study. J. Cognit. Neurosci. 14, 1125–1138.

Hyvärinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis.

Indefrey, P., 2011. The spatial and temporal signatures of word production components: a critical update. Front. Psychol. 2, 255.

Kawamoto, A.H., Liu, Q., Kello, C.T., 2015. The segment as the minimal planning unit in speech production and reading aloud: evidence and implications. Front. Psychol. 6, 1457.

Keitel, A., Gross, J., Kayser, C., 2018. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol. 16, e2004473.

Liu, Y., Fan, H., Li, J., Jones, J.A., Liu, P., Zhang, B., Liu, H., 2018. Auditory-motor control of vocal production during divided attention: behavioral and ERP correlates. Front. Neurosci. 12, 113.

Luo, H., Poeppel, D., 2007. Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54, 1001–1010.

Marty, B., Bourguignon, M., Jousmäki, V., Wens, V., Op de Beeck, M., Van Bogaert, P., Goldman, S., Hari, R., De Tiège, X., 2015. Cortical kinematic processing of executed and observed goal-directed hand actions. Neuroimage 119, 221–228.

Marty, B., Bourguignon, M., Op de Beeck, M., Wens, V., Goldman, S., Van Bogaert, P., Jousmäki, V., De Tiège, X., 2015. Effect of movement rate on corticokinematic coherence. Neurophysiol. Clin. 45, 469–474.

Molinaro, N., Lizarazu, M., 2018. Delta(but not theta)-band cortical entrainment involves speech-specific processing. Eur. J. Neurosci. 48, 2642–2650.

Molinaro, N., Lizarazu, M., Lallier, M., Bourguignon, M., Carreiras, M., 2016. Out-of-synchrony speech entrainment in developmental dyslexia. Hum. Brain Mapp. 37, 2767–2783.

Muthukumaraswamy, S.D., 2013. High-frequency brain activity and muscle artifacts in MEG/EEG: a review and recommendations. Front. Hum. Neurosci. 7, 138.

Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1–25.

Nozari, N., Dell, G.S., Schwartz, M.F., 2011. Is comprehension necessary for error detection? A conflict-based account of monitoring in speech production. Cognit. Psychol. 63, 1–33.

Numminen, J., Curio, G., 1999. Differential effects of overt, covert and replayed speech on vowel-evoked responses of the human auditory cortex. Neurosci. Lett. 272, 29–32.

Numminen, J., Salmelin, R., Hari, R., 1999. Subject's own speech reduces reactivity of the human auditory cortex. Neurosci. Lett. 265, 119–122.

Oldfield, R.C., 1971. Edinburgh Handedness Inventory. PsycTESTS Dataset.

Park, H., Kayser, C., Thut, G., Gross, J., 2016. Lip movements entrain the observers' low-frequency brain oscillations to facilitate speech intelligibility. Elife 5, e14521.

Park, H., Thut, G., Gross, J., 2018. Predictive entrainment of natural speech through two fronto-motor top-down channels. Lang. Cognit. Neurosci. https://doi.org/10.1080/23273798.2018.1506589.

Peelle, J.E., Gross, J., Davis, M.H., 2013a. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebr. Cortex 23, 1378–1387.

Peelle, J.E., Gross, J., Davis, M.H., 2013b. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cerebr. Cortex 23, 1378–1387.

Piitulainen, H., Bourguignon, M., De Tiège, X., Hari, R., Jousmäki, V., 2013. Coherence between magnetoencephalography and hand-action-related acceleration, force, pressure, and electromyogram. Neuroimage 72, 83–90.

Piitulainen, H., Bourguignon, M., De Tiège, X., Hari, R., Jousmäki, V., 2013. Corticokinematic coherence during active and passive finger movements. Neuroscience 238, 361–370.
Piitulainen, H., Bourguignon, M., Hari, R., Jousmäki, V., 2015. MEG-compatible pneumatic stimulator to elicit passive finger and toe movements. Neuroimage 112, 310–317.
Poeppel, D., 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as "asymmetric sampling in time.". Speech Commun. 41, 245–255.
Reuter, M., Schmansky, N.J., Rosas, H.D., Fischl, B., 2012. Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage 61, 1402–1418.
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.
Schelter, B., Timmer, J., Eichler, M., 2009. Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. J. Neurosci. Methods 179, 121–130.
Schelter, B., Winterhalder, M., Eichler, M., Peifer, M., Hellwig, B., Guschlbauer, B., Lücking, C.H., Dahlhaus, R., Timmer, J., 2006. Testing for directed influences among neural signals using partial directed coherence. J. Neurosci. Methods 152, 210–219.
Schneider, T., Neumaier, A., 2001. Algorithm 808: ARfit—a matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. ACM Trans. Math Software 27, 58–65.
Sekihara, K., Nagarajan, S.S., 2008. Adaptive Spatial Filters for Electromagnetic Brain Imaging. Springer Science & Business Media.
Shiller, D.M., Sato, M., Gracco, V.L., Baum, S.R., 2009. Perceptual recalibration of speech sounds following speech motor learning. J. Acoust. Soc. Am. 125, 1103–1113.
Simmonds, A.J., Leech, R., Collins, C., Redjep, O., Wise, R.J.S., 2014. Sensory-motor integration during speech production localizes to both left and right plana temporale. J. Neurosci. 34, 12963–12972.
Sulpizio, S., Kinoshita, S., 2016. Editorial: bridging reading aloud and speech production. Front. Psychol. 7, 661.
Taulu, S., Simola, J., 2006. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys. Med. Biol. 51, 1759–1768.
Taulu, S., Simola, J., Kajola, M., 2005. Applications of the signal space separation method. IEEE Trans. Signal Process. 53, 3359–3372.
Tremblay, S., Shiller, D.M., Ostry, D.J., 2003. Somatosensory basis of speech production. Nature 423, 866–869.
Vander Ghinst, M., Bourguignon, M., Niesen, M., Wens, V., Hassid, S., Choufani, G., Jousmäki, V., Hari, R., Goldman, S., De Tiège, X., 2019. Cortical tracking of speech-in-noise develops from childhood to adulthood. J. Neurosci. 39, 2938–2950.
Vander Ghinst, M., Bourguignon, M., Op de Beeck, M., Wens, V., Marty, B., Hassid, S., Choufani, G., Jousmäki, V., Hari, R., Van Bogaert, P., Goldman, S., De Tiège, X., 2016a. Left superior temporal gyrus is coupled to attended speech in a cocktail-party Auditory scene. J. Neurosci. 36, 1596–1606.
Vander Ghinst, M., Bourguignon, M., Op de Beeck, M., Wens, V., Marty, B., Hassid, S., Choufani, G., Jousmäki, V., Hari, R., Van Bogaert, P., Goldman, S., De Tiège, X., 2016b. Left superior temporal gyrus is coupled to attended speech in a cocktail-party Auditory scene. J. Neurosci. 36, 1596–1606.
Vigario, R., Sarela, J., Jousmiki, V., Hamalainen, M., Oja, E., 2000. Independent component approach to the analysis of EEG and MEG recordings. IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 47, 589–593.
Wens, V., Marty, B., Mary, A., Bourguignon, M., Op de Beeck, M., Goldman, S., Van Bogaert, P., Peigneux, P., De Tiège, X., 2015. A geometric correction scheme for spatial leakage effects in MEG/EEG seed-based functional connectivity mapping. Hum. Brain Mapp. 36, 4604–4621.
Zion Golumbic, E.M., Poeppel, D., Schroeder, C.E., 2012. Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. Brain Lang. 122, 151–161.
Zion Golumbic, E.M., Zion Golumbic, E.M., Ding, N., Bickel, S., Lakatos, P., Schevon, C.A., McKhann, G.M., Goodman, R.R., Emerson, R., Mehta, A.D., Simon, J.Z., Poeppel, D., Schroeder, C.E., 2013. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party. Neuron 77, 980–991.