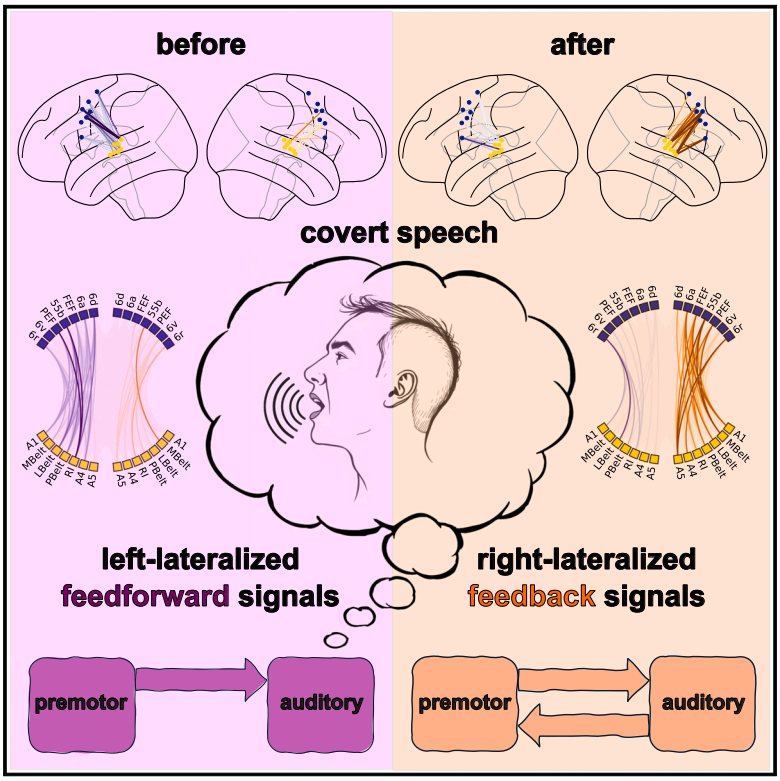
Cell Reports

Article

Time-resolved hemispheric lateralization of audiomotor functional connectivity during covert speech production

Graphical abstract



Authors

Francesco Mantegna, Joan Orpella, David Poeppel

Correspondence

fm1672@nyu.edu

In brief

Mantegna et al. investigated interareal communication during covert speech using magnetoencephalography. They observed distinct hemispheric lateralization patterns before and after expected covert production. Their findings suggest that covert speech involves a parallel neural architecture similar to overt speech, with leftward and rightward lateralization corresponding to feedforward and feedback signals, respectively.

Highlights

- Covert speech involves interareal communication between sensory and motor regions
- Hemispheric lateralization of functional connectivity can be measured using MEG
- Connectivity is left lateralized before and right lateralized after covert production
- Covert speech involves a parallel neural architecture similar to overt speech

Mantegna et al., 2025, Cell Reports *44*, 115137
January 28, 2025 Published by Elsevier Inc.
https://doi.org/10.1016/j.celrep.2024.115137

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Time-resolved hemispheric lateralization of audiomotor functional connectivity during covert speech production

Francesco Mantegna,[1,2,5,*] Joan Orpella,[1,3] and David Poeppel[1,4]
[1]Department of Psychology, New York University, New York, NY 10003, USA
[2]Department of Engineering Science, Oxford University, Oxford, Oxfordshire, UK
[3]Department of Neuroscience, Georgetown University Medical Center, Washington, DC 20007, USA
[4]Center for Language, Music and Emotion (CLaME), New York University, New York, NY 10003, USA
[5]Lead contact
*Correspondence: fm1672@nyu.edu
https://doi.org/10.1016/j.celrep.2024.115137

SUMMARY

Covert speech involves the internal generation of articulatory movements and their sensory consequences. While overt speech involves a combination of feedforward and feedback signals, feedback signals may be substantially different, or even absent, during covert speech. Despite the differences, we conjectured that sensorimotor interareal communication during covert speech is implemented through the same channels recruited during overt speech. An influential overt speech model proposed that feedforward and feedback signals are segregated to the left and right hemispheres, respectively. Here, we used magnetoencephalography to investigate the lateralization of functional connectivity before and after covert speech production. The data reveal leftward lateralization preceding and rightward lateralization following predicted covert speech onset. This alternating lateralization pattern is observed only in the connection between premotor and auditory regions and in the alpha frequency band. The electrophysiological data, derived entirely from covert speech, add a provocative perspective to adjudicate between overt speech motor control models.

INTRODUCTION

Speech production requires the rapid coordination of multiple articulators in the vocal tract. Despite its inherent complexity, speech production typically feels effortless, because it is supported by extensive previous experience. Speakers leverage the practice accumulated throughout a lifetime to generate a blueprint of upcoming movements and their sensory consequences. This blueprint or plan can be used to imagine movements and their consequences with remarkable precision when covertly speaking utterances,[1] i.e., in the absence of motor execution. Although the vividness of imagined speech is subject to individual differences, most individuals report internal auditory and somatosensory percepts associated with covert speech production, except in rare cases.[2] These internal percepts are consistent with the notion of "efference copies"[3]—internal copies of planned movements and their consequences. The neurophysiological bases of efference copies have been investigated across different motor domains, e.g., eye[4] or limb movement.[5] Previous studies have investigated the neural dynamics of speech efference copies using time-resolved methods.[6–8] However, the information flow (e.g., as measured by functional connectivity analyses) associated with speech efference copies and the functional anatomical organization of its underlying neural architecture are not understood.

Different hypotheses have been advanced about the dynamics of information flow associated with the internal generation of covert speech. One possibility is that covert speech involves only feedforward (i.e., top-down) signals. That is, in the absence of sensory feedback, the precision of the efference copy cannot be evaluated (cf. the simulation theory of motor imagery[9]). Another possibility is that covert speech involves a combination of feedforward (top-down) and feedback (i.e., bottom-up) signals. In this case, the efference copy is evaluated against an internally generated sensory target, which generates feedback. There is theoretical and experimental evidence for this view (cf. the emulation theory of motor imagery[10]). For example, there exists indirect evidence for the ability to internally evaluate efference copies from instances of motor learning without movement.[11,12] This is most compelling for motor skills associated with high expertise, such as during silent practice in musicians.[13] Another example is the decreased reliance on feedback that comes with extensive motor practice.[14–16] Overall, the existence of feedback signals during covert speech remains an open question. Even if we put to the test the feedforward plus feedback account, the feedback signals associated with covert speech may be substantially different from the ones associated with overt speech because they operate in the absence of external input.

These two differing accounts of covert speech production have implications for the neural architecture supporting

interareal communication between motor and sensory areas. Critically, a feedforward plus feedback account requires that information is transferred in non-interfering channels, which is not the case in a feedforward-only account. Overt speech studies point to a functional anatomic segregation of feedforward and feedback signals, but covert speech data are lacking. Because of the compelling similarities between covert and overt speech, we conjecture that the same neural architecture would be recruited during covert speech as in overt speech. Indeed, despite the differences between covert and overt speech—mostly related to the inhibitory control[17] and self-monitoring mechanisms[18]—previous research points to fundamental similarities in the underlying neural operations.[19]

Functional segregation of feedforward and feedback signals can be accomplished in various ways, but two possibilities have garnered empirical support in the overt speech literature: separation between hemispheres and separation within a hemisphere. The DIVA model of speech motor control[20] proposes a parallel neural architecture whereby feedforward signals primarily occur in the left hemisphere, while feedback signals predominantly take place in the right. Support for this hypothesis comes from clinical,[21,22] functional magnetic resonance imaging (fMRI),[23–25] and non-human animal studies.[26] However, while these studies successfully isolate these two distinct operation types, it is not known whether differences in hemispheric lateralization can be observed within a realistic timescale for speech production (that is, on the millisecond scale). On the other hand, intracortical studies of speech production in humans[27] and vocalization in non-human animals[28] point to a reversal of the information flow between motor and auditory areas around the time of vocal production within a single hemisphere, suggesting that forward and feedback signals may be resolved (albeit arguably through different anatomic projections) within the same hemisphere. While these studies capture rapid temporal dynamics, they could not investigate hemispheric lateralization due to limited electrode coverage that allowed them to measure only a single hemisphere. Magnetoencephalography (MEG) offers a balance between spatial and temporal resolution to examine the hypothesis of hemispheric lateralization for forward and feedback signals within a realistic timescale during covert speech.

Here, we characterize functional connectivity dynamics between motor and sensory areas during covert speech. Based on overt speech, we hypothesized that feedforward and feedback signals are lateralized to the left and right hemispheres, respectively. Participants were visually presented with consonant-vowel syllables (e.g., pa, ta, and ka) and were instructed to articulate the syllables either overtly (i.e., speaking) or covertly (i.e., imagined). The task was performed by participants in a behavioral pretest and in an MEG experiment. In the behavioral pretest, participants performed the task overtly (with movement and sound). We estimated the mean production latency across participants based on burst release onsets that were annotated by visual inspection of sound recordings. It was assumed based on existing literature[29,30] that the same amount of time would be necessary to internally initiate the covert articulation of the syllables. In the MEG experiment, participants performed the task covertly (no movement and no sound). Functional connectivity between motor and sensory regions of interest (ROIs) was estimated. Systematic phase delays between ROI pairs were measured as an index of functional connectivity. The relative difference in functional connectivity between the left and the right hemispheres over time was measured as an index of hemispheric lateralization. We conjectured that feedforward and feedback signals are present before and after the expected covert speech onset, respectively. Therefore, we compared hemispheric lateralization measured in these two different time windows. The results motivate the conclusion that interareal communication between motor and sensory regions preceding and following the predicted covert speech onset lateralizes to left and right hemispheres, respectively.

RESULTS

Fifty-three right-handed volunteers took part in the behavioral study and 40 in the MEG study. Participants were instructed to covertly speak syllables that were visually prompted on the screen. We presented three different consonant-vowel syllables (/pa/, /ta/, and /ka/). For each syllable, we kept the vowel constant and varied the consonant. One syllable per trial was presented in a randomized order. We used different syllables to ensure that participants retrieve a motor plan for each trial instead of reusing the same motor plan over and over. Likewise, the associated auditory target differed.

Measuring temporal landmarks in an overt speech pretest

Covert speech involves a significant amount of variability within and across participants. Such variability is problematic, because the neural data analyses are sensitive to the temporal misalignment of brain responses. It is not possible to measure this temporal misalignment here, because the task does not involve overt behavior. Nevertheless, to provide temporal landmarks, we tested participants' performance on an additional task. This task was designed both to collect behavioral measurements that can be used as a reference to interpret the neural data and to establish an exclusion criterion for the neuroimaging experiment.

Participants performed an overt speech behavioral pretest (Figure 1A). Participants were asked to produce the syllable presented on the screen aloud. For each trial, we annotated speech sound onsets by inspection of sound recordings. The speech sound onset annotation criterion coincided with the release burst onset (i.e., the earliest acoustic energy that can be measured during speech production, corresponding to consonant onset). (See Figure S1 for vowel onset annotation.) The results of the behavioral pretest were used to exclude participants from the MEG experiment. If the (25%–75%) interquartile range was larger than 400 ms or the median was lower than 200 ms or higher than 600 ms, participants were excluded from the MEG experiment. Thirteen participants were excluded based on behavioral pretest results and therefore did not participate in the MEG study. We parameterized the remaining participants' speech latencies distribution using a Gamma probability density function (Figure 1B), which adequately captures the basic features of the distribution (i.e., asymmetrical with a long right tail

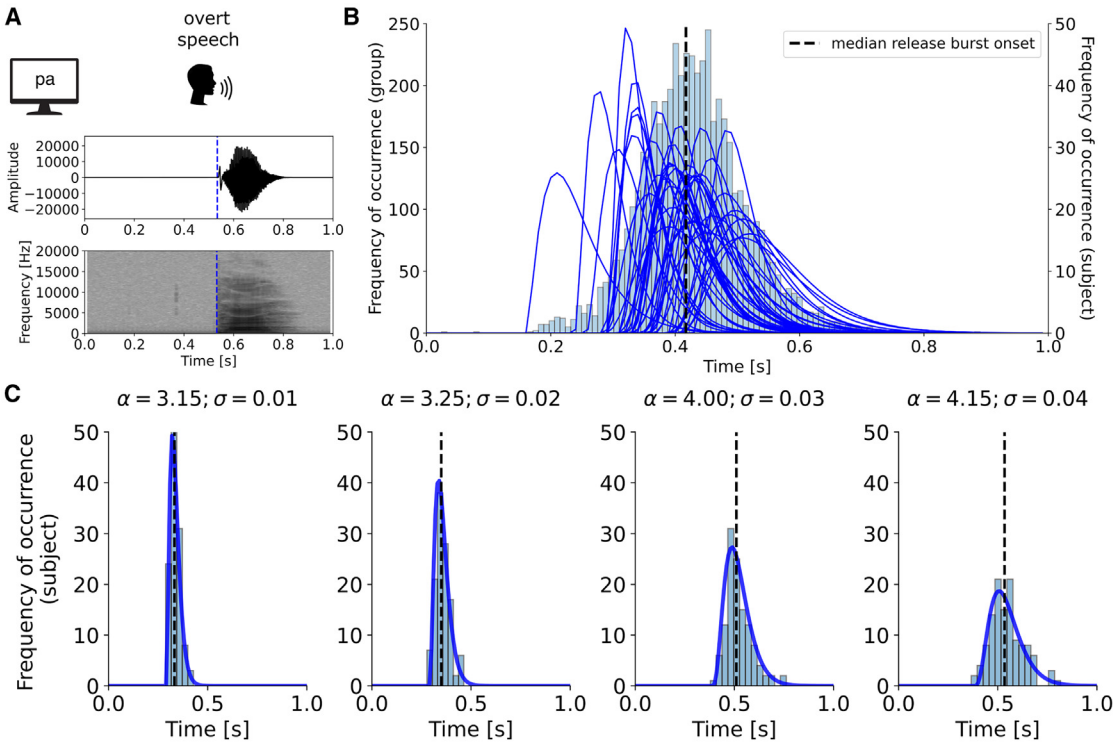


Figure 1. Overt speech production latencies measured in the behavioral pretest

(A) Experimental paradigm (top) and a waveform and a spectrogram for a speech sound (bottom). Speech sound onsets were annotated by visual inspection of the release burst onset for each trial (blue vertical dashed line).

(B) Gamma probability density function fit of the speech sound onset distribution for each participant (blue lines) and a histogram of speech sound onsets across participants. The black dashed vertical line marks the median speech sound onset across participants.

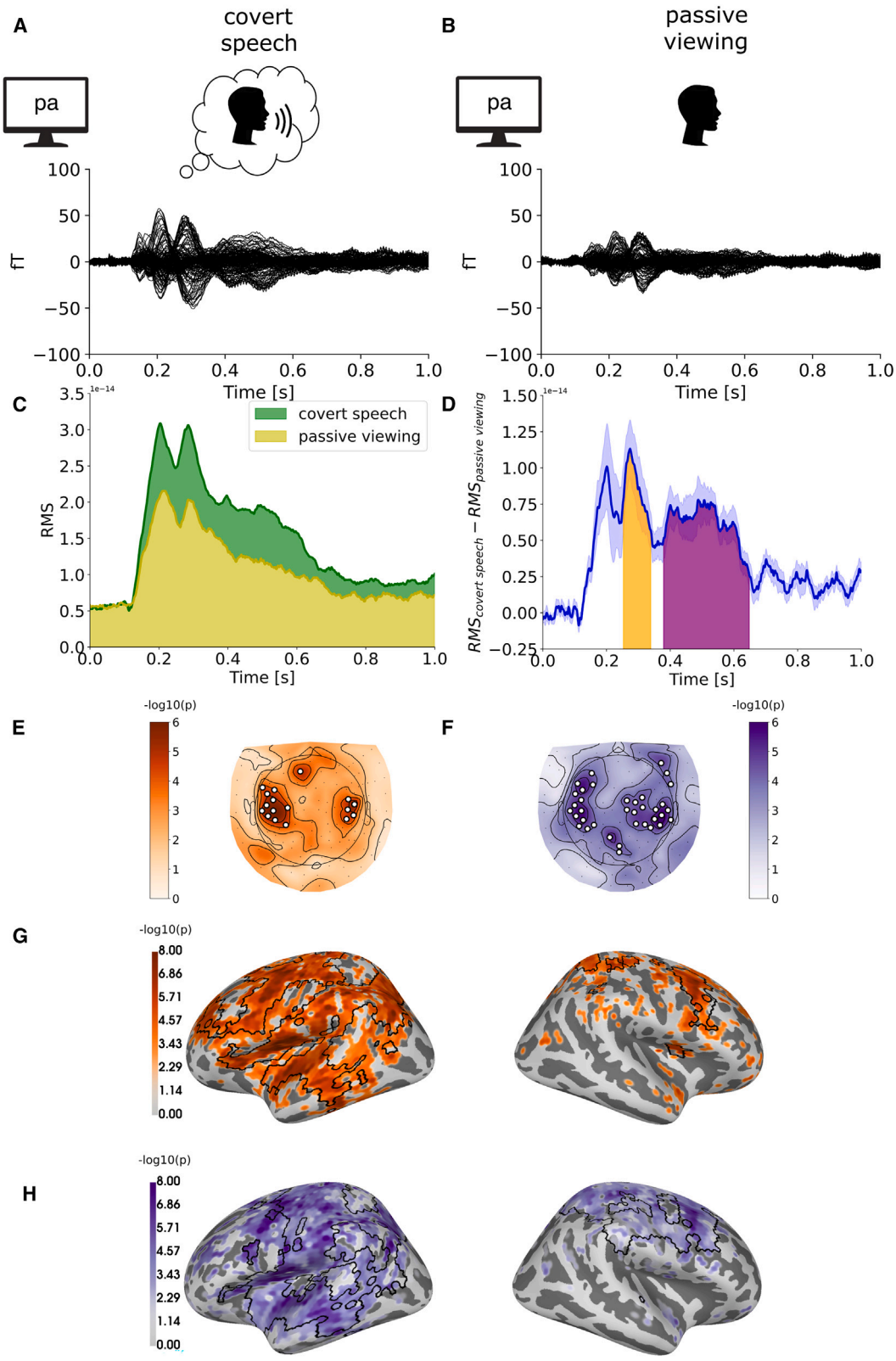
(C) Four representative individual fit examples illustrate how different combinations of the shape (α) and scale (σ) parameters capture the distinctive features of the speech onset distribution.

and leftward skew). The Gamma function is based on two parameters: a shape (α) and a scale (σ) parameter. The shape parameter represents the skewness of the distribution and the scale parameter represents the width of the distribution (see Figure 1C for individual fit examples). The distribution of speech onsets after exclusion is shown in Figure 1B. The median speech onset across participants was 417 ms (std 51 ms) after visual presentation. Based on previous research, we assumed that the production latency for covert and overt speech should be approximately the same.[29,30] Therefore, we used the median overt speech sound onset as a reference to interpret neural data recorded during covert speech.

MEG and electromyography event-related responses

The main experiment featured two conditions (Figures 2A and 2B, top): covert speech and passive viewing. In the covert speech condition, participants internally simulated speaking the visually presented syllable. In the passive viewing condition, participants were asked to view the visual stimulus but were requested not to read or speak. The passive viewing condition served as a control condition. The contrast between the two was designed to test whether the neural responses associated with covert speech are distinct from those associated with passive viewing. We expected this to be the case, because the neural responses associated with covert speech should be both exogenously elicited by the visual stimulus and endogenously generated by covert speech planning and its (covert) consequences. The neural responses elicited by passive viewing should be primarily exogenously elicited, although some degree of planning might still be present.

We collected MEG data while participants performed the task. To illustrate the data, the grand average ($n = 40$) event-related responses are shown in Figures 2A and 2B. First, we established that high temporal precision in the overt speech pretest translated into more stable brain responses during covert speech. For an explicit analysis of the relationship between behavioral and neural variability, see Figure S2. Then, we compared event-related responses elicited during the covert speech and the passive viewing conditions. The stacked area chart in Figure 2C shows that the magnitude of the evoked response measured during covert speech is higher than the magnitude of the evoked response measured during passive viewing from ~200 to ~600 ms. We ran a cluster-based permutation test[31] to assess the differences between covert speech and passive viewing conditions in Figure 2D. We observed two clusters where the evoked responses measured during the covert speech condition were significantly larger ($p < 0.005$) than during the passive viewing condition. The two clusters show different



(legend on next page)

temporal profiles: we observed one cluster in an earlier time window from ~250 to ~360 ms and one cluster in a later time window from ~390 to ~650 ms. Sensor space visualization in Figures 2E and 2F reveals that the differences between covert speech and passive viewing conditions are most pronounced in frontal and temporal sensors in both clusters. Source space projection in Figures 2G and 2H reveals that the differences between covert speech and passive viewing conditions are most pronounced in premotor, motor, insular, somatosensory, and temporal regions in both clusters (largely overlapping with the peri-Sylvian language network). Overall, formal comparisons reveal similarities and differences between conditions. In the ~0–250 ms time window, the evoked responses are similar in both conditions, because they are exogenously elicited by the visual stimulus. In the ~250–650 ms time window, the evoked responses are larger in the covert speech condition, suggesting a differential engagement of the brain areas typically recruited during speech production.

To isolate the endogenously generated responses from exogenously elicited responses, we used MEG-MRI co-registration to project the signals to source space and to select the signals coming from motor and sensory areas. T1 anatomical scans were used for some participants and a template brain was used for the rest of the participants to guide source reconstruction. For source reconstruction, we used a dynamic imaging of coherent sources (DICS) beamformer filter[32] to ensure that the phase relationship existing in sensor space was preserved in source space. We selected *a priori* ROIs that are hypothesized to be the neural generators of feedforward and feedback signals in prevailing speech motor control models. We used a brain atlas[33] to define the spatial location of six ROIs (three in each hemisphere) pertaining to premotor, somatosensory, and auditory cortex. As is established in the literature,[34] we extracted one source reconstructed MEG time series for each ROI.

We also measured electromyography (EMG) while participants performed the task. Muscle movements were analyzed to ensure that participants were not moving their articulators. This control analysis is crucial to safeguard that the measurements are not contaminated by motor artifacts. We recorded muscle movements from electrodes placed on the jaw and lower lip. Previous research has shown that jaw and lip movement is sufficient to distinguish stop consonants during overt speech.[35] The grand average event-related EMG response shows microdeflections in both conditions (Figures 3A and 3B, bottom traces). These microdeflections may reflect muscle contractions that are not inhibited when the motor output is suppressed (i.e., micromovements[36]). In line with this view, the covert speech microdeflections should be larger than the passive viewing microdeflections. We compared event-related responses elicited during the covert speech and the passive viewing conditions. The stacked area chart in Figure 3C shows that the magnitude of the evoked response measured during covert speech is higher than the magnitude of the evoked response measured during passive viewing from ~200 to ~500 ms. However, we observed no significant differences between conditions in Figure 3D, suggesting that microdeflections may reflect involuntary muscle contractions instead. In line with this view, previous studies have shown that microdeflections do not discriminate phonetic features during covert speech[37] (see Figure S3 for an explicit analysis of MEG signal contamination during covert vs. overt articulation).

Consistent phase shifts index functional connectivity between ROIs

Next, we measured functional connectivity between sensory and motor areas. These analyses served to capture the crosstalk between these areas associated with feedforward and feedback signals. Previous electrophysiological evidence as well as biophysical models suggests that systematic phase shifts in the neural oscillations measured from different brain areas reflect interactions between those brain areas.[38] The analysis pipeline is shown in Figure 4E. For each time point across the trial epoch, we estimated systematic phase shifts between time series extracted from ROI pairs. We used the weighted phase lag index (wPLI) to measure functional connectivity.[39] This functional connectivity measure is well suited to MEG data for methodological reasons. wPLI does not capture spurious connections due to volume conduction. wPLI is an undirected functional connectivity measure—it provides a value between 0 and 1 indicating the probability of the absence (0) or presence (>0) of a phase shift between two time series. We used the expected covert speech onset as a reference to distinguish between phase shifts associated with feedforward and feedback signals. We assumed that feedforward and feedback signals happen before and after expected covert speech production, respectively.

The functional connectivity analysis is frequency specific. Phase synchronization estimation is sensitive to frequency range and amplitude modulation within that range. Previous studies combining modeling and experimental data have shown that phase synchronization estimation is more accurate in narrow than in broad frequency bands.[40] An empirical investigation of the spectral signatures in the neural data for each subject allowed us to identify the most prominent frequencies in the covert speech task. We measured the power spectral density from

Figure 2. Cross-condition comparison of MEG evoked responses

(A and B) Schema of the experimental paradigm (top) and the MEG (bottom) grand average (n = 40) of event-related responses during the covert speech (A) and the passive viewing (B) conditions. The MEG waveform consists of one line for each of the 157 measured sensors.

(C) The stacked area chart allows for a comparison between the covert speech (green curve) and the passive viewing (yellow curve) MEG evoked responses.

(D) The difference between covert speech and passive viewing evoked responses (blue line). The shaded area around the mean (blue line) represents the standard error of the mean. The orange and the purple shaded areas represent the two time windows where the difference between conditions is statistically significant ($p < 0.005$).

(E and F) The topographies of the earlier (E) (orange) and later (F) (purple) significant time windows. White dots represent the sensors showing significant differences across participants.

(G and H) The source reconstruction of the earlier (G) (orange) and later (H) (purple) significant time windows. Black outlines represent the vertices showing significant differences across participants.

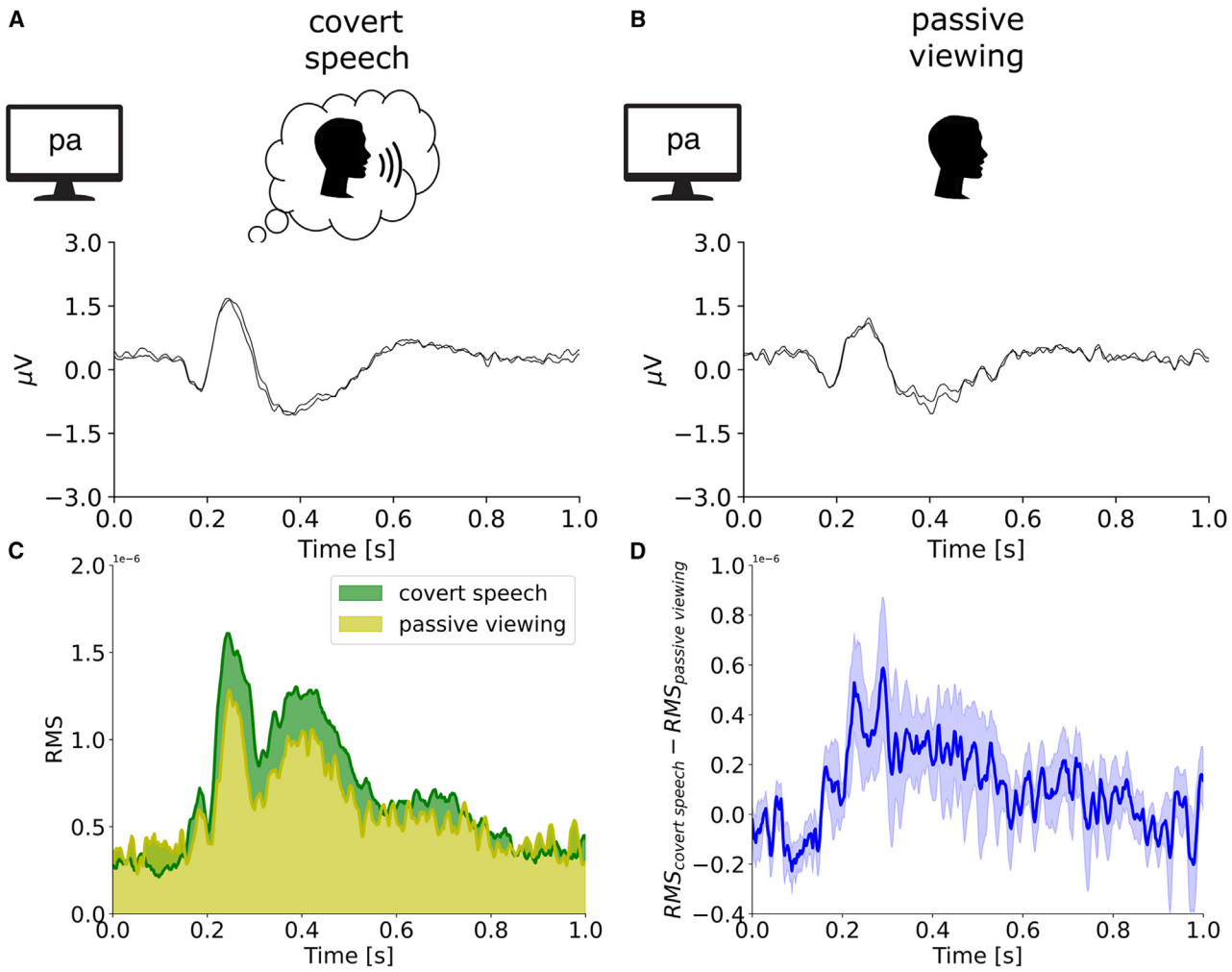


Figure 3. Cross-condition comparison of EMG evoked responses

(A and B) Schema of the experimental paradigm (top) and EMG (bottom) grand average (n = 40) of event-related responses during the covert speech (A) and the passive viewing (B) conditions. The EMG waveform consists of two lines (lip and jaw electrodes).

(C) The stacked area chart allows for a comparison between the covert speech (green curve) and the passive viewing (yellow curve) EMG evoked responses.

(D) The difference between covert speech and passive viewing evoked responses (blue line). The shaded area around the mean (blue line) represents the standard error of the mean.

frontal, parietal, and temporal MEG sensors (Figure 4A, blue dots) capturing the signal recorded from sensory and motor areas (i.e., to minimize visual alpha). The parameterization of the spectrum procedure is illustrated in Figure 4A. We fitted a model to parameterize the power spectrum into periodic and aperiodic components.[41] The periodic component is task related, while the aperiodic component is task irrelevant. Figure 4C illustrates the overall occurrence of frequency peaks across the cohort. The peaks were concentrated around 10, 20, and 30 Hz. Since the center peak frequencies in the low beta and high beta frequency bands are multiples of 10 Hz (the highest measured peak), it is possible that those are harmonics.[42] To assess functional connectivity, we focused on the frequency ranges that show a good signal-to-noise ratio (i.e., sufficient power to reliably estimate phase), regardless of whether those frequencies are independent or harmonic. Although there are some differences in power and frequency of occurrence across participants (Figures 4B and 4D), alpha (8–12 Hz), low beta (18–22 Hz), and high beta (28–32 Hz) are the most informative frequencies. Therefore, we restricted the functional connectivity analysis to these frequency bands.

Hemispheric lateralization of feedforward and feedback signals

We measured hemispheric lateralization over time to test whether the temporal dynamics of functional connectivity associated with feedforward and feedback signals is segregated into hemispheres. For each frequency band and for each pairwise connection (between premotor, somatosensory, and auditory regions), we measured the lateralization index (LI) by computing the normalized difference between phase shifts in different hemispheres for every time point in the trial. A negative lateralization

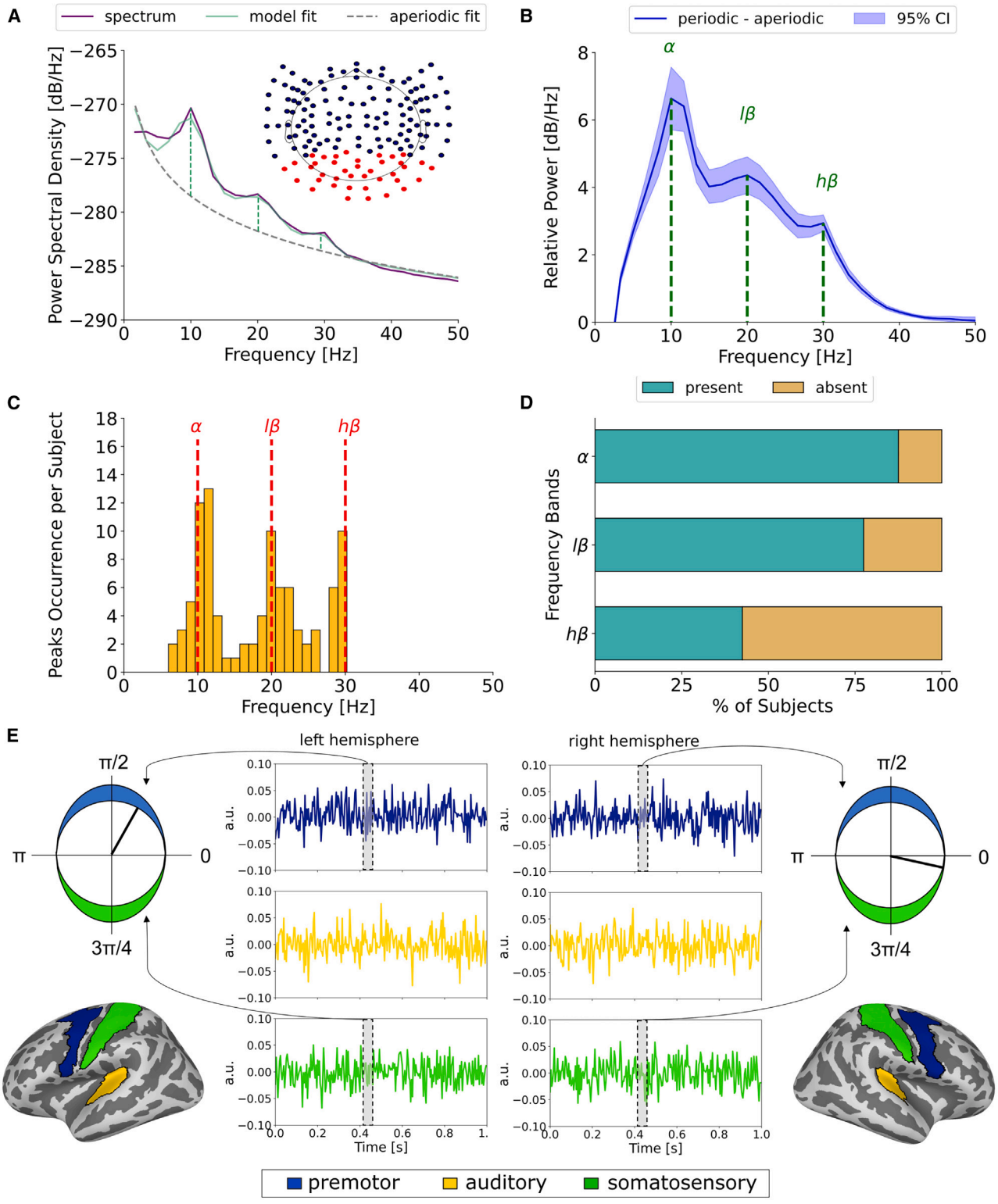


Figure 4. Parameterization of the spectrum and functional connectivity analysis

(A) Illustration of the parameterization of the spectrum procedure. The power spectrum (purple line) is parameterized into periodic (green line) and aperiodic (gray dashed line) components. The center peak frequencies are marked with vertical green dashed lines. A topographic map of MEG sensors is also represented. Frontal, parietal, and temporal sensors (blue dots) were used to estimate the power spectrum, while occipital sensors (red dots) were not used.

(legend continued on next page)

index value reflects higher probability of phase shifts between the ROIs in the left hemisphere and a positive lateralization index in the right.

Figures 5A–5C show the functional connectivity results obtained for the alpha band (8–12 Hz) and for the connection between premotor and auditory ROIs in the covert speech condition (for an analysis in different frequencies and brain regions, see Figures S4–S6). Figures 5D–5F show the analogous results for the passive viewing condition. In the covert speech condition, there is a negative-to-positive switch in the lateralization index precisely around the expected covert speech onset (Figure 5A). We observe a clear lateralization pattern between ~200 and ~400 ms. We interpret this as reflecting feedforward signals because it *precedes* the expected covert speech onset. Between ~400 and ~600 ms, there is a rightward lateralization of functional connectivity; given that it *follows* the expected covert speech onset, we interpret it as reflecting feedback signals. Critically, when looking at the two hemispheres separately (Figure 5B), we observe stronger evidence for connectivity in one hemisphere accompanied by weaker evidence for connectivity in the other hemisphere. The relative difference between hemispheres is more pronounced in the time window preceding compared to the time window following the expected covert production. One-tailed related-samples t tests across subjects show that the phase shifts were significantly left lateralized (more negative than zero) in the time window before covert speech onset ($t(39) = -1.803$, $p < 0.05$) and significantly right lateralized (more positive than zero) after covert speech onset ($t(39) = 2.312$, $p < 0.05$). Moreover, a one-tailed related-samples t test shows that the phase shifts in the time window before covert speech onset were significantly more negative than the phase shifts in the time window after covert speech onset ($t(39) = -2.714$, $p < 0.005$) (Figure 5C). Although the results are robust at the group level, we observed interindividual variability (see connecting lines in Figure 5C). Despite the variability, the left-to-right lateralization is predominant in the sample (for an explicit analysis of interindividual variability, see Figure S7). We replicated the same results when realigning the neural data to individual speech latencies (see Figures S8 and S9). Moreover, as expected, the passive viewing condition did not exhibit the same negative-to-positive switch in the lateralization index (Figures 5D–5F). We investigated the time course of hemispheric lateralization differences between covert speech and passive viewing conditions with a time-resolved related-samples t test. The results show that in the covert speech condition, lateralization index is significantly more negative (i.e., left lateralized) in a time window before the expected covert production and significantly more positive (i.e., right lateralized) in two time windows after the expected covert production (Figure S10). Taken together, these results indicate that an active engagement of the participant is necessary to observe the lateralization switch pattern and that passive viewing of the syllables alone is not sufficient to elicit the effect.

Next, we tested the reliability of hemispheric lateralization and its resilience to background noise. In other words, we asked whether the left-to-right lateralization pattern observed in the main analysis is reliable despite interindividual variability. To answer this question, we ran a surrogate test[43] at the individual subject level to investigate whether we can still observe the same lateralization pattern when separating the signal from background noise (which is responsible for interindividual variability). A surrogate signal is a randomized copy of the original signal that is generated by perturbing the phase relationship while preserving the spectral properties of the original signal (Figure S11). It is important that the spectral properties of the original signal are preserved, such that the signal-to-noise ratio is similar and the comparison between the original signal and the surrogate signal is fair. We took the conservative approach of generating 1,000 surrogates. Then, we estimated the wPLI for each time point in the trial in the original signal and the surrogate signals. If the observed value in the original signal falls beyond the 95% percentile of the surrogate distribution, then we can conclude that it significantly exceeds the noise ceiling (Figure S12A). We assessed hemispheric lateralization over time considering only the validated subjects (i.e., the subjects that passed the surrogate test). Negative and positive values reflect leftward and rightward hemispheric lateralization, respectively. A negative-to-positive switch can be observed even if we measure count-based asymmetry over time considering only the validated subjects (Figures S12B and S12C). The proportion of validated subjects showing leftward vs. rightward hemispheric lateralization indicates that there is a leftward asymmetry before the expected covert production and a rightward asymmetry after the expected covert production even if we take into account only the phase shifts that stand out from background noise.

To characterize with higher resolution the spatial arrangement of the lateralization switch pattern, we measured functional connectivity using a finer parcellation scheme. In this step, the original brain regions were further divided into smaller regions based on anatomical and functional criteria.[33] The names and locations of the ROIs defined using the finer parcellation are displayed in Figure 6 (and see Figure S13). All the results presented were restricted to the connections between premotor and auditory areas in the alpha frequency band (for an analysis in different brain regions, see Figures S14 and S15). The aim of this analysis was to investigate whether the lateralization switch pattern is uniformly distributed across parcels or spuriously driven by a small number of parcels. The finer parcellation scheme is associated with higher spatial uncertainty because the brain regions are smaller in size and more susceptible to anatomical variability

(B) The flat spectrum (periodic-aperiodic component) averaged across subjects. The shaded area around the mean (blue line) represents the 95% confidence interval.
(C) A histogram of peak occurrence per subject. The center peak frequencies are concentrated around 10, 20, and 30 Hz (α, lβ, and hβ; red vertical dashed lines).
(D) The percentage of participants in the cohort showing α, lβ, or hβ peaks is represented in the horizontal bar plot.
(E) Source reconstructed MEG time series extracted from three color-coded regions of interest (ROIs; premotor in blue, somatosensory in green, and auditory in yellow) for each hemisphere. The positions of the ROIs on a template brain are represented for each hemisphere. The circular plots represent the phase shift between two ROIs (premotor in blue and somatosensory in green) in a time point across the trial window.

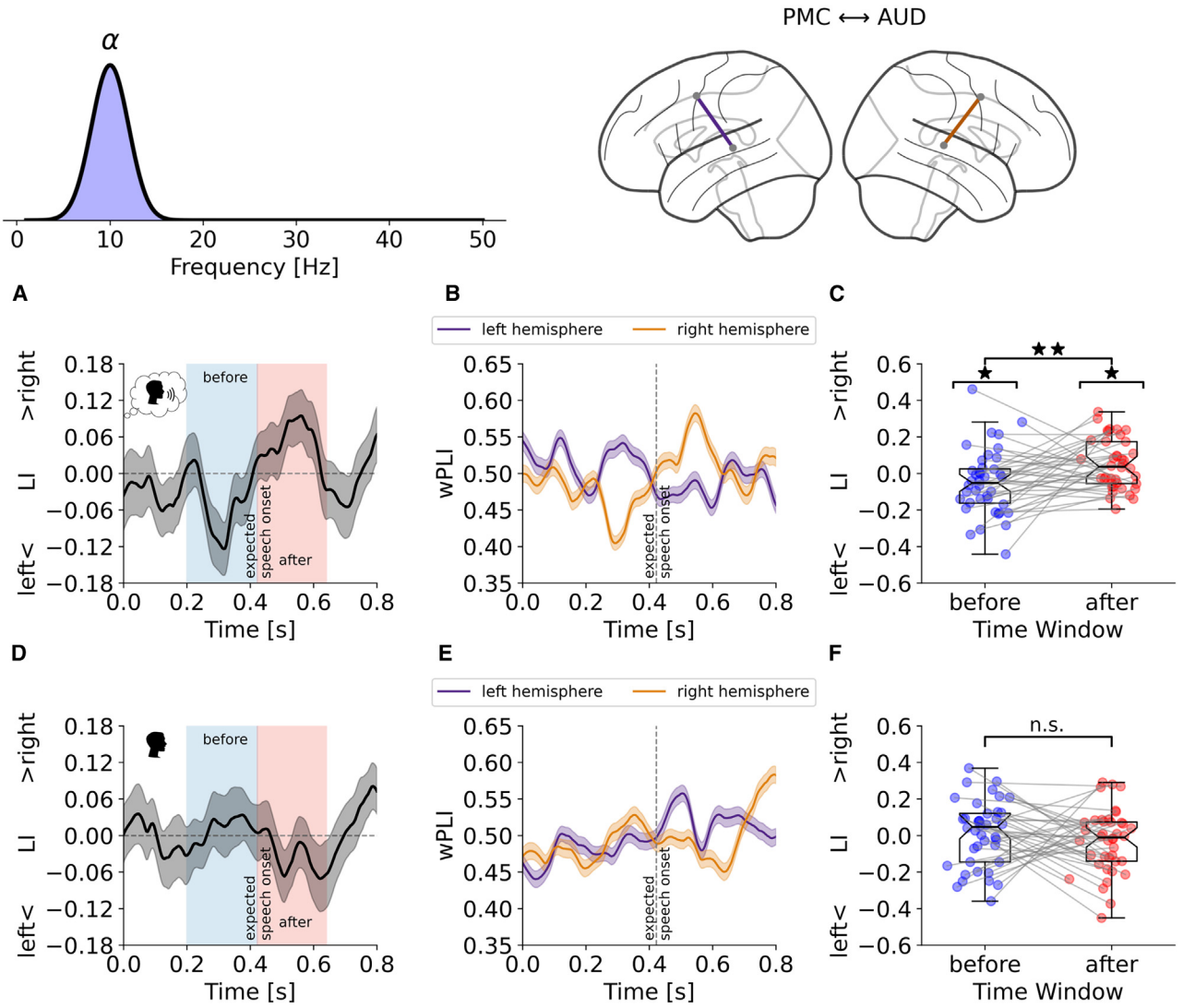


Figure 5. Hemispheric lateralization of functional connectivity over time

The results of the covert speech condition are shown in (A), (B), and (C), and the results of the passive viewing condition are shown in (D), (E), and (F). (A) and (D) show the mean lateralization index over time in the connection between premotor and auditory ROIs in the alpha frequency band (8–12 Hz). A negative lateralization index value reflects leftward lateralization, and a positive lateralization value reflects rightward lateralization. The black line represents the mean across subjects, the shaded gray area around the black line represents the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the overt behavioral pretest. The shaded blue and red areas indicate the time windows before and after the expected covert speech onset, respectively. (B) and (E) show the weighted phase lag index (wPLI) for each hemisphere, separately. Purple and orange represent the left and right hemispheres. The lines and the shaded area around the lines represent the mean and the standard error of the mean, respectively. (C) and (F) represent the mean and standard deviation, as well as the single data points, of the lateralization index in the time windows before (in blue) and after (in red) the expected covert speech onset. In the covert speech condition, the time window before the expected covert speech onset is significantly left-latetalized, the time window after the expected covert speech onset is significantly right-lateralized, and there is a significant difference in hemispheric lateralization between the two time windows (*$p < 0.05$; **$p < 0.01$). In the passive viewing condition, there is no signficant difference between the two time windows (n.s., not significant). The connecting lines show individual increase or decrease in lateralization index from the time window before to the time window after the expected covert production for each subject.

across subjects. For this reason, we used single-subject statistics instead of group-level statistics to make inferences about the population. We ran a surrogate test for each subject and ROI, and we used the false discovery rate (FDR) method to correct for multiple comparisons. Then, we estimated the population prevalence[44] for each hemisphere (i.e., γ_{right} and γ_{left}), which is a measure of consensus across single-subject statistics based on the proportion of the sample showing a statistically significant difference from background noise (as estimated in the surrogate test). For each ROI and each time point, we computed the prevalence difference between rightward and leftward lateralization (i.e., $\gamma_{right} - \gamma_{left}$). Then, we averaged the prevalence difference in the time windows before (Figure 6A; ~200–400 ms) and after (Figure 6D; ~400–600 ms) the expected speech onset. In the

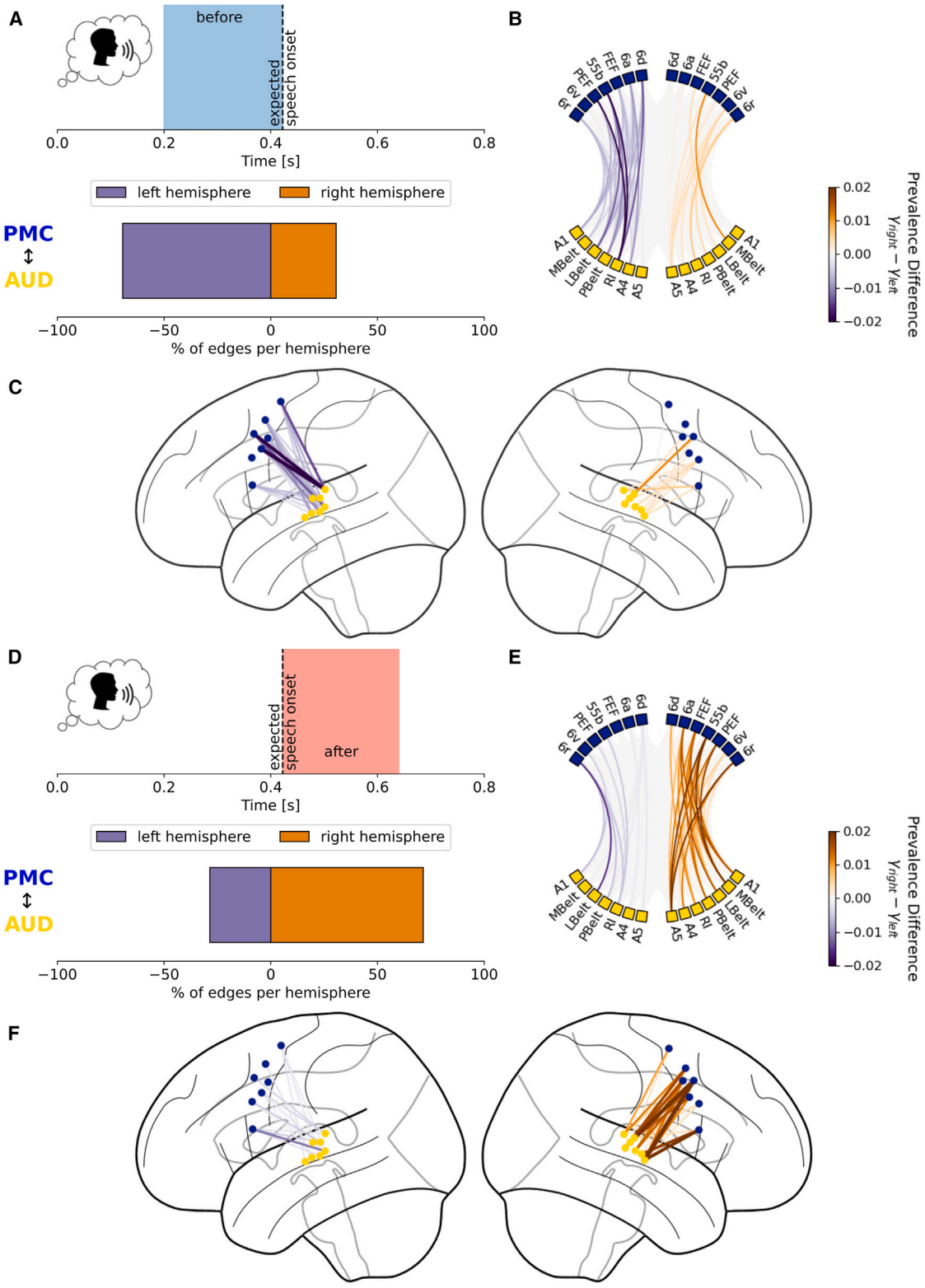


Figure 6. Hemispheric lateralization of functional connectivity using the finer parcellation scheme

The results obtained in the time window before and after the expected speech onset are shown in (A), (B), and (C) and in (D), (E), and (F), respectively. (A and D) A schematic representation of the selected time window is shown at the top, and the percentage of edges showing hemispheric lateralization (left hemisphere in

(legend continued on next page)

time window before the expected speech onset, we observed a negative prevalence difference reflecting leftward lateralization in most of the connections (Figures 6B and 6C). In the time window after the expected speech onset, we observed a positive prevalence difference reflecting rightward lateralization in most of the connections (Figures 6E and 6F). In both cases, prevalence difference was uniformly distributed across the ROIs in the lateralized hemisphere, confirming that the lateralization pattern was not spuriously caused by a small number of connections.

Finally, we investigated the information flow associated with feedforward and feedback signals. Information can either flow persistently in one direction or alternately flow in both directions (i.e., bidirectional). We hypothesized that, similar to overt speech, feedforward and feedback signals would be associated with sustained directed information flow from premotor to sensory and from sensory to premotor areas, respectively. We used the directed phase lag index (dPLI) to capture the directionality of the phase shift.[45] This functional connectivity measure provides an empirical estimate of whether the signals are consistently leading or lagging based on the number of trials showing a phase shift in one or the other direction (Figures S16A–S16D). The time course of the dPLI in the left and right hemisphere is shown in Figure S16E. In the time window preceding the expected covert speech onset, we measured a sustained left-lateralized information flow from premotor to auditory areas (Figure S16F). A time-resolved related-samples t test shows that the phase shifts from premotor to auditory areas were significantly more pronounced in the left than in the right hemisphere in the time window between ~200 and 260 ms ($p < 0.05$). In the time window following the expected covert speech onset, we did not observe a sustained directed information flow in either of the hemispheres (Figure S16G). Taken together, these results suggest that, while the feedforward stage involves mainly top-down information transfer, the feedback stage involves a mix of top-down and bottom-up information transfer. This finding provides additional evidence for the fact that feedback signals observed during covert speech are substantially different from the ones expected during overt speech. One possible interpretation is that during covert speech (i.e., in the absence of external input), the internally generated percept must be maintained "vivid" while its features are processed. Thus, a combination of top-down and bottom-up information flow is observed.

DISCUSSION

We tested the hypothesis that feedforward and feedback signals in covert speech production take place in the left and right hemispheres, respectively. To this end, we investigated the lateralization and dynamics associated with the covert production of single syllables. We used MEG to measure how the hemispheric lateralization of functional connectivity between sensory and motor areas changes over time. In particular, we focused on the critical time windows immediately preceding and following the expected covert speech onset. The data reveal that, within a realistic timescale for speech production, there is a discrete switch from leftward to rightward lateralization in functional connectivity around the expected covert speech onset. This effect occurs specifically in the alpha frequency band (8–12 Hz). Although this effect is robust at the group level, it is variable across subjects. Nevertheless, an investigation of interindividual variability reveals that the left-to-right lateralization is predominant in the sample.

We assumed that feedforward signals are generated during covert speech, even if sensory feedback is unavailable. In line with this, evidence from limb motor control research shows that feedforward signals can be observed even when somatosensory feedback is temporarily or permanently blocked. For instance, deafferented patients who have lost sensory myelinated fibers below the neck can still perform basic limb movements necessary for locomotion, relying solely on feedforward signals.[46] In addition, hemodynamic responses in the somatosensory cortex have been detected in subjects whose proprioceptive feedback was temporarily suppressed using an ischemic nerve block.[47] In speech motor control studies, feedforward signals could similarly be present even when auditory feedback is temporarily or permanently blocked. For example, the role of feedforward processing without feedback can be investigated by deactivating cochlear implants in patients with permanent hearing loss. Behavioral studies show that cochlear implant users rely exclusively on feedforward processing when auditory feedback is unavailable.[48] Furthermore, electrocorticography data from cochlear implant users indicate that speech-related auditory responses can still be recorded from auditory regions even when the implant is switched off.[49]

We observed a leftward lateralization of audiomotor functional connectivity before the expected onset of covert speech. One possible interpretation of this pattern is that it reflects feedforward signaling associated with speech planning. Studies in both humans and non-human animals suggest that left-hemispheric lesions in motor areas selectively impair motor planning. This selectivity is evident in cases where the structure of the output is compromised (e.g., substitutions or omissions), while the ability to vocalize remains intact. For example, in a study[26] examining the effects of unilateral transections of the nerve linking motor brain-stem areas to the syrinx, the avian equivalent of the mammalian larynx, transecting this nerve in the right hemisphere caused only mild singing impairment. In contrast, transecting it in the left hemisphere left canaries unable to produce structured vocal output, despite retaining the ability to perform basic vocal tract movements. The behavioral effects of left-hemispheric lesions suggest an impairment of the brain structures involved in motor planning. Similarly, neuropsychological data from humans indicate that left-hemispheric lesions disrupt the ability to formulate speech motor plans without impairing

purple, right hemisphere in orange) is shown in a horizontal bar plot at the bottom. (B and E) The prevalence difference in the connections between premotor and auditory areas is represented in a functional connectivity circle plot. Purple lines represent leftward lateralization, and orange lines represent rightward lateralization. The premotor ROIs are sorted from the more ventral to the more dorsal. The auditory ROIs are sorted from primary to secondary auditory cortex. (C and F) The prevalence difference in the connections between premotor and auditory areas is represented in a glass brain plot.

the ability to articulate speech. For instance, ideomotor apraxia of speech, associated with lesions in the ventral premotor cortex and/or anterior insula of the left hemisphere,[50,51] primarily affects speech motor planning[21,22] by limiting the capacity to combine basic speech units into complex sequences. Overall, these studies support the idea that the left hemisphere is primarily engaged in generating feedforward signals associated with speech planning during overt speech. We hypothesize that the same neural circuitry activated during overt speech may also be engaged during covert speech.

We assumed that covert speech outcomes are evaluated and revised when necessary, possibly resulting in an internal model update. In line with this, there is evidence suggesting that errors are detected and corrected during covert speech. Psycholinguistic models, in particular, propose that covert speech undergoes monitoring to anticipate or prevent errors that might occur in upcoming utterances.[52] Anticipating mismatches enables quicker error detection and correction, providing the system with significant computational advantages.[53] This monitoring process is evident during covert speech of error-prone material, which leads to longer covert speech latencies.[54] Linguistic information is internally generated and reviewed at various levels of abstraction during covert speech. Previous studies indicate that experimental manipulations that increase errors at both high (e.g., phonemic[55]) and low (e.g., featural[56,57]) levels of abstraction affect the monitoring process, leading to longer covert speech latencies. This process also triggers neural error signals; for instance, distinct electrophysiological responses precede speech production, differing based on whether an error is anticipated or unanticipated in forthcoming utterances.[58] In addition, hemodynamic responses from auditory brain areas have been recorded during the silent recitation of tongue-twister word lists, suggesting that error-prone material is linked to stronger auditory activity, even without external sensory input.[59]

We observed a rightward lateralization of audiomotor functional connectivity following the expected onset of covert speech. One possible interpretation of this pattern is that it reflects feedback signaling associated with speech monitoring. The neural mechanisms underlying feedback processing have been investigated in feedback perturbation studies, where the output is altered, triggering an error signal when the input is perceived. Previous research has demonstrated that functional connectivity associated with auditory feedback perturbation is predominantly right lateralized. For instance, one study measured brain responses to first-formant-shifted versions of spoken words, finding that pitch perturbation led to an increase in functional connectivity between the right primary auditory cortex and the right ventral premotor cortex.[23] However, these findings may pertain specifically to a particular type of feedback perturbation (in spectral features), which would be more consistent with the time-based hemispheric lateralization of auditory perception. Converging evidence from multiple studies[60,61] suggests a hemispheric specialization for auditory processing at various timescales: the left hemisphere is typically more sensitive to shorter timescale modulations and the right to longer ones. In support of this, evidence suggests that the left hemisphere may be more responsive to temporal feedback, whereas the right hemisphere may be more attuned to spectral feedback.[62] This perspective is consistent with our findings. We observed that the highest peak of right lateralization approximately coincided with the timing of vowel onset. Thus, spectral properties (such as the sustained vocal resonance of vowels) may be more relevant than temporal properties (like the transient closure and release of consonants) in the context of this study, where we focused on audiomotor, rather than somatomotor, functional connectivity. Overall, these studies support the interpretation that the right hemisphere is primarily engaged in generating feedback signals associated with speech monitoring during overt speech, particularly when spectral features are most relevant. We suggest that the same neural circuitry activated during overt speech may also be recruited during covert speech.

The results add a provocative perspective on speech motor control models. Our work was inspired by the parallel neural architecture proposed by the DIVA model.[20] The notion that feedforward signals are left lateralized, while feedback signals are right lateralized, is supported by neuropsychological and neuroimaging studies that have carefully isolated these operations to explore their neural underpinnings. In this study, we observed a *leftward lateralization before the expected onset of covert speech*, followed by a *rightward lateralization after the onset of covert speech*. One possible interpretation of these findings is that feedforward and feedback signals during covert speech may engage the same neural pathways utilized during overt speech. Although the DIVA model does not make predictions regarding covert speech, other models, such as the state feedback control (SFC) model,[63,64] account for feedforward and feedback signals even in the absence of motor execution. The most significant difference between the DIVA and the SFC models is that, in the SFC model, feedforward and feedback processes rely not only on external sensory consequences but also on internal sensory targets, which are the perceptual goals associated with upcoming utterances. While further research is needed to examine the similarities and differences of feedforward and feedback signals in covert and overt speech, our findings show that integrating various aspects of overt speech motor control models is desirable and may provide valuable insights for their potential application in the context of covert speech.

Limitations of the study

The assumption that covert and overt speech involves similar signals responsible for interareal communication between motor and sensory regions is debatable. While theoretical and indirect empirical evidence suggests the presence of feedback signals during covert speech,[10–16] their existence is far from being conclusive and requires further investigation. Even if we assume that such feedback signals exist during covert speech, they are likely to differ from those observed during overt speech, given the absence of external input.[63,64] The directed connectivity analysis confirms this distinction by showing right-lateralized bidirectional information flow following the expected covert production. This pattern reflects a combination of top-down and bottom-up signals, in contrast to the strictly bottom-up signals expected during overt speech feedback. Importantly, these differences do not invalidate the conclusions about the overlapping neural architecture. In fact, even though the predictions of the DIVA model[20] are based entirely on overt speech, the data

indicate that covert speech likely engages, at least partially, the proposed parallel neural circuitry underlying overt speech. However, a direct comparison between covert and overt speech is needed to clarify the similarities and differences in the signals passing through these neural circuits.

We observed interindividual variability in the left-to-right lateralization switch. This variability is not uncommon in MEG connectivity studies, which often rely on large samples, averaging, and clustering methods to address such differences (cf. Schoffelen et al.[65]). However, we considered several factors that may contribute to this variability. First, there are potential confounds that are difficult to quantify. Although participants reported right-handedness, leading us to assume left-lateralized language functions, there are cases in the literature where individuals exhibit reversed laterality.[66] In addition, despite our efforts to improve participants' temporal precision, some participants may have struggled to imagine speaking within the instructed temporal range, resulting in unpredictable temporal shifts. Moreover, hidden variables may have contributed to the observed variability. For instance, we expect that the vividness of internally generated speech would be associated with a better signal-to-noise ratio in the neural response, similar to what is observed in visual imagery studies.[67] Finally, confounding variables such as fatigue and alertness could also unpredictably influence the variability of the neural response.

RESOURCE AVAILABILITY

Lead contact

Requests for further information should be directed to and will be fulfilled by the lead contact, Francesco Mantegna (fmantegna93@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data reported in this paper will be shared by the lead contact upon request.
- All original code has been deposited at https://doi.org/10.5281/zenodo.13292069 and is publicly available as of the date of publication.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (https://www.nsf.gov/) grant 2043717 (D.P.). The authors would like to thank Yuri Antonacci, Andrew Chang, Adeen Flinker, Amirhossein Khalilian-Gourtani, Natalie Schaworonkow, and Arianna Zuanazzi for their suggestions and comments on a preliminary version of the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, F.M., J.O., and D.P.; methodology, F.M. and J.O.; software, F.M.; validation, F.M.; formal analysis, F.M.; investigation, F.M. and J.O.; resources, D.P.; data curation, F.M.; writing – original draft, F.M.; writing – review & editing, F.M., J.O., and D.P.; visualization, F.M.; supervision, J.O. and D.P.; project administration, D.P. and J.O.; funding acquisition, D.P.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Participants
  - Handedness assessment
- METHOD DETAILS
  - Experimental design and procedure
  - Overt speech pre-test: Speech sound onset annotation and analysis
  - MEG data acquisition and pre-processing
  - EMG data acquisition and pre-processing
  - MRI-MEG co-registration and source reconstruction
  - ROI definition using brain atlas
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Estimation of MEG/EMG response magnitude
  - Correlation between behavioral and neural variability
  - Outliers definition and exclusion criteria
  - Cluster-based permutation test
  - Estimation and parametrization of the power spectrum
  - Time-resolved frequency representation
  - Functional connectivity analysis
  - Sigmoidal normalization
  - Lateralization index
  - Surrogate analysis
  - Population prevalence

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2024.115137.

Received: March 15, 2024
Revised: October 28, 2024
Accepted: December 11, 2024

REFERENCES

1. Nalborczyk, L., Longcamp, M., Bonnard, M., Serveau, V., Spieser, L., and Alario, F.X. (2023). Distinct neural mechanisms support inner speaking and inner hearing. Cortex *169*, 161–173.
2. Hinwar, R.P., and Lambert, A.J. (2021). Anauralia: The silent mind and its association with aphantasia. Front. Psychol. *12*, 744213.
3. Von Holst, E., and Mittelstaedt, H. (1950). Das reafferenzprinzip: wechselwirkungen zwischen zentralnervensystem und peripherie. Naturwissenschaften *37*, 464–476.
4. Bridgeman, B. (1995). A review of the role of efference copy in sensory and oculomotor control systems. Ann. Biomed. Eng. *23*, 409–422.
5. Latash, M.L. (2021). Efference copy in kinesthetic perception: a copy of what is it? J. Neurophysiol. *125*, 1079–1094.
6. Tian, X., and Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. Front. Psychol. *1*, 166.
7. Tian, X., and Poeppel, D. (2012). Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation. Front. Hum. Neurosci. *6*, 314.
8. Niziolek, C.A., Nagarajan, S.S., and Houde, J.F. (2013). What does motor efference copy represent? Evidence from speech production. J. Neurosci. *33*, 16110–16116.
9. Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. Neuroimage *14*, S103–S109.

10. Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. Behav. Brain Sci. *27*, 377–442.
11. Gentili, R., Han, C.E., Schweighofer, N., and Papaxanthis, C. (2010). Motor learning without doing: trial-by-trial improvement in motor performance during mental training. J. Neurophysiol. *104*, 774–783.
12. Kim, O.A., Forrence, A.D., and McDougle, S.D. (2022). Motor learning without movement. Proc. Natl. Acad. Sci. USA *119*, e2204379119.
13. Loimusalo, N., Huovinen, E., and Puurtinen, M. (2019). Successful approaches to mental practice: A case study of four pianists. Music Perform. Res. *9*.
14. Proteau, L. (1992). On the specificity of learning and the role of visual information for movement control. Adv. Psychol. *85*, 67–103.
15. Proteau, L., Marteniuk, R.G., and Lévesque, L. (1992). A sensorimotor basis for motor learning: Evidence indicating specificity of practice. Q. J. Exp. Psychol. *44*, 557–575.
16. Krigolson, O., Van Gyn, G., Tremblay, L., and Heath, M. (2006). Is there" feedback" during visual imagery? Evidence from a specificity of practice paradigm. Can. J. Exp. Psychol. *60*, 24–32.
17. Nalborczyk, L., Debarnot, U., Longcamp, M., Guillot, A., and Alario, F.X. (2022). The role of motor inhibition during covert speech production. Front. Hum. Neurosci. *16*, 804832.
18. Perrone-Bertolotti, M., Rapin, L., Lachaux, J.P., Baciu, M., and Lœvenbruck, H. (2014). What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. Behav. Brain Res. *261*, 220–239.
19. Soroush, P.Z., Herff, C., Ries, S.K., Shih, J.J., Schultz, T., and Krusienski, D.J. (2023). The nested hierarchy of overt, mouthed, and imagined speech activity evident in intracranial recordings. Neuroimage *269*, 119913.
20. Tourville, J.A., and Guenther, F.H. (2011). The DIVA model: A neural theory of speech acquisition and production. Lang. Cognit. Process. *26*, 952–981.
21. Ogar, J., Willock, S., Baldo, J., Wilkins, D., Ludy, C., and Dronkers, N. (2006). Clinical and anatomical correlates of apraxia of speech. Brain Lang. *97*, 343–350.
22. Laganaro, M., Croisier, M., Bagou, O., and Assal, F. (2012). Progressive apraxia of speech as a window into the study of speech planning processes. Cortex *48*, 963–971.
23. Tourville, J.A., Reilly, K.J., and Guenther, F.H. (2008). Neural mechanisms underlying auditory feedback control of speech. Neuroimage *39*, 1429–1443.
24. Golfinopoulos, E., Tourville, J.A., Bohland, J.W., Ghosh, S.S., Nieto-Castanon, A., and Guenther, F.H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. Neuroimage *55*, 1324–1338.
25. Segawa, J.A., Tourville, J.A., Beal, D.S., and Guenther, F.H. (2015). The neural correlates of speech motor sequence learning. J. Cognit. Neurosci. *27*, 819–831.
26. Nottebohm, F., and Nottebohm, M.E. (1976). Left hypoglossal dominance in the control of canary and white-crowned sparrow song. J. Comp. Physiol. *108*, 171–192.
27. Khalilian-Gourtani, A., Wang, R., Chen, X., Yu, L., Dugan, P., Friedman, D., Doyle, W., Devinsky, O., Wang, Y., Flinker, A., and Flinker, A. (2024). A corollary discharge circuit in human speech. Proc. Natl. Acad. Sci. USA *121*, e2404121121.
28. García-Rosales, F., López-Jury, L., González-Palomares, E., Wetekam, J., Cabral-Calderín, Y., Kiai, A., Kössl, M., Hechavarría, J.C., and Hechavarría, J.C. (2022). Echolocation-related reversal of information flow in a cortical vocalization network. Nat. Commun. *13*, 3642.
29. Decety, J., Jeannerod, M., and Prablanc, C. (1989). The timing of mentally represented actions. Behav. Brain Res. *34*, 35–42.
30. Landauer, T.K. (1962). Rate of implicit speech. Percept. Mot. Skills *15*, 646.
31. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. J. Neurosci. Methods *164*, 177–190.
32. Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., and Salmelin, R. (2001). Dynamic imaging of coherent sources: studying neural interactions in the human brain. Proc. Natl. Acad. Sci. USA *98*, 694–699.
33. Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al. (2016). A multi-modal parcellation of human cerebral cortex. Nature *536*, 171–178.
34. Schoffelen, J.M., and Gross, J. (2009). Source connectivity analysis with MEG and EEG. Hum. Brain Mapp. *30*, 1857–1865.
35. Gracco, V.L., and Lofqvist, A. (1994). Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. J. Neurosci. *14*, 6585–6597.
36. Nikulin, V.V., Hohlefeld, F.U., Jacobs, A.M., and Curio, G. (2008). Quasi-movements: A novel motor–cognitive phenomenon. Neuropsychologia *46*, 727–742.
37. Nalborczyk, L., Grandchamp, R., Koster, E.H.W., Perrone-Bertolotti, M., and Lœvenbruck, H. (2020). Can we decode phonetic features in inner speech using surface electromyography? PLoS One *15*, e0233282.
38. Vinck, M., Uran, C., Spyropoulos, G., Onorato, I., Broggini, A.C., Schneider, M., and Canales-Johnson, A. (2023). Principles of large-scale neural interactions. Neuron *111*, 987–1002.
39. Vinck, M., Oostenveld, R., Van Wingerden, M., Battaglia, F., and Pennartz, C.M.A. (2011). An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. Neuroimage *55*, 1548–1565.
40. Gupta, D., and James, C.J. (2007). Narrowband vs. broadband phase synchronization analysis applied to independent components of ictal and interictal EEG. In 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE), pp. 3864–3867.
41. Donoghue, T., Haller, M., Peterson, E.J., Varma, P., Sebastian, P., Gao, R., Noto, T., Lara, A.H., Wallis, J.D., Knight, R.T., et al. (2020). Parameterizing neural power spectra into periodic and aperiodic components. Nat. Neurosci. *23*, 1655–1665.
42. Schaworonkow, N. (2023). Overcoming harmonic hurdles: genuine beta-band rhythms vs. contributions of alpha-band waveform shape. Imag. Neurosci. *1*, 1–8.
43. Schreiber, T., and Schmitz, A. (2000). Surrogate time series. Phys. Nonlinear Phenom. *142*, 346–382.
44. Ince, R.A., Paton, A.T., Kay, J.W., and Schyns, P.G. (2021). Bayesian inference of population prevalence. Elife *10*, e62461.
45. Stam, C.J., and van Straaten, E.C.W. (2012). Go with the flow: use of a directed phase lag index (dPLI) to characterize patterns of phase relations in a large-scale model of brain dynamics. Neuroimage *62*, 1415–1428.
46. Lajoie, Y., Teasdale, N., Cole, J.D., Burnett, M., Bard, C., Fleury, M., Forget, R., Paillard, J., Lamarre, Y., and Lamarre, Y. (1996). Gait of a deafferented subject without large myelinated sensory fibers below the neck. Neurology *47*, 109–115.
47. Christensen, M.S., Lundbye-Jensen, J., Geertsen, S.S., Petersen, T.H., Paulson, O.B., and Nielsen, J.B. (2007). Premotor cortex modulates somatosensory cortex during voluntary movements without proprioceptive feedback. Nat. Neurosci. *10*, 417–419.
48. Matthies, M.L., Svirsky, M., Perkell, J., and Lane, H. (1996). Acoustic and articulatory measures of sibilant production with and without auditory feedback from a cochlear implant. J. Speech Hear. Res. *39*, 936–946.
49. Miller, C.A., Behroozmand, R., Etler, C.P., Nourski, K.V., Reale, R.A., Oya, H., Kawasaki, H., Greenlee, J.D.W., and Greenlee, J.D. (2021). Neural correlates of vocal auditory feedback processing: Unique insights from electrocorticography recordings in a human cochlear implant user. Eneuro *8*.
50. Dronkers, N.F. (1996). A new brain region for coordinating speech articulation. Nature *384*, 159–161.

51. Hillis, A.E., Work, M., Barker, P.B., Jacobs, M.A., Breese, E.L., and Maurer, K. (2004). Re-examining the brain regions crucial for orchestrating speech articulation. Brain *127*, 1479–1487.

52. Levelt, W.J. (1983). Monitoring and self-repair in speech. Cognition *14*, 41–104.

53. Hartsuiker, R.J., and Kolk, H.H. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. Cognit. Psychol. *42*, 113–157.

54. Wheeldon, L.R., and Morgan, J.L. (2002). Phoneme monitoring in internal and external speech. Lang. Cognit. Process. *17*, 503–535.

55. Oppenheim, G.M., and Dell, G.S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. Cognition *106*, 528–537.

56. Oppenheim, G.M., and Dell, G.S. (2010). Motor movement matters: The flexible abstractness of inner speech. Mem. Cognit. *38*, 1147–1160.

57. Corley, M., Brocklehurst, P.H., and Moat, H.S. (2011). Error biases in inner and overt speech: evidence from tongue twisters. J. Exp. Psychol. Learn. Mem. Cogn. *37*, 162–175.

58. Möller, J., Jansma, B.M., Rodriguez-Fornells, A., and Münte, T.F. (2007). What the brain does before the tongue slips. Cerebr. Cortex *17*, 1173–1178.

59. Okada, K., Matchin, W., and Hickok, G. (2018). Neural evidence for predictive coding in auditory cortex during speech production. Psychon. Bull. Rev. *25*, 423–430.

60. Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. Cognit. Sci. *25*, 679–693.

61. Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. Speech Commun. *41*, 245–255.

62. Floegel, M., Fuchs, S., and Kell, C.A. (2020). Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. Nat. Commun. *11*, 2839.

63. Houde, J.F., and Nagarajan, S.S. (2011). Speech production as state feedback control. Front. Hum. Neurosci. *5*, 82.

64. Hickok, G., Houde, J., and Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. Neuron *69*, 407–422.

65. Schoffelen, J.M., Hultén, A., Lam, N., Marquand, A.F., Uddén, J., and Hagoort, P. (2017). Frequency-specific directed interactions in the human brain network for language. Proc. Natl. Acad. Sci. USA *114*, 8083–8088.

66. Fischer, R.S., Alexander, M.P., Gabriel, C., Gould, E., and Milione, J. (1991). Reversed lateralization of cognitive functions in right handers: exceptions to classical aphasiology. Brain *114*, 245–261.

67. Dijkstra, N., Bosch, S.E., and van Gerven, M.A.J. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. J. Neurosci. *37*, 1367–1373.

68. Gramfort, A., Luessi, M., Larson, E., Engemann, D.A., Strohmeier, D., Brodbeck, C., Goj, R., Jas, M., Brooks, T., Parkkonen, L., et al. (2013). MEG and EEG data analysis with MNE-Python. Front. Neurosci. *7*, 267.

69. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

70. Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al. (2020). Array programming with NumPy. Nature *585*, 357–362.

71. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods *17*, 261–272.

72. Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis: I. Segmentation and surface reconstruction. Neuroimage *9*, 179–194.

73. Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. *9*, 90–95.

74. Peirce, J., Gray, J.R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., Lindeløv, J.K., and Lindeløv, J.K. (2019). PsychoPy2: Experiments in behavior made easy. Behav. Res. Methods *51*, 195–203.

75. Boersma, P. (2001). Praat, a system for doing phonetics by computer. Glot Int. *5*, 341–345.

76. Seabold, S., and Perktold, J. (2010). Statsmodels: econometric and statistical modeling with python. SciPy *7*.

77. Oldfield, R.C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia *9*, 97–113.

78. Kent, R.D., and Read, C. (2002). The Acoustic Analysis of Speech, 2nd ed. (Singular Publishing Group).

79. Byrd, R.H., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. *16*, 1190–1208.

80. Adachi, Y., Shimogawara, M., Higuchi, M., Haruta, Y., and Ochiai, M. (2001). Reduction of non-periodic environmental magnetic noise in MEG measurement by continuously adjusted least squares method. IEEE Trans. Appl. Supercond. *11*, 669–672.

81. De Cheveigné, A., and Simon, J.Z. (2007). Denoising based on time-shift PCA. J. Neurosci. Methods *165*, 297–305.

82. Makeig, S., Bell, A., Jung, T.P., and Sejnowski, T.J. (1995). Independent component analysis of electroencephalographic data. Adv. Neural Inf. Process. Syst. *8*.

83. Douw, L., Nieboer, D., Stam, C.J., Tewarie, P., and Hillebrand, A. (2018). Consistency of magnetoencephalographic functional connectivity and network reconstruction using a template versus native MRI for co-registration. Hum. Brain Mapp. *39*, 104–119.

84. Fischl, B., Sereno, M.I., and Dale, A.M. (1999). Cortical surface-based analysis: II: inflation, flattening, and a surface-based coordinate system. Neuroimage *9*, 195–207.

85. Fischler, M.A., and Bolles, R.C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM *24*, 381–395.

86. Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. *15*, 70–73.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
Python 3.8	Python Software Foundation	RRID:SCR_008394
MNE Python	Gramfort et al. [68]	RRID:SCR_005972
Scikit-learn	Pedregosa et al. [69]	RRID:SCR_002577
Numpy	Harris et al. [70]	RRID:SCR_008633
Scipy	Virtanen et al. [71]	RRID:SCR_008058
Freesurfer	Dale et al. [72]	RRID:SCR_001847
Matplotlib	Hunter et al. [73]	RRID:SCR_008624
Nilearn	https://github.com/nilearn/nilearn	RRID:SCR_001362
PsychoPy	Peirce et al. [74]	RRID:SCR_006571
Praat	Boersma [75]	RRID:SCR_016564
Statsmodels	Seabold et al. [76]	RRID:SCR_016074
Custom code	This paper	https://doi.org/10.5281/zenodo.13292069

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Participants

Forty participants (mean age 27.18, range 20-47, 12 male) with no history of psychiatric or neurological disorders took part in this MEG experiment. All of them reported normal hearing and normal or corrected-to-normal vision. Participants signed an informed consent form. Ethical approval to conduct the study was provided by New York University Institutional Review Board (IRB) for human subjects research.

Handedness assessment

Participants' hand preference was assessed using the Edinburgh Handedness Inventory[77] (EHI). This questionnaire was designed to ascertain hand preference in daily life activities. Participants were asked to rate hand preference for the following list of items and activities: writing, drawing, throwing, scissors, comb, toothbrush, spoon, hammer, screwdriver, tennis racket, knife without fork, knife with fork, cricket bat, golf club, broom, rake, striking a match, opening a box (lid), dealing cards, threading a needle. If the preference is for either the left or the right hand, then a "+" is marked on that side. If the preference for a particular hand is so strong that one wouldn't use the other hand unless forced, then a "++" is marked on that side. If there is no preference for any hand, then a "+" is marked on both sides. R and L refer to the total number of "+" marked for the right and left hand, respectively. The laterality quotient was estimated using the following formula: LQ = ((R-L)/(R+L))*100. Scores below -40 reflect left-handedness. Scores between -40 and +40 reflect ambidexterity. Scores above +40 reflect right-handedness. Only right-handed participants with a laterality quotient greater than +40 were admitted to the experiment.

METHOD DETAILS

Experimental design and procedure

The stimuli consisted of a set of three consonant-vowel (CV) syllables: /pa/, /ta/ and /ka/. We varied the consonant and we kept the vowel constant. The syllables were selected to be similar in duration and overall acoustics but well distinct in motor space. According to the international phonetic alphabet (IPA) these unvoiced stop consonants have different places of articulation. /p/, /t/, and /k/ are bilabial, alveolar, and velar consonants that recruit articulators in the front, mid, and back parts of the upper vocal tract, respectively. The number of the syllables was chosen to be neither too small nor too large to guarantee the presence and the consistent timing of the forward motor plan. On the one hand, we designed the experiment such that planning always preceded covert speech production. Thus, we used different syllables to ensure that participants retrieve the desired motor plan and its sensory consequences for each trial instead of reusing the same target over and over. On the other hand, we designed the experiment such that planning was consistent across trials in both onset and duration. Thus, we used only three syllables instead of many to minimize the variability in the forward motor plan that is intrinsic to specific syllable attributes, e.g., frequency of usage or syllable complexity.

Participants went through two experimental conditions: a covert speech condition and a passive viewing condition. The experimental procedure was structured as follows. There was a one second baseline period. Then, one syllable was visually presented on the screen for one second. In the covert speech condition, participants were instructed to imagine speaking as soon as they saw the syllable. In the passive viewing condition, participants were instructed to passively staring at the syllable. One single participant also performed an overt speech condition in which the syllables were produced aloud. This dataset was solely used to evaluate the impact of articulatory movements on electromyographic and magnetoencephalographic measurements. Syllable presentation order was randomized. Participants completed 4 blocks for each condition. Each block consisted of 120 trials; therefore, participants performed 480 trials per condition. We used PsychoPy toolbox[74] version 2, Python release 3.8.3, for stimulus delivery.

Overt speech pre-test: Speech sound onset annotation and analysis

Participants performed a behavioral pre-test that was designed to collect temporal landmarks and familiarize participants with the task. Participants were seated in front of a computer screen in a soundproof booth. We used a 16-channel, 8-preamp, 24-bit/96kHz, MOTU system for audio input/output. Participants' utterances were recorded using a microphone. The experimental procedure was structured as follows. There was a one second baseline period. Then, one syllable was visually presented on the screen for one second. Participants were instructed to pronounce the syllable aloud as soon as they saw the syllable. The behavioral pre-test was used to collect an estimate of the production latency for each participant. As a first screening, we used a custom MATLAB script to compute the envelope of the speech sound waveform and we automatically extracted the rising of the peak for each trial. These speech sound onsets measures were used to estimate participants' temporal precision. If participants were temporally imprecise we asked them to practice more. Participants who were unable to meet certain criteria were excluded from the experiment. The exclusion criteria were: a median speech onset between 200-600 ms and an interquartile range smaller than 400 ms.

The data collected from the behavioral pre-test was subsequently used to annotate speech sound onsets for each participant. In order to do that, we used the Praat software[75]. Speech sound onsets were determined by visual inspection based on the characteristic spectrogram signatures of unvoiced stop consonants used (i.e., /p/, /t/, and /k/). Unvoiced stop consonants (e.g., /p/, /t/, /k/) involve occlusions in the vocal tract which have specific acoustic signatures corresponding to a silent period (closure) and a noise burst (release). In particular, unvoiced stop consonants show clear differences in the shape of the noise burst: /p/ shows a short-lived wide range burst across all the spectrogram that has lower intensity; /t/ shows a more prolonged burst in the upper part of the spectrogram that has higher intensity; /k/ shows an even longer burst in the lower part of the spectrogram that also has higher intensity[78]. Additionally, as an alternative annotation criterion we detected the vowel onset which is characterized by a transition from non-vocal to vocal energy corresponding to the onset of vocal folds vibration. For a subset of participants (N=20), vowel onsets were determined by visual inspection of both the waveform and the spectrogram of sound recordings. In the waveform, vowel onset was identified by detecting the point at which there is a noticeable increase in amplitude. In the spectrogram, vowel onset was identified by observing the emergence of distinct formant patterns and increased energy in the lower frequency bands, which are characteristic of vowel sounds (e.g., /a/).

We parameterized participants' speech latencies distribution using a Gamma probability density function. The Gamma function is based on two parameters: a shape (α) and a scale (σ) parameter. The shape parameter represents the skewness of the distribution and the scale parameter represents the width of the distribution. The optimal shape and scale parameters were estimated using Maximum Likelihood Estimation (MLE). We used the quasi-Newton non-smooth optimization Broyden, Fletcher, Goldfarb, and Shanno (L-BFGS-B) method for bound constrained minimization[79]. The initial values were set to $\alpha = 4$ and $\sigma = 0.03$, the bounds were set to $\alpha = [2, 6]$ and $\sigma = [0.005, 0.05]$. For the algorithm implementation, we used the Scipy software[71].

MEG data acquisition and pre-processing

Individual head shapes and fiducial landmarks (nasion, right and left pre-auricular points), were digitized using a 3D laser scanning hardware (Polhemus, FastSCAN COBRA 3D) and a 3D digitizer software (Source Signal Imaging, Inc.). Five Head Position Indicator (HPI) coils were placed on participant's mastoid bones and forehead to keep track of participant's head position inside the dewar through electromagnetic induction. We measured head position before and after each recording block. When the maximum difference between head positions before and after each block was above 1 cm, data was excluded from further analysis. Prior to data acquisition, all metal and other potential sources of electromagnetic interference were removed. Prior to running the experiment we recorded 2 minutes of empty room.

MEG recordings were obtained in a magnetically shielded room (Vacuum Schmelze, Hanau, Germany) using a 157-channel whole-head axial gradiometer system (KIT, Kanazawa Institute of Technology, Japan) and 3 orthogonally oriented reference magnetometers. Participants performed the task in a supine position. When positioning participants in the MEG scanner, we ensured tight contact with the dewar. The MEG signal was sampled at 1 kHz. Two filters were applied during data acquisition: a zero-phase two-pass Butterworth band-pass filter 1-200 Hz and a notch filter at 60 Hz. Visual stimuli were presented using a CP-X8150 LCD projector (Hitachi America LTD). Images were projected on a first-surface mirror (Edmund Scientific, Barrington, NJ) suspended from the ceiling and fixed at a 45° angle; the incoming image hit the mirror and was reflected 90° straight down. A Cedrus StimTracker was used to keep track of trigger delivery with high temporal precision.

MEG preprocessing was applied using MNE-Python[68] (v0.20.7), Python release 3.8.3, combined with custom routines. First, we removed external and internal sources of noise from the recorded MEG signal. External noise (e.g., stationary noise, environmental

noise) was removed offline from the MEG recordings using two denoising algorithms run in sequence. First, we used a continuously adjusted least-squares method[80] using the noise recorded during the empty room. This method consists in estimating regression coefficients from reference magnetometers recorded during the empty room to regress out environmental noise from axial gradiometers recorded during the experiment. Then, the remaining environmental noise was removed using a time-shifted PCA[81] using the noise recorded during the experiment. Noise magnetic fields measured by reference magnetometers are filtered and subtracted from axial gradiometers. The filters (one per reference magnetometer/axial gradiometer pair) are obtained by delaying the reference signals, orthogonalizing them to obtain a basis, projecting the brain sensors onto the noise-derived basis, and removing the projections to obtain clean data. Internal noise (e.g., heartbeat, muscular activity, eye blinks) was reduced using Independent Component Analysis (ICA)[82]. We used a fixed-point algorithm to estimate 30 independent components on epoched data. Up to 10 components were excluded based on visual inspection of spatial topographies and latent sources.

MEG signals were zero-meaned and detrended. Time series were downsampled to 250 Hz. Next, we segmented the raw signal into trial epochs starting at -1 s before and ending 1 s after syllable visual presentation (2 seconds in total). We standardized the MEG signals by centering and scaling. Then, we subtracted the event-related response from the single-trial data to attenuate the effects of evoked response on the estimation and subsequent interpretation of functional connectivity. Evoked responses violate the underlying assumption of stationarity, due to systematic variability in the latency of the peak of the transient signals across brain regions.

EMG data acquisition and pre-processing

Participants were instructed to avoid head, body and limb movements, to keep their eyes open and to maintain eye fixation during visual cue presentation. Participants were instructed to avoid jaw and lips movements in the covert speech condition and the passive viewing condition. Articulatory movements were continuously monitored using electromyography (EMG). We recorded the EMG signal from four electrodes: one reference electrode placed on the right mastoid, one ground electrode placed on the right wrist, one electrode placed below the cheekbone in order to record jaw movements and one electrode placed between the lower lip and the chin in order to record lips movements. Previous studies have shown that jaw and lip movements are rich enough to capture the information necessary to distinguish stop consonants [35]. Electrode impedance was kept below 25 kΩ.

EMG electrodes were connected to an MEG-compatible BrainAmp DC amplifier (Brain Products GmbH, Gilching, Germany). The EMG was recorded at a sample rate of 500 Hz. A 60 Hz notch filter was applied to remove power line noise. Data was referenced online to the right mastoid. Data was re-referenced offline. A zero-phase, two-pass Butterworth bandpass filter with a 1 Hz high-pass frequency cut-off and a 50 Hz lowpass frequency cut-off was applied. We segmented the raw signal into epochs between -1 s before 1 s after the visual syllable presentation. Time series were down-sampled to 250 Hz and the EMG signal was zero-meaned and detrended. A baseline correction was applied by computing the mean of the 1s baseline period preceding syllable cue onset and subtracting this mean from the entire trial epoch. We used an auto-reject algorithm for automatic artifact rejection. This algorithm defines a threshold for artifact rejection that is specific for each participant based on a cross-validation procedure. This individualized threshold was motivated by the variability across subjects in the EMG recordings introduced by differences in skin conductance, muscle artifacts, heartbeat artifacts, etc…

MRI-MEG co-registration and source reconstruction

T1-weighted anatomical scans were acquired for 14 participants. When the anatomical scans were not available we used a template average brain to perform source reconstruction[83]. The shape of the template average brain was adjusted to match participants' head shape that was measured using the 3D laser scanner. To perform group-level analyses in a common reference frame, we computed a linear interpolation (i.e., morphing) between the decimated individual source model and a template average brain.

The anatomical scans were 3D reconstructed using the Freesurfer software[72,84]. A Boundary Element Model (BEM) was estimated using the watershed algorithm. MRI and MEG coordinate systems were co-registered by matching digitized anatomical fiducial landmarks to participant's T1 scan. The resulting whole brain 3D mesh (5124 vertices; 6.2 mm average source spacing), the BEM model and the aligned coordinate frames were used to compute the forward model for source reconstruction.

The inverse problem consists in the mapping from a low dimensional sensor space (157 axial gradiometers) to a high dimensional source space (5124 vertices). The solution of the inverse problem consists in finding a set of sensor weights for each source location. Non-adaptive distributed source imaging methods (e.g., MNE, dSPM) use matrix multiplication. The data matrix is multiplied by the weighting matrix. Weights are based on the sensor locations with respect to the brain. Thus, the weights are fixed and do not change over time or over frequency. Matrix multiplication involves a linear mixing of the sources resulting in the alteration of the original phase relationship. This alteration must be avoided because it will no longer be possible to tease apart phase shifts caused by the brain generated magnetic signals from phase shifts caused by source reconstruction. Adaptive distributed source imaging methods, like beamformer filters, can be used to preserve the phase relationship. This approach differs from the previous one mainly in how the weights are computed. Weights are computed not only according to physical sensor positions but also according to time, frequency, condition, and subject. We used a Dynamic Imaging of Coherent Sources (DICS) beamformer filter[32]. This method consists in the localization of oscillatory brain activity and the detection of its phase relationships across brain regions using a frequency domain implementation of a spatial filter. This information is included in the weights such that the spatial distribution of power and phase in a chosen frequency band is preserved after projection. Source reconstruction was implemented using MNE-Python[68] (v0.20.7), Python release 3.8.3.

ROI definition using brain atlas

We *a priori* selected regions of interest (ROIs) in the peri-Sylvian language network. We used a cortical parcellation scheme obtained using a brain atlas[33]. We followed authors' guidelines to select brain regions. The ROIs were defined according to functional and anatomical criteria. We used two different ROIs parcellation schemes. On the one hand, we used a coarse-grained parcellation scheme including larger ROIs in the peri-Sylvian language network. In this case, vertices were grouped into 3 ROIs: premotor regions (i.e., premotor cortex), auditory regions (i.e., primary and secondary auditory cortex), somatosensory regions (i.e., primary somatosensory cortex). On the other hand, we used a fine-grained parcellation scheme including smaller ROIs in the peri-Sylvian language network. In this case, vertices were grouped into 18 ROIs that can be grouped into three super categories: premotor regions (i.e., 6r, 6v, FEF, 55b, PEF, 6a, 6d), auditory regions (i.e., A1, LBelt, MBelt, PBelt, RI, A4, A5), somatosensory regions (green clusters; i.e., 1, 2, 3a, 3b).

For each ROI, we extracted time courses using the "pca flip" method. Singular value decomposition was applied to each vertex time course in the ROI, followed by the extraction of the first singular vector. Next, each ROI's time course was scaled and sign flipped. The sign flip was applied to account for the fact that MEG dipoles are positive and negative.

QUANTIFICATION AND STATISTICAL ANALYSIS

Estimation of MEG/EMG response magnitude

The root mean square (RMS) of the MEG and EMG data was computed as a measure of signals' overall strength. For each timepoint, the RMS was calculated by first squaring the amplitude values of the time series data, then averaging these squared values sensors or channels, and finally taking the square root of this average. The formula of RMS is the following:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \quad \text{(Equation 1)}$$

This calculation was repeated for each timepoint across the trial, yielding a single RMS value that represents the magnitude of the signal in that timepoint. RMS is a measure of response magnitude regardless of its sign (i.e., including positive or negative deflections). The RMS values were used to quantify the strength of neural and electro-muscular responses to compare signal amplitudes across different conditions. This method provides a robust measure of signal amplitude while minimizing the impact of noise and outliers.

Correlation between behavioral and neural variability

Behavioral variability was measured as the standard deviation of the speech latencies across trials. Neural variability was measured as the coefficient of variation of the brain response magnitude across trials. The response brain magnitude was measured as the Root Mean Square (RMS) across sensors averaged across all timepoints. The Coefficient of Variation (CoV) is given by the ratio between the standard deviation normalized by the mean multiplied by 100 (yielding a percentage). CoV is a suitable measure of dispersion for brain response magnitude because the means are significantly different across participants (due to various reasons irrelevant to the purpose of this analysis, e.g., age differences, anatomical differences, measurement noise). We measured Pearson correlation to assess the relationship between behavioral and neural variability. The statistical significance of the correlation was estimated using a permutation test. We generated 1000 permutations of the data and we tested whether the observed value fell beyond the 95% percentile of the permuted values distribution.

Outliers definition and exclusion criteria

To robustly estimate the best linear relationship between behavioral and neural variability, we employed the RANSAC (RANdom SAmple Consensus) algorithm[85] to fit a linear regression model. RANSAC is particularly effective in the presence of outliers, as it iteratively selects random subsets of the data, fits a linear model to these subsets, and then determines the consensus set, or inliers, which are points that closely fit the model. By doing so, RANSAC prioritizes the inliers, providing a more reliable estimate of the underlying linear relationship while minimizing the influence of outliers.

Then, we calculated the 95% and 99% prediction intervals. These intervals define the range within which we expect future observations to fall, given the model and the variability in the data. To identify outliers, we examined which observations fell outside the 99% prediction interval. These observations, being highly unlikely under the fitted model, were classified as outliers. The prediction interval for a new observation in a linear regression model is calculated using the following formula:

$$\hat{y}_{new} \pm t_{\alpha/2,n-2} \cdot \sqrt{\hat{\sigma}^2 \left(1 + \frac{1}{N} + \frac{(x_{new} - \underline{x})^2}{\sum_{i=1}^{N} (x_i - \underline{x})^2}\right)} \quad \text{(Equation 2)}$$

This formula accounts for both the uncertainty in the prediction for a new observation and the variability in the data, $\hat{y}_{new}$ is the predicted value of the new observation based on the fitted linear model with the t-value scaling the prediction interval to reflect the desired confidence level.

Cluster-based permutation test

We used a cluster-based permutation test[31] (CBPT; 1000 permutations, one-tailed) to identify time windows showing significant differences between conditions. CBPT is a nonparametric statistical test consisting of two different stages: a cluster formation stage and an inferential stage. In the cluster formation stage, the unit-level statistic is computed for each time point in the trial. We used a t-test as a unit-level statistic. We computed the unit-level statistic 1000 times. For each iteration, the data points in the time series were shuffled. Then, the original t values were compared to permuted t values yielding uncorrected p-values. Significant time points were selected according to an *a priori* defined alpha criterion (i.e., $p < 0.05$), and adjacent time points not exceeding this value were grouped together into clusters. Finally, all the t values within each cluster were summed (i.e., 'maxsum'). Minimum cluster size was set to 25 time points (~100 ms). In the inferential stage, the summed unit-level permutation values within each cluster were used to compute the cluster-level statistical distribution under the null hypothesis of exchangeability. We calculated the percentage of clusters for which the un-permuted cluster-level statistic was larger than the permuted cluster-level statistic. If the cluster p-value was smaller than $p < 0.05$ then we assumed that the data in the two conditions were significantly different.

Estimation and parametrization of the power spectrum

To investigate in which frequency bands there is a power modulation we measured the Power Spectral Density (PSD) from frontal, parietal and temporal sensors for each subject. We used Welch's method[86] for PSD estimation. Welch's method is preferred when computing the PSD in situations where the signal is non-stationary, like the MEG signal. Welch's method involves dividing the signal into overlapping segments, applying a window function to each segment, and then averaging the resulting periodograms to obtain the power spectral density estimate. This approach is particularly useful for analyzing signals with time-varying characteristics.

Taking a data-driven approach, we identified the features of the PSD that allow us to determine in which frequency bands the power is concentrated across subjects[41]. The PSD can be divided into aperiodic (broad region across the frequency spectrum) and periodic (narrow region in the frequency spectrum) components. The aperiodic component is associated with task irrelevant brain activity. It is related to individual differences including age, anatomical structure, etc.. It can be mathematically described as an inverse exponential function. The periodic component is associated with task relevant brain activity. It is related to behavioral and physiological changes. It can be quantified by looking at the center peaks and bandwidth in the power spectrum. We fitted a model to parametrize the power spectrum into aperiodic and periodic components. The aperiodic component was fitted using a 1/f function. The periodic component was fitted using multiple Gaussian functions.

Time-resolved frequency representation

To estimate power and phase for each time point in the trial epoch, it is necessary to capture rapid changes in the frequency domain over time. A method that is well-suited for that purpose is Morlet wavelet convolution. A Morlet wavelet is a sine wave windowed with a Gaussian envelope. Its shape resembles a sine wave in the middle but then tapers off to zero at both ends. The sine wave is designed to capture the characteristics of the signal in a specific frequency band. The Morlet wavelet can be used as a sliding kernel to run a convolution. The convolution operation can be used to estimate power and phase for each time point in the trial epoch.

We ran a continuous wavelet transform using Morlet wavelet convolution to estimate time-resolved cross spectral densities (CSDs) across ROIs for each time point in the trial epoch. Cross spectral density is the frequency domain analogue of the time-domain covariance matrix. S_{xy} is the cross spectral density between activities at ROIs X and Y.

Functional connectivity analysis

Functional connectivity between electrophysiological signals can be estimated using methods assessing whether one signal is a linear transformation (e.g., shifting and/or scaling) of another signal. Functional connectivity measures can be undirected or directed (e.g., unidirectional, bidirectional). Undirected measures capture the intensity of the connection between two brain regions without considering the directionality of the information flow. Directional measures capture the directionality of the information flow in the connection between two brain regions with or without considering the intensity of the connection.

An undirected functional connectivity measure based on phase synchrony is the weighted Phase Lag Index[39]. wPLI captures non-instantaneous phase synchronization between time series. The formula for wPLI estimation is the following:

$$wPLI_{xy}(\omega) = \frac{\left|Im\left(E\left[S_{xy}(\omega)\right]\right)\right|}{Im\left(\left|E\left[S_{xy}(\omega)\right]\right|\right)} \quad \text{(Equation 3)}$$

Where the numerator is the absolute value of the imaginary part of the expected value of the cross-spectrum between signal x and y at frequency ω. The denominator is the imaginary part of the absolute value of the expected value of signal x and y at frequency. The role of the denominator is to downweight the smallest phase shifts resulting in a more stable measure that is more resistant to noise. By design, wPLI captures non-instantaneous synchronization with a phase shift corresponding to one quarter or three quarter cycle

(90° or 270°) because the imaginary part of the cross-spectrum is considered. In contrast, wPLI does not capture instantaneous synchronization when the signals are in phase (0°) or in phase opposition (180°) because the real part of the cross-spectrum is not considered. wPLI is a normalized quantity between 0 and 1. Values around 0 indicate a low probability that there is a phase shift. Values around 1 indicate a high probability that there is a phase shift.

A directional functional connectivity measure based on phase synchrony is the directed Phase Lag Index[45]. dPLI provides an estimate of whether one signal is leading or lagging. The formula for dPLI estimation is the following:

$$dPLI_{xy}(\omega) = E\left[H\left(Im\left(S_{xy}\right)\right)\right] \quad \text{(Equation 4)}$$

Where H is a Heaviside function. This step function returns 1 if the imaginary part of the cross-spectrum is positive and 0 if the imaginary part of the cross-spectrum is negative. This value is estimated for each trial. Then, the sum is divided by the number of trials to find out whether the proportion is biased towards positive or negative values. Values around 0.5 reflect the absence of directionality. Values above 0.5 reflect that S_x is leading and S_y is lagging. Values below 0.5 reflect that S_y is leading and S_x is lagging.

Both wPLI and dPLI are designed to be resistant to volume conduction of uncorrelated sources that is present in MEG signals. We used the MNE connectivity toolbox (https://mne.tools/mne-connectivity/dev/) to estimate the wPLI and dPLI.

Sigmoidal normalization

Some subjects and some brain regions may be systematically associated with higher or lower functional connectivity. We wanted to avoid that individual differences across subjects and brain regions could influence the group-level analysis. Therefore, we applied a normalization to minimize differences across subjects and brain regions. Given that it is possible that there is some outlier due to noise, normalization by the mean or the maximum is not ideal in this case. Thus, we used a sigmoidal normalization to minimize the effect of extreme wPLI values (approximately above 80% and below 20%). The sigmoidal normalization formula is the following:

$$nPLI(x_i) = 1 \Big/ \left(1 + exp\left(\frac{x_i - \bar{x}}{s}\right)\right) \quad \text{(Equation 5)}$$

To maintain the same sign after the sigmoidal normalization we apply the following transformation:

$$nPLI(x_i) = |nPLI(x_i) - 1| \quad \text{(Equation 6)}$$

Lateralization index

We measured hemispheric lateralization of functional connectivity over time. For each time point in the trial epoch, we estimated wPLI between ROIs in the left and right hemispheres. Then, we measured the Lateralization Index (LI) using the following formula:

$$LI = \frac{wPLI_{right} - wPLI_{left}}{wPLI_{right} + wPLI_{left}} \quad \text{(Equation 7)}$$

A negative value reflects leftward lateralization, a positive value reflects rightward lateralization.

Surrogate analysis

Surrogate data can be used for statistical nonlinearity testing[43]. A surrogate is a time series that is generated by randomizing one property of the original signal while leaving all the other properties as much as possible unaltered. In this case, we wanted to randomize the phase while leaving the spectrum (as well as the temporal autocorrelation function) unaltered. We used an Iterative Amplitude Adjusted Fourier Transformed (IAAFT) algorithm to generate surrogates. This algorithm consists of three steps: randomization, spectral adjustment and amplitude adjustment. The second and the third steps are repeated iteratively until convergence is reached. In the randomization step, the data is shuffled such that the phase relationship is altered and the power spectrum is also altered. Following the initial randomization the power spectrum approximates white noise. In the spectral adjustment step, the power spectrum is adjusted by replacing the magnitudes of the Fourier coefficients with those of the original time series. Another time series having a power spectrum identical to the original time series but with different phase distribution is obtained by inverse Fourier transform. In the amplitude adjustment step, the amplitudes in the surrogate time series are adjusted based on the amplitudes in the original time series. All the values in the surrogate and original time series are sorted by their amplitude. Then, the amplitudes of the surrogates are replaced by the amplitudes of the original signal with the same ranking value. For example, the highest value of the surrogate time series is substituted by the highest value of the original time series. If the amplitude adjustment step alters the power spectrum, both the spectral and the amplitude adjustment steps will be repeated until convergence is reached.

Population prevalence

We used a Bayesian population prevalence method[44] to test how single-subject level results generalize to the population. This method looks at effects within each subject in the study and asks how likely it would be to see the same result if the experiment was repeated with a new subject chosen from the population at random. Using this method, it is possible to quantify how typical or atypical an observed effect is in the population, and the uncertainty around this estimate. Bayesian prevalence consists of a two-level procedure. At the first level, a statistical test is applied within each participant, the result of which can be binarized using

a within-participant null hypothesis statistical test. In this case, we used the surrogate test (i.e., a non-parametric statistical test) for within-participants statistical evaluation controlling for the family-wise error rate at α=0.05. We used the False Discovery Rate (FDR) method to correct for multiple comparisons. At the second level, the binary results from the first level (i.e. the counts of significant participants) are the input to the Bayesian population prevalence computation. Often the research question might involve comparing an effect between different experimental conditions in the same set of participants. In this case, we want to test whether the phase shifts measured using wPLI are above background noise in the left and right hemisphere. Within the population prevalence framework, the question would be whether the prevalence of true positive results differs between the two hemispheres. To estimate the difference in prevalence of true positive results between two different tests applied to the same sample of participants (i.e., surrogate test in the connection in the left and right hemisphere), the input parameters are the number of participants rejecting the null hypothesis in both tests, only in each of the two tests, and the total number of participants. Bayesian population prevalence can be applied to any neuroimaging technique (EEG, MEG, fMRI, ECoG) to infer the proportion of responsive individual neurons (e.g., sensors, vertices, parcels) within a brain region.

Cell Reports, Volume *44*

Supplemental information

Time-resolved hemispheric lateralization of audiomotor functional connectivity during covert speech production

Francesco Mantegna, Joan Orpella, and David Poeppel

Supplementary Information

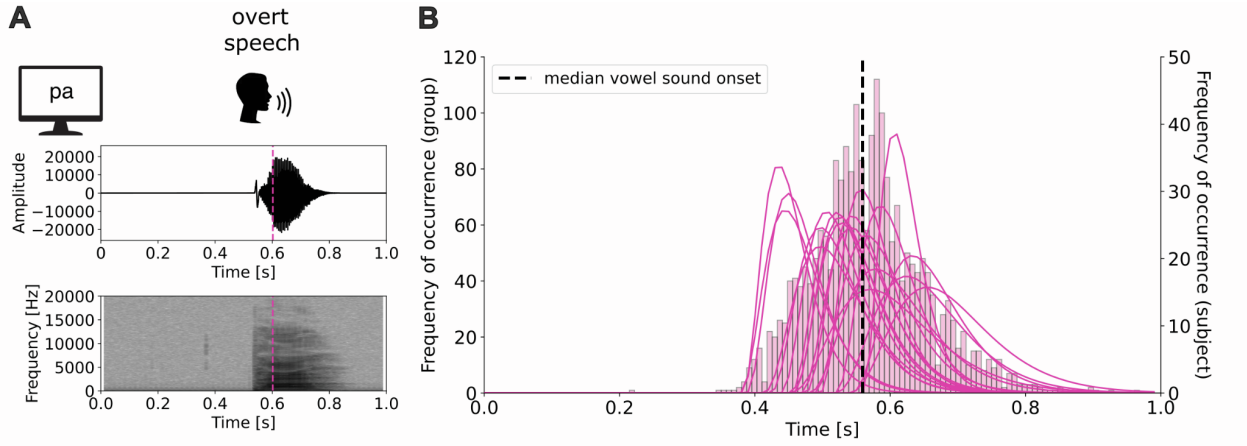


Figure S1. *Vowel onset annotation of speech sounds in the behavioral pre-test*. (**A**) Experimental paradigm (top), a waveform and a spectrogram for a speech sound (bottom). For a subset of participants (N=20), speech sound onsets were annotated by visual inspection of the vowel onset for each trial (pink vertical dashed line). (**B**) Gamma probability density function fit of the vowel onset distribution for each participant (pink lines) and a histogram of vowel onsets across participants. The black dashed vertical line marks the median vowel onset across participants. The median vowel onset across participants was 560 ms (std 49 ms) after visual presentation.

Methods S1: The relationship between behavioral and neural variability, related to Figure 1. To further corroborate the assumption that participants' performance in the overt speech task is predictive of the brain responses measured in the covert speech task, we investigated the relationship between behavioral and neural variability. Specifically, we asked whether the variability of the speech latencies across trials is related to the variability of the evoked response magnitude across trials. To answer this question, we measured the correlation between the standard deviation across trials in the behavioral data and the coefficient of variation across trials in the neural data. We observed no significant correlation between behavioral variability and neural variability ($r(39) = 0.17$, $p > 0.1$). To investigate whether hidden (e.g., vividness) and confounding (e.g., fatigue, vividness) variables dramatically influence the relationship between behavioral and neural data in some extreme cases, we designed a statistical method to define outliers. We fitted a linear regression model robust against outliers. The observations falling outside the 99% prediction interval of the linear regression model (i.e., the interval in which a future observation will fall with a certain probability, given what has already been observed) were considered outliers. When we excluded one outlier, we observed a significant correlation between behavioral and neural variability ($r(38) = 0.33$, $p < 0.05$). The correlation between behavioral and neural data is represented in Fig. S2. The results suggest that the temporal precision of speech latencies in the overt speech task is predictive of the temporal alignment of evoked responses in the covert speech task, except for some extreme cases.

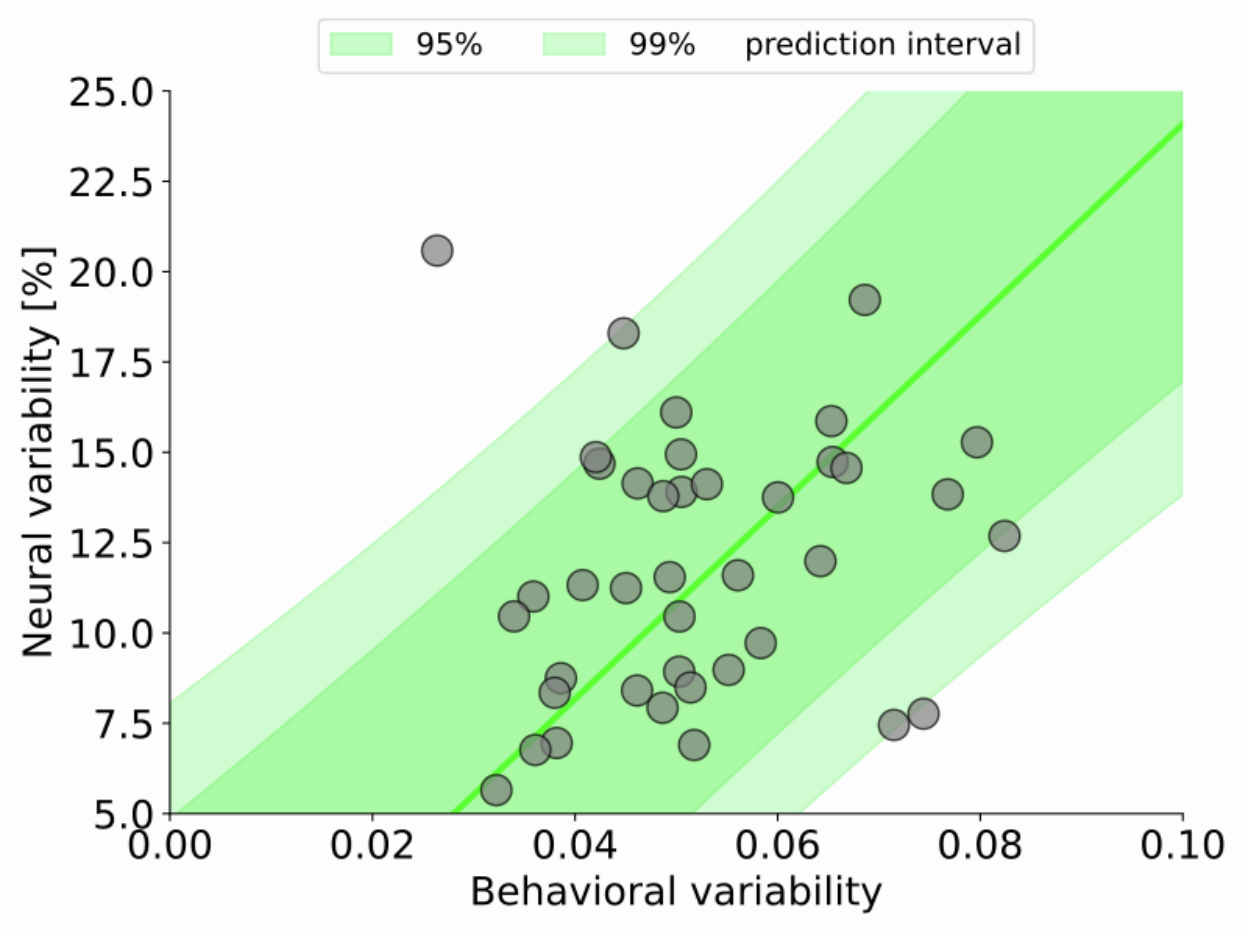


Figure S2. *Correlation between behavioral and neural variability*. The green line represents the best fit of a linear regression model robust to outliers. The shaded area represents the 95% (dark green) and 99% (light green) prediction intervals. These intervals define the range within which we expect future observations to fall, given the model and the variability in the data. The points falling outside the prediction interval, being highly unlikely under the fitted model, were considered outliers.

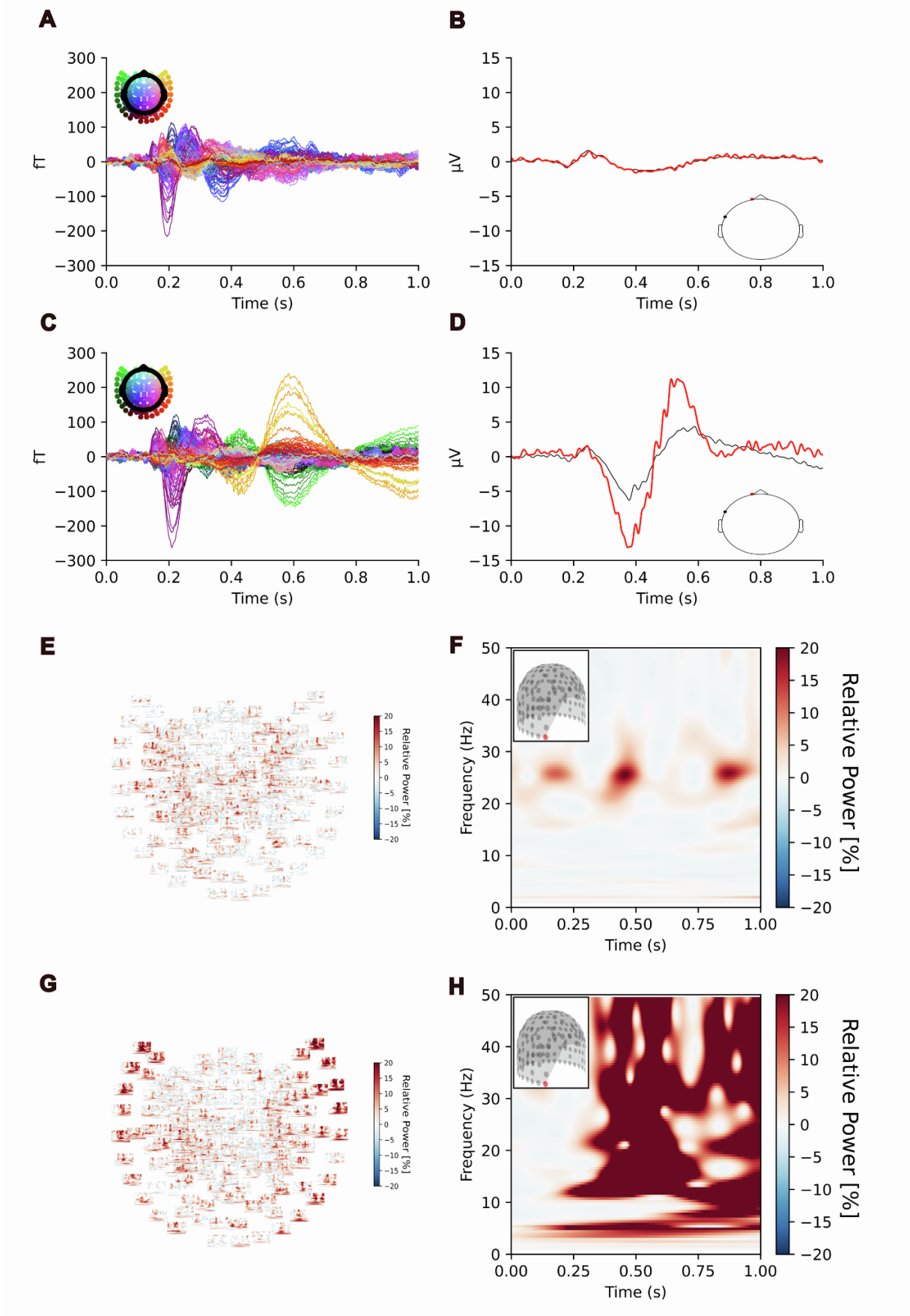

A
B
C
D
E
F
G
H
fT
μV
Time (s)
Frequency (Hz)
Relative Power [%]

Figure S3. *Contaminated vs. uncontaminated MEG signals during overt vs. covert articulation*. A comparison between MEG and EMG signals measured during covert (**A**, **B**, **D**, **E**) and overt (**C**, **D**, **G**, **H**) speech production for one representative subject. Panels (**A**) and (**C**) show the event-related evoked MEG response associated with covert and overt speech production. Every line represents a different MEG sensor. MEG sensors are color-coded according to their spatial location that is shown in the topography in the upper left. Panels (**B**) and (**D**) show the event-related EMG response associated with covert and overt speech production. The red and the black lines represent the signal measured from the electrodes placed on the participant's lip and jaw, respectively. Panels (**E**) and (**G**) show an overview of the induced time-frequency representation across MEG sensors during covert and overt speech production. The induced time-frequency representations estimated for overt speech from the lateral anterior sensors is more noisy as compared to those estimated for covert speech. Panels (**F**) and (**H**) show the induced time-frequency representation for one representative anterior MEG channel that is located near participants' mouth.

Methods S2: MEG/EMG comparison between covert and overt speech, related to Figure 3. We investigated how micromovements compare to overt jaw and lip movements when producing the same syllables. We collected EMG and MEG data during overt speech for one participant to compare the evoked responses between conditions. Fig. S3AC shows MEG event-related responses for a representative subject during covert and overt speech production. Speaking-related artifacts are evident during overt production. Time-frequency analyses show that this artifact is more pronounced in the MEG sensors close to the participant's mouth (i.e., frontal-most sensors; Fig. S3EG and Fig. S3FH). Fig. S3bd shows the EMG event-related responses for a representative subject during covert and overt speech production. Muscle movements measured during overt production are orders of magnitude larger than those measured during covert production. Micromovements measured during covert speech do not contaminate the MEG signal as compared to the actual movements measured during overt speech.

Methods S3: ***Frequency and brain region specificity, related to Figure 5.*** The functional connectivity analysis revealed a lateralization switch pattern in a specific frequency band and connection between brain regions. In particular, we observed an inversion in hemispheric lateralization when the functional connectivity analysis is restricted to the alpha frequency band (8-12 Hz) - and to the connection between premotor and auditory regions. If this lateralization switch is really frequency specific and brain region dependent, then other frequencies and connections between other brain regions should not show the same pattern. We ran additional control analyses to test for frequency specificity and brain region specificity. To test frequency specificity, we measured hemispheric asymmetry in other frequency bands. When we ran the functional connectivity analysis in the same connection between premotor and auditory regions but in the beta frequency band (low beta 18-22 Hz, high beta 28-32 Hz) we did not observe the lateralization switch pattern (see Fig. S4-S5). To test brain region specificity, we measured hemispheric asymmetry in the connection between different brain regions. When we ran the functional connectivity analysis in the alpha frequency band but in the connection between premotor and somatosensory regions, we did not observe the lateralization switch pattern neither using the coarser (Fig. S6) nor the finer parcellation scheme (Fig. S14). Overall, these controls confirm that the lateralization switch pattern is frequency and area specific. However, when using the first EMG micro-deflection (instead of the expected covert onset) as a reference, we observed a left lateralization between premotor and somatosensory regions in the alpha band from ~200 to 230 ms both using the coarser and the finer parcellation scheme (Fig. S15). This finding suggests that different references may be appropriate for auditory and somatosensory processing.

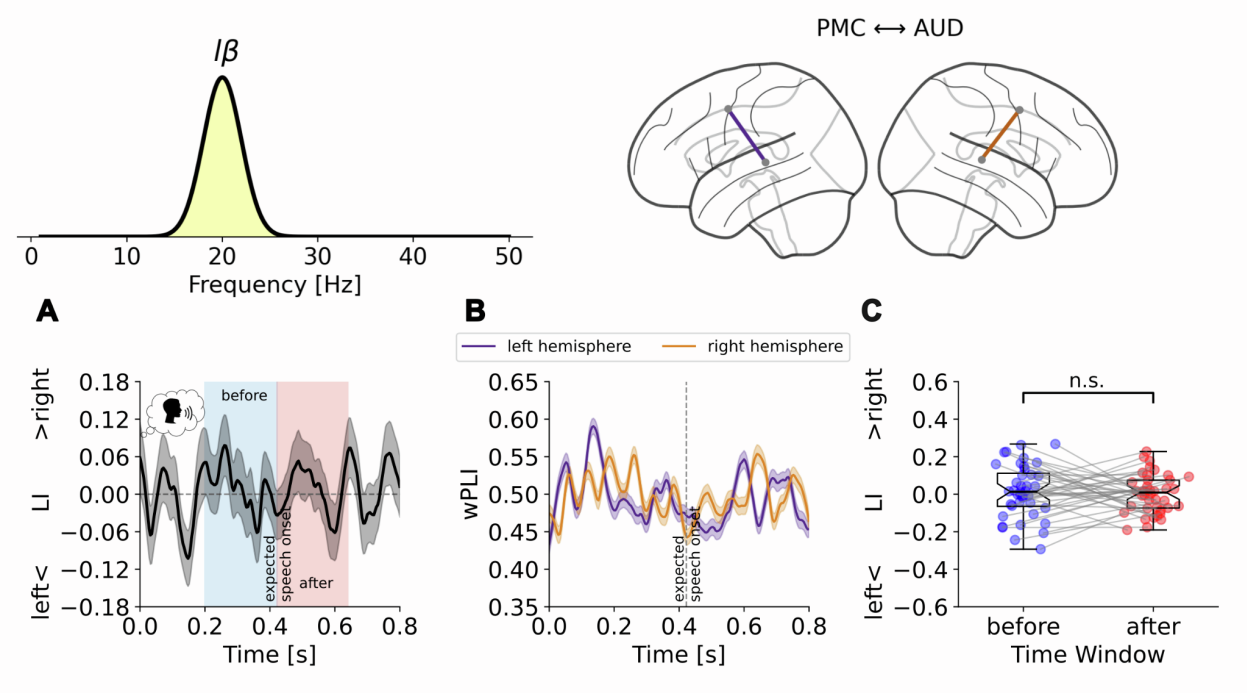


Figure S4. *Hemispheric lateralization over time in the connection between premotor and auditory regions in the low beta frequency band (18-22 Hz)*. The results in the covert speech condition are shown in panels (**A**, **B**, **C**). Panel (**A**) shows the mean lateralization index over time. A negative lateralization index value reflects leftward lateralization, a positive lateralization value reflects rightward lateralization. The black line represents the mean across subjects, the shaded gray area around the black line represents the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the overt behavioral pre-test. The shaded blue and red areas indicate the time windows before and after the expected covert speech onset, respectively. Panel (**B**) shows the weighted phase lag index (wPLI) for each hemisphere, separately. Purple and orange colors represent the left and right hemisphere. The lines and the shaded area around the lines represent the mean and the standard error of the mean, respectively. Panel (**C**) shows a bar plot representing the mean and standard deviation, as well as the single data points, of the lateralization index in the time windows before (in blue) and after (in red) the expected covert speech onset. There is no significant difference in hemispheric lateralization between time windows (n.s., not significant). The connecting lines show individual increase or decrease in lateralization index from the time window before to the time window after expected covert production for each subject.

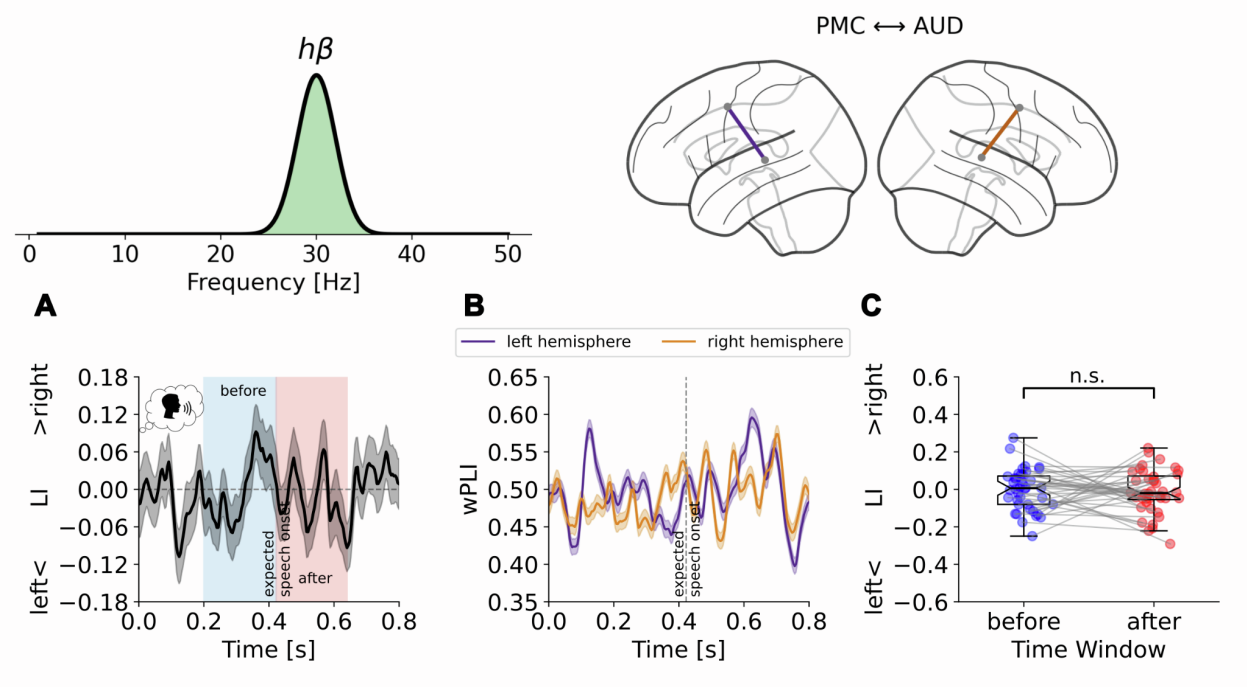


Figure S5. *Hemispheric lateralization over time in the connection between premotor and auditory regions in the high beta frequency band (28-30 Hz)*. The results in the covert speech condition are shown in panels (**A**, **B**, **C**). Panel (**A**) shows the mean lateralization index over time. A negative lateralization index value reflects leftward lateralization, a positive lateralization value reflects rightward lateralization. The black line represents the mean across subjects, the shaded gray area around the black line represents the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the overt behavioral pre-test. The shaded blue and red areas indicate the time windows before and after the expected covert speech onset, respectively. Panel (**B**) shows the weighted phase lag index (wPLI) for each hemisphere, separately. Purple and orange colors represent the left and right hemisphere. The lines and the shaded area around the lines represent the mean and the standard error of the mean, respectively. Panel (**C**) shows a bar plot representing the mean and standard deviation, as well as the single data points, of the lateralization index in the time windows before (in blue) and after (in red) the expected covert speech onset. There is no significant difference in hemispheric lateralization between time windows (n.s., not significant). The connecting lines show individual increase or decrease in lateralization index from the time window before to the time window after expected covert production for each subject.

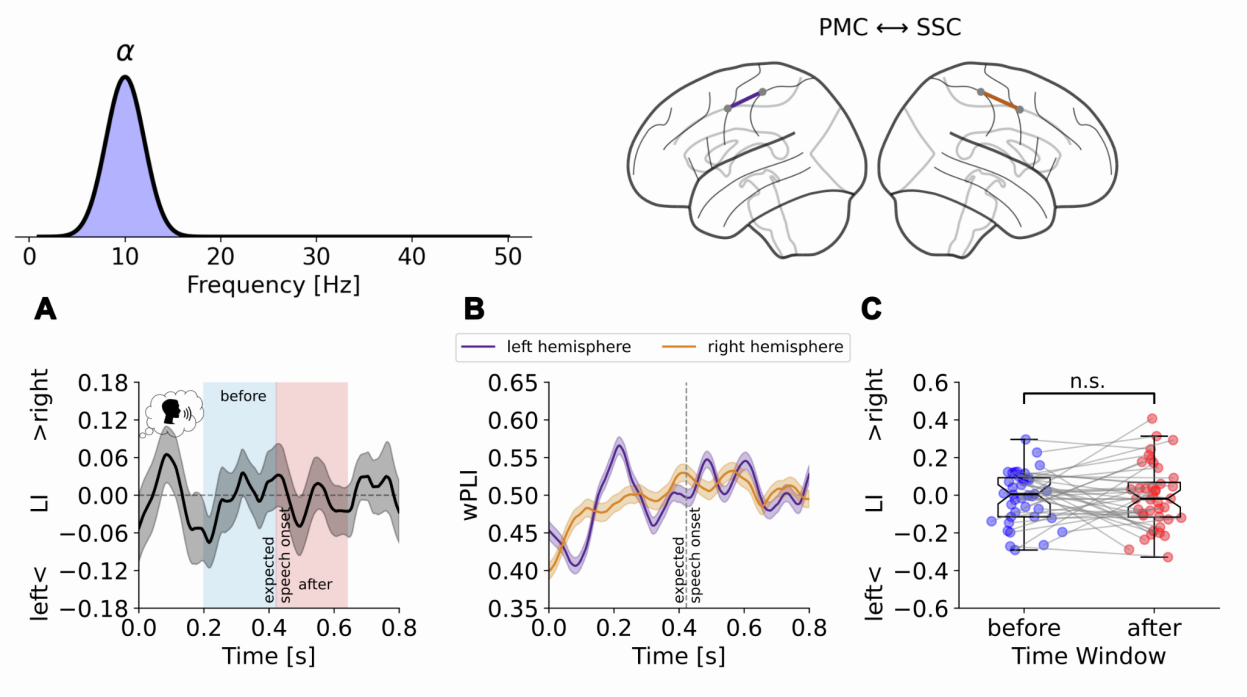


Figure S6. *Hemispheric lateralization over time in the connection between premotor and somatosensory in the alpha frequency band (8-12 Hz)*. The results in the covert speech condition are shown in panels (**A**, **B**, **C**). Panel (**A**) shows the mean lateralization index over time. A negative lateralization index value reflects leftward lateralization, a positive lateralization value reflects rightward lateralization. The black line represents the mean across subjects, the shaded gray area around the black line represents the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the overt behavioral pre-test. The shaded blue and red areas indicate the time windows before and after the expected covert speech onset, respectively. Panel (**B**) shows the weighted phase lag index (wPLI) for each hemisphere, separately. Purple and orange colors represent the left and right hemisphere. The lines and the shaded area around the lines represent the mean and the standard error of the mean, respectively. Panel (**C**) shows a bar plot representing the mean and standard deviation, as well as the single data points, of the lateralization index in the time windows before (in blue) and after (in red) the expected covert speech onset. There is no significant difference in hemispheric lateralization between time windows (n.s., not significant). The connecting lines show individual increase or decrease in lateralization index from the time window before to the time window after expected covert production for each subject.

Methods S4: Inter-individual variability analysis, related to Figure 5. We observed inter-individual variability in the lateralization reversal dynamics. The connecting lines in Figure 5C show individual differences in the variation of the lateralization index from the time window before to the time window after the expected covert production for each subject. Subjects can be grouped in two subsamples depending on the direction of the lateralization reversal. A LI increase (i.e., left-to-right lateralization) can be observed for the majority of the subjects (***Δ*** LI = LI after - LI before > 0, N = 24), while a LI decrease (i.e., right-to-left lateralization) can be observed for fewer subjects (***Δ*** LI = LI after - LI before < 0, N = 16). We further investigated the differences between the two subsamples by measuring the lateralization reversal magnitude (|***Δ*** LI|), that is the absolute strength of the LI increase or LI decrease pre- vs. post- covert speech production. The median |***Δ*** LI| of the subjects showing a LI increase (0.242) is higher than the median |***Δ*** LI| of the subjects showing a LI decrease (0.083). A two-sided unequal sample size Wilcoxon rank-sum test reveals that the |***Δ*** LI| of the subsample showing a LI increase (i.e., left-to-right lateralization) is significantly larger ($w = 2.31$, $p < 0.05$) as compared to the |***Δ*** LI| of the subsample showing a LI decrease (i.e., right-to-left lateralization). Moreover, a visualization of the |***Δ*** LI| distribution (Fig. S7) reveals that most of the subjects in the LI decrease subsample (7/16) show a stable lateralization pattern (neither increase nor decrease, |***Δ*** LI| < 0.05). The remaining subjects showing the opposite trend may be affected by different sources of noise (e.g., temporal misalignment, vividness, fatigue). Overall, despite some subjects showing the opposite trend, the predominant lateralization reversal pattern in the data is the same as the one emerging at the group level (i.e., left-to-right lateralization).

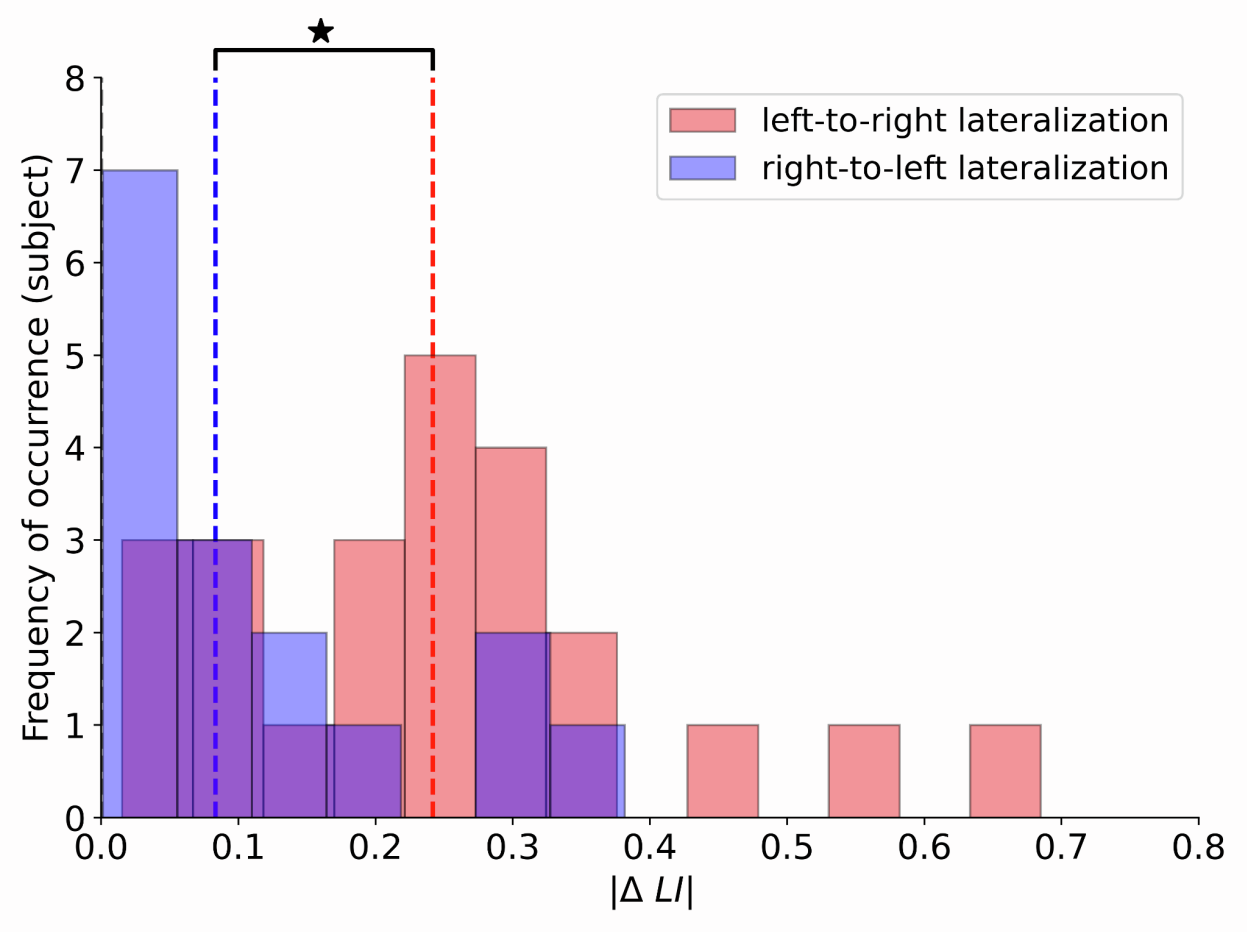


Figure S7. *Prevalence of the lateralization reversal direction (left-to-right vs. right-to-left) across subjects*. The histogram represents the frequency of occurrence of the lateralization reversal magnitude in the subsamples showing a LI increase (i.e., left-to-right lateralization; red) and a LI decrease (i.e., right-to-left lateralization; blue) pre- vs. post- covert speech production. The vertical dashed lines represent the median estimated for each distribution. The magnitude of the hemispheric lateralization (|***Δ*** LI|) is significantly higher in the group of subjects showing left-to-right as compared to right-to-left lateralization (*p < 0.05).

Methods S5: Re-alignment to individual speech latencies, related to Figure 5. To provide support for the conjecture that covert speech latency is similar to overt speech latency, we investigated the impact of a realignment procedure on the temporal dynamics of hemispheric lateralization. The realignment procedure consisted in rearranging the phase shifts measured during covert speech using the temporal landmarks measured during overt speech in the behavioral pre-test. The aim of the realignment procedure was to minimize potential temporal misalignments across subjects caused by individual differences in speech latency. We computed the difference between the median speech onset within and across subjects, and we rearranged the wPLI time series such that individual median speech onsets were realigned to the group median speech onset (Fig. S8AB). Then we measured the lateralization index (Fig. S8CD) and the wPLI for each hemisphere separately (Fig. S8EF). We investigated the time course of hemispheric lateralization differences between realigned and non-realigned data with a time-resolved relative samples t-test. The results show that: when data is not realigned, the LI is significantly more negative (i.e., left-lateralized) in the time-window ranging from ~280 to ~330 ms (before the expected covert production); when data is realigned, the LI is significantly more positive (i.e., right-lateralized) in the time-window ranging from ~490 to ~510 ms (after the expected covert production) (Fig. S9). One possible interpretation for the twofold effect of the realignment procedure is that feedforward and feedback signals are time-locked to different events. Feedforward signals are time-locked to visual stimulus presentation, while feedback signals are time-locked to speech production onset. Therefore, the leftward lateralization pattern is more pronounced without the realignment (centered at 0 ms), while the rightward lateralization pattern is more pronounced with the realignment (centered at 417 ms).

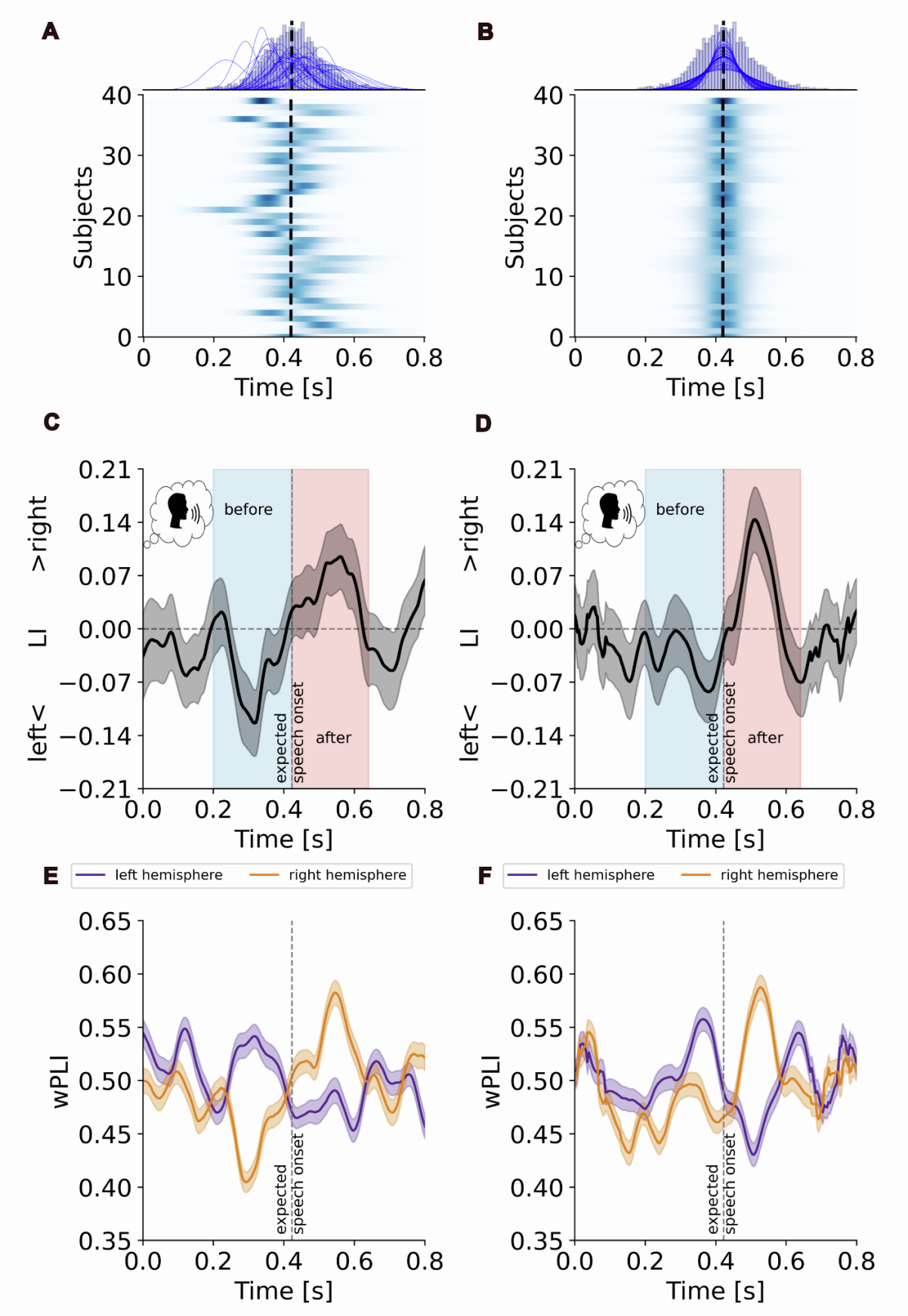

A
B
C
D
E
F
Subjects
Time [s]
LI
>right
left<
before
after
expected
speech onset
wPLI
left hemisphere
right hemisphere

Figure S8. *Hemispheric lateralization over time when the phase shifts are realigned with respect to individual overt speech onsets*. (**A**) Distribution of speech onsets before alignment as measured in the behavioral overt speech pre-test. (**B**) Distribution of speech onsets after alignment. For each subject, we measured the difference between individual median speech onset and group median speech onset. Then, the phase shifts were realigned to the group median speech onset (412 ms). Panels (**C**, **D**) show the mean lateralization index over time in the connection between premotor and auditory regions in the alpha frequency band (8-12 Hz). A negative lateralization index value reflects leftward lateralization, a positive lateralization value reflects rightward lateralization. The black line represents the mean across subjects, the shaded gray area around the black line represents the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the overt behavioral pre-test. The shaded blue and red areas indicate the time windows before and after the expected covert speech onset, respectively. Panels (**E**, **F**) show the weighted phase lag index (wPLI) for each hemisphere, separately. Purple and orange colors represent the left and right hemisphere. The lines and the shaded area around the lines represent the mean and the standard error of the mean, respectively. The LI and the wPLI were measured using the phase shifts without applying the realignment (**C**, **E**) and applying the realignment (**D**, **F**).

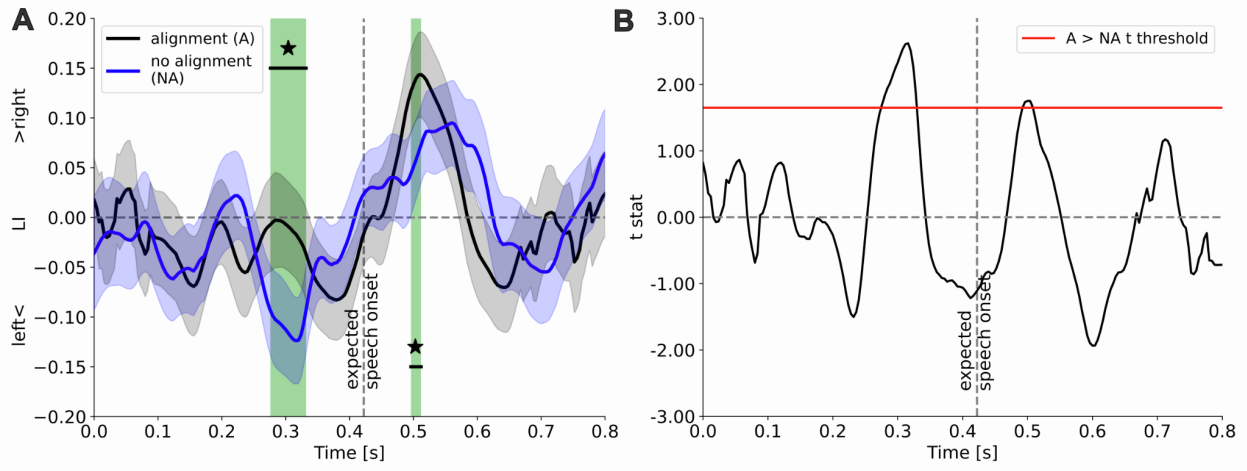


Figure S9. *Time course of hemispheric lateralization differences between alignment vs. no-alignment procedures*. (**A**) lateralization index over time in the connection between premotor and auditory ROIs in the alpha frequency band. A negative lateralization index value reflects leftward lateralization, a positive lateralization value reflects rightward lateralization. The blue and black lines represent the mean across subjects in the alignment and no-alignment procedures, respectively. The shaded areas around the lines represent the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the over behavioral pre-test. The shaded green areas represent the time windows in which the lateralization index was significantly different between procedures (*p < 0.05). (**B**) time-resolved relative samples t-test. The black line represents the t statistic for the difference between procedures. The horizontal red line represents the positive (alignment > no alignment) t-threshold (alpha 5%) based on t distribution and degrees of freedom.

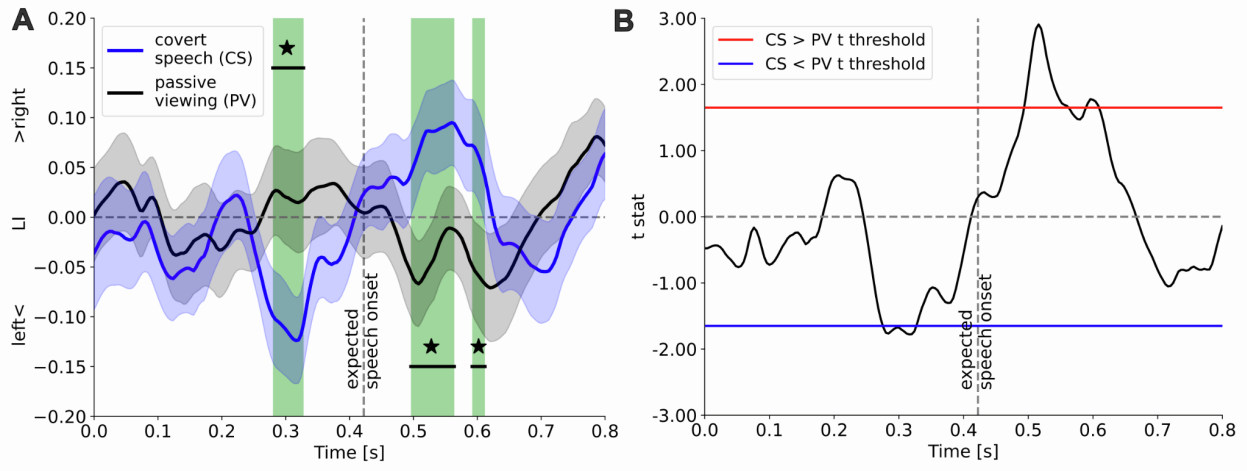


Figure S10. *Time course of hemispheric lateralization differences between covert speech and passive viewing conditions*. (**A**) lateralization index over time in the connection between premotor and auditory ROIs in the alpha frequency band. A negative lateralization index value reflects leftward lateralization, a positive lateralization value reflects rightward lateralization. The blue and black lines represent the mean across subjects in the covert speech and passive viewing conditions, respectively. The shaded areas around the lines represent the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the over behavioral pre-test. The shaded green areas represent the time windows in which the lateralization index was significantly different between conditions (*p < 0.05). (**B**) time-resolved relative samples t-test. The black line represents the t statistic for the difference between conditions. The horizontal red and blue lines represent the positive (covert speech > passive viewing) and negative (covert speech < passive viewing) t-threshold (alpha 5%) based on t distribution and degrees of freedom.

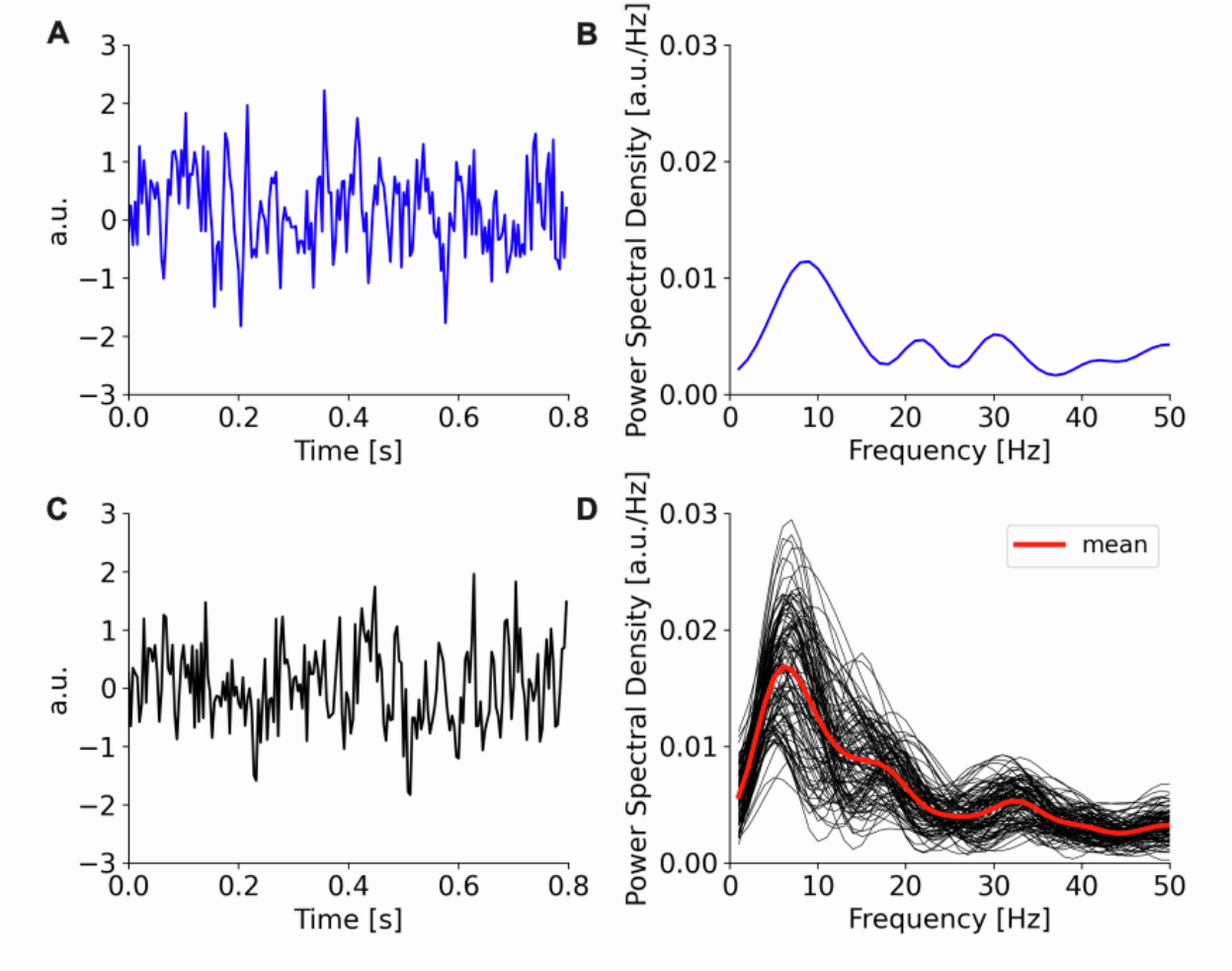


Figure S11. *Iteratively Amplitude Adjusted Fourier Transformed (IAAFT) surrogates*. One representative example of the time series and the power spectral density associated with the data (panels **A**, **B**, blue line) and the surrogates (panels **C**, **D**, black line) in one single trial in one single ROI. The data was projected in source space using a beamformer filter in the alpha frequency band (8-12 Hz). Each surrogate was generated by randomizing the phase of the original signal while preserving as much as possible its original spectral profile. The black lines in panel (**D**) represent the power spectral density estimated for 100 surrogates. The red thicker line represents the mean across 100 surrogates.

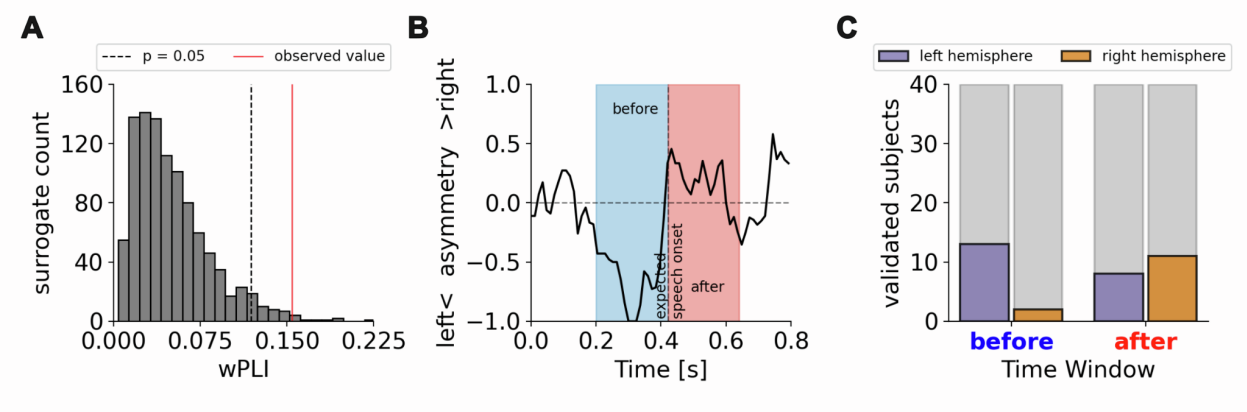


Figure S12. *Surrogate test for single-subject statistical validation of hemispheric lateralization*. For each subject and for each timepoint, we estimated the phase lag index from both the data and 1000 surrogates. Panel (**A**) shows the non-parametric statistical test that we used to evaluate the significance of the phase lag index estimated from the data. The statistical significance is determined based on whether the observed value falls beyond the 95th percentile in the surrogate distribution. Panel (**B**) shows the proportion of validated subjects showing either leftward (negative values) or rightward (positive values) hemispheric asymmetry for each timepoint. The dotted vertical line represents the expected covert speech onset based on the overt behavioral pre-test. The shaded blue and red areas indicate the time windows before and after the expected covert speech onset, respectively. Panel (**C**) shows the absolute number of validated subjects in the time windows before and after the expected covert speech onset.

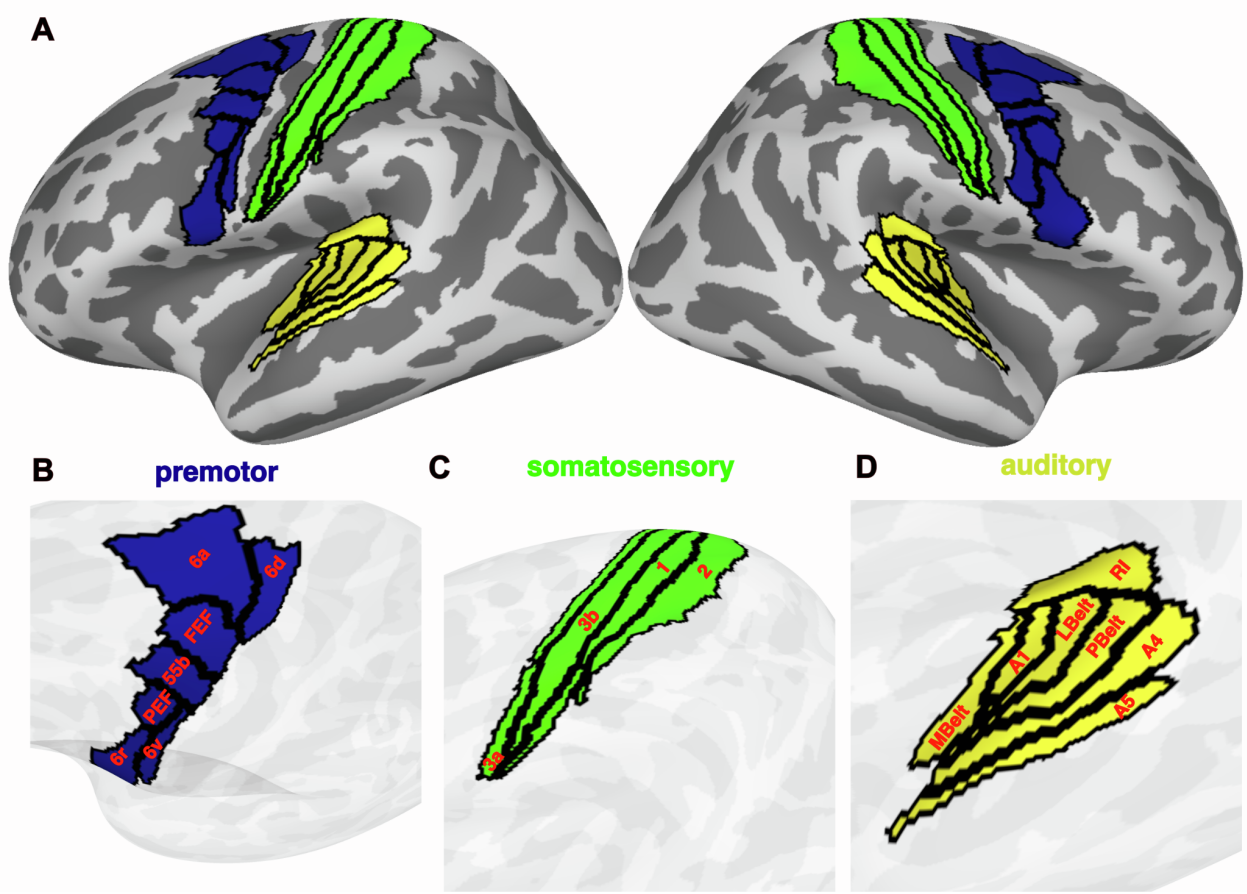


Figure S13. *ROIs definition using a finer parcellation scheme*. (**A**) Spatial location of premotor (blue), somatosensory (green) and auditory (yellow) ROIs. Premotor ROIs are located in the precentral gyrus. (**B**) Premotor ROIs names. 6d: 6 dorsal, 6a: 6 anterior, FEF: frontal eye field, 55b, PEF: premotor eye field, 6v: 6 ventral, 6r: 6 rostral. Somatosensory ROIs are located in the postcentral gyrus (**C**) Somatosensory ROIs names. 3a, 3b, 1, 2. Auditory ROIs are located in Heschl's gyrus and superior temporal gyrus. (**D**) Auditory ROIs names. A1: primary auditory cortex, LBelt: lateral belt, MBelt: medial belt, PBelt: para-belt, RI: retro-insular cortex, A4, A5.

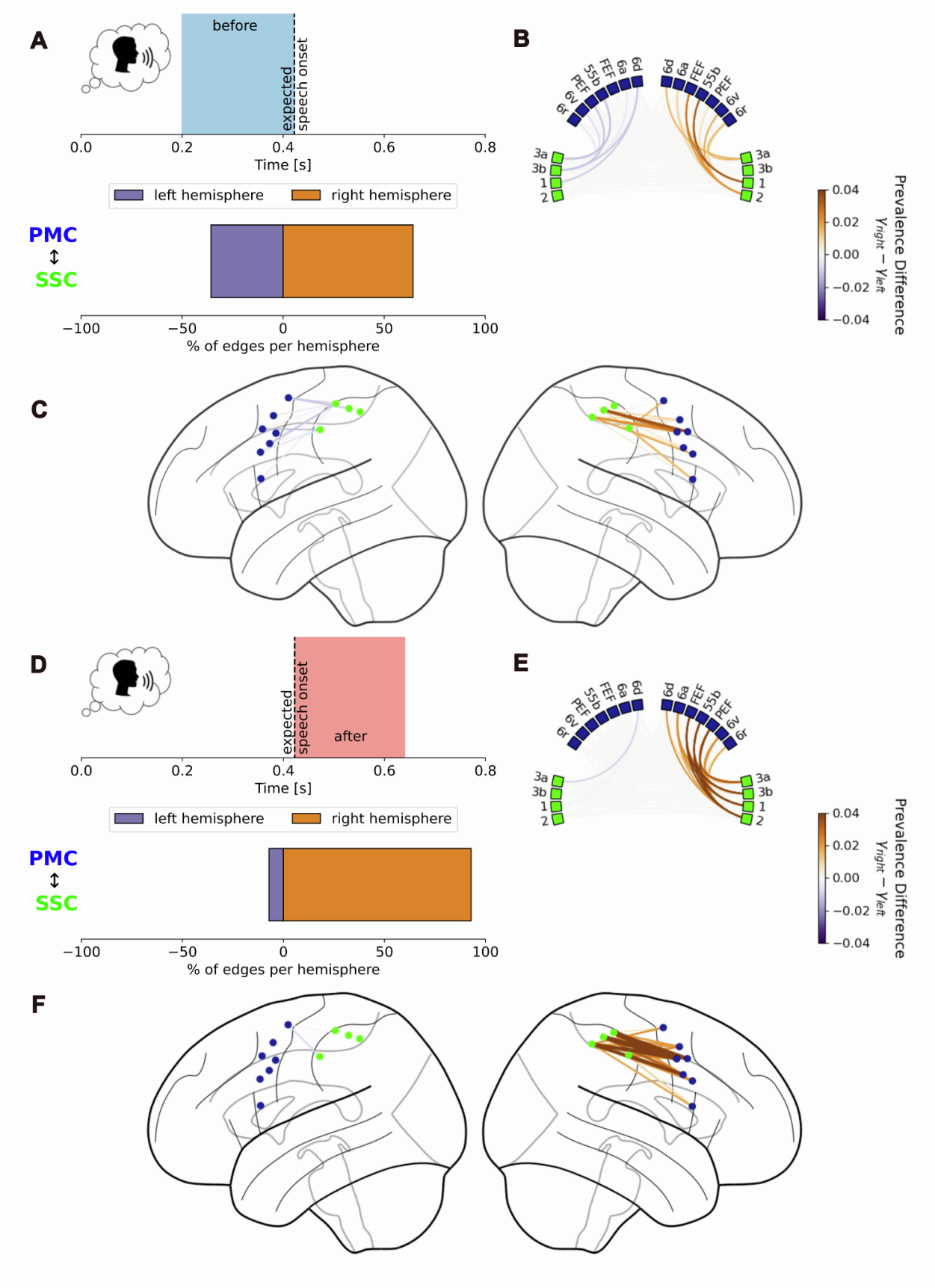

A
before
expected
speech onset
0.0
0.2
0.4
0.6
0.8
Time [s]
left hemisphere
right hemisphere
PMC
↕
SSC
−100
−50
0
50
100
% of edges per hemisphere
B
6d
6a
FEF
55b
PEF
6v
6r
3a
3b
1
2
0.04
0.02
0.00
−0.02
−0.04
Prevalence Difference
$Y_{right} - Y_{left}$
C
D
after
E
F

Figure S14. *Hemispheric lateralization of functional connectivity between premotor and somatosensory regions defined using finer parcellation scheme.* The results obtained in the time window before and after the expected speech onset are shown in panels (**A**, **B**, **C**) and in panels (**D**, **E**, **F**), respectively. (**A**, **D**) A schematic representation of the selected time window is shown in the upper panels, the percentage of edges showing hemispheric lateralization (left hemisphere in purple, right hemisphere in orange) is shown in a horizontal bar plot in the lower panels. (**B**, **E**) the prevalence difference in the connections between premotor and somatosensory areas is represented in a functional connectivity circle plot. Purple lines represent leftward lateralization, orange lines represent rightward lateralization. The premotor ROIs are sorted from the more ventral to the more dorsal. The somatosensory ROIs are sorted from the more anterior to the more posterior. (**C**, **F**) the prevalence difference in the connections between premotor and somatosensory areas is represented in a glass brain plot.

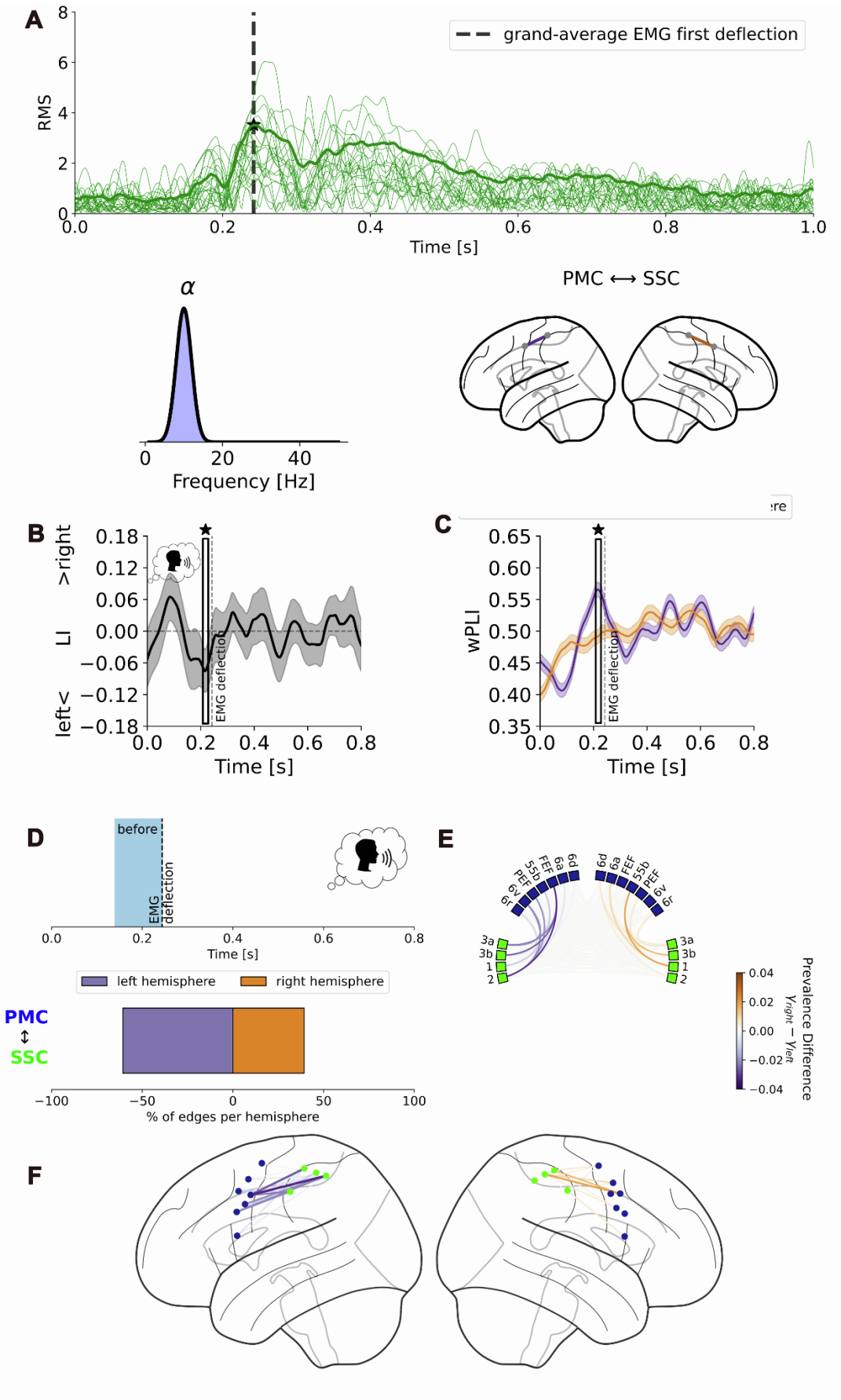

A
RMS
grand-average EMG first deflection
Time [s]
α
Frequency [Hz]
PMC ⟷ SSC
B
>right
LI
left<
EMG deflection
Time [s]
C
wPLI
EMG deflection
Time [s]
D
before
EMG deflection
Time [s]
left hemisphere
right hemisphere
PMC
SSC
% of edges per hemisphere
E
Prevalence Difference
F

Figure S15. *Hemispheric lateralization of functional connectivity between premotor and somatosensory regions with respect to the first EMG micro-deflection.* (**A**) The grand-average (N=40) EMG event-related response (thick green line) superimposed on individual participants' event-related responses (thin green lines). The vertical dashed line indicates the first and most prominent peak in the EMG event-related response corresponding to micromovements onset. (**B, C**) Hemispheric lateralization results obtained using the coarser parcellation scheme in the time window before the first EMG micro-deflection are represented. Panel (**B**) shows the mean lateralization index over time. A negative lateralization index value reflects leftward lateralization, a positive lateralization value reflects rightward lateralization. The black line represents the mean across subjects, the shaded gray area around the black line represents the standard error of the mean. The dotted vertical line represents the first EMG micro-deflection. The black rectangle indicates the time window where the lateralization index is significantly more negative than 0 (*p < 0.05). Panel (**C**) shows the weighted phase lag index (wPLI) for each hemisphere, separately. Purple and orange colors represent the left and right hemisphere. The lines and the shaded area around the lines represent the mean and the standard error of the mean, respectively. The black rectangle indicates the time window where there is a significant difference between hemispheres (*p < 0.05). (**D**, **E**, **F**) Hemispheric lateralization results obtained using the finer parcellation scheme in the time window before the first EMG micro-deflection are represented. (**D**) A schematic representation of the selected time window is shown in the upper panels, the percentage of edges showing hemispheric lateralization (left hemisphere in purple, right hemisphere in orange) is shown in a horizontal bar plot in the lower panels. (**E**) the prevalence difference in the connections between premotor and somatosensory areas is represented in a functional connectivity circle plot. Purple lines represent leftward lateralization, orange lines represent rightward lateralization. The premotor ROIs are sorted from the more ventral to the more dorsal. The somatosensory ROIs are sorted from the more anterior to the more posterior. (**F**) the prevalence difference in the connections between premotor and somatosensory areas is represented in a glass brain plot.

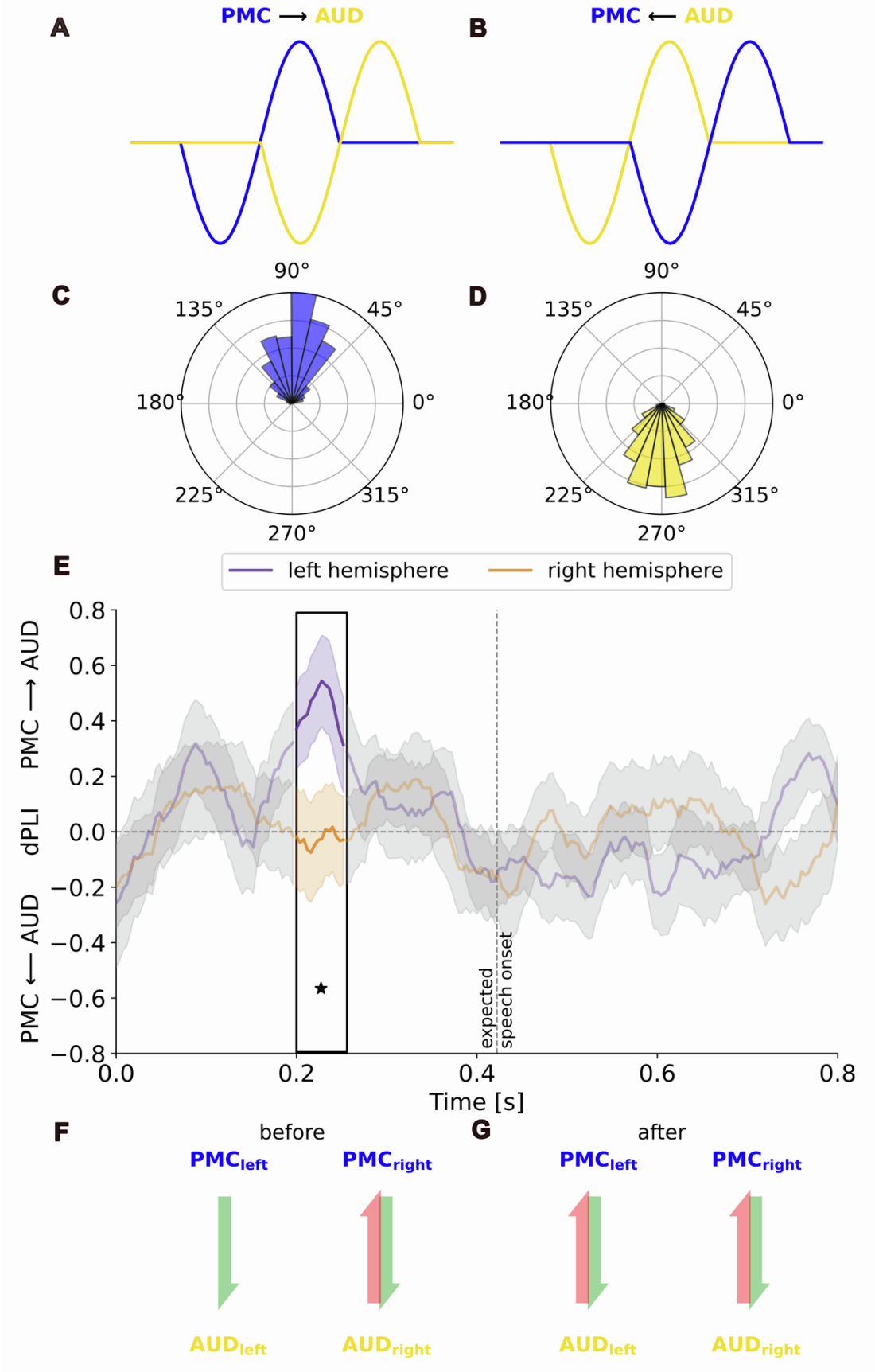

A
PMC ⟶ AUD
B
PMC ⟵ AUD
C
D
90°
135°
45°
180°
0°
225°
315°
270°
E
left hemisphere
right hemisphere
PMC ⟶ AUD
dPLI
PMC ⟵ AUD
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
expected
speech onset
0.0
0.2
0.4
0.6
0.8
Time [s]
F
before
PMCleft
PMCright
AUDleft
AUDright
G
after
PMCleft
PMCright
AUDleft
AUDright

Figure S16. *Hemispheric lateralization of the directionality of the information flow in the connection between premotor and auditory regions in the alpha band*. (**A**) Schematic representation of the scenario in which the premotor time series is leading and the auditory time series is lagging. (**B**) Schematic representation of the scenario in which the auditory time series is leading and the premotor time series is lagging. (**C**) Schematic representation of the distribution of the phase shifts when the premotor time series is leading and the auditory time series is lagging. (**D**) Schematic representation of the distribution of the phase shifts when the auditory time series is leading and the premotor time series is lagging. (**E**) directed Phase Lag Index (dPLI) over time in the left (purple) and right (orange) hemisphere. The line represents the mean across subjects, the shaded area around the line represents the standard error of the mean. The dotted vertical line represents the expected covert speech onset based on the overt behavioral pre-test. A negative dPLI is measured when the premotor time series is leading and the auditory time series is lagging. A positive dPLI is measured when the auditory time series is leading and the premotor time series is lagging. The black rectangle indicates the time window where there is a significant difference between hemispheres (*$p < 0.05$). (**F**, **G**) Schematic representation of the directionality of the information flow in the left and right hemispheres in time windows before (**F**) and after (**G**) the expected covert speech onset. The green arrow represents top-down information flow, the red arrow represents bottom-up information flow.