

# Frequency-specific cortico-subcortical interaction in continuous speaking and listening

Reviewed Preprint

v1 • May 17, 2024

Not revised

Omid Abbasi , Nadine Steingraber, Nikos Chalas, Daniel S. Kluger, Joachim Gross

Institute for Biomagnetism and Biosignal Analysis, University of Münster, Münster, Germany • Otto-Creutzfeldt-Center for Cognitive and Behavioral Neuroscience, University of Münster, Münster, Germany

 [https://en.wikipedia.org/wiki/Open\\_access](https://en.wikipedia.org/wiki/Open_access) Copyright information

## Abstract

Speech production and perception involve complex neural dynamics in the human brain. Using magnetoencephalography (MEG), our study explores the interaction between cortico-cortical and cortico-subcortical connectivities during these processes. Our connectivity findings during speaking revealed a significant connection from the right cerebellum to the left temporal areas in low frequencies, which displayed an opposite trend in high frequencies. Notably, high-frequency connectivity was absent during the listening condition. These findings underscore the vital roles of cortico-cortical and cortico-subcortical connections within the speech production and perception network. The results of our new study enhance our understanding of the complex dynamics of brain connectivity during speech processes, emphasizing the distinct frequency-based interactions between various brain regions.

### eLife assessment

Abbasi and colleagues use Granger causality to explore the cortico-subcortical dynamics during speaking and listening. They find **valuable** evidence for bi-directional connectivity in distinct frequency bands as a function of behaviour, but currently offer **incomplete** support for the validity of their analyses and the predictive coding interpretation of their results.

<https://doi.org/10.7554/eLife.97083.1.sa3>

## Introduction

Human communication can be described as two dynamic systems (i.e., speaker and listener) that are coupled via sensory information (Silbert et al., 2014 ) and operate according to principles of active inference by minimising prediction errors (Friston and Frith, 2015 ). Within this framework, the speaker's predictive processing model steers speech production, allowing adjustments in volume, speed, or articulation based on proprioceptive and auditory feedback. Similarly, the listener's predictive processing model generates anticipations about upcoming

speech which are continually updated and compared with incoming sensory data. The implementation of this model in the human brain was shown to be associated with brain rhythms (Abbasi et al., 2023 [DOI](#); Arnal and Giraud, 2012 [DOI](#); Park et al., 2015 [DOI](#)).

Brain rhythms have been extensively investigated during continuous speech perception. A consistent finding in these studies is the synchronization of frequency-specific brain activity, with the rhythmic amplitude modulation in continuous speech (Giraud and Poeppel, 2012 [DOI](#); Gross et al., 2013b [DOI](#)). The exploration of the brain networks that underpins speech perception has also revealed a stronger causal influence from higher-order brain regions (such as the left inferior frontal gyrus and left motor areas) to the auditory cortex for intelligible speech compared to unintelligible speech (Park et al., 2015 [DOI](#)). In the domain of speech production, both invasive (Llorens et al., 2011 [DOI](#); Ozker et al., 2022 [DOI](#); Riès et al., 2017 [DOI](#)) and non-invasive (Ganushchak et al., 2011 [DOI](#); Janssen et al., 2020 [DOI](#)) electrophysiological studies have significantly contributed, capitalizing on their superior temporal resolution compared to fMRI. However, it's important to note that invasive recordings offer only limited spatial coverage from specific recording sites in patients. Magnetoencephalography (MEG) has been used by several research teams to explore the intricate relationship between brain frequency-specific dynamics and speech production. Ruspantini and colleagues observed notable coherence between MEG activity originating from the motor cortex and EMG activity in lip muscles, with a peak frequency of approximately 2-3 Hz (Ruspantini et al., 2012 [DOI](#)). In a separate investigation conducted by Alexandrou and colleagues, they identified high-frequency band modulation (60-90 Hz) within bilateral temporal, frontal, and parietal brain regions (Alexandrou et al., 2017 [DOI](#)). In our recent study, we also explored speech production and perception using MEG to map the cortical networks engaged during these two experimental conditions (Abbasi et al., 2023 [DOI](#)). We reported significant connectivity from the motor area to STG in lower frequency bands (up to beta) during speaking, and the reverse pattern in high gamma frequencies (Abbasi et al., 2023 [DOI](#)).

Subcortical areas also contribute in updating the brain's internal model during speech production and perception. Thalamic nuclei serve as central nodes for circuits associated with language processing, interacting with the frontal cortex, basal ganglia, cerebellum, and dopaminergic groups. Notably, the motor-related thalamic nuclei link the basal ganglia and cerebellum with the frontal cortex, contributing to both motor and cognitive aspects of language (Barbas et al., 2013 [DOI](#); Silveri, 2021 [DOI](#)). Relatedly, several previous studies have reported the cerebellum's role in several cognitive functions such as speech production and perception (Giraud et al., 2007 [DOI](#); Skipper and Lametti, 2021 [DOI](#)) movement coordination, timing, motor programming, speech motor control, and sensory prediction (Parrell et al., 2017 [DOI](#)). Lesion studies, neuroimaging investigations, and brain stimulation studies have consistently linked the cerebellum to critical aspects of timing and predictive processing in speech perception (Ackermann, 2008 [DOI](#); Manto et al., 2012 [DOI](#); Parrell et al., 2017 [DOI](#)). Temporo-cerebellar coupling has been suggested to support the continuous updating of internal models of the auditory environment (Kotz et al., 2014 [DOI](#)). In a more recent study, Stockert and colleagues reported bidirectional temporo-cerebellar anatomical connectivity (Stockert et al., 2021 [DOI](#)). The authors suggested that these anatomical temporo-cerebellar connections may support the encoding and modeling of rapidly modulated auditory spectro-temporal patterns.

However, to the best of our knowledge no study has so far investigated spectrally resolved, directed functional coupling between cortical and subcortical brain areas during listening and speaking. In this MEG study, we intend to investigate the specific frequency-related connections between both cortical and subcortical brain regions while participants engaged in speaking and listening tasks. Participants answered seven questions, each lasting 60 seconds (speaking condition), and also listened to recordings of their own speech from the previous speaking session (listening condition; see Methods for details). Initially, we identified the cortical and subcortical brain regions involved in speech perception and production through meta-analysis studies. Subsequently, we examined the brain network and directed connections within it to shed light on the cognitive processes underlying listening and speaking. This involved a direct examination of

frequency-specific communication channels for both feedforward and feedback signaling during speech production and perception using multivariate Granger causality. Our connectivity findings confirmed the participation of subcortical regions such as the thalamus and cerebellum in both speech production and perception. We found that feedback signals, connecting the cerebellum to the auditory area, occur at slower rhythms (below 40 Hz), while feedforward signals (in the reverse direction) occur at faster rhythms (above 40 Hz).

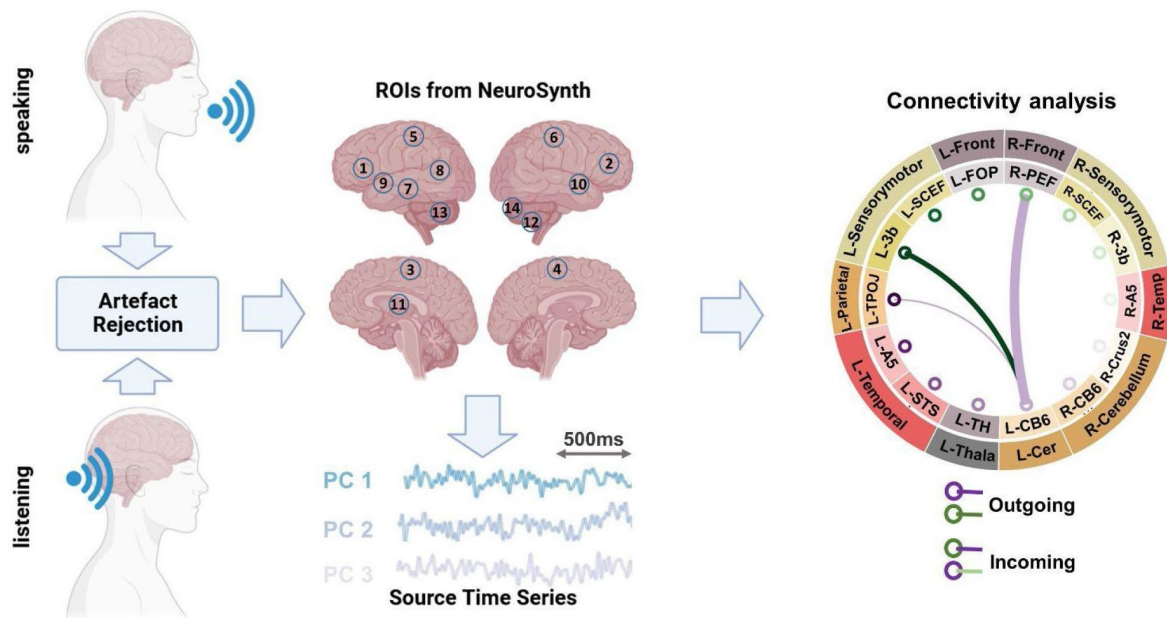
## Results

In this study, we investigated the connectivity between brain regions that are involved in speech production and perception. Our ROI-based analysis utilized the automatic meta-analysis provided by [neurosynth.org](https://neurosynth.org) using the term ‘speech’ that resulted in 642 fMRI studies highlighting the brain areas involved in speech production and perception. We extracted the MNI coordinates of active areas from the resulting neurosynth statistical map, resulting in 14 parcels. For each voxel in a given parcel, the beamformer-derived time-series for all the three source orientations were subjected to singular value decomposition (SVD). In every parcel, we used the three strongest components for further analysis. For estimating pairwise functional connectivity between all parcels ( $n=14$ ), we used multivariate nonparametric Granger causality [mGC; ([Schaum et al., 2021](#))] and related them to the hallmarks of predictive processing models during continuous speaking and listening (see **Fig. 1**). The method uses three-dimensional representations of parcel activity and thus estimates connectivity more accurately than traditional one-dimensional representations ([Chalas et al., 2022](#)).

The calculation of mGC between two parcels yielded two spectra reflecting both directions (A→B and B→A). Next, we computed the directed asymmetry index (DAI) for each pair of spectra ([Bastos et al., 2015](#)). This measure captures the predominant direction of mGC between two parcels based on the relative difference between both directions (see ‘Methods’ for more information). A positive DAI indicates the dominant directionality from A toward B, while a negative DAI indicates the opposite directionality. First, we used group statistics to identify brain areas where DAI values in specific frequency bands in the speaking as well as listening conditions differed significantly from zero. We performed group statistics from 0 to 100 Hz. Next, we defined the following canonical frequency bands: Delta/Theta (0-7 Hz), alpha (7-13 Hz), beta (15-30 Hz), gamma (30-60 Hz), high gamma (60-90 Hz).

Using connectogram plots, we visualized the connectivity between the selected 14 ROIs and examined neural mechanisms associated with speech production and perception, as well as predictive processing models. The number of outgoing and incoming edges for each node was also calculated using the brain connectivity toolbox and represented as the node strength ([Rubinov and Sporns, 2010](#)). The upper row of **Fig. 2** illustrates the strength for three different nodes in the speaking condition. In L-SCEF (pre-supplementary motor area) and R-CB6 (right cerebellum lobule VI), the number of outgoing edges in low frequencies is higher compared to incoming edges. However, we observed a reverse pattern in high frequencies. Interestingly, in the left STG (L-A5), there are more incoming connections compared to outgoing in low frequencies, and this changes in the higher frequencies, with more outgoing connections seen in this area. This result indicates that in low frequencies, the sensory-motor area and cerebellum predominantly transmit information, while in higher frequencies, they are more involved in receiving it. On the other hand, the left STG receives information in low frequencies and transmits it in high frequencies.

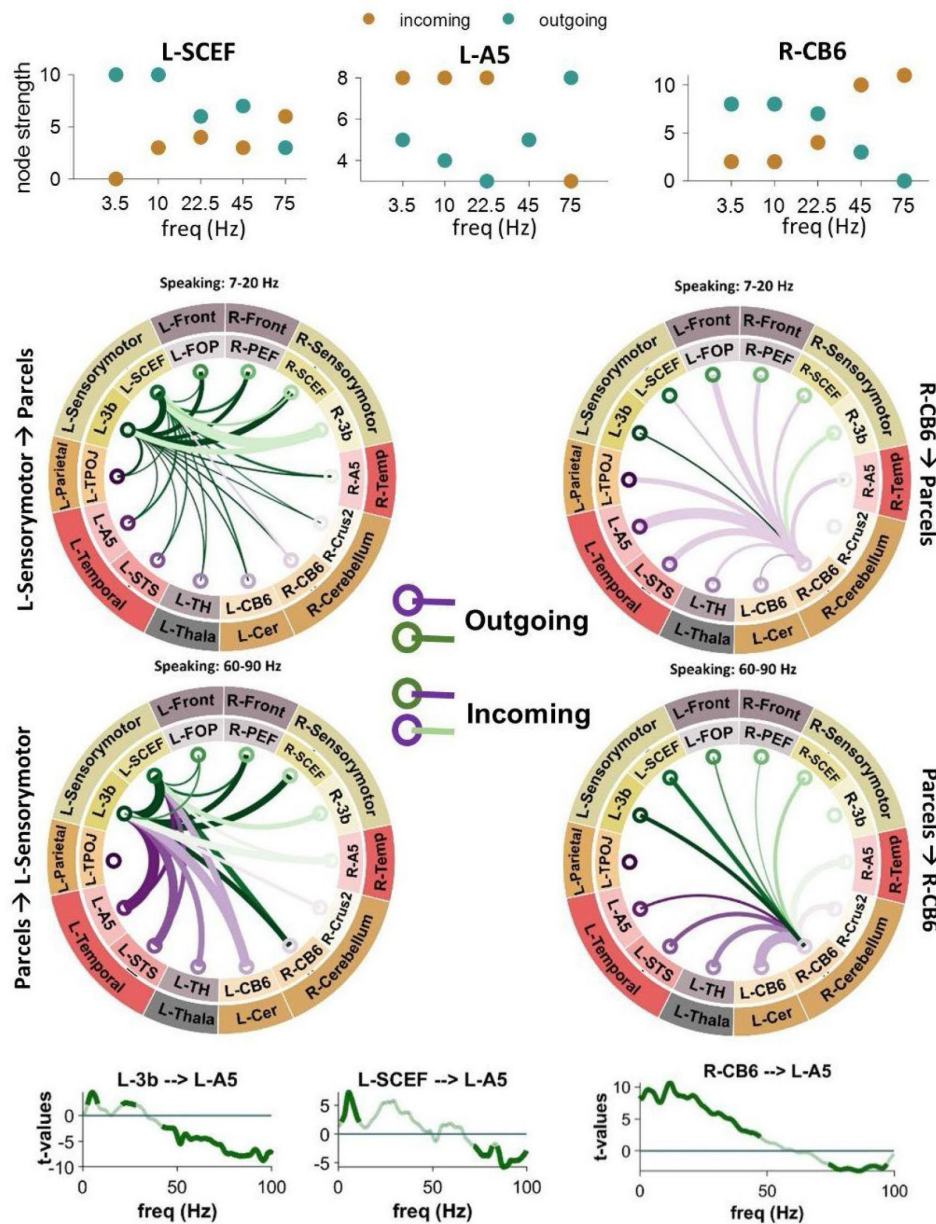
The connectograms in the middle parts of **Fig. 2** illustrate the connections of the left sensory-motor area (left plots; L-3b and L-SCEF) and the right cerebellum (right plots) to other cortical and subcortical parcels in low (top: 7-20 Hz) and high (bottom: 60-90 Hz) frequencies during the speaking condition. In low frequencies, both the sensory-motor area and the right cerebellum primarily send information to other parcels, including the left and right temporal regions, while in



**Fig. 1.**

### Methodological pipeline.

Thirty participants answered 7 given questions (60 seconds each; *speaking* condition) as well as listened to audio-recordings of their own voice from previous sessions (*listening* condition) while MEG data were recorded. Artefacts were removed from the recorded MEG data (Abbasi et al., 2021). Individual MRIs were used to estimate source models per participant which were interpolated to a template volumetric grid. Relevant areas in speech production and perception networks were identified from *Neurosynth.org* platform. Fourteen corresponding anatomical parcels in HCP and AAL atlases were identified: L-FOP (1), R-PEF+6v (2), L-SCEF (3), R-SCEF (4), L-3b (5), R-3b (6), L-STs (7), L-TPOJ1 (8), L-A5 (9), R-A5 (10), L-TH (11), R-CB-Cruss2 (12), L-CB-6 (13), R-CB-6 (14). For each identified parcel, estimated source time-series were extracted. Next, using a blockwise approach, we considered the first three SVD components of each parcel as a block and estimated the connectivity between each pair of parcels using a multivariate nonparametric Granger causality approach (mGC; (Schaum et al., 2021)). In this study, the connectivity results are presented using connectogram plots. In the connectograms, nodes represent the brain areas and edges represent the strength and direction of the connections between them. The thickness of the edges indicates the magnitude of the t-values, while color indicates the directionality of the connectivity. In other words, when node A connects to node B, the edge will have the same color as node A, and vice versa when node B connects to node A. Note that only significant connections are shown in the connectograms ( $p < .05$ , cluster correction). For instance in the illustrated connectogram, the purple edge between L-CB6 and R-PEF shows significant connectivity from L-CB6 to R-PEF. R=Right, L=Left.



**Fig. 2.**

### Connectivity analysis between the sensory-motor, cerebellum, and STG.

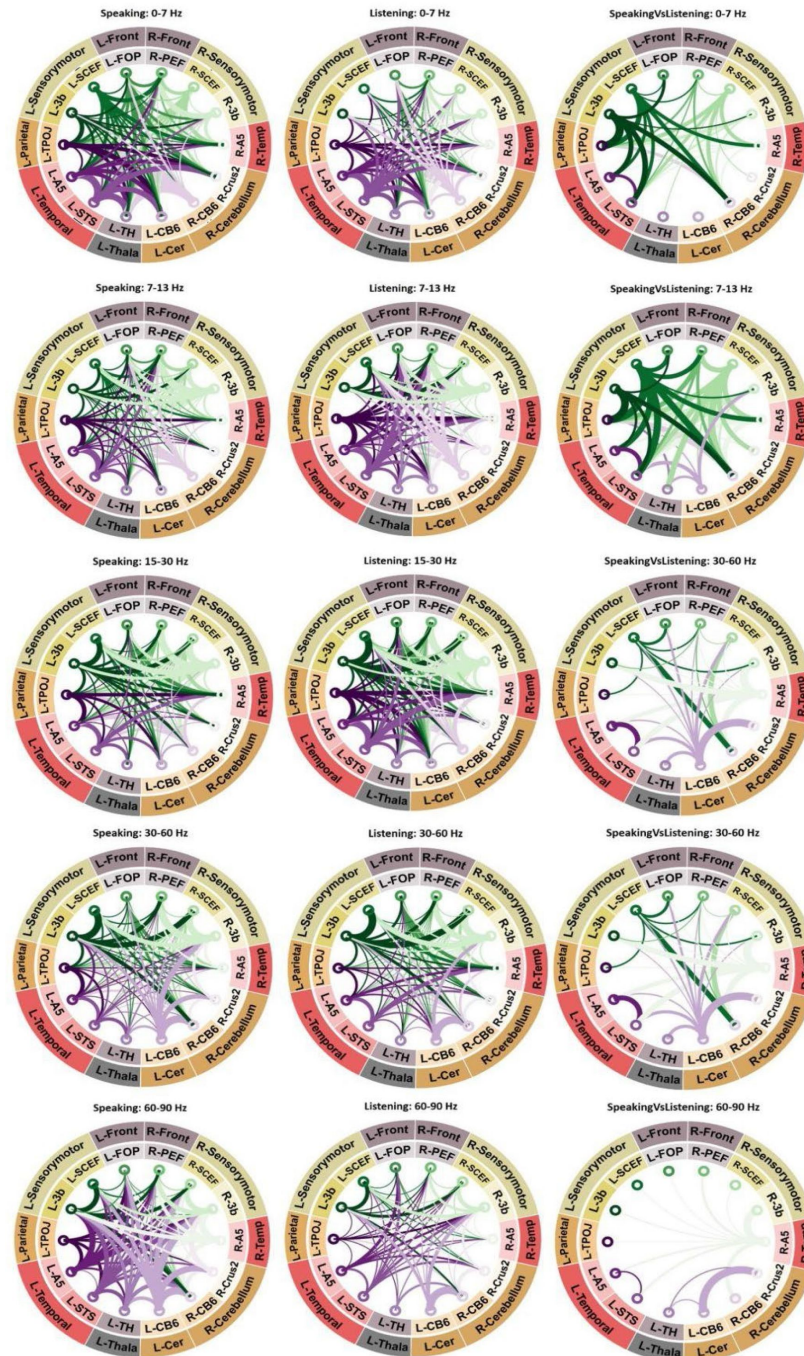
Upper panels illustrate the strength for L-SCEF, L-A5, and R-CB6 nodes in the speaking condition. In the middle part, connectograms illustrate connections between the left sensory-motor area (left plots) and the right cerebellum (right plots) during speaking to other cortical and subcortical parcels at low frequencies (top: 7-20 Hz) and high frequencies (bottom: 60-90 Hz). Lower panels illustrate spectrally-resolved DAI from L-3b (left spectra), LSCEF (middle spectra), and R-CB6 (right spectra) to L-A5. The results represented in the whole figure are significant connectivity patterns that passed a cluster-based permutation test ( $p < .05$ , cluster correction).

the high frequencies, these nodes primarily receive information. Interestingly, our connectivity results illustrate that the top-down effects from higher-order cortical areas to lower-order areas during speaking occur in distinct frequency bands. **Fig. 2** (lower panels) depicts that the strongest top-down connectivity from the left pre-supplementary motor area (pre-SMA; L-SCEF) to lower-order cortical areas such as L-A5 occur in the theta band (peak at 5-7 Hz). However, the top-down connectivity from L-3b to L-A5 was strongest in two distinct theta (peak at 7 Hz) and beta (peak at 21 Hz) frequency bands. Moreover, these results revealed significant bottom-up connectivity from lower-order cortical areas, namely L-A5 to left sensory-motor area (L-3b and L-SCEF) in high frequency bands. As we previously showed (Abbasi et al., 2023), this striking reversal indicates a dissociation of bottom-up and top-down information flow during speaking in distinct frequency bands. Signals communicated top-down are predominantly transmitted in low-frequency ranges, while those communicated bottom-up are transmitted in high-frequency ranges. Notably, **Fig. 2** (bottom-right panel) also depicts directional connectivity from the right cerebellum lobule VI (R-CB6) to the left STG (L-A5) in frequencies below 50 Hz. In contrast, in frequencies above 50 Hz, we observed reverse signaling from the left STG to the right cerebellum during speaking.

The connectograms in **Fig. 3** provide a comprehensive overview of significant connectivities across various frequency bands among the selected fourteen brain areas during speaking and listening conditions. Specifically highlighting the first column's connectograms during speaking, we observe comparable connectivity patterns to those of the left sensory-motor area. Connections from the right higher-order cortical regions, such as R-SCEF and R-3b, towards lower-order areas like R-A5, are noticeable in low-frequency bands (<30 Hz). Conversely, for higher frequencies (>30 Hz), we noticed the reverse directional connections, from lower-order to higher-order brain regions. Furthermore, **Fig. 4A** reveals strong top-down connectivity from bilateral 3b and left FOP to bilateral A5 and left STS, particularly prominent in two distinct frequency bands: theta (peak at 7 Hz) and beta (peak at 21 Hz). Additionally, similar to the left hemisphere, our results in **Fig. 4A** show significant bottom-up connectivity from lower-order cortical areas (bilateral A5 and left STS) to higher-order cortical regions (bilateral 3b and bilateral SCEF) across both low and high frequency bands.

We observed a similar pattern during the listening condition compared to the speaking condition, including reversed directionality between low and high frequencies (**Fig. 3**, second column). However, both top-down effects in the low frequency band and bottom-up effects in the high frequency band were less pronounced during listening compared to speaking. The third column in **Fig. 3** as well as the third panel of **Fig. 4** confirm these results by showing the direct comparisons of DAI connectivity between speaking and listening. This figure illustrates significant differences in connectivity between speaking and listening conditions in different frequency bands. Specifically, stronger top-down signals originating from higher-order cortical areas (such as bilateral SCEF and 3b) towards lower-order brain areas, such as bilateral A5 and left STS were observed during speaking compared to listening. Conversely, the individual spectrally-resolved DAI plots for the listening conditions, along with the connectograms, indicate the presence of significant connectivity solely from low-order areas (such as bilateral A5) towards higher-order cortical areas (bilateral SCEF and 3b) during the speaking condition, as opposed to the listening condition. More detailed spectrally-resolved DAI between all ROIs and both conditions are illustrated in **Fig. 4**.

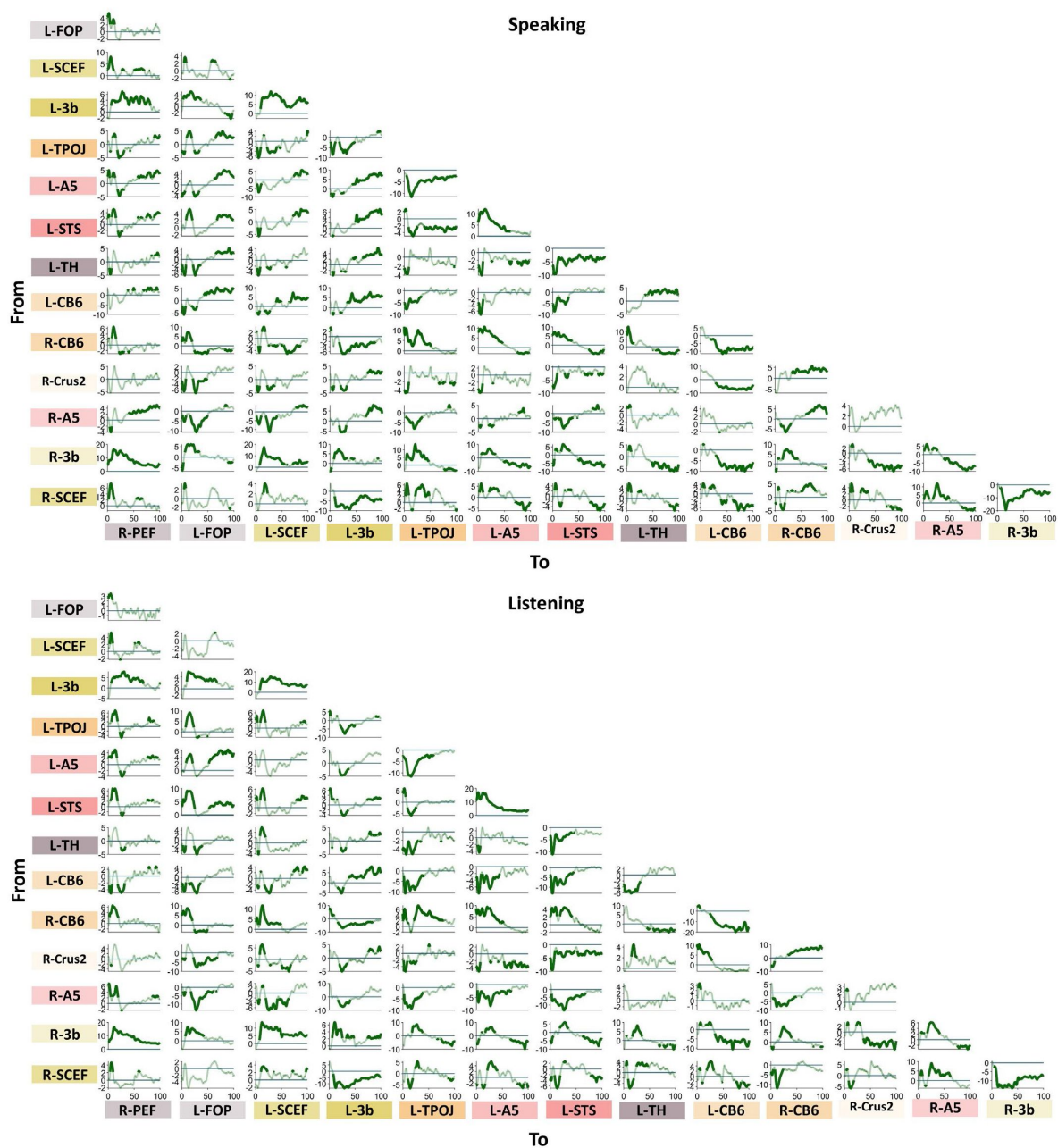
Notably, we found significant connections between subcortical and cortical areas, indicating their involvement in speech production and perception networks. **Fig. 4** depicts the connectivity from the right cerebellum to other cortical and subcortical parcels. In contrast with what we reported for the speaking condition, during listening, there is only a significant connectivity in low frequency to the left temporal area but not a reverse connection in the high frequencies. Our analysis also revealed robust connectivity between the right cerebellum and the left parietal cortex, evident in both speaking and listening conditions, with stronger connectivity observed



**Fig. 3.**

### Connectivity analysis.

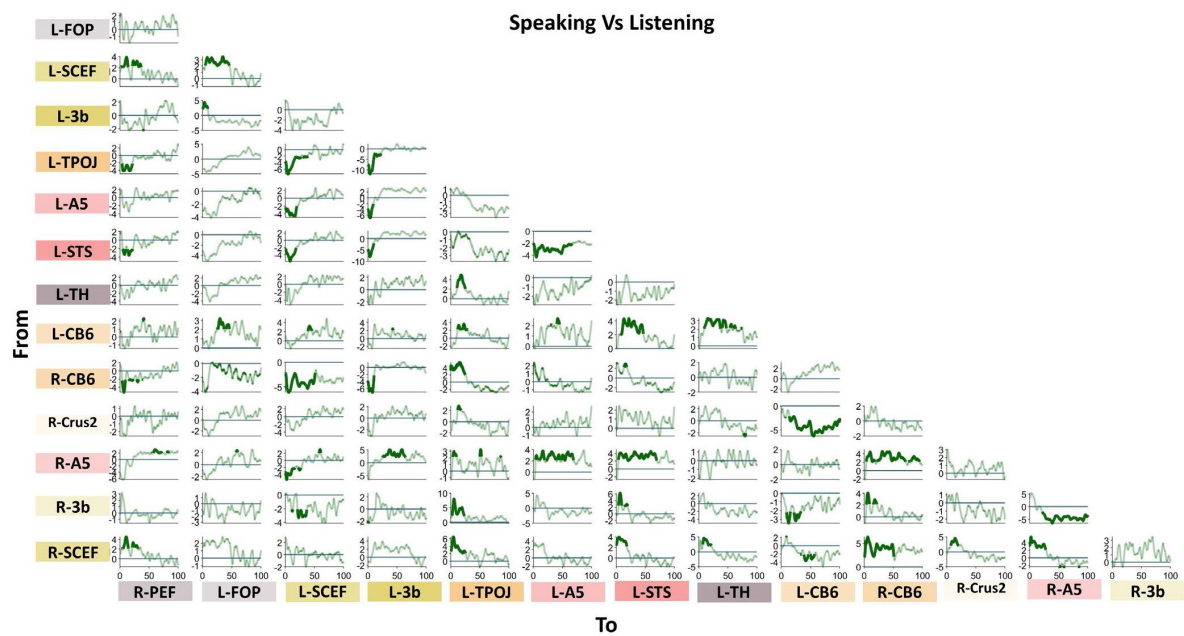
Significant connectivity between 14 ROIs involved in speech production and perception. A cluster-based permutation test was used to detect significant connectivity patterns. The results of statistical analysis revealed significant connectivities between different brain areas during speaking (first column of connectograms), listening (second column of connectograms), and the comparison between speaking and listening conditions (third column of connectograms) across various frequency bands. In the connectograms, nodes represent the brain areas and edges represent the strength and direction of the connections between them. The thickness of the edges indicates the magnitude of the t-values, while color indicates the directionality of the connectivity. In other words, when node A connects to node B, the edge will have the same color as node A, and vice versa when node B connects to node A. Note that only significant connections are shown ( $p < .05$ , cluster correction).



**Fig. 4.**

### Connectivity results.

Spectrally-resolved DAI between all pairs of ROIs. A cluster-based permutation test was used to detect significant connectivity patterns. The results of statistical analysis revealed significant connectivities between different brain areas during speaking (top), listening (middle), and the comparison between speaking and listening conditions (bottom). The directionality is from y-axis to x-axis. X-axis represents frequency (Hz) and y-axis represents DAI values. Note that significant values are highlighted with increased line width ( $p < .05$ , cluster correction).



**Fig. 4.** (continued)

during speaking. Notably, **Figure 4** depicts a prominent frequency peak in the alpha band, illustrating the specific frequency range through which information flows from the cerebellum to the parietal areas. There is also an intriguing connection between the left thalamus and right cerebellum, which reveals a distinct pattern of connectivity. Specifically, we found significant connectivity from the cerebellum to the thalamus in the low-frequency range during speaking, while the opposite pattern was observed in the high-frequency range (**Fig. 4**).

Finally, we examined the relationship between the connectivity patterns from the selected cortical parcels to the right cerebellum and the coupling of the speech envelope with the oscillations in the superior temporal gyrus (STG). The speech-brain coupling was computed using a multivariate mutual information (MI) approach presented in our recently published study (Abbasi et al., 2023). We investigated the correlations between the directional connectivity indices and MI values across participants for each parcel and frequency band in speaking condition. Our analysis showed significant negative correlations between the RCB6 to L\_A5 connectivity and the speech-STG coupling in the theta band (at 130 ms lag) for speaking condition (**Fig. 5**;  $p < .05$ ).

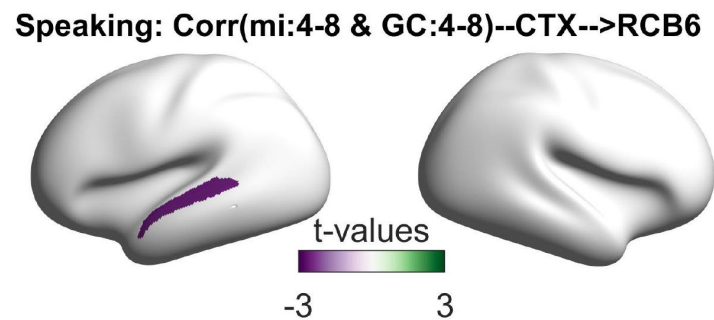
## Discussion

The present study utilized multivariate Granger causality analysis to investigate the frequency-specific brain networks during continuous speaking and listening. Our results revealed significant connectivity from bilateral higher-order cortical areas to bilateral auditory areas (STG and STS) in lower frequency bands (up to beta) during speaking, and in the opposite direction in gamma frequencies. Notably, subcortical areas also contributed to the speech network, with directional communication observed from the right cerebellum lobule VI (R-CB6) to the left auditory area below 30 Hz. Conversely, in frequencies above 30 Hz, we observed reverse signaling from the left auditory areas to the right cerebellum.

### Directed connectivity in speech production and perception

In this study, we aimed to investigate directional connectivity in speech production and perception. Using the automatic meta-analysis provided by neurosynth.org, we included cortical and subcortical brain regions involved in these processes. Our multivariate Granger causality analysis revealed the contribution of distinct frequency channels in both continuous speaking and listening in the auditory-motor-cerebellum domain. Specifically, feedback signals were communicated via slow rhythms below 40 Hz, whereas feedforward signals were communicated via faster rhythms. Our study builds upon our previous work, which demonstrated top-down signaling in low frequencies from higher-order cortical areas to STG and STS and the reverse pattern in higher frequencies. Our findings are also supported by previous studies showing distinct frequency channels for feedforward and feedback communication between two hierarchically different auditory areas (Fontolan et al., 2014) and demonstrating that prediction errors are represented in gamma power while predictions are represented in lower frequency beta power (Sedley et al., 2016).

Based on our meta-analysis results, several subcortical areas such as cerebellum and thalamus are involved in speech production and perception. Previous studies have supported the cerebellum's role in speech production and perception (Giraud et al., 2007; Skipper and Lametti, 2021). Our connectivity analysis findings demonstrate the involvement of the cerebellum within speech networks. Specifically, our results indicate directed connectivity from the right cerebellum to left temporal areas (L-A5 and L-STS) in lower frequencies, and the reverse direction in higher frequencies. Similar to the connections between higher-order cortical areas and temporal areas, feedback signaling between the cerebellum and temporal areas occurs in low frequencies (below 40 Hz), that could potentially play a role in conveying timing information for upcoming speech. In contrast, feedforward signaling occurs in gamma frequency (above 60 Hz) from temporal areas to



**Fig. 5.**

**Correlation between speech-brain coupling and granger causality.**

Speech-STG coupling in theta range (positively lagged: 130ms) is negatively correlated with the directional connectivity from the right cerebellum lobule VI to the left temporal areas in speaking condition in theta frequency band.

the cerebellum, facilitating the update of sensory predictions. Comparison between speech production and perception conditions revealed stronger feedback signaling from the cerebellum to temporal areas during speech production, aligning with the nature of updating sensory predictions in this context. Our findings closely align with recent studies that have investigated the role of the cerebellum in predictive internal modeling mechanisms in speech production and perception (Stockert et al., 2021 [↗](#); Todorović et al., 2023 [↗](#)). These studies suggest that the combined yet distinct activation of temporal, parietal, and cerebellar regions during internal and external monitoring points toward their involvement in auditory and somatosensory targets and continuous updating of auditory environmental models. Moreover, temporo-cerebellar coupling may underlie the precise encoding of temporal structure and support the ongoing optimization of spectro-temporal models of the auditory environment within a network comprising the prefrontal cortex, temporal cortex, and cerebellum (Kotz et al., 2014 [↗](#); Stockert et al., 2021 [↗](#); Todorović et al., 2023 [↗](#)).

Our findings of significant negative correlations between right cerebellar connectivity to the left temporal area and speech-STG coupling in the theta band during speech production resonate with our prior work (Abbasi et al., 2023 [↗](#)). In our previous study, we reported a similar negative correlation involving theta-range speech-brain coupling in the left auditory area and top-down beta connectivity from motor areas. These observations parallel existing research on sensory attenuation, where the brain predicts the sensory outcomes of self-generated actions, resulting in reduced cortical responses (Martikainen et al., 2005 [↗](#)). In earlier studies, we associated this phenomenon with the modulation of beta-range directional couplings originating from motor cortices toward bilateral auditory regions, indicating predictive processes (Abbasi and Gross, 2020 [↗](#)). Now, our new results introduce the cerebellum as another key node in this sensory attenuation mechanism, expanding our understanding of how the brain anticipates and minimizes sensory responses during speech production.

The similarities observed between the feedback and feedforward signaling from the cerebellum and higher-order cortical areas to the temporal areas suggest a shared contribution in predicting the sensory consequences of generated speech. It is proposed that cortico-cerebellar and corticocortical predictions interact in speech networks, with the cerebellum potentially involved in predicting well-learned speech, while the cortex flexibly applies predictions in novel contexts (Ackermann, 2008 [↗](#); Ackermann et al., 1998 [↗](#); Skipper and Lametti, 2021 [↗](#)). These findings deepen our understanding of the role of the cerebellum in speech production and perception and align with previous research highlighting its involvement in predictive processing mechanisms.

We also observed significant connectivity between the right cerebellum and the left parietal cortex, specifically in the low-frequency ranges, which peaked in the alpha range. This result is consistent with previous research demonstrating the alpha band connectivity between parietal and temporal areas during speech production and perception (Abbasi et al., 2023 [↗](#)). We and others previously suggested that the parietal cortex might control and modulate the alpha rhythms in the early auditory areas which serves as a multi-dimensional filter (Abbasi et al., 2023 [↗](#); Lakatos et al., 2019 [↗](#)). Moreover, our current results reveal that during speaking and listening, the parietal cortex not only receives inputs from higher cortical areas such as the motor cortex, but also receives inputs from the cerebellum. These inputs enable the parietal cortex to encode predicted sensory consequences during speaking and provide top-down signals to early auditory areas. Additionally, our new findings demonstrate stronger connectivity from the cerebellum to the parietal areas during the speaking condition compared to the listening condition. This enhanced connectivity may be attributed to the selective inhibition of auditory signals during speaking, as predicted by the internal predictive coding model (Cao et al., 2017 [↗](#); Floegel et al., 2020 [↗](#)).

The thalamus is another subcortical structure involved in speech perception and production which appears to play a significant role as it serves as a crucial intermediary in the cortico-subcortical neural network involving the cerebellum (Barbas et al., 2013 [↗](#); Silveri, 2021 [↗](#)). Our findings reinforce this notion, revealing significant connectivity from the right cerebellum to the left thalamus in the low-frequency band, with the opposite direction observed in the high-frequency band during speaking. These connectivity patterns align with the results observed between the right cerebellum and other cortical areas, providing further support for the thalamus's role in interconnecting cortical and subcortical structures, facilitating communication and coordination in various stages of verbal production (Crosson, 2013 [↗](#)).

This study identifies separate frequency bands used in both speaking and listening. The findings highlight the importance of feedback and feedforward communication in these processes, suggesting potential involvement in optimizing neural processing during speech tasks. Notably, we have also unveiled the anticipated role of the cerebellum within this framework. However, definitive conclusions should be drawn with caution given recent studies raising concerns about the notion that top-down and bottom-up signals can only be transmitted via separate frequency channels (Ferro et al., 2021 [↗](#); Schneider et al., 2021 [↗](#); Vinck et al., 2023 [↗](#)). Therefore, further investigation is warranted to thoroughly assess the alignment between our results and the contribution of the predictive coding framework in speech production and perception.

## Methods

### Participants

Thirty native German-speaking participants (15 males, mean age  $25.1 \pm 2.8$  years [ $M \pm SD$ ], range 20-32 years) were recruited for this study. Prior written informed consent was obtained before measurements and participants received monetary compensation after partaking. The study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki.

### Recordings

Magnetoencephalography, EMG, and speech signals were recorded simultaneously. The speech recording had a sampling rate of 44.1 kHz, whereas the MEG, a 275 channel, whole-head sensor system (OMEGA 275, VSM Medtech Ltd., Vancouver, BC, Canada) was sampled with 1200Hz.

In order not to cause any artefacts by the microphone used for capturing audio data, it was placed at a distance of 155cm from the participants mouth. Three pairs of EMG surface electrodes were placed after tactile inspection to find the correct location to capture muscle activity from the m. genioglossus, m. orbicularis oris, and m. zygomaticus major (for exact location see **Fig. 1** [↗](#) in (Abbasi et al., 2021 [↗](#))). One pair of electrodes was used for each muscle with about 1 cm between electrodes. A low-pass online filter with a 300Hz cut-off was applied to the recorded MEG and EMG data.

### Paradigm

Participants were asked to sit relaxed while performing the given tasks and to keep their eyes on a white fixation cross. The experiment was split in two separate parts: The first one consisted of answering given questions, each for 60 s, thus recording overt speech. Participants had to answer seven questions covering neutral topics, such as 'What does a typical weekend look like for you?'. A color change from white to blue fixation cross indicated the beginning of the time period in which participants should speak and the end was marked by a color change back to white. The

second part focused on perceiving speech in the way that participants listened to their own answers from part one. The list of questions as well as further details of the paradigm presented can be found in (Abbasi et al., 2023 [↗](#)).

## Preprocessing and data analysis

Prior to data analysis, MEG data were visually inspected. No jump artefacts or bad channels were detected. A discrete Fourier transform (DFT) filter was applied to eliminate 50 Hz line noise from the continuous MEG and EMG data. Moreover, EMG data was highpass-filtered at 20 Hz and rectified. Continuous head position and rotation were extracted from the fiducial coils placed at anatomical landmarks (nasion, left, and right ear canals). MEG, EMG, and head movement signals were downsampled to 256 Hz and segmented to non-overlapping 60s trials corresponding to each of their overt answers. In the preprocessing and data analysis steps, custom-made scripts in Matlab R2020 (The Mathworks, Natick, MA, USA) in combination with the Matlab-based FieldTrip toolbox (Oostenveld et al., 2011 [↗](#)) were used in accord with current MEG guidelines (Gross et al., 2013a [↗](#)).

## Artefact rejection

For removing the speech-related artefacts we used the pipeline presented in (Abbasi et al., 2021 [↗](#)). In a nutshell, the artefact rejection comprises four major steps: (i) Head movement-related artefact was initially reduced by incorporating the head position time-series into the general linear model (GLM) using regression analysis (Stolk et al., 2013 [↗](#)). (ii) To further remove the residual artefact, singular value decomposition (SVD) was used to estimate the spatial subspace (components) containing the speech-related artefact from the MEG data. (iii) Artefactual components were detected via visual inspections and mutual information (MI) analysis and then removed from the single-trial data (Abbasi et al., 2016 [↗](#)). (iv) Finally, all remaining components were back-transformed to the sensor level.

## Source localization

For source localisation we aligned individual T1-weighted anatomical MRI scans with the digitised head shapes using the iterative closest point algorithm. Then, we segmented the MRI scans and generated single-shell volume conductor models (Nolte, 2003 [↗](#)), and used this to create forward models. For group analyses, individual MRIs were linearly transformed to a MNI template provided by Fieldtrip. Next, the linearly constrained minimum variance (LCMV) algorithm was used to compute time-series for each voxel on a 5-mm grid. The time-series were extracted for each dipole orientation, resulting in three time-series per voxel. The reduced version of the HCP brain atlas as well as AAL atlas were applied on the source space time-series in order to reduce the dimensionality of the data, resulting in 230 cortical parcels (Tait et al., 2020 [↗](#)) and 116 subcortical parcels respectively. Since the HCP atlas only covers cortical areas we used the AAL atlas for subcortical areas. Finally, we extracted the first three components of a singular value decomposition (SVD) of time-series from all dipoles in this parcel, explaining most of the variance.

## ROI selection

In this study, we focused on the brain regions that are involved in speech production and perception. In order to find the areas involved in speech production and perception network, we utilized the automatic meta-analysis provided by *neurosynth.org* [↗](#) using the term ‘speech’ that resulted in a meta-analysis of 642 fMRI studies highlighting the brain areas involved in speech production and perception. We extracted the MNI coordinates of active areas from the presented map, found their corresponding voxels and identified the respective parcels on HCP and AAL atlases where these voxels are located. This resulted in 14 cortical and subcortical parcels (see **Fig. 1** [↗](#)).

## Connectivity analysis

We performed connectivity analysis by using a multivariate nonparametric Granger causality approach (mCG; (Schaum et al., 2021 [DOI](#))). We computed the mCG to determine the directionality of functional coupling between all the detected involved parcels, in pairwise steps, during speech production and perception. Initially, the source signals were divided into trials of four seconds, with 500 milliseconds overlap. We used the fast Fourier transform in combination with multitapers (2 Hz smoothing) to compute the cross-spectral density (CSD) matrix of the trials. Next, using a blockwise approach, we considered the first three SVD components of each parcel as a block and estimated the connectivity between STG and other parcels. Finally, we computed the directed influence asymmetry index (DAI) defined by (Bastos et al., 2015 [DOI](#)) as

$$DAI = \frac{mGC(parcel \rightarrow STG) - mGC(STG \rightarrow parcel)}{mGC(parcel \rightarrow STG) + mGC(STG \rightarrow parcel)}$$

Therefore, a positive DAI for a given frequency indicates that the selected parcel conveys feedforward influences to STG in this frequency, and a negative DAI indicates feedback influences. Note that for the connectivity analysis, we used MEG data with 1200 Hz sampling rate without downsampling.

## Statistical analysis

We determined significant connectivity patterns (DAI values) in both speaking and listening conditions using non-parametric cluster-based permutation tests (Maris and Oostenveld, 2007 [DOI](#)). First, we estimated the statistical contrast of connectivities during speaking and listening compared to zero for each parcel and participant. Second, the DAI values in the speaking condition were contrasted with DAI values in the listening condition at the group level. The statistical analysis was conducted for different frequency bands (Delta/Theta (0-7 Hz), alpha (7-13 Hz), beta (15-30 Hz), gamma (30-60 Hz), high gamma (60-90 Hz)) using a dependent-samples t-test. We used a cluster-based correction to account for multiple comparisons across frequencies and parcels. We performed five thousand permutations and set the critical alpha value at 0.05.

## Correlation analysis

To examine the potential relationship between our connectivity findings from all the parcels to L-CB6 and speech-STG coupling, we conducted non-parametric cluster-based permutation tests. Our first step was to calculate top-down connectivity values for selected parcels and frequency bands, then to compute speech-STG couplings for each frequency band. To account for multiple comparisons across parcels, we employed the Pearson method implemented in the `ft_statfun_correlationT` function in Fieldtrip, with cluster-based correction. Our analysis was repeated for different frequency bands. Therefore, our results are not corrected across frequencies. We performed five thousand permutations and set the critical alpha value at 0.05.

## Data availability

An example data is available at [https://osf.io/9fq47/?view\\_only=e6bad57efb854f069474c2d6f93a00a6](https://osf.io/9fq47/?view_only=e6bad57efb854f069474c2d6f93a00a6) [DOI](#). Raw data, however, are protected by data privacy laws and cannot be made widely available, but may be made available upon reasonable request (subject to these privacy laws).

## References

- Abbasi O., Gross J. (2020) **Beta-band oscillations play an essential role in motor-auditory interactions** *Hum. Brain Mapp* **41**:656–665 <https://doi.org/10.1002/hbm.24830>
- Abbasi O., Hirschmann J., Schmitz G., Schnitzler A., Butz M. (2016) **Rejecting deep brain stimulation artefacts from MEG data using ICA and mutual information** *J. Neurosci. Methods* **268**:131–141 <https://doi.org/10.1016/j.jneumeth.2016.04.010>
- Abbasi O., Steingraber N., Chalas N., Kluger D.S., Gross J. (2023) **Spatiotemporal dynamics characterise spectral connectivity profiles of continuous speaking and listening** *PLoS Biol* **21** <https://doi.org/10.1371/journal.pbio.3002178>
- Abbasi O., Steingraber N., Gross J. (2021) **Correcting MEG artifacts caused by overt speech** *Front. Neurosci* **15** <https://doi.org/10.3389/fnins.2021.682419>
- Ackermann H., Wildgruber D., Daum I., Grodd W. (1998) **Does the cerebellum contribute to cognitive aspects of speech production? A functional magnetic resonance imaging (fMRI) study in humans** *Neurosci. Lett* **247**:187–190 [https://doi.org/10.1016/s0304-3940\(98\)00328-0](https://doi.org/10.1016/s0304-3940(98)00328-0)
- Ackermann H. (2008) **Cerebellar contributions to speech production and speech perception: psycholinguistic and neurobiological perspectives** *Trends Neurosci* **31**:265–272 <https://doi.org/10.1016/j.tins.2008.02.011>
- Alexandrou A.M., Saarinen T., Mäkelä S., Kujala J., Salmelin R. (2017) **The right hemisphere is highlighted in connected natural speech production and perception** *Neuroimage* **152**:628–638 <https://doi.org/10.1016/j.neuroimage.2017.03.006>
- Arnal L.H., Giraud A.-L. (2012) **Cortical oscillations and sensory predictions** *Trends Cogn Sci (Regul Ed)* **16**:390–398 <https://doi.org/10.1016/j.tics.2012.05.003>
- Barbas H., García-Cabezas M.Á., Zikopoulos B. (2013) **Frontal-thalamic circuits associated with language** *Brain Lang* **126**:49–61 <https://doi.org/10.1016/j.bandl.2012.10.001>
- Bastos A.M., Vezoli J., Bosman C.A., Schoffelen J.-M., Oostenveld R., Dowdall J.R., De Weerd P., Kennedy H., Fries P. (2015) **Visual areas exert feedforward and feedback influences through distinct frequency channels** *Neuron* **85**:390–401 <https://doi.org/10.1016/j.neuron.2014.12.018>
- Cao L., Thut G., Gross J. (2017) **The role of brain oscillations in predicting self-generated sounds** *Neuroimage* **147**:895–903 <https://doi.org/10.1016/j.neuroimage.2016.11.001>
- Chalas N., Daube C., Kluger D.S., Abbasi O., Nitsch R., Gross J. (2022) **Multivariate analysis of speech envelope tracking reveals coupling beyond auditory cortex** *Neuroimage* **258** <https://doi.org/10.1016/j.neuroimage.2022.119395>
- Crosson B. (2013) **Thalamic mechanisms in language: a reconsideration based on recent findings and concepts** *Brain Lang* **126**:73–88 <https://doi.org/10.1016/j.bandl.2012.06.011>

- Ferro D., van Kempen J., Boyd M., Panzeri S., Thiele A. (2021) **Directed information exchange between cortical layers in macaque V1 and V4 and its modulation by selective attention** *Proc Natl Acad Sci USA* **118** <https://doi.org/10.1073/pnas.2022097118>
- Floegel M., Fuchs S., Kell C.A. (2020) **Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control** *Nat. Commun* **11** <https://doi.org/10.1038/s41467-020-16743-2>
- Fontolan L., Morillon B., Liegeois-Chauvel C., Giraud A.-L. (2014) **The contribution of frequency-specific activity to hierarchical information processing in the human auditory cortex** *Nat. Commun* **5** <https://doi.org/10.1038/ncomms5694>
- Friston K.J., Frith C.D. (2015) **Active inference, communication and hermeneutics** *Cortex* **68**:129–143 <https://doi.org/10.1016/j.cortex.2015.03.025>
- Ganushchak L.Y., Christoffels I.K., Schiller N.O. (2011) **The use of electroencephalography in language production research: a review** *Front. Psychol* **2** <https://doi.org/10.3389/fpsyg.2011.00208>
- Giraud A.-L., Kleinschmidt A., Poeppel D., Lund T.E., Frackowiak R.S.J., Laufs H. (2007) **Endogenous cortical rhythms determine cerebral specialization for speech perception and production** *Neuron* **56**:1127–1134 <https://doi.org/10.1016/j.neuron.2007.09.038>
- Giraud A.-L., Poeppel D. (2012) **Cortical oscillations and speech processing: emerging computational principles and operations** *Nat. Neurosci* **15**:511–517 <https://doi.org/10.1038/nn.3063>
- Gross J. *et al.* (2013) **Good practice for conducting and reporting MEG research** *Neuroimage* **65**:349–363 <https://doi.org/10.1016/j.neuroimage.2012.10.001>
- Gross J., Hoogenboom N., Thut G., Schyns P., Panzeri S., Belin P., Garrod S. (2013) **Speech rhythms and multiplexed oscillatory sensory coding in the human brain** *PLoS Biol* **11** <https://doi.org/10.1371/journal.pbio.1001752>
- Janssen N., van der Meij M., López-Pérez P.J., Barber H.A. (2020) **Exploring the temporal dynamics of speech production with EEG and group ICA** *Sci. Rep* **10** <https://doi.org/10.1038/s41598-020-60301-1>
- Kotz S.A., Stockert A., Schwartze M. (2014) **Cerebellum, temporal predictability and the updating of a mental model** *Philos. Trans. R. Soc. Lond. B Biol. Sci* **369** <https://doi.org/10.1098/rstb.2013.0403>
- Lakatos P., Gross J., Thut G. (2019) **A new unifying account of the roles of neuronal entrainment** *Curr. Biol* **29**:R890–R905 <https://doi.org/10.1016/j.cub.2019.07.075>
- Llorens A., Trébuchon A., Liégeois-Chauvel C., Alario F.-X. (2011) **Intra-cranial recordings of brain activity during language production** *Front. Psychol* **2** <https://doi.org/10.3389/fpsyg.2011.00375>
- Manto M. *et al.* (2012) **Consensus paper: roles of the cerebellum in motor control--the diversity of ideas on cerebellar involvement in movement** *Cerebellum* **11**:457–487 <https://doi.org/10.1007/s12311-011-0331-9>

- Maris E., Oostenveld R. (2007) **Nonparametric statistical testing of EEG- and MEG-data** *J. Neurosci. Methods* **164**:177–190 <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Martikainen M.H., Kaneko K., Hari R. (2005) **Suppressed responses to self-triggered sounds in the human auditory cortex** *Cereb. Cortex* **15**:299–302 <https://doi.org/10.1093/cercor/bhh131>
- Nolte G. (2003) **The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors** *Phys. Med. Biol* **48**:3637–3652 <https://doi.org/10.1088/0031-9155/48/22/002>
- Oostenveld R., Fries P., Maris E., Schoffelen J.-M. (2011) **FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data** *Comput. Intell. Neurosci* <https://doi.org/10.1155/2011/156869>
- Ozker M., Doyle W., Devinsky O., Flinker A. (2022) **A cortical network processes auditory error signals during human speech production to maintain fluency** *PLoS Biol* **20** <https://doi.org/10.1371/journal.pbio.3001493>
- Park H., Ince R.A.A., Schyns P.G., Thut G., Gross J. (2015) **Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners** *Curr. Biol* **25**:1649–1653 <https://doi.org/10.1016/j.cub.2015.04.049>
- Parrell B., Agnew Z., Nagarajan S., Houde J., Ivry R.B. (2017) **Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration** *J. Neurosci* **37**:9249–9258 <https://doi.org/10.1523/JNEUROSCI.3363-16.2017>
- Riès S.K. *et al.* (2017) **Spatiotemporal dynamics of word retrieval in speech production revealed by cortical high-frequency band activity** *Proc Natl Acad Sci USA* **114**:E4530–E4538 <https://doi.org/10.1073/pnas.1620669114>
- Rubinov M., Sporns O. (2010) **Complex network measures of brain connectivity: uses and interpretations** *Neuroimage* **52**:1059–1069 <https://doi.org/10.1016/j.neuroimage.2009.10.003>
- Ruspantini I., Saarinen T., Belardinelli P., Jalava A., Parviainen T., Kujala J., Salmelin R. (2012) **Corticomuscular coherence is tuned to the spontaneous rhythmicity of speech at 2-3 Hz** *J. Neurosci* **32**:3786–3790 <https://doi.org/10.1523/JNEUROSCI.3191-11.2012>
- Schaum M., Pinzuti E., Sebastian A., Lieb K., Fries P., Mobascher A., Jung P., Wibral M., Tüscher O. (2021) **Right inferior frontal gyrus implements motor inhibitory control via beta-band oscillations in humans** *eLife* **10** <https://doi.org/10.7554/eLife.61679>
- Schneider M., Broggin A.C., Dann B., Tzanou A., Uran C., Sheshadri S., Scherberger H., Vinck M. (2021) **A mechanism for inter-areal coherence through communication based on connectivity and oscillatory power** *Neuron* **109**:4050–4067 <https://doi.org/10.1016/j.neuron.2021.09.037>
- Sedley W., Gander P.E., Kumar S., Kovach C.K., Oya H., Kawasaki H., Howard M.A., Griffiths T.D. (2016) **Neural signatures of perceptual inference** *eLife* **5** <https://doi.org/10.7554/eLife.11476>
- Silbert L.J., Honey C.J., Simony E., Poeppel D., Hasson U. (2014) **Coupled neural systems underlie the production and comprehension of naturalistic narrative speech** *Proc Natl Acad Sci USA* **111**:E4687–96 <https://doi.org/10.1073/pnas.1323812111>

Silveri M.C. (2021) **Contribution of the Cerebellum and the Basal Ganglia to Language Production: Speech, Word Fluency, and Sentence Construction-Evidence from Pathology** *Cerebellum* **20**:282–294 <https://doi.org/10.1007/s12311-020-01207-6>

Skipper J.I., Lametti D.R. (2021) **Speech Perception under the Tent: A Domain-general Predictive Role for the Cerebellum** *J. Cogn. Neurosci* **33**:1517–1534 [https://doi.org/10.1162/jocn\\_a\\_01729](https://doi.org/10.1162/jocn_a_01729)

Stockert A., Schwartze M., Poeppel D., Anwender A., Kotz S.A. (2021) **Temporo-cerebellar connectivity underlies timing constraints in audition** *eLife* **10** <https://doi.org/10.7554/eLife.67303>

Stolk A., Todorovic A., Schoffelen J.-M., Oostenveld R. (2013) **Online and offline tools for head movement compensation in MEG** *Neuroimage* **68**:39–48 <https://doi.org/10.1016/j.neuroimage.2012.11.047>

Tait L., Özkan A., Szul M.J., Zhang J. (2020) **Cortical source imaging of resting-state MEG with a high resolution atlas: An evaluation of methods** *BioRxiv* <https://doi.org/10.1101/2020.01.12.903302>

Todorović S., Anton J.-L., Sein J., Nazarian B., Chanoine V., Rauchbauer B., Kotz S.A., Runnqvist E. (2023) **Cortico-Cerebellar Monitoring of Speech Sequence Production** *Neurobiology of Language* :1–21 [https://doi.org/10.1162/nol\\_a\\_00113](https://doi.org/10.1162/nol_a_00113)

Vinck M., Uran C., Spyropoulos G., Onorato I., Broggin A.C., Schneider M., Canales-Johnson A. (2023) **Principles of large-scale neural interactions** *Neuron* **111**:987–1002 <https://doi.org/10.1016/j.neuron.2023.03.015>

## Editors

Reviewing Editor

**Andrea Martin**

Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

Senior Editor

**Floris de Lange**

Donders Institute for Brain, Cognition and Behaviour, Nijmegen, Netherlands

## Reviewer #1 (Public Review):

Abbasi et al. assess in this MEG study the directed connectivity of both cortical and subcortical regions during continuous speech production and perception. The authors observed bidirectional connectivity patterns between speech-related cortical areas as well as subcortical areas in production and perception. Interestingly, they found in speaking low-frequency connectivity from subcortical (the right cerebellum) to cortical (left superior temporal) areas, while connectivity from the cortical to subcortical areas was in the high frequencies. In listening a similar cortico-subcortical connectivity pattern was observed for the low frequencies, but the reversed connectivity in the higher frequencies was absent.

The work by Abbasi and colleagues addresses a relevant, novel topic, namely understanding the brain dynamics between speaking and listening. This is important because traditionally production and perception of speech and language are investigated in a modality-specific manner. To have a more complete understanding of the neurobiology underlying these different speech behaviors, it is key to also understand their similarities and differences.

Furthermore, to do so, the authors utilize state-of-the-art directed connectivity analyses on MEG measurements, providing a quite detailed profile of cortical and subcortical interactions for the production and perception of speech. Importantly, and perhaps most interesting in my opinion, is that the authors find evidence for frequency-specific directed connectivity, which is (partially) different between speaking and listening. This could suggest that both speech behaviors rely (to some extent) on similar cortico-cortical and cortico-subcortical networks, but different frequency-specific dynamics.

These elements mentioned above (investigation of both production and perception, both cortico-cortical and cortico-subcortical connectivity is considered, and observing frequency-specific connectivity profiles within and between speech behaviors), make for important novel contributions to the field. Notwithstanding these strengths, I find that they are especially centered on methodology and functional anatomical description, but that precise theoretical contributions for neurobiological and cognitive models of speech are less transparent. This is in part because the study compares speech production and perception in general, but no psychophysical or psycholinguistic manipulations are considered. I also have some critical questions about the design which may pose some confounds in interpreting the data, especially with regard to comparing production and perception.

(1) While the cortico-cortical and cortico-subcortical connectivity profiles highlighted in this study and the depth of the analyses are impressive, what these data mean for models of speech processing remains on the surface. This is in part due, I believe, to the fact that the authors have decided to explore speaking and listening in general, without targeting specific manipulations that help elucidate which aspects of speech processing are relevant for the particular connectivity profiles they have uncovered. For example, the frequency-specific directed connectivity is it driven by low-level psychophysical attributes of the speech or by more cognitive linguistic properties? Does it relate to the monitoring of speech, timing information, and updating of sensory predictions? Without manipulations trying to target one or several of these components, as some of the referenced work has done (e.g., Floegel et al., 2020; Stockert et al., 2021; Todorović et al., 2023), it is difficult to draw concrete conclusions as to which representations and/or processes of speech are reflected by the connectivity profiles. An additional disadvantage of not having manipulations within each speech behavior is that it makes the comparison between listening and speaking harder. That is, speaking and listening have marked input-output differences which likely will dominate any comparison between them. These physically driven differences (or similarities for that matter; see below) can be strongly reduced by instead exploring the same manipulations/variables between speaking and listening. If possible (if not to consider for future work), it may be interesting to score psychophysical (e.g., acoustic properties) or psycholinguistic (e.g., lexical frequency) information of the speech and see whether and how the frequency-specific connectivity profiles are affected by it.

(2) Recent studies comparing the production and perception of language may be relevant to the current study and add some theoretical weight since their data and interpretations for the comparisons between production and perception fit quite well with the observations in the current work. These studies highlight that language processes between production and perception, specifically lexical and phonetic processing (Fairs et al., 2021), and syntactic processing (Giglio et al., 2024), may rely on the same neural representations, but are differentiated in their (temporal) dynamics upon those shared representations. This is relevant because it dispenses with the classical notion in neurobiological models of language where production and perception rely on (partially) dissociable networks (e.g., Price, 2010). Rather those data suggest shared networks where different language behaviors are dissociated in their dynamics. The speech results in this study nicely fit and extend those studies and their theoretical implications.

(3) The authors align the frequency-selective connectivity between the right cerebellum and left temporal speech areas with recent studies demonstrating a role for the right cerebellum for the internal modelling in speech production and monitoring (e.g., Stockert et al., 2021; Todorović et al., 2023). This link is indeed interesting, but it does seem relevant to point out that at a more specific scale, it does not concern the exact same regions between those studies and the current study. That is, in the current study the frequency-specific connectivity with temporal regions concerns lobule VI in the right cerebellum, while in the referenced work it concerns Crus I/II. The distinction seems relevant since Crus I/II has been linked to the internal modelling of more cognitive behavior, while lobule VI seems more motor-related and/or contextual-related (e.g., D'Mello et al., 2020; Runnqvist et al., 2021; Runnqvist, 2023).

(4) On the methodological side, my main concern is that for the listening condition, the authors have chosen to play back the speech produced by the participants in the production condition. Both the fixed order as well as hearing one's own speech as listening condition may produce confounds in data interpretation, especially with regard to the comparison between speech production and perception. Could order effects impact the observed connectivity profiles, and how would this impact the comparison between speaking and listening? In particular, I am thinking of repetition effects present in the listening condition as well as prediction, which will be much more elevated for the listening condition than the speaking condition. The fact that it also concerns their own voice furthermore adds to the possible predictability confound (e.g., Heinks-Maldonado et al., 2005). In addition, listening to one's speech which just before has been articulated may, potentially strategically even, enhance inner speech and "mouthing" in the participants, hereby thus engaging the production mechanism. Similarly, during production, the participants already hear their own voice (which serves as input in the subsequent listening condition). Taken together, both similarities or differences between speaking and listening connectivity may have been due to or influenced by these order effects, and the fact that the different speech behaviors are to some extent present in both conditions.

(5) The ability of the authors to analyze the spatiotemporal dynamics during continuous speech is a potentially important feat of this study, given that one of the reasons that speech production is much less investigated compared to perception concerns motor and movement artifacts due to articulation (e.g., Strijkers et al., 2010). Two questions did spring to mind when reading the authors' articulation artifact correction procedure: If I understood correctly, the approach comes from Abbasi et al. (2021) and is based on signal space projection (SSP) as used for eye movement corrections, which the authors successfully applied to speech production. However, in that study, it concerned the repeated production of three syllables, while here it concerns continuous speech of full words embedded in discourse. The articulation and muscular variance will be much higher in the current study compared to three syllables (or compared to eye movements which produce much more stable movement potentials compared to an entire discourse). Given this, I can imagine that corrections of the signal in the speaking condition were likely substantial and one may wonder (1) how much signal relevant to speech production behavior is lost?; (2) similar corrections are not necessary for perception, so how would this marked difference in signal processing affect the comparability between the modalities?

#### References:

- Abbasi, O., Steingraber, N., & Gross, J. (2021). Correcting MEG artifacts caused by overt speech. *Frontiers in Neuroscience*, 15, 682419.
- D'Mello, A. M., Gabrieli, J. D., & Nee, D. E. (2020). Evidence for hierarchical cognitive control in the human cerebellum. *Current Biology*, 30(10), 1881-1892.
- Fairs, A., Michelas, A., Dufour, S., & Strijkers, K. (2021). The same ultra-rapid parallel brain dynamics underpin the production and perception of speech. *Cerebral Cortex Communications*, 2(3), tgab040.

- Floegel, M., Fuchs, S., & Kell, C. A. (2020). Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. *Nature Communications*, 11(1), 2839.
- Giglio, L., Ostarek, M., Sharoh, D., & Hagoort, P. (2024). Diverging neural dynamics for syntactic structure building in naturalistic speaking and listening. *Proceedings of the National Academy of Sciences*, 121(11), e2310766121.
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42(2), 180-190.
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the new York Academy of Sciences*, 1191(1), 62-88.
- Runnqvist, E., Chanoine, V., Strijkers, K., Pattamadilok, C., Bonnard, M., Nazarian, B., ... & Alario, F. X. (2021). Cerebellar and cortical correlates of internal and external speech error monitoring. *Cerebral Cortex Communications*, 2(2), tgab038.
- Runnqvist, E. (2023). Self-monitoring: The neurocognitive basis of error monitoring in language production. In *Language production* (pp. 168-190). Routledge.
- Stockert, A., Schwartz, M., Poeppel, D., Anwander, A., & Kotz, S. A. (2021). Temporo-cerebellar connectivity underlies timing constraints in audition. *Elife*, 10, e67303.
- Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral cortex*, 20(4), 912-928.
- Todorović, S., Anton, J. L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., ... & Runnqvist, E. (2023). Cortico-cerebellar monitoring of speech sequence production. *Neurobiology of Language*, 1-21.

<https://doi.org/10.7554/eLife.97083.1.sa2>

## Reviewer #2 (Public Review):

### Summary:

The authors re-analyse MEG data from a speech production and perception study and extend their previous Granger causality analysis to a larger number of cortical-cortical and in particular cortical-subcortical connections. Regions of interest were defined by means of a meta-analysis using Neurosynth.org and connectivity patterns were determined by calculating directed influence asymmetry indices from the Granger causality analysis results for each pair of brain regions. Abbasi et al. report feedforward signals communicated via fast rhythms and feedback signals via slow rhythms below 40 Hz, particularly during speaking. The authors highlight one of these connections between the right cerebellum lobule VI and auditory association area A5, where in addition the connection strength correlates negatively with the strength of speech tracking in the theta band during speaking (significant before multiple comparison correction). Results are interpreted within a framework of active inference by minimising prediction errors.

While I find investigating the role of cortical-subcortical connections in speech production and perception interesting and relevant to the field, I am not yet convinced that the methods employed are fully suitable to this endeavour or that the results provide sufficient evidence to make the strong claim of dissociation of bottom-up and top-down information flow during speaking in distinct frequency bands.

### Strengths:

The investigation of electrophysiological cortical-subcortical connections in speech production and perception is interesting and relevant to the field. The authors analyse a valuable dataset, where they spent a considerable amount of effort to correct for speech production-related artefacts. Overall, the manuscript is well-written and clearly structured.

## Weaknesses:

The description of the multivariate Granger causality analysis did not allow me to fully grasp how the analysis was performed and I hence struggled to evaluate its appropriateness. Knowing that (1) filtered Granger causality is prone to false positives and (2) recent work demonstrates that significant Granger causality can simply arise from frequency-specific activity being present in the source but not the target area without functional relevance for communication (Schneider et al. 2021) raises doubts about the validity of the results, in particular with respect to their frequency specificity. These doubts are reinforced by what I perceive as an overemphasis on results that support the assumption of specific frequencies for feedforward and top-down connections, while findings not aligning with this hypothesis appear to be underreported. Furthermore, the authors report some main findings that I found difficult to reconcile with the data presented in the figures. Overall, I feel the conclusions with respect to frequency-specific bottom-up and top-down information flow need to be moderated and that some of the reported findings need to be checked and if necessary corrected.

## Major points

(1) I think more details on the multivariate GC approach are needed. I found the reference to Schaum et al., 2021 not sufficient to understand what has been done in this paper. Some questions that remained for me are:

(i) Does multivariate here refer to the use of the authors' three components per parcel or to the conditioning on the remaining twelve sources? I think the latter is implied when citing Schaum et al., but I'm not sure this is what was done here?

If it was not: how can we account for spurious results based on indirect effects?

(ii) Did the authors check whether the GC of the course-target pairs was reliably above the bias level (as Schaum et. al. did for each condition separately)? If not, can they argue why they think that their results would still be valid? Does it make sense to compute DAIs on connections that were below the bias level? Should the data be re-analysed to take this concern into account?

(iii) You may consider citing the paper that introduced the non-parametric GC analysis (which Schaum et al. then went on to apply): Dhamala M, Rangarajan G, Ding M. Analyzing Information Flow in Brain Networks with Nonparametric Granger Causality. *Neuroimage*. 2008; 41(2):354-362. <https://doi.org/10.1016/j.neuroimage.2008.02.020>

(2) GC has been discouraged for filtered data as it gives rise to false positives due to phase distortions and the ineffectiveness of filtering in the information-theoretic setting as reducing the power of a signal does not reduce the information contained in it (Florin et al., 2010; Barnett and Seth, 2011; Weber et al. 2017; Pinzuti et al., 2020 - who also suggest an approach that would circumvent those filter-related issues). With this in mind, I am wondering whether the strong frequency-specific claims in this work still hold.

(3) I found it difficult to reconcile some statements in the manuscript with the data presented in the figures:

(i) Most notably, the considerable number of feedforward connections from A5 and STS that project to areas further up the hierarchy at slower rhythms (e.g. L-A5 to R-PEF, R-Crus2, L CB6 L-Tha, L-FOP and L-STs to R-PEF, L-FOP, L-TOPJ or R-A5 as well as R-STs both to R-Crus2, L-CB6, L-Th) contradict the authors' main message that 'feedback signals were communicated via slow rhythms below 40 Hz, whereas feedforward signals were communicated via faster

rhythms'. I struggled to recognise a principled approach that determined which connections were highlighted and reported and which ones were not.

(ii) "Our analysis also revealed robust connectivity between the right cerebellum and the left parietal cortex, evident in both speaking and listening conditions, with stronger connectivity observed during speaking. Notably, Figure 4 depicts a prominent frequency peak in the alpha band, illustrating the specific frequency range through which information flows from the cerebellum to the parietal areas." There are two peaks discernible in Figure 4, one notably lower than the alpha band (rather theta or even delta), the other at around 30 Hz. Nevertheless, the authors report and discuss a peak in the alpha band.

(iii) In the abstract: "Notably, high-frequency connectivity was absent during the listening condition." and p.9 "In contrast with what we reported for the speaking condition, during listening, there is only a significant connectivity in low frequency to the left temporal area but not a reverse connection in the high frequencies."

While Fig. 4 shows significant connectivity from R-CB6 to A5 in the gamma frequency range for the speaking, but not for the listening condition, interpreting comparisons between two effects without directly comparing them is a common statistical mistake (Makin and Orban de Xivry). The spectrally-resolved connectivity in the two conditions actually look remarkably similar and I would thus refrain from highlighting this statement and indicate clearly that there were no significant differences between the two conditions.

(iv) "This result indicates that in low frequencies, the sensory-motor area and cerebellum predominantly transmit information, while in higher frequencies, they are more involved in receiving it."

I don't think that this statement holds in its generality: L-CB6 and R-3b both show strong output at high frequencies, particularly in the speaking condition. While they seem to transmit information mainly to areas outside A5 and STS these effects are strong and should be discussed.

(4) "However, definitive conclusions should be drawn with caution given recent studies raising concerns about the notion that top-down and bottom-up signals can only be transmitted via separate frequency channels (Ferro et al., 2021; Schneider et al., 2021; Vinck et al., 2023)."

I appreciate this note of caution and think it would be useful if it were spelled out to the reader why this is the case so that they would be better able to grasp the main concerns here. For example, Schneider et al. make a strong point that we expect to find Granger-causality with a peak in a specific frequency band for areas that are anatomically connected when the sending area shows stronger activity in that band than the receiving one, simply because of the coherence of a signal with its own linear projection onto the other area. The direction of a Granger causal connection would in that case only indicate that one area shows stronger activity than the other in the given frequency band. I am wondering to what degree the reported connectivity pattern can be traced back to regional differences in frequency-specific source strength or to differences in source strength across the two conditions.

<https://doi.org/10.7554/eLife.97083.1.sa1>

### Reviewer #3 (Public Review):

In the current paper, Abbasi et al. aimed to characterize and compare the patterns of functional connectivity across frequency bands (1 Hz - 90 Hz) between regions of a speech network derived from an online meta-analysis tool (Neurosynth.org) during speech production and perception. The authors present evidence for complex neural dynamics from which they highlight directional connectivity from the right cerebellum to left superior

temporal areas in lower frequency bands (up to beta) and between the same regions in the opposite direction in the (lower) high gamma range (60-90 Hz). Abbasi et al. interpret their findings within the predictive coding framework, with the cerebellum and other "higher-order" (motor) regions transmitting top-down sensory predictions to "lower-order" (sensory) regions in the lower frequencies and prediction errors flowing in the opposite direction (i.e., bottom-up) from those sensory regions in the gamma band. They also report a negative correlation between the strength of this top-down functional connectivity and the alignment of superior temporal regions to the syllable rate of one's speech.

#### Strengths:

- (1) The comprehensive characterization of functional connectivity during speaking and listening to speech may be valuable as a first step toward understanding the neural dynamics involved.
- (2) The inclusion of subcortical regions and connectivity profiles up to 90Hz using MEG is interesting and relatively novel.
- (3) The analysis pipeline is generally adequate for the exploratory nature of the work.

#### Weaknesses:

- (1) The work is framed as a test of the predictive coding theory as it applies to speech production and perception, but the methodological approach is not suited to this endeavor.
- (2) Because of their theoretical framework, the authors readily attribute roles or hierarchy to brain regions (e.g., higher- vs lower-order) and cognitive functions to observed connectivity patterns (e.g., feedforward vs feedback, predictions vs prediction errors) that cannot be determined from the data. Thus, many of the authors' claims are unsupported.
- (3) The authors' theoretical stance seems to influence the presentation of the results, which may inadvertently misrepresent the (otherwise perfectly valid; cf. Abbasi et al., 2023) exploratory nature of the study. Thus, results about specific regions are often highlighted in figures (e.g., Figure 2 top row) and text without clear reasons.
- (4) Some of the key findings (e.g., connectivity in opposite directions in distinct frequency bands) feature in a previous publication and are, therefore, interesting but not novel.
- (5) The quantitative comparison between speech production and perception is interesting but insufficiently motivated.
- (6) Details about the Neurosynth meta-analysis and subsequent selection of brain regions for the functional connectivity analyses are incomplete. Moreover, the use of the term 'Speech' in Neurosynth seems inappropriate (i.e., includes irrelevant works, yielding questionable results). The approach of using separate meta-analyses for 'Speech production' and 'Speech perception' taken by Abbasi et al. (2023) seems more principled. This approach would result, for example, in the inclusion of brain areas such as M1 and the BG that are relevant for speech production.
- (7) The results involving subcortical regions are central to the paper, but no steps are taken to address the challenges involved in the analysis of subcortical activity using MEG. Additional methodological detail and analyses would be required to make these results more compelling. For example, it would be important to know what the coverage of the MEG system is, what head model was used for the source localization of cerebellar activity, and if specific preprocessing or additional analyses were performed to ensure that the localized subcortical activity (in particular) is valid.

(8) The results and methods are often detailed with important omissions (a speech-brain coupling analysis section is missing) and imprecisions (e.g., re: Figure 5; the Connectivity Analysis section is copy-pasted from their previous work), which makes it difficult to understand what is being examined and how. (It is also not good practice to refer the reader to previous publications for basic methodological details, for example, about the experimental paradigm and key analyses.) Conversely, some methodological details are given, e.g., the acquisition of EMG data, without further explanation of how those data were used in the current paper.

(9) The examination of gamma functional connectivity in the 60 - 90 Hz range could be better motivated. Although some citations involving short-range connectivity in these frequencies are given (e.g., within the visual system), a more compelling argument for looking at this frequency range for longer-range connectivity may be required.

(10) The choice of source localization method (linearly constrained minimum variance) could be explained, particularly given that other methods (e.g. dynamic imaging of coherent sources) were specifically designed and might potentially be a better alternative for the types of analyses performed in the study.

(11) The mGC analysis needs to be more comprehensively detailed for the reader to be able to assess what is being reported and the strength of the evidence. Relatedly, first-level statistics (e.g., via estimation of the noise level) would make the mGC and DAI results more compelling.

(12) Considering the exploratory nature of the study, it is essential for other researchers to continue investigating and validating the results presented in the current manuscript. Thus, it is concerning that data and scripts are not fully and openly available. Data need not be in its raw state to be shared and useful, which circumvents the stated data privacy concerns.

<https://doi.org/10.7554/eLife.97083.1.sa0>

#### Author response:

##### **Reviewer #1 (Public Review):**

*Abbasi et al. assess in this MEG study the directed connectivity of both cortical and subcortical regions during continuous speech production and perception. The authors observed bidirectional connectivity patterns between speech-related cortical areas as well as subcortical areas in production and perception. Interestingly, they found in speaking low-frequency connectivity from subcortical (the right cerebellum) to cortical (left superior temporal) areas, while connectivity from the cortical to subcortical areas was in the high frequencies. In listening a similar cortico-subcortical connectivity pattern was observed for the low frequencies, but the reversed connectivity in the higher frequencies was absent.*

*The work by Abbasi and colleagues addresses a relevant, novel topic, namely understanding the brain dynamics between speaking and listening. This is important because traditionally production and perception of speech and language are investigated in a modality-specific manner. To have a more complete understanding of the neurobiology underlying these different speech behaviors, it is key to also understand their similarities and differences. Furthermore, to do so, the authors utilize state-of-the-art directed connectivity analyses on MEG measurements, providing a quite detailed profile of cortical and subcortical interactions for the production and perception of speech. Importantly, and perhaps most interesting in my opinion, is that the authors find evidence for frequency-specific directed connectivity, which is (partially) different between*

*speaking and listening. This could suggest that both speech behaviors rely (to some extent) on similar cortico-cortical and cortico-subcortical networks, but different frequency-specific dynamics.*

*These elements mentioned above (investigation of both production and perception, both cortico-cortical and cortico-subcortical connectivity is considered, and observing frequency-specific connectivity profiles within and between speech behaviors), make for important novel contributions to the field. Notwithstanding these strengths, I find that they are especially centered on methodology and functional anatomical description, but that precise theoretical contributions for neurobiological and cognitive models of speech are less transparent. This is in part because the study compares speech production and perception in general, but no psychophysical or psycholinguistic manipulations are considered. I also have some critical questions about the design which may pose some confounds in interpreting the data, especially with regard to comparing production and perception.*

*(1) While the cortico-cortical and cortico-subcortical connectivity profiles highlighted in this study and the depth of the analyses are impressive, what these data mean for models of speech processing remains on the surface. This is in part due, I believe, to the fact that the authors have decided to explore speaking and listening in general, without targeting specific manipulations that help elucidate which aspects of speech processing are relevant for the particular connectivity profiles they have uncovered. For example, the frequency-specific directed connectivity is it driven by low-level psychophysical attributes of the speech or by more cognitive linguistic properties? Does it relate to the monitoring of speech, timing information, and updating of sensory predictions? Without manipulations trying to target one or several of these components, as some of the referenced work has done (e.g., Floegel et al., 2020; Stockert et al., 2021; Todorović et al., 2023), it is difficult to draw concrete conclusions as to which representations and/or processes of speech are reflected by the connectivity profiles. An additional disadvantage of not having manipulations within each speech behavior is that it makes the comparison between listening and speaking harder. That is, speaking and listening have marked input-output differences which likely will dominate any comparison between them. These physically driven differences (or similarities for that matter; see below) can be strongly reduced by instead exploring the same manipulations/variables between speaking and listening. If possible (if not to consider for future work), it may be interesting to score psychophysical (e.g., acoustic properties) or psycholinguistic (e.g., lexical frequency) information of the speech and see whether and how the frequency-specific connectivity profiles are affected by it.*

We thank the reviewer for pointing this out. The current study is indeed part of a larger project investigating the role of the internal forward model in speech perception and production. In the original, more comprehensive study, we also included a masked condition where participants produced speech as usual, but their auditory perception was masked. This allowed us to examine how the internal forward model behaves when it doesn't receive the expected sensory consequences of generated speech. However, for the current study, we focused solely on data from the speaking and listening conditions due to its specific research question. We agree that further manipulations would be interesting. However, for this study our focus was on natural speech and we avoided other manipulations (beyond masked speech) so that we can have sufficiently long recording time for the main speaking and listening conditions.

*(2) Recent studies comparing the production and perception of language may be relevant to the current study and add some theoretical weight since their data and interpretations for the comparisons between production and perception fit quite well with the observations in the current work. These studies highlight that language*

*processes between production and perception, specifically lexical and phonetic processing (Fairs et al., 2021), and syntactic processing (Giglio et al., 2024), may rely on the same neural representations, but are differentiated in their (temporal) dynamics upon those shared representations. This is relevant because it dispenses with the classical notion in neurobiological models of language where production and perception rely on (partially) dissociable networks (e.g., Price, 2010). Rather those data suggest shared networks where different language behaviors are dissociated in their dynamics. The speech results in this study nicely fit and extend those studies and their theoretical implications.*

We thank the reviewer for the suggestion and we will include these references and the points made by the reviewer in our revised manuscript.

*(3) The authors align the frequency-selective connectivity between the right cerebellum and left temporal speech areas with recent studies demonstrating a role for the right cerebellum for the internal modelling in speech production and monitoring (e.g., Stockert et al., 2021; Todorović et al., 2023). This link is indeed interesting, but it does seem relevant to point out that at a more specific scale, it does not concern the exact same regions between those studies and the current study. That is, in the current study the frequency-specific connectivity with temporal regions concerns lobule VI in the right cerebellum, while in the referenced work it concerns Crus I/II. The distinction seems relevant since Crus I/II has been linked to the internal modelling of more cognitive behavior, while lobule VI seems more motor-related and/or contextual-related (e.g., D'Mello et al., 2020; Runqvist et al., 2021; Runqvist, 2023).*

We thank the reviewer for their insightful comment. The reference was intended to provide evidence for the role of the cerebellum in internal modelling in speech. We do not claim that we have the spatial resolution with MEG to reliably spatially resolve specific parts of the cerebellum.

*(4) On the methodological side, my main concern is that for the listening condition, the authors have chosen to play back the speech produced by the participants in the production condition. Both the fixed order as well as hearing one's own speech as listening condition may produce confounds in data interpretation, especially with regard to the comparison between speech production and perception. Could order effects impact the observed connectivity profiles, and how would this impact the comparison between speaking and listening? In particular, I am thinking of repetition effects present in the listening condition as well as prediction, which will be much more elevated for the listening condition than the speaking condition. The fact that it also concerns their own voice furthermore adds to the possible predictability confound (e.g., Heinks-Maldonado et al., 2005). In addition, listening to one's speech which just before has been articulated may, potentially strategically even, enhance inner speech and "mouthing" in the participants, hereby thus engaging the production mechanism. Similarly, during production, the participants already hear their own voice (which serves as input in the subsequent listening condition). Taken together, both similarities or differences between speaking and listening connectivity may have been due to or influenced by these order effects, and the fact that the different speech behaviors are to some extent present in both conditions.*

This is a valid point raised by the reviewer. By listening to their own previously produced speech, our participants might have anticipated and predicted the sentences easier. However, during designing our experiment, we tried to lower the chance of this anticipation by several steps. First, participants were measured in separate sessions for speech production and perception tasks. There were always several days' intervals between performing these two

conditions. Secondly, our questions were mainly about a common/general topic. Consequently, participants may not remember their answers completely.

Importantly, using the same stimulus material for speaking and listening guaranteed that there was no difference in the low-level features of the material for both conditions that could have affected the results of our statistical comparison.

Due to bone conduction, hearing one's unaltered own speech from a recording may seem foreign and could lead to unwanted emotional reactions e.g. embarrassment, so participants were asked whether they heard their own voice in a recording already (e.g. from a self-recorded voice-message in WhatsApp) which most of them confirmed. Participants were also informed that they were going to hear themselves during the measurement to further reduce unwanted psychophysiological responses.

*(5) The ability of the authors to analyze the spatiotemporal dynamics during continuous speech is a potentially important feat of this study, given that one of the reasons that speech production is much less investigated compared to perception concerns motor and movement artifacts due to articulation (e.g., Strijkers et al., 2010). Two questions did spring to mind when reading the authors' articulation artifact correction procedure: If I understood correctly, the approach comes from Abbasi et al. (2021) and is based on signal space projection (SSP) as used for eye movement corrections, which the authors successfully applied to speech production. However, in that study, it concerned the repeated production of three syllables, while here it concerns continuous speech of full words embedded in discourse. The articulation and muscular variance will be much higher in the current study compared to three syllables (or compared to eye movements which produce much more stable movement potentials compared to an entire discourse). Given this, I can imagine that corrections of the signal in the speaking condition were likely substantial and one may wonder (1) how much signal relevant to speech production behavior is lost?; (2) similar corrections are not necessary for perception, so how would this marked difference in signal processing affect the comparability between the modalities?*

One of the results of our previous study (Abbasi et al., 2021) was that the artefact correction was not specific to individual syllables but generalised across syllables. Also, the repeated production of syllables was associated with substantial movements of the articulators mimicking those observed during naturalistic speaking. We therefore believe that the artefact rejection is effective during speaking. We also checked this by investigating speech related coherence in brain parcels in spatial proximity to the articulators. In our previous study we also show that the correction method retains neural activity to a very large degree. We are therefore confident that speaking and listening conditions can be compared and that the loss of true signals from correcting the speaking data will be minor.

#### References:

- Abbasi, O., Steingraber, N., & Gross, J. (2021). Correcting MEG artifacts caused by overt speech. *Frontiers in Neuroscience*, 15, 682419.
- D'Mello, A. M., Gabrieli, J. D., & Nee, D. E. (2020). Evidence for hierarchical cognitive control in the human cerebellum. *Current Biology*, 30(10), 1881-1892.
- Fairs, A., Michelas, A., Dufour, S., & Strijkers, K. (2021). The same ultra-rapid parallel brain dynamics underpin the production and perception of speech. *Cerebral Cortex Communications*, 2(3), tgab040.

- Floegel, M., Fuchs, S., & Kell, C. A. (2020). Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. *Nature Communications*, 11(1), 2839.
- Giglio, L., Ostarek, M., Sharoh, D., & Hagoort, P. (2024). Diverging neural dynamics for syntactic structure building in naturalistic speaking and listening. *Proceedings of the National Academy of Sciences*, 121(11), e2310766121.
- Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, 42(2), 180-190.
- Price, C. J. (2010). The anatomy of language: a review of 100 fMRI studies published in 2009. *Annals of the new York Academy of Sciences*, 1191(1), 62-88.
- Runnqvist, E., Chanoine, V., Strijkers, K., Pattamadilok, C., Bonnard, M., Nazarian, B., ... & Alario, F. X. (2021). Cerebellar and cortical correlates of internal and external speech error monitoring. *Cerebral Cortex Communications*, 2(2), tgab038.
- Runnqvist, E. (2023). Self-monitoring: The neurocognitive basis of error monitoring in language production. In *Language production* (pp. 168-190). Routledge.
- Stockert, A., Schwartze, M., Poeppel, D., Anwander, A., & Kotz, S. A. (2021). Temporo-cerebellar connectivity underlies timing constraints in audition. *Elife*, 10, e67303.
- Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: electrophysiological correlates of word frequency and cognate effects. *Cerebral cortex*, 20(4), 912-928.
- Todorović, S., Anton, J. L., Sein, J., Nazarian, B., Chanoine, V., Rauchbauer, B., ... & Runnqvist, E. (2023). Cortico-cerebellar monitoring of speech sequence production. *Neurobiology of Language*, 1-21.

#### **Reviewer #2 (Public Review):**

##### *Summary:*

The authors re-analyse MEG data from a speech production and perception study and extend their previous Granger causality analysis to a larger number of cortical-cortical and in particular cortical-subcortical connections. Regions of interest were defined by means of a meta-analysis using Neurosynth.org and connectivity patterns were determined by calculating directed influence asymmetry indices from the Granger causality analysis results for each pair of brain regions. Abbasi et al. report feedforward signals communicated via fast rhythms and feedback signals via slow rhythms below 40 Hz, particularly during speaking. The authors highlight one of these connections between the right cerebellum lobule VI and auditory association area A5, where in addition the connection strength correlates negatively with the strength of speech tracking in the theta band during speaking (significant before multiple comparison correction). Results are interpreted within a framework of active inference by minimising prediction errors.

*While I find investigating the role of cortical-subcortical connections in speech production and perception interesting and relevant to the field, I am not yet convinced that the methods employed are fully suitable to this endeavour or that the results provide sufficient evidence to make the strong claim of dissociation of bottom-up and top-down information flow during speaking in distinct frequency bands.*

*Strengths:*

*The investigation of electrophysiological cortical-subcortical connections in speech production and perception is interesting and relevant to the field. The authors analyse a valuable dataset, where they spent a considerable amount of effort to correct for speech production-related artefacts. Overall, the manuscript is well-written and clearly structured.*

*Weaknesses:*

*The description of the multivariate Granger causality analysis did not allow me to fully grasp how the analysis was performed and I hence struggled to evaluate its appropriateness. Knowing that (1) filtered Granger causality is prone to false positives and (2) recent work demonstrates that significant Granger causality can simply arise from frequency-specific activity being present in the source but not the target area without functional relevance for communication (Schneider et al. 2021) raises doubts about the validity of the results, in particular with respect to their frequency specificity. These doubts are reinforced by what I perceive as an overemphasis on results that support the assumption of specific frequencies for feedforward and top-down connections, while findings not aligning with this hypothesis appear to be underreported. Furthermore, the authors report some main findings that I found difficult to reconcile with the data presented in the figures. Overall, I feel the conclusions with respect to frequency-specific bottom-up and top-down information flow need to be moderated and that some of the reported findings need to be checked and if necessary corrected.*

*Major points*

*(1) I think more details on the multivariate GC approach are needed. I found the reference to Schaum et al., 2021 not sufficient to understand what has been done in this paper. Some questions that remained for me are:*

*(i) Does multivariate here refer to the use of the authors' three components per parcel or to the conditioning on the remaining twelve sources? I think the latter is implied when citing Schaum et al., but I'm not sure this is what was done here?*

*If it was not: how can we account for spurious results based on indirect effects?*

Yes, multivariate refers to the three components.

*(ii) Did the authors check whether the GC of the course-target pairs was reliably above the bias level (as Schaum et. al. did for each condition separately)? If not, can they argue why they think that their results would still be valid? Does it make sense to compute DAIs on connections that were below the bias level? Should the data be re-analysed to take this concern into account?*

We performed statistics on DAI and believe that this is a valid approach. We argue that random GC effects would not survive our cluster-corrected statistics.

*(iii) You may consider citing the paper that introduced the non-parametric GC analysis (which Schaum et al. then went on to apply): Dhamala M, Rangarajan G, Ding M.*

Analyzing Information Flow in Brain Networks with Nonparametric Granger Causality. *Neuroimage*. 2008; 41(2):354-362. <https://doi.org/10.1016/j.neuroimage.2008.02.020>

Thanks, we will add this reference in the revised version.

*(2) GC has been discouraged for filtered data as it gives rise to false positives due to phase distortions and the ineffectiveness of filtering in the information-theoretic setting as reducing the power of a signal does not reduce the information contained in it (Florin et al., 2010; Barnett and Seth, 2011; Weber et al. 2017; Pinzuti et al., 2020 - who also suggest an approach that would circumvent those filter-related issues). With this in mind, I am wondering whether the strong frequency-specific claims in this work still hold.*

This must be a misunderstanding. We are aware of the problem with GC on filtered data. But GC was here computed on broadband data and not in individual frequency bands.

*(3) I found it difficult to reconcile some statements in the manuscript with the data presented in the figures:*

*(i) Most notably, the considerable number of feedforward connections from A5 and STS that project to areas further up the hierarchy at slower rhythms (e.g. L-A5 to R-PEF, R-Crus2, L CB6 L-Tha, L-FOP and L-STs to R-PEF, L-FOP, L-TOPJ or R-A5 as well as R-STs both to R-Crus2, L-CB6, L-Th) contradict the authors' main message that 'feedback signals were communicated via slow rhythms below 40 Hz, whereas feedforward signals were communicated via faster rhythms'. I struggled to recognise a principled approach that determined which connections were highlighted and reported and which ones were not.*

*(ii) "Our analysis also revealed robust connectivity between the right cerebellum and the left parietal cortex, evident in both speaking and listening conditions, with stronger connectivity observed during speaking. Notably, Figure 4 depicts a prominent frequency peak in the alpha band, illustrating the specific frequency range through which information flows from the cerebellum to the parietal areas." There are two peaks discernible in Figure 4, one notably lower than the alpha band (rather theta or even delta), the other at around 30 Hz. Nevertheless, the authors report and discuss a peak in the alpha band.*

*(iii) In the abstract: "Notably, high-frequency connectivity was absent during the listening condition." and p.9 "In contrast with what we reported for the speaking condition, during listening, there is only a significant connectivity in low frequency to the left temporal area but not a reverse connection in the high frequencies."*

*While Fig. 4 shows significant connectivity from R-CB6 to A5 in the gamma frequency range for the speaking, but not for the listening condition, interpreting comparisons between two effects without directly comparing them is a common statistical mistake (Makin and Orban de Xivry). The spectrally-resolved connectivity in the two conditions actually look remarkably similar and I would thus refrain from highlighting this statement and indicate clearly that there were no significant differences between the two conditions.*

*(iv) "This result indicates that in low frequencies, the sensory-motor area and cerebellum predominantly transmit information, while in higher frequencies, they are more involved in receiving it."*

*I don't think that this statement holds in its generality: L-CB6 and R-3b both show strong output at high frequencies, particularly in the speaking condition. While they seem to transmit information mainly to areas outside A5 and STS these effects are strong and should be discussed.*

We appreciate the reviewer's thoughtful comments. We acknowledge that not all connectivity patterns strictly adhere to the initial observation regarding feedback and feedforward communication. It's true that our primary focus was on interactions between brain regions known to be crucial for speech prediction, including auditory, somatosensory, and cerebellar areas. However, we also presented connectivity patterns across other regions to provide a more comprehensive picture of the speech network. We believe this broader perspective can be valuable for future research directions.

Regarding the reviewer's observation about the alpha band peak in Figure 4, we agree that a closer examination reveals the connectivity from right cerebellum to the left parietal is in a wider low frequency range. We will refrain from solely emphasizing the alpha band and acknowledge the potential contribution of lower frequencies to cerebellar-parietal communication.

We also appreciate the reviewer highlighting the need for a more nuanced interpretation of the listening condition connectivity compared to the speaking condition. The reviewer is correct in pointing out that while Figure 4 suggests a high-frequency connectivity from L-A5 to R-CB only in the speaking condition, a direct statistical comparison between conditions might not reveal a significant difference. We will revise the manuscript to clarify this point.

Finally, a closer examination of Figure 3 revealed that the light purple and dark green edges in the speaking condition for R-CB6 and L-3b suggest outgoing connections at low frequencies, while other colored edges indicate information reception at high frequencies. We acknowledge that exceptions to this directional pattern might exist and warrant further investigation in future studies.

*(4) "However, definitive conclusions should be drawn with caution given recent studies raising concerns about the notion that top-down and bottom-up signals can only be transmitted via separate frequency channels (Ferro et al., 2021; Schneider et al., 2021; Vinck et al., 2023)."*

*I appreciate this note of caution and think it would be useful if it were spelled out to the reader why this is the case so that they would be better able to grasp the main concerns here. For example, Schneider et al. make a strong point that we expect to find Granger-causality with a peak in a specific frequency band for areas that are anatomically connected when the sending area shows stronger activity in that band than the receiving one, simply because of the coherence of a signal with its own linear projection onto the other area. The direction of a Granger causal connection would in that case only indicate that one area shows stronger activity than the other in the given frequency band. I am wondering to what degree the reported connectivity pattern can be traced back to regional differences in frequency-specific source strength or to differences in source strength across the two conditions.*

This is indeed an important point. That is why we are discussing our results with great caution and specifically point the reader to the relevant literature. We are indeed thinking about a future study where we investigate this connectivity using other connectivity metrics and a detailed consideration of power.

#### **Reviewer #3 (Public Review):**

*In the current paper, Abbasi et al. aimed to characterize and compare the patterns of functional connectivity across frequency bands (1 Hz - 90 Hz) between regions of a speech network derived from an online meta-analysis tool (Neurosynth.org) during speech production and perception. The authors present evidence for complex neural dynamics from which they highlight directional connectivity from the right cerebellum to*

*left superior temporal areas in lower frequency bands (up to beta) and between the same regions in the opposite direction in the (lower) high gamma range (60-90 Hz). Abbasi et al. interpret their findings within the predictive coding framework, with the cerebellum and other "higher-order" (motor) regions transmitting top-down sensory predictions to "lower-order" (sensory) regions in the lower frequencies and prediction errors flowing in the opposite direction (i.e., bottom-up) from those sensory regions in the gamma band. They also report a negative correlation between the strength of this top-down functional connectivity and the alignment of superior temporal regions to the syllable rate of one's speech.*

*Strengths:*

*(1) The comprehensive characterization of functional connectivity during speaking and listening to speech may be valuable as a first step toward understanding the neural dynamics involved.*

*(2) The inclusion of subcortical regions and connectivity profiles up to 90Hz using MEG is interesting and relatively novel.*

*(3) The analysis pipeline is generally adequate for the exploratory nature of the work.*

*Weaknesses:*

*(1) The work is framed as a test of the predictive coding theory as it applies to speech production and perception, but the methodological approach is not suited to this endeavor.*

We agree that we cannot provide definite evidence for predictive coding in speech production and perception and we believe that we do not make that claim in the manuscript. However, our results are largely consistent with what can be expected based on predictive coding theory.

*(2) Because of their theoretical framework, the authors readily attribute roles or hierarchy to brain regions (e.g., higher- vs lower-order) and cognitive functions to observed connectivity patterns (e.g., feedforward vs feedback, predictions vs prediction errors) that cannot be determined from the data. Thus, many of the authors' claims are unsupported.*

We will revise the manuscript to more clearly differentiate our results (e.g. directed Granger-Causality from A to B) from their interpretation (potentially indicating feedforward or feedback signals).

*(3) The authors' theoretical stance seems to influence the presentation of the results, which may inadvertently misrepresent the (otherwise perfectly valid; cf. Abbasi et al., 2023) exploratory nature of the study. Thus, results about specific regions are often highlighted in figures (e.g., Figure 2 top row) and text without clear reasons.*

Our connectograms reveal a multitude of results that we hope is interesting to the community. At the same time the wealth of findings poses a problem for describing them. We did not see a better way then to highlight specific connections of interest.

*(4) Some of the key findings (e.g., connectivity in opposite directions in distinct frequency bands) feature in a previous publication and are, therefore, interesting but not novel.*

We actually see this as a strength of the current manuscript. The computation of connectivity is here extended to a much larger sample of brain areas. It is reassuring to see that the

previously reported results generalise to other brain areas.

*(5) The quantitative comparison between speech production and perception is interesting but insufficiently motivated.*

We thank the reviewer for this comment. We have addressed that in detail in response to the point (1&4) of reviewer 1.

*(6) Details about the Neurosynth meta-analysis and subsequent selection of brain regions for the functional connectivity analyses are incomplete. Moreover, the use of the term 'Speech' in Neurosynth seems inappropriate (i.e., includes irrelevant works, yielding questionable results). The approach of using separate meta-analyses for 'Speech production' and 'Speech perception' taken by Abbasi et al. (2023) seems more principled. This approach would result, for example, in the inclusion of brain areas such as M1 and the BG that are relevant for speech production.*

We agree that there are inherent limitations in automated meta-analysis tools such as Neurosynth. Papers are used in the meta-analysis that might not be directly relevant. However, Neurosynth has proven its usefulness over many years and has been used in many studies. We also agree that our selection of brain areas is not complete. But Granger Causality analysis of every pair of ROIs leads to complex results and we had to limit our selection of areas.

*(7) The results involving subcortical regions are central to the paper, but no steps are taken to address the challenges involved in the analysis of subcortical activity using MEG. Additional methodological detail and analyses would be required to make these results more compelling. For example, it would be important to know what the coverage of the MEG system is, what head model was used for the source localization of cerebellar activity, and if specific preprocessing or additional analyses were performed to ensure that the localized subcortical activity (in particular) is valid.*

There is a large body of evidence demonstrating that MEG can record signals from deep brain areas such as thalamus and cerebellum including Attal & Schwarz 2013, Andersen et al, Neuroimage 2020; Piastra et al., 2020; Schnitzler et al., 2009. These and other studies provide evidence that state-of-the-art recording (with multichannel SQUID systems) and analysis is sufficient to allow reconstruction of subcortical areas. However, spatial resolution is clearly reduced for these deep areas. We will add a statement in the revised manuscript to acknowledge this limitation.

*(8) The results and methods are often detailed with important omissions (a speech-brain coupling analysis section is missing) and imprecisions (e.g., re: Figure 5; the Connectivity Analysis section is copy-pasted from their previous work), which makes it difficult to understand what is being examined and how. (It is also not good practice to refer the reader to previous publications for basic methodological details, for example, about the experimental paradigm and key analyses.) Conversely, some methodological details are given, e.g., the acquisition of EMG data, without further explanation of how those data were used in the current paper.*

We will revise the relevant sections of the manuscript.

*(9) The examination of gamma functional connectivity in the 60 - 90 Hz range could be better motivated. Although some citations involving short-range connectivity in these frequencies are given (e.g., within the visual system), a more compelling argument for looking at this frequency range for longer-range connectivity may be required.*

Given previous evidence of connectivity in the gamma band we think that it would be a weakness to exclude this frequency band from analysis.

*(10) The choice of source localization method (linearly constrained minimum variance) could be explained, particularly given that other methods (e.g. dynamic imaging of coherent sources) were specifically designed and might potentially be a better alternative for the types of analyses performed in the study.*

Both LCMV and DICS are beamforming methods. We used LCMV because we wanted used Granger Causality which requires broadband signals. DICS would only provide frequency-specific band-limited signals.

*(11) The mGC analysis needs to be more comprehensively detailed for the reader to be able to assess what is being reported and the strength of the evidence. Relatedly, first-level statistics (e.g., via estimation of the noise level) would make the mGC and DAI results more compelling.*

We perform group-level cluster-based statistics on mGC while correcting for multiple comparisons across frequency bands and brain parcels and report only significant results. This is an established approach that is routinely used in this type of studies.

*(12) Considering the exploratory nature of the study, it is essential for other researchers to continue investigating and validating the results presented in the current manuscript. Thus, it is concerning that data and scripts are not fully and openly available. Data need not be in its raw state to be shared and useful, which circumvents the stated data privacy concerns.*

We acknowledge the reviewer's concern regarding the full availability of the dataset. Due to privacy limitations on the collected data, we are unable to share it publicly at this time. However, to promote transparency and enable further exploration, we have provided the script used for data analysis and an example dataset. This example dataset should provide a clear understanding of the data structure and variables used in the analysis. Additionally, we are happy to share the complete dataset upon request from research teams interested in performing in-depth secondary analyses.

<https://doi.org/10.7554/eLife.97083.1.sa4>