Bridging verbal coordination and neural dynamics

Isaïh Schwab-Mohamed[1,2], Manuel R. Mercier[1], Agnès Trébuchon[1,3], Benjamin Morillon[1,2], Leonardo Lancia[2,4], Daniele Schön[1,2]

[1]Aix Marseille University, Inserm, INS, Inst Neurosci Syst, Marseille, France.
[2]Aix-Marseille Univ, Institute of Language, Communication and the Brain, France.
[3]APHM, Hôpital de la Timone, Service de Neurophysiologie Clinique, Marseille, France.
[4]Aix Marseille University, CNRS, Laboratoire Parole et Langage (LPL), Aix-en-Provence, France.

***Conflict of interests**:* The authors declare no competing interests.

***Acknowledgments**:* We thank all patients for their willingful participation and all the personnel from the epileptology unit.

***Funding sources**:* ANR-21-CE28-0010 (to D.S), ANR-20-CE28-0007-01 & ERC, SPEEDY, ERC-CoG-101043344 (to B.M), ANR-17-EURE-0029 (NeuroMarseille). This work, carried out within the Institute of Convergence ILCB, was also supported by grants from France 2030 (ANR-16-CONV-0002), the French government under the Programme «Investissements d'Avenir», and the Excellence Initiative of Aix-Marseille University (A*MIDEX, AMX-19-IET-004).

***Author contributions**:* I.S.M, L.L and D.S. designed research; L.L. wrote the code for the experimental setup;I.S.M., L.L, A.T. and M.M. acquired data; I.S.M, L.L and D.S. analyzed data; I.S.M and D.S. wrote the paper; I.S.M, A.T., M.M., B.M., L.L and D.S. edited the paper.

Data availability statement: The conditions of our ethics approval do not permit public archiving of anonymised study data. Readers seeking access to the data should contact Dr. Daniele Schön (daniele.schon@univ-amu.fr). Access will be granted to named individuals in accordance with ethical procedures governing the reuse of clinical data, including completion of a formal data sharing agreement.

Code availability statement: Data analyses were performed using custom scripts in Matlab & Python, and will be available upon publication on Github.

Abstract

Our use of language, which is profoundly social in nature, essentially takes place in interactive contexts and is shaped by precise coordination dynamics that interlocutors must observe. Thus language interaction is high demanding on fast adjustment of speech production. Here, we developed a real-time coupled-oscillators virtual partner that allows - by changing the coupling strength parameters - to modulate the ability to synchronise speech with a speaker. Then, we recorded the intracranial brain activity of 16 patients with drug-resistant epilepsy while they performed a verbal coordination task with the virtual partner (VP). More precisely, patients had to repeat short sentences synchronously with the VP. This synchronous speech task is efficient to highlight both the dorsal and ventral language pathways. Importantly, combining time-resolved verbal coordination and neural activity shows more spatially differentiated patterns and different types of neural sensitivity along the dorsal pathway. More precisely, high-frequency activity in secondary auditory regions is highly sensitive to verbal coordinative dynamics, while primary regions are not. Finally, the high-frequency activity of the IFG BA44 seems to specifically index the online coordinative adjustments that are continuously required to compensate deviation from synchronisation. These findings illustrate the possibility and value of using a fully dynamic, adaptive and interactive language task to gather deeper understanding of the subtending neural dynamics involved in speech perception, production as well as their interaction.

Introduction

Language has been most frequently studied by separately assessing perception and production in an isolated context, by contrast with the interactive context that characterizes best its daily usage. The use of language in interactive context, particularly during conversations, calls upon our numerous predictive and adaptive processes (Pickering & Gambi, 2018). Importantly, analyses of conversational data have highlighted the phenomenon of interactive alignment (Pickering & Garrod, 2004), illustrating that during a verbal exchange interlocutors tend to imitate each other and to align their linguistic representations on several levels including phonetic, syntactic, or semantic (Garrod & Pickering, 2009). This is an unconscious and dynamic phenomenon that possibly renders exchanges between speakers more fluid (Marsh et al., 2009). It consists in mutual anticipation (prediction) and coordination of speech production, leading, for instance, to a reduction of turn-taking durations (Levinson, 2016 ; Corps et al., 2018).

Recently, in an effort to assess speech and language in more ecological contexts, researchers in neuroscience have used interactive paradigms to study some of these coordinative phenomena. These consisted of turn-taking behaviors such as alternating naming tasks (Mukherjee et al., 2019), questions and answers investigating motor preparations (Bögels et al., 2015), or manipulating the turn predictability in end-of-turn detection tasks (Magyari et al., 2014 ; for a review see, Bögels et al., 2017). However, the synchronous speech paradigm has been overlooked. This paradigm requires the simultaneous and synchronized production of the same word, sentence, or text between two people. Interestingly, this task is remarkably well performed without any particular training, between a speaker and a recording as well as between two speakers (Cummins, 2002 ; Cummins, 2003 ; Assaneo et al., 2019). As in most joint tasks, individuals must mutually adjust their behavior (here speech production) to optimize coordination (Cummins, 2009). Furthermore, synchronous speech favors the emergence of alignment phenomena, for instance of the fundamental frequency or the syllable onset (Assaneo et al., 2019 ; Bradshaw & McGettigan, 2021 ; Bradshaw et al., 2023). Overall synchronous speech represents a good interactive framework with a sufficient level of experimental control. It offers several possibilities for neurophysiological investigation of both speech perception and production and is an interesting case to consider for models of speech motor control.

Synchronous speech resembles to a certain extent to delayed auditory feedback tasks, which involve real-time perturbations in the speech production signal (such as changes in fundamental frequency and delay). These tasks can induce speech errors as well as modulations in speech and voice features (Stuart et al., 2002 ; Yamamoto & Kawabata, 2014 ; Karlin et al., 2021). Additionally, these tasks provide insights into predictive models of speech motor control, where the brain generates an internal estimate of production and corrects errors when auditory feedback deviates from the estimate (Hickok et al., 2011 ; Houde & Nagarajan, 2011 ; Tourville & Guenther, 2011 ; Ozerk et al., 2022 ; Floegel et al., 2023). Previous studies have revealed increased responses in the superior temporal regions compared to normal feedback conditions (Hirano et al., 1997 ; Hashimoto & Sakai, 2003 ; Takaso et al., 2010 ; Ozerk et al., 2022 ; Floegel et al., 2020). However, synchronous speech paradigms allow investigating the neural bases of coordinative behavior rather than of error correction. Importantly, the precise spectro-temporal dynamics and spatial distribution of the cortical networks underlying speech coordination remain unknown.

To address this issue, we first developed a real-time coupled-oscillators virtual partner that allows - by changing the coupling strength parameters - to modulate the ability to synchronise speech with a speaker. The virtual partner (Lancia et al.,2017) and the synchronous speech task were first tested on a control group to ensure the ability of the virtual agent to coordinate its speech production in real time with the participants. In certain

interactions, the agent was programmed to actively cooperate with the participants by synchronizing its syllables with theirs. In other interactions, the agent was programmed to deviate from synchronization by producing its syllables between those of the participants. As a result, participants are constantly required to adapt their verbal productions in order to maintain synchronization. Appropriate tuning of the coupling parameters of the virtual agent enabled us to create a variable context of coordination yielding a broad distribution of phase delays between the speaker and agent productions. Then, we leveraged the excellent spatial specificity and temporal resolution of stereotactic depth electrodes recordings and acquired neural activity from 16 patients with drug-resistant epilepsy while they performed the adaptive synchronous speech task with the virtual partner.

Materials and methods

Control – participants

30 participants (17 women, mean age 24.7 y, range 19-42 y) took part in the study. All were French native speakers with normal hearing and no neurological disorders. Participants provided written informed consent prior to the experimental session and the experimental protocol was approved by the Institutional Review board of the French Institute of Health (IRB00003888). 5 participants (2 women) were excluded from analysis for poor signal-to-noise ratio in speech recordings.

Patients – participants

16 patients (7 women, mean age 29.8 y, range 17 - 50 y) with pharmacoresistant epilepsy took part in the study. They were included if their implantation map covered at least partially the Heschl's gyrus and had sufficiently intact diction to support relatively sustained language production. All patients were French native speakers. Neuropsychological assessments carried out before stereotactic EEG (sEEG) recordings indicated that they had intact language functions and met the criteria for normal hearing. In none of them were the auditory areas part of their epileptogenic zone as identified by experienced epileptologists. Recordings took place at the Hôpital de La Timone (Marseille, France). Patients provided written informed consent prior to the experimental session and the experimental protocol was approved by the Institutional Review board of the French Institute of Health (IRB00003888).

Data acquisition

The speech signal was recorded using a microphone (RODE NT1) adjusted on a stand so that it was positioned in front of the participant’s mouth. Etymotic insert earphones (Etymotic Research E-A-R-TONE gold) fitted with 10mm foam eartips were used for sound presentation. The parameters and sound adjustment were set using an external low-latency sound card (RME Babyface Pro Fs), allowing a tailored configuration for each participant. A calibration was made to find a comfortable volume and an optimal balance for both the sound of the participant's own voice, which was fed back through the headphones, and the sound of the stimuli.

The sEEG signal was recorded using depth electrodes shafts of 0.8 mm diameter containing 5 to 18 electrode contacts (Dixi Medical or Alcis, Besançon, France). The contacts were 2 mm long and were spaced from each other by 1.5 mm. The placement of the electrode implantations were determined solely on clinical grounds. The cohort consists of 16 unilateral implantations (9 left, 7 bilateral implantations, yielding a total of 236 electrodes and 3185 contacts (see Figure 1). Due to insufficient cover of the right hemisphere we only focus here on the left hemisphere in all presented analyses. All Patients were recorded in a sound-proof Faraday cage using a 256-channels amplifier (Brain Products), sampled at 1kHz and high-pass filtered at 0.16 Hz.

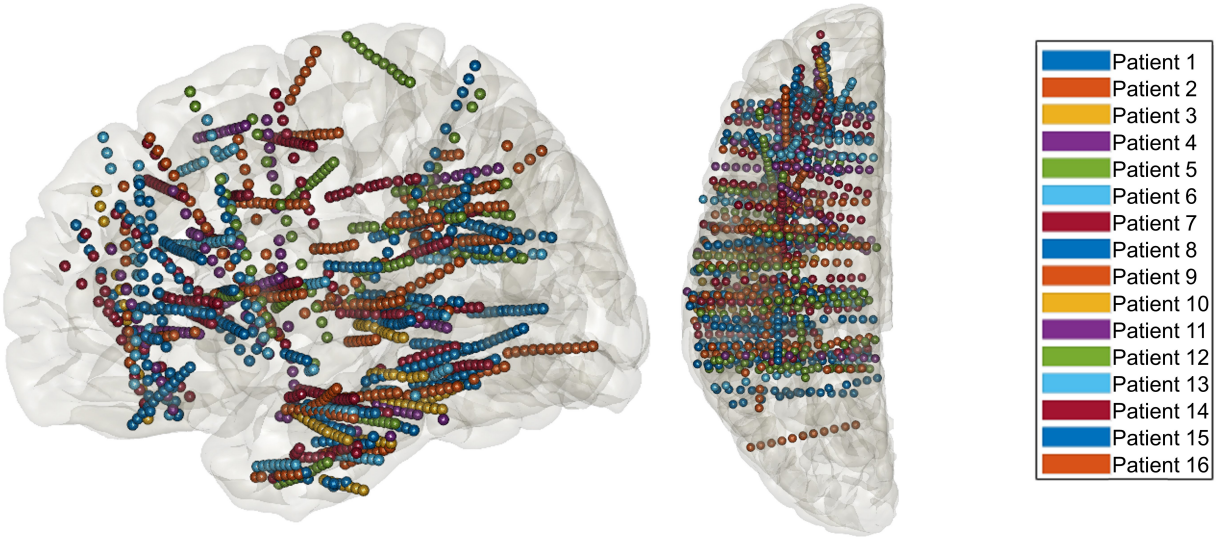


Figure 1 – Anatomical localization of the sEEG electrodes for each patient (N=16).

Stimuli

Four stimuli corresponding to four short sentences were pre-recorded by a woman and a man. This allowed to adapt to the natural gender differences in fundamental frequency. All stimuli were normalised in amplitude.

Stimuli consisted of four sentences: "papi mamie" (/*papi mami*/, grandpa grandma) / "papi m'a dit" (/*papi ma di*/, grandpa told me) / "mamie lavait ma main" (/*mami lavɛ ma mɛ̃*/, grandma washed my hand) / "mamie manie ma main" (/*mami mani ma mɛ̃*/, grandma handles my hand). The four sentences purposely differed in terms of number of syllables (4-4-6-6). Moreover two of them contained deviations from an otherwise repeating phonological pattern : both in /*papi ma di*/ and in /*mami mani ma mɛ̃*/ , the repeated opening/closing of the lips is substituted by the formation and release of a tongue constriction. These manipulations made the sentences more or less easy to articulate (easy-medium-easy-medium). Sentence durations were 2.07, 2.11, 3.16 and 2.87s, respectively with a syllable rate of ~3Hz.

Experimental design

Participants, comfortably seated on a medical chair, performed a real-time interactive synchronous speech task with an artificial agent (Virtual Partner, henceforth VP, see next section) who can modulate and adapt to the participant's speech in realtime.

The experiment required 3 steps (Figure 2A). First, the sentences were presented in written form and the experimenter verified that each participant could pronounce them correctly. Second, a training phase took place. This required repeating each stimulus over and over together with the VP for ~14 seconds. More precisely, the sentence was first presented on a screen. When the participant pressed the "space" bar, the visual stimulus went off, and the VP started to « speak ». The participant was instructed to repeat the stimuli as synchronously as possible with the VP for the whole trial duration. In the training phase the VP did not adapt to the participant.

The training allowed participants to familiarise with the synchronous speech task but also to build a personalised virtual partner model incorporating the articulatory variability of each participant (see section below).

The training was followed by a resting phase, allowing the recording of resting state activity for a period of 5 min. This time also allowed to build, for each participant, the virtual partner model.

The third step was the actual experiment. This was identical to the training but consisted of 24 trials (14s long, speech rate ~3Hz, yielding ~1000 syllables). Importantly, the VP varied its coupling behaviour to the participant. More precisely, for a third of the sequences the VP had a neutral behaviour (close to zero coupling : k = +/- 0.01). For a third it had a moderate coupling, meaning that the VP synchronised more to the participant speech (k = -0.09). And for the last third of the sequences the VP had a moderate coupling but with a phase shift of pi/2, meaning that it moderately aimed to speak in between the participant syllables (k = + 0.09). Depending on patient fatigue, a second experimental session with 24 extra sequences, was proposed (6/16 patients). Control group participants run a single session of 48 trials.

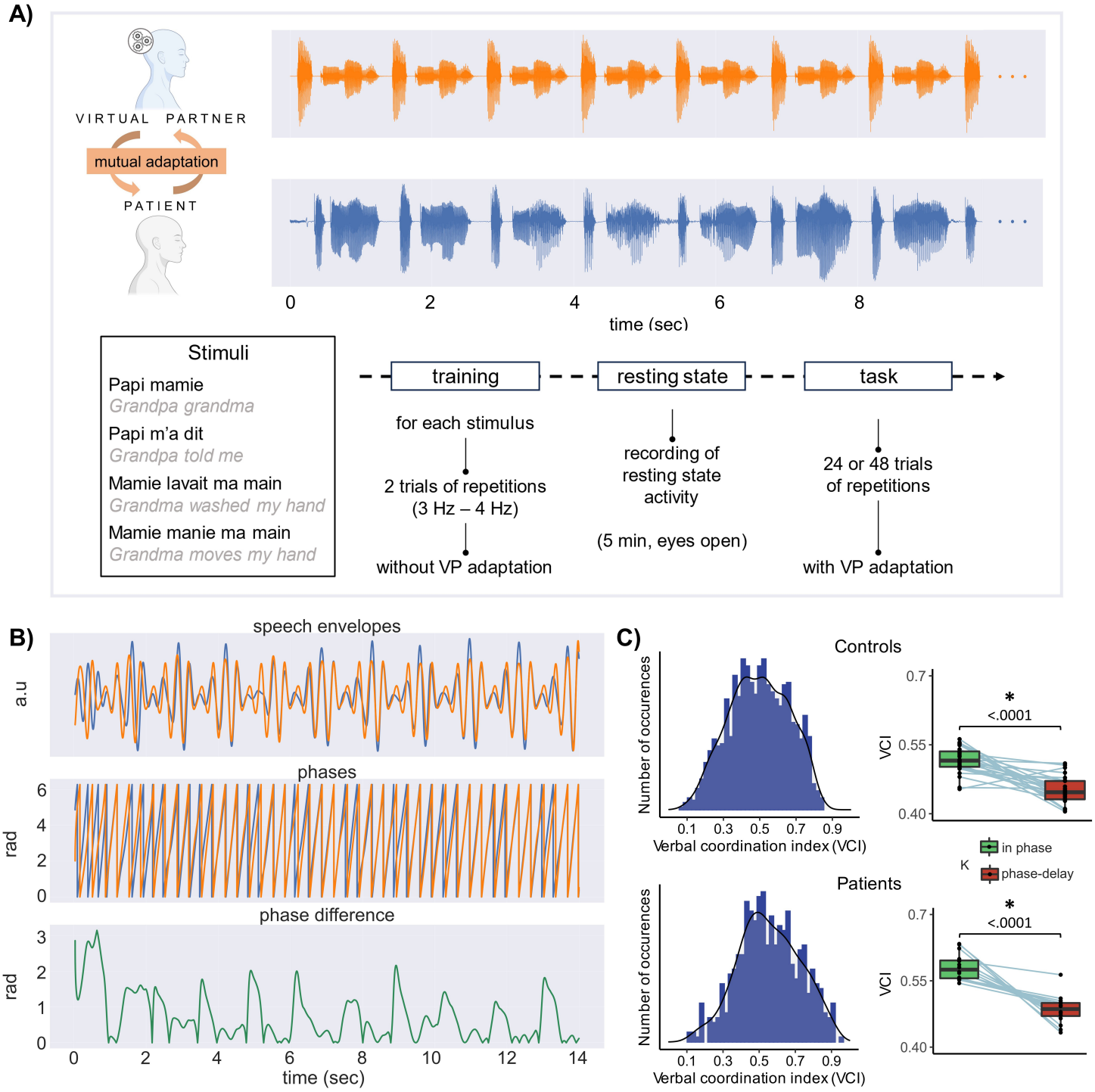


Figure 2 – Paradigm and coordination indexes. (A) Top : *illustration of one trial of the interactive synchronous speech repetition task (orange: virtual partner speech; blue: participant speech; stimulus papi m'a dit repeated 10 times ; only the 10 first seconds are represented).* ***Bottom :*** *the four speech utterances used in the task and the experimental procedure.* ***(B)*** *Speech signals processing stages. The top panel corresponds to the speech envelope, the second to the phase of speech envelope and the third panel to the phase difference between VP and participant speech envelopes, illustrating the coordination dynamics along one trial.* ***(C)*** *Left: distributions of verbal coordination index (phase locking values between VP and participant speech envelopes, for each trial) for all participants (top) and patients . Right: boxplots for control participants (top) and patients showing the trial-averaged verbal coordination index as a function of the virtual partner parameters (in-phase coupling vs coupling with a 180° shift) .*

Virtual partner (principles)

The virtual partner (henceforth VP) used for the experiment allows to generate speech (words or short utterances) while adapting it in real-time to the concurrent speech input. The VP environment is built upon the Psychtoolbox-3 program and operates within MATLAB, leveraging custom C subroutines to boost its performance. Its operation revolves around a loop whose iterations are executed at consistent intervals of $\Delta t = 4$ ms. During each iteration, the program analyzes the latest segment (25 ms) of speech produced by the participant and streams a portion of speech to the output device. More precisely, at each iteration of the main loop, the functioning of the VP can be described in four steps:

1. A feature vector is extracted from the last chunk of the input signal.

2. The phase value of the input signal chunk is calculated by mapping the input feature vector onto the corresponding vector of the stimulus signal and retrieving the associated phase value. This step is performed using a dynamic-time-warping algorithm (Dixon, 2005). To enhance precision, the input chunk is mapped onto several model utterances (all time-aligned with the signal used as a stimulus) that are tailored to the characteristics of the participant's speech in the training phase.

3. A chunk of stimulus signal is chosen to be sent to the output device. The selection is guided by applying the Kuramoto (1975) equation to the difference between the phase values representing the current positions of the participant and the VP in their syllabic cycles. This enables real-time lengthening or shortening of speech chunks as needed to adjust towards the preferred phase (0 or $\pi/2$). Significantly, in applying this equation, we assign a specific value to the coupling strength parameter « k », linking the behavior of the VP to that of the participant. Values can vary from trial to trial and could be close to 0 ($k = +/- 0.01$) resulting in a neutral behavior, negative values ($k = - 0.09$) equivalent to a moderate coupling behavior (tendency to synchronize), or positive values ($k = + 0.09$) corresponding to a moderate coupling behavior with a phase shift of $\pi/2$ (tendency to speak in between participant syllables).

4. The selected chunk of stimulus signal is integrated into the output stream via WSOLA synthesis (Waveform Similarity Overlap-Add).

Of note, the coupling strength were chosen to be rather weak and thus do not allow to reach 0 or $\pi/2$ phase synchrony, but rather yield the desired large panel of phase delays in the VP-participant coordinative behaviour (see Figure1).

Data analysis

speech signal

Speech signals of the participants and the VP were processed using Praat and Python scripts.

First, raw speech signals were downsampled from 48 KHz to 16 KHz. Then speech envelope was extracted using a pass-band filter between 2.25 and 5 Hz . The phase was computed using the Hilbert transform.

To quantify the degree of coordination of the verbal interaction (verbal coordination index, VCI) we computed the phase locking value using mne_connectivity python function spectral_connectivity_time with the method 'plv' based on the phase coherence model developed by Lachaux et al. (1999) . Phase locking was computed for

224 each trial on speech temporal envelope (Fig 1.B), resulting in 24 or 48 verbal coordination indexes (VCI) per
225 patient depending on the number of sessions performed. To assess the effect of the coupling parameter we
226 computed a linear mixed-model contrasting *in-phase coupling trials and trials with a coupling set towards a*
227 *180° shift* (lmer(VCI ~ K + (1|participant)). As expected, because of the varying adaptive behaviour of the VP, VCI
228 varies across trials, indexing more or less efficient coordinative behaviour (Figure 1C). Moreover, in order to
229 estimate whether the level of performance was greater than chance level, we computed, for each patient and
230 each trial, a null distribution obtained by randomly shifting the phase between the VP and the patient speech
231 (500 times, see Figure S1).

232

233– **sEEG signal**

234 **General preprocessing related to electrodes localisation**

235 To increase spatial sensitivity and reduce passive volume conduction from neighboring regions (Mercier et al.,
236 2017), the signal was offline re-referenced using bipolar montage. That is, for a pair of adjacent electrode
237 contacts, the referencing led to a virtual channel located at the midpoint locations of the original contacts. To
238 precisely localize the channels, a procedure similar to the one used in the iELVis toolbox was applied (Groppe et
239 al., 2017). First, we manually identified the location of each channel centroid on the post-implant CT scan using
240 the Gardel software (Medina Villalon et al., 2018). Second, we performed volumetric segmentation and cortical
241 reconstruction on the pre-implant MRI with the Freesurfer image analysis suite (documented and freely
242 available for download online http://surfer.nmr.mgh.harvard.edu/). This segmentation of the pre-implant MRI
243 with SPM12 provides us with both the tissue probability maps (i.e. gray, white, and cerebrospinal fluid (CSF)
244 probabilities) and the indexed-binary representations (i.e., either gray, white, CSF, bone, or soft tissues). This
245 information allowed us to reject electrodes not located in the brain. Third, the post-implant CT scan was
246 coregistered to the pre-implant MRI via a rigid affine transformation and the pre-implant MRI was registered to
247 the MNI template (MNI 152 Linear), via a linear and a non-linear transformation from SPM12 methods (Penny
248 et al., 2011), through the FieldTrip toolbox (Oostenveld et al., 2011). Fourth, applying the corresponding
249 transformations, we mapped channel locations to the pre-implant MRI brain that was labeled using the volume-
250 based Human Brainnetome Atlas (Fan et al., 2016).

251 **Signal preprocessing**

252 Continous signal was filtered using 1) a notch filter at 50 Hz and harmonics up to 300 Hz to remove power line
253 artifacts, 2) a bandpass filter between 0.5 Hz and 300 Hz.

254 To define artifacted channel we used the broadband (raw) signal delimited on the experimental task recording.
255 Channels with a variance greater than 2*IQR (interquartile range, i.e. a non-parametric estimate of the standard
256 deviation) were tagged as artifacted channels (6% of the channels). Channels defined as artifacted were
257 excluded from subsequent analysis.

258 The continuous signal during the task was then epoched from -0.2 to 14 s relative to the onset of the first
259 stimulus (repetition) of each trial. Generating 24 or 48 epochs depending on the number of sessions each patient
260 had completed. A baseline correction was applied from -0.1 to 0 s. The first 500 ms were cut from the epoched
261 data to avoid the activity burst generated at the beginning of the stimulus.

262 The continuous signal from the resting state session was also epoched in seventeen 14-second non-overlapping
263 epochs.

Power spectral density

The power spectral density (PSD) computation was conducted for each channels, with the MNE-python function Epochs.compute_psd, trial-by-trial (epochs) in a range of 125 frequencies, logarithmically scaled, ranging from 0.5 to 125 Hz. Six canonical frequency bands were investigated by averaging the power spectra between their respective boundaries: delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), low gamma (30-50 Hz) and high frequency activity or HFa (70-125 Hz).

Global effect

The global effect of the task (versus rest) was computed on each frequency band by susbtracting first the mean resting state activity from the mean experimental activity and then dividing by the mean resting state activity. This approach has the advantage of centering the magnitude before expressing it in percentage (Mercier et al., 2022). For each frequency band and channel, the statistical difference between task activity and the baseline (resting) was estimated with permutation tests (N=1000) using the scipy library.

Coupling behavioural and neurophysiological data

Behavioural speech data and neurophysiological data were jointly analysed using two approaches. In the first approach a two step procedure was used. First, for each frequency band, channel and trial, we computed the mean power. Then, for each frequency band and channel, we computed the correlation across trials using a non-parametric correlation metric (Spearman) between the power and the verbal coordination index (VCI between VP and patient speech). A rho value was thus attributed to each channel and frequency band. The significance was assessed using a permutation approach, similar to the one used for the global effect (see above). In the second approach, we computed the phase-amplitude coupling (PAC) between speech phase and high frequency neural activity. More precisely, we used two regressors for the phase: 1) the phase of the speech envelope of the VP, corresponding to the speech input, 2) the instantaneous phase difference between VP and patient phases, corresponding to the instantaneous coordination of participant and virtual partner (see Figure 2B). As for power, we used the high frequency activity from 70 to 125 Hz. The computation was performed using Tensorpac, an open-source Python toolbox for tensor-based phase-amplitude coupling (PAC) measurement in electrophysiological brain signals (Combrisson et al. 2020).

Clustering analysis

A spatial unsupervised clustering analysis (k-means) was conducted on all significant channels, separately for the global effect (task versus rest) and brain-behaviour correlation analyses. Precisely, we used the silhouette score method on the k-means result (Rousseeuw, 1987 ; Shahapure & Nicholas, 2020). This provides a measure of consistency within clusters of data (or alternatively the goodness of clusters separation). Scores were computed for difference numbers of clusters (from two to ten). The highest silhouette score indicates the optimal number of clusters. Clustering and slihouette scores were computed using the Scikit-learn's Kmeans and silhouette score function (Pedregosa et al., 2011).

Results

Speech coordinative behaviour

The present synchronous speech task allowed us to create a more or less predictable context of coordination , with the objective of obtaining a wide range of coordination variability. Indeed, the intrinsic nature of the task (speech coordination) on one side and the variable coupling parameter of the VP on the other, require continuous subtle adjustments of the participants speech production. The degree of coordination between speech signals (VP and participant) was assessed at the syllabic level computing the phase locking value between the speech temporal envelopes, a measure of the strength of the interaction (or coupling) of two signals. This metric, the verbal coordination index (VCI), is a proxy of the quality of the performance in coordinating speech production with the VP. Overall, the coordinative behaviour was affected, for both controls and patients, by the coupling parameters of the model with a better coordination when the VP was set to synchronize with participants compared to when it was set to speak with a 180° syllabic shift (controls: t ratio = 4.55, p = <.0001; patients: t ratio = 6.53, p = <.0001, see boxplots in Figure 2C). This produced, as desired, a rather large coupling variability (controls: range : 0.06 - 0.84 ; mean : 0.49 ; median : 0.49 ; patients range : 0.11 - 0.94 ; mean : 0.55 ; median : 0.54, see Fig 2C). This variability was also present at the individual level (see, for patients only, Figure S1). Nonetheless, while variable, the coordinative behaviour was significantly better than chance for every patient (see Figure S1).

Synchronous speech strongly activates the language network from delta to high gamma range

To investigate the sensitivity of synchronous speech in generating spectrally-resolved neural responses, we first analyzed the neural responses in both a spatially and spectrally resolved manner with respect to a resting-state baseline condition. Overall, neural responses are present in all six canonical frequency bands, from the delta range (1-4 Hz) up to high gamma activity (HFa, 70-125 Hz, see Fig 3A) with medium to large modulation (increase or decrease) in activity compared to baseline (Figure 3A). More precisely, while theta, alpha and beta bands show massive desynchronisation, in particular in the STG BA41/42 (primary auditory cortex), STG BA22 (secondary auditory cortex), and IFG BA44 (Broca's area), the low gamma and HFa bands are dominated by power increase in particular in the auditory cortex (STG BA 41/42) and in the inferior frontal gyrus (IFG BA44).

As expected, the whole language network is strongly involved, including both dorsal and ventral pathways (Fig 3A). More precisely, in the temporal lobe the superior, middle and inferior temporal gyri, in the parietal lobe the inferior parietal lobule (IPL) and in the frontal lobe the inferior frontal gyrus (IFG) and the middle frontal gyrus (MFG).

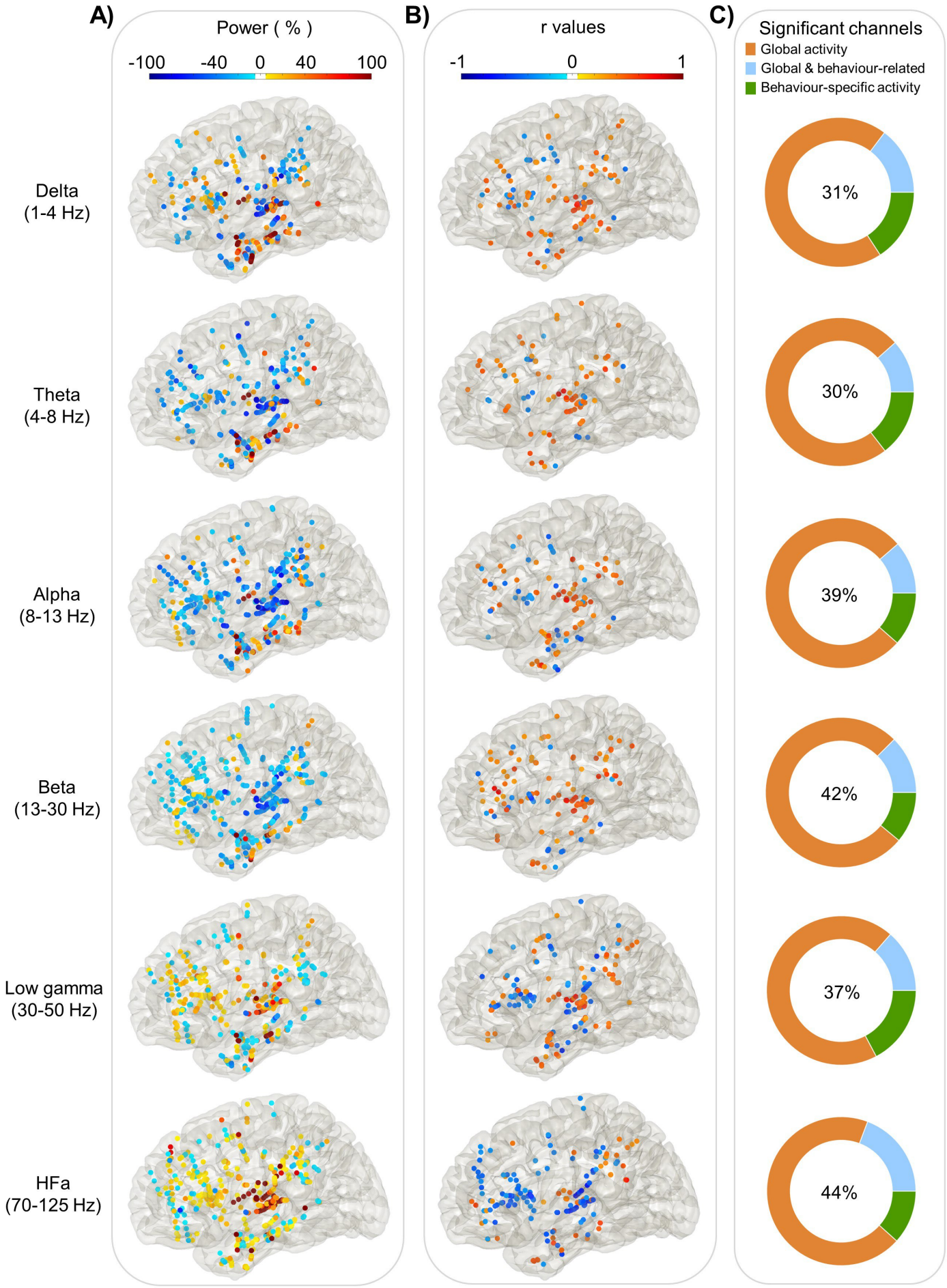
The brain needs behaviour: global vs behaviour-specific neural activity

Comparing the overall activity during the task to activity during rest gives a broad view of the network involved in the task. However, as stated above, the task was conceived to engender a rather wide range of verbal coordination across trials for each participant. This variety of coordinative behaviours allows to explore the link between verbal and simultaneous neural activity, by computing the correlation across trials between the verbal coordination index and the mean power (see methods). This analysis allows us to estimate the extent to which neural activity in each frequency band is modulated as a function of the quality of verbal coordination.

Figure 3B shows the significant r values (Spearman correlation) for each frequency band. The first observation is a gradual transition in the direction of correlations as we move up frequency bands, from positive correlations at low frequencies to negative ones at high frequencies. This is possibly due to the reversed desynchronization/synchronization process in low and high frequency bands reported above. In other words, while in the low frequency bands stronger desynchronization goes along with weaker verbal coordination, in the HFa stronger activity is associated with weaker verbal coordination.

Importantly, compared to the global activity (task vs rest, Fig 3A), the neural profile of the behaviour-related activity (Fig 3B) is more clustered, in particular in the low gamma and HFa ranges. Indeed, silhouette scores are systematically greater for behaviour-related activity compared to global activity, indicating a better consistency of the clustering independently of the frequency band of interest (see Figure S2). Moreover, the clusters of the behaviour-related activity are most consistent in the HFa range and are located in the IFG BA44, the IPL BA 40 and the STG BA 41/42 and BA22 (see Figure S2).

Comparing global activity and behaviour-related activities shows that, for each frequency band, approximately 2/3 of the channels are only significant in the global activity (Figure 3C, orange part). Of the remaining third, half of the significant channels show a modulation of power compared to baseline that also significantly correlates with the quality of behavioral synchronisation (Figure 3C, blue part). The other half are only visible in the brain-behaviour correlation analysis (Figure 3C, green part, behaviour-specific).


A)
Power (%)
-100
-40
0
40
100
B)
r values
-1
0
1
C)
Significant channels
Global activity
Global & behaviour-related
Behaviour-specific activity
Delta
(1-4 Hz)
Theta
(4-8 Hz)
Alpha
(8-13 Hz)
Beta
(13-30 Hz)
Low gamma
(30-50 Hz)
HFa
(70-125 Hz)
31%
30%
39%
42%
37%
44%

361

Figure 3 – Power spectrum analyses. *Each dot represents a significant channel (permutation test).* ***(A)*** *Global activity* ***(****Task versus Rest) for each frequency band. The activity is expressed in % of power change compared to resting.* ***(B)*** *Behaviour-related activity: r values of the Spearman correlation across trials between the iEEG power and the verbal coordination index (VCI).* ***(C)*** *The proportion of significant channels in the task vs rest (orange), in the brain-behaviour correlation (green) and in both (blue). The percentage in the center indicates the overall proportion of significant channels from the three categories (with respect to the total number of channels).*

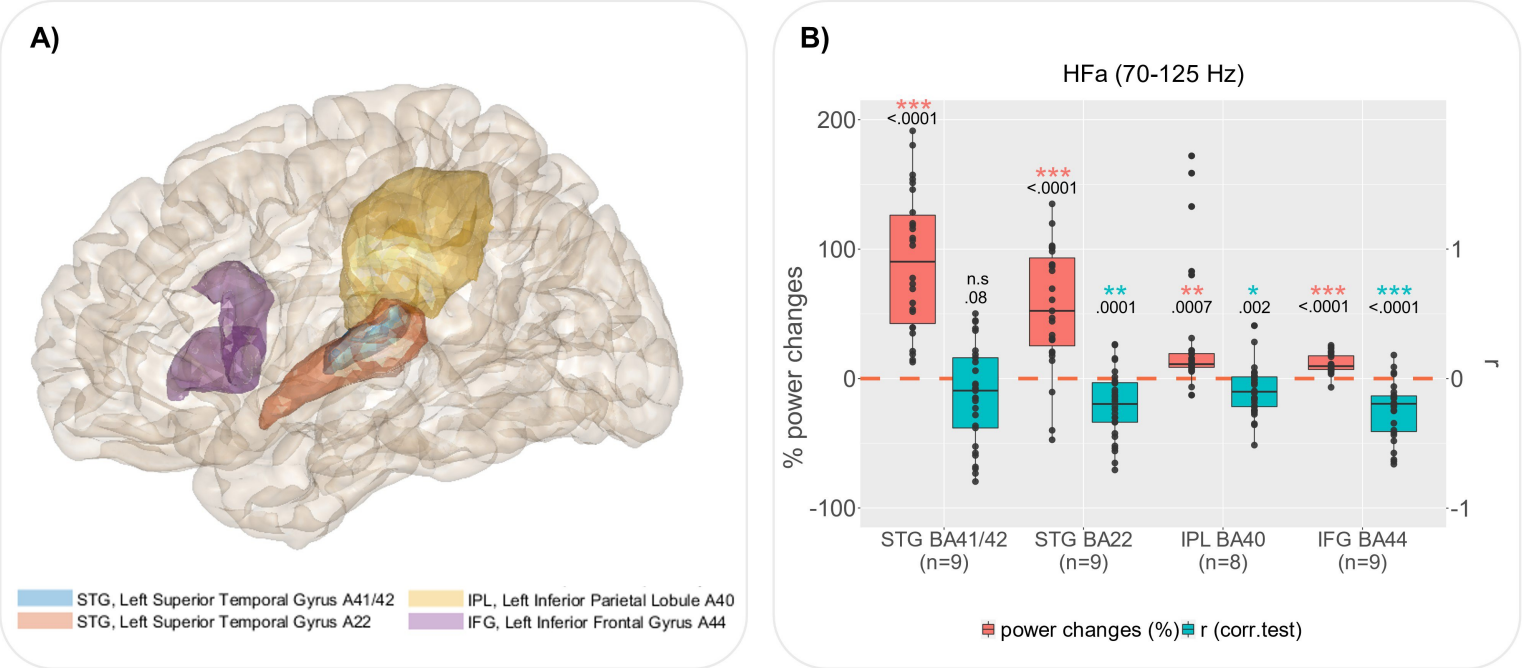


Figure 4 – Group analysis by regions of interest. ***(A)*** *Regions of interest (ROI) defined according to the cluster analysis (see Figure S2), the delimitation of regions is based on the Brainnetome atlas.* ***(B)*** *For each ROI, boxplots illustrate, in red, channels with significant global power changes (HFa, task vs rest) and, in blue, their corresponding r values (correlation between HFa power and verbal coordination index, VCI). Red and blue stars indicate a significant difference from a null distribution. Dots represent independent iEEG channels. The « n » below each region of interest specifies the number of patients. STG : superior temporal gyrus ; IPL : inferior parietal lobule ; IFG : inferior frontal gyrus ; BA : brodmann area.*

Spectral profiles in the language network are nuanced by behaviour.

In order to further explore the brain-behaviour relation in an anatomically language-relevant network, we focused on the frequency band yielding the most consistent clustering (HFa) and on those regions (ROI, in the left hemisphere) recorded in at least 7 patients: STG BA41/42 (primary auditory cortex), STG BA22 (secondary auditory cortex), IPL BA40 (inferior parietal gyrus) and IFG BA44 (inferior frontal gyrus, Broca's region). Within each ROI, using spearman correlation, we quantified the presence and strength of the link between neural activity and the degree of behavioral coordination.

Figure 4B shows a dramatic decrease of HFa along the dorsal pathway. However, while STG BA41/42 presents the strongest power increase (compared to baseline), it shows no significant correlation with verbal coordination (t=-1.81, ns). By contrast, the STG BA22 shows a significant negative correlation in the HFa band (t=-4.40, p<0.001), marking a fine distinction between primary and secondary auditory cortex. Finally, the brain-behavior correlation is maximal in the IFG BA44 (t=-5.60, p<0.001).

The IFG is sensitive to speech coordination dynamics.

To model the relation between verbal coordination dynamics and neural activity we conducted a behavior-brain phase amplitude coupling (PAC) analysis at the single trial level, using the power of high frequency neural activity (HFa: 70-125 Hz). For the low-frequency behavioral phase signal, this analysis investigated either the phase of the speech signals of the virtual partner (VP) or the verbal coordination dynamics (phase difference).

When looking at the whole brain analysis (Figure 5A), coupling is strongest as expected in the auditory regions (STG BA41/42 and STG BA22), but it is also present in the left IPL and IFG. Notably, when comparing - within the regions of interest previously described (Figure 5B) - the PAC with the virtual partner speech and the PAC with the phase difference, the coupling relationship changes when moving along the dorsal pathway: a stronger coupling in the auditory regions with the speech input, no difference between speech and coordination dynamics in the IPL and a stronger coupling for the coordinative dynamics compared to speech signal in the IFG (Figure 5B). Similar results were obtained when using the phase of the patient speech rather than the VP speech .

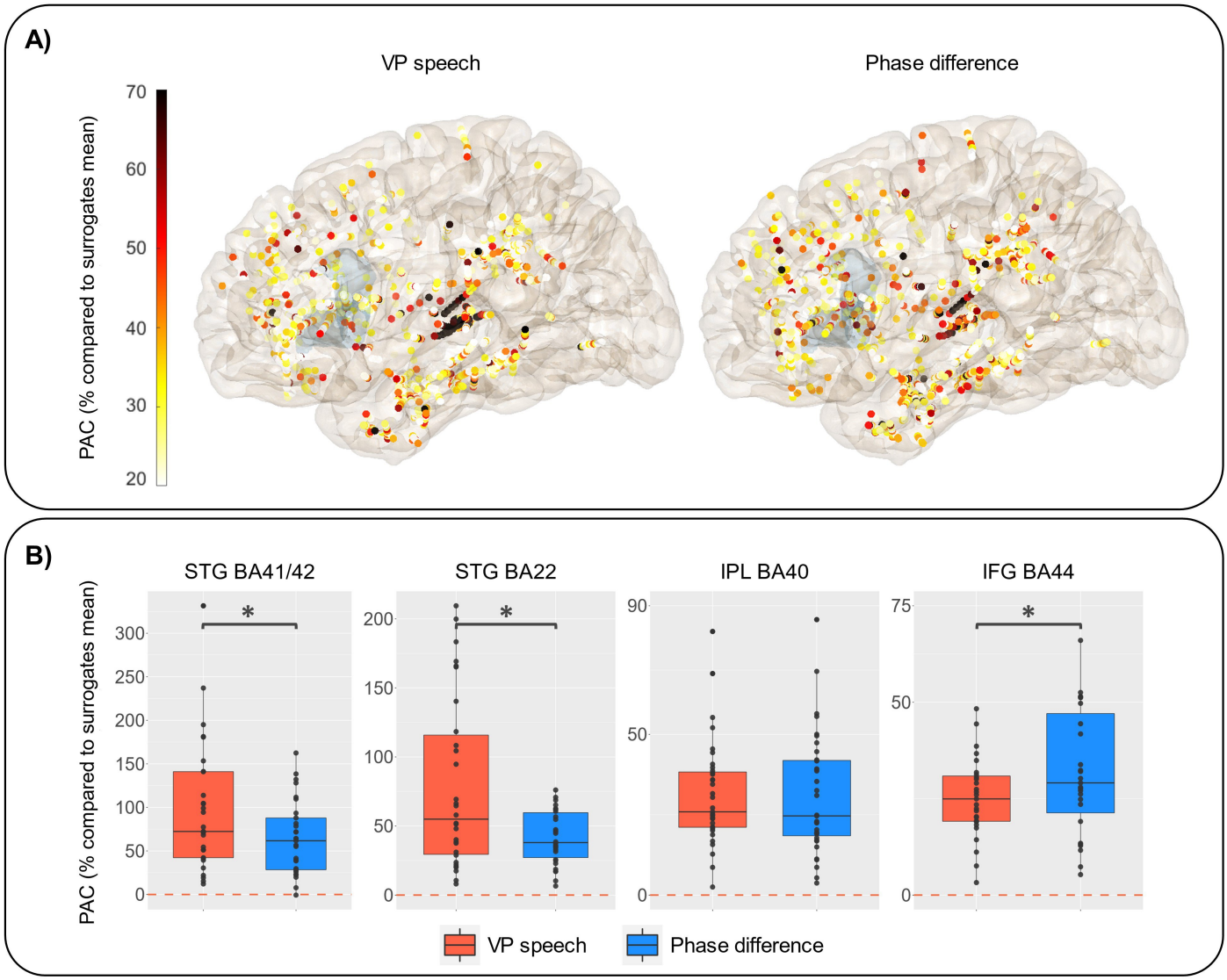


Figure 5 – Phase-amplitude coupling between virtual partner speech signal or coordination dynamics and high frequency activity (HFa). A) *Representation of the increase in PAC expressed in % compared to surrogates when using the virtual partner speech (left) or the coordination dynamics (phase difference between VP and patient, right).The shaded (slight blue) area corresponds to the location of the IFG BA44.* ***B)*** *PAC values for VP (in red) and phase difference (in blue) by regions of interest. Statistical difference calculated using paired wilcoxon test (STG BA41/42 : p=0.01 ; STG BA22 : p=0.004 ; IPL BA40 : p=0.6 ; IFG BA44 : p=0.02). Y-axis range has been adjusted to better illustrate the contrast between VP speech and coordination dynamics.*

Discussion

In this study, we investigated the speech coordinative adjustments using a novel interactive synchronous speech task (Lancia et al., 2017). To assess the relation between speech coordination and neural dynamics, we capitalized on the excellent spatiotemporal sensitivity of human stereotactic recordings (sEEG) from 16 patients with drug-resistant epilepsy while they produced short sentences along with a virtual partner (VP). Critically, the virtual partner is able to coordinate and adapt in real-time its verbal production with those of the participants, thus enabling to create a variable context of coordination yielding a broad distribution of phase delays between participant and VP productions. Several interesting findings can be emphasized. Firstly, the task involving both speech production and perception is efficient to highlight both the dorsal and ventral pathways, from low to high frequency activity (Figure 3). Secondly, spectral profiles of neural responses in the language network are nuanced when combined with behaviour, highlighting the fact that some regions are « simply » involved in the task, in a general manner, while others are truly sensitive to the quality of verbal coordination (Figure 3). Thirdly, high-frequency activity in secondary auditory regions shows a stronger sensitivity to behaviour (coordination success) compared to primary auditory regions (Figure 4). Finally, the high-frequency activity of the IFG BA44 specifically indexes the online coordinative adjustments that are continuously required to compensate the deviation from synchronisation (Figure 5).

Secondary auditory regions are more sensitive to coordinative behaviour.

The increase compared to baseline observed in high frequency activity (HFa) in both the STG BA41/42 (primary auditory cortex) and STG BA22 (secondary auditory cortex) is associated to different cognitive functions. Indeed, when considering whether such activity correlates with the behavioural coordination index – here the phase locking value between the patient and speaker speech – only the STG BA22 shows a significant correlation with a reduced HFa in strongly coordinated trials. This spatial distinction made between primary and secondary auditory regions in terms of their sensitivity to task demands has already been observed in the auditory cortex in particular in the high frequency activity (Nourski, 2017). The use of auditory target detection paradigms – phonemic categorisation (Chang et al., 2011), tone detection (Steinschneider et al., 2014), semantic categorisation tasks (Nourski et al., 2015) – wherein participants are asked to press a button when they hear the target, revealed task-dependent modulations only in the STG (posterolateral superior temporal gyrus or PLST) and not in Heschl's gyrus. In these studies, modulations corresponded to increases in activity in response to targets compared to non-targets, which has been interpreted in terms of selective attention or a bias toward behaviorally relevant stimuli (Petkov et al., 2004 ; Mesgarani & Chang, 2012). Here we extend these findings to a more complex and interactive context. The negative correlation observed in STG BA22 can be interpreted through the prism of internal forward model theory, suggesting that, during actions, a copy of the motor signal is transmitted to sensory regions to be used as a predictive signal to modulate perception and sensory processing. Thereby, the decrease in high-frequency activity as speech coordination increases could be explained by the closer correspondence between the speech signal perceived from the VP and that produced by the patient. This phenomenon has been previously described as speaker-induced suppression (SIS), characterised by a suppression of the neural response corresponding to the speech signal of one's own voice compared to the speech signals of others (Niziolek et al. 2013 ; Kurteff et al., 2023). In the present task, auditory feedback from the on-going speech stream is used to maintain fluent speech production. Moreover, when the two speech signals come close enough in time, the patient possibly perceives them as its own voice. Although this is not in line with previously described absence of SIS during synchronous speech (Jasmin et al., 2016), this echoes more recent findings of Ozerk et al. (2022) who showed, using an auditory feedback task with different fixed delay conditions, a significant response enhancement in auditory cortex that scaled according to the

duration of feedback delay. Overall, our result well fit within the framework of prediction errors and error maps (Friston et al., 2020; Tourville & Guenther, 2011), extending the use of synchronous speech to the study of internal forward models of speech in a more dynamic and less strict context.

Inferior frontal gyrus (BA44) as a seat of speech coordination planning in dynamic context.

Our results highlight the involvement of the inferior frontal gyrus (IFG), in particular the BA44 region, in speech coordination. First, trials with a weak verbal coordination (VCI) are accompanied by more prominent high frequency activity (HFa, Fig.4). Secondly, when considering the within-trial time-resolved dynamics, the phase-amplitude coupling (PAC) reveals a tight relation between the low frequency behavioral dynamics (phase) and the modulation of high-frequency neural activity (amplitude, Fig.5B). This relation is strongest when considering the phase adjustments rather than the phase of speech of the VP per se: larger deviation in verbal coordination are accompanied by increase in high-gamma activity. Overall, these findings are in line with the importance of higher-level frontal mechanisms for behavioural flexibility and their role in the hierarchical generative models underlying speech perception and production (Cope et al., 2017). More precisely, they are in line with works redefining the role of Broca's area (BA44 and BA45) in speech production as being more associated to speech planning rather than articulation per se (Flinker et al., 2015 ; Basilakos et al., 2018). Indeed, electrodes covering Broca's area show a greatest activity before the onset of articulation and not during speech production. This has been interpreted in favour of a role at a prearticulatory stage rather than an « on-line » coordination of the speech articulators at least in picture naming (Schuhmann et al., 2009), word repetition (Flinker et al., 2015 ; Ferpozzi et al., 2018) and turn-taking (Castellucci et al., 2022). According to these studies Broca's area may be a « functional gate » at a preacticulatory stage, allowing the phonetic translation before speech articulation (Ferpozzi et al., 2018). Our use of a synchronous speech task allows to refine this view, by showing that these prearticulatory commands are of continuous rather than discrete nature.

In other terms, the discrete (on-off) and ignition-like behaviour of neuronal populations in Broca's area gating prearticulatory commands before speech may be due to the discrete nature of the tasks used to assess speech production. Notably, picture naming, word repetition, word reading and even turn-onsets imply that speech production is preceded by a silent period during which the speaker listens to speech or watches pictures. By contrast, the synchronous speech task requires continuous temporal adjustments of verbal productions in order to reach synchronisation with the virtual partner. Relatedly, the involvement of IFG in accurate speech timing has been previously shown via thermal manipulation (Long et al., 2016).

Of note, temporal adjustements (prediction error corrections) are also needed for fluent speech in general, beyond synchronous speech, and give rise to the rhythmic nature of speech. Temporal adjustments possibly take advantage of the auditory input and are referred to as audio-motor interactions that can be modeled as a coupled oscillator (Poeppel & Assaneo , 2020). Interestingly, during speech perception the coupling of theta and gamma bands in the auditory cortex reflects tracking of slow speech fluctuations to spiking gamma (Morillon et al., 2010 ; Morillon et al., 2012 ; Hyafil et al., 2015; Lizarazu et al. 2019 ; Oganian & Chang, 2019 ; Leonard et al., 2024), similarly to what we describe in the auditory cortex. By contrast, in the inferior frontal gyrus, the coupling in the high frequency activity is strongest with the input-output phase difference (input of the VP - output of the speaker), a metric that reflects the amount of error in the internal computation to reach optimal coordination, which indicates that this region optimises the predictive and coordinative behaviour required by the task. This well fits with the anatomical connectivity that has been described between Broca's and Wernicke's territory via the long segment of the arcuate fasciculus, possibly setting the base for a mapping from speech representations in Wernicke's area to the predictive proprioceptive adjustments processed in Broca's area (Catani & Ffytche, 2005 ; Oestreich et al., 2018).

Finally, while the case of synchronous speech may seem quite far away from real-life conversational contexts, the models describing language interaction consider that listeners covertly imitate the speaker's speech and timely construct a representation of the underlying communicative intention which allows early fluent turn-taking (Pickering & Gambi, 2018 ; Levinson, 2016). Moreover, synchronous speech has recently gained interest in the neuroscience field due to important results showing a relation between anatomo-functional features and synchronization abilities. More precisely, Assaneo and collaborators (2019) used a spontaneous speech synchronization test wherein participants produce the syllable /tah/ while listening to a random syllable sequence of predefined pace. The authors identified two groups of participants (high and low synchronisers) characterised by their ability to naturally synchronise their productions more or less easily with the auditory stimulus. Importantly, the ability to synchronise correlates to the degree of lateralization of the arcuate fasciculus, that connects the inferior frontal gyrus and the auditory temporal regions, high synchronizers showing greater lateralization to the left than the low synchronizers. The more this structural connectivity of the arcuate fasiculus is lateralised in the left hemisphere, the more the activity of the IFG is synchronised with the envelope of the audio stimulus of the SSS-test, during a passive listening task. A major limitation of EEG and MEG studies is that they are very sensitive to speech production artefacts, which is not the case of iEEG. Thus, the full dynamics of speech interaction are difficult to assess with surface recordings. Our findings extend these results in several manners. First, speech production is not limited to a single syllable, but to complex utterances. Secondly, the auditory stimulus is not preset but adapts and changes behaviour in real-time as a function of the dynamic of the « dyad », here patient-virtual partner. Thirdly, and most importantly, iEEG recording allowing speech artefact-free data, we could extend the relation between coordination abilities and anatomical circuitry of the IFG to the neural dynamics of this same region, showing that it plays an important role in the temporal adjustement of speech that are necessary to synchronize to external speech.

To conclude, the present study illustrates the possibility and interest of using a fully dynamic, adaptive and interactive language task to gather deeper understanding of the subtending neural dynamics involved in speech perception, production as well as their interaction.

Supplementary Material

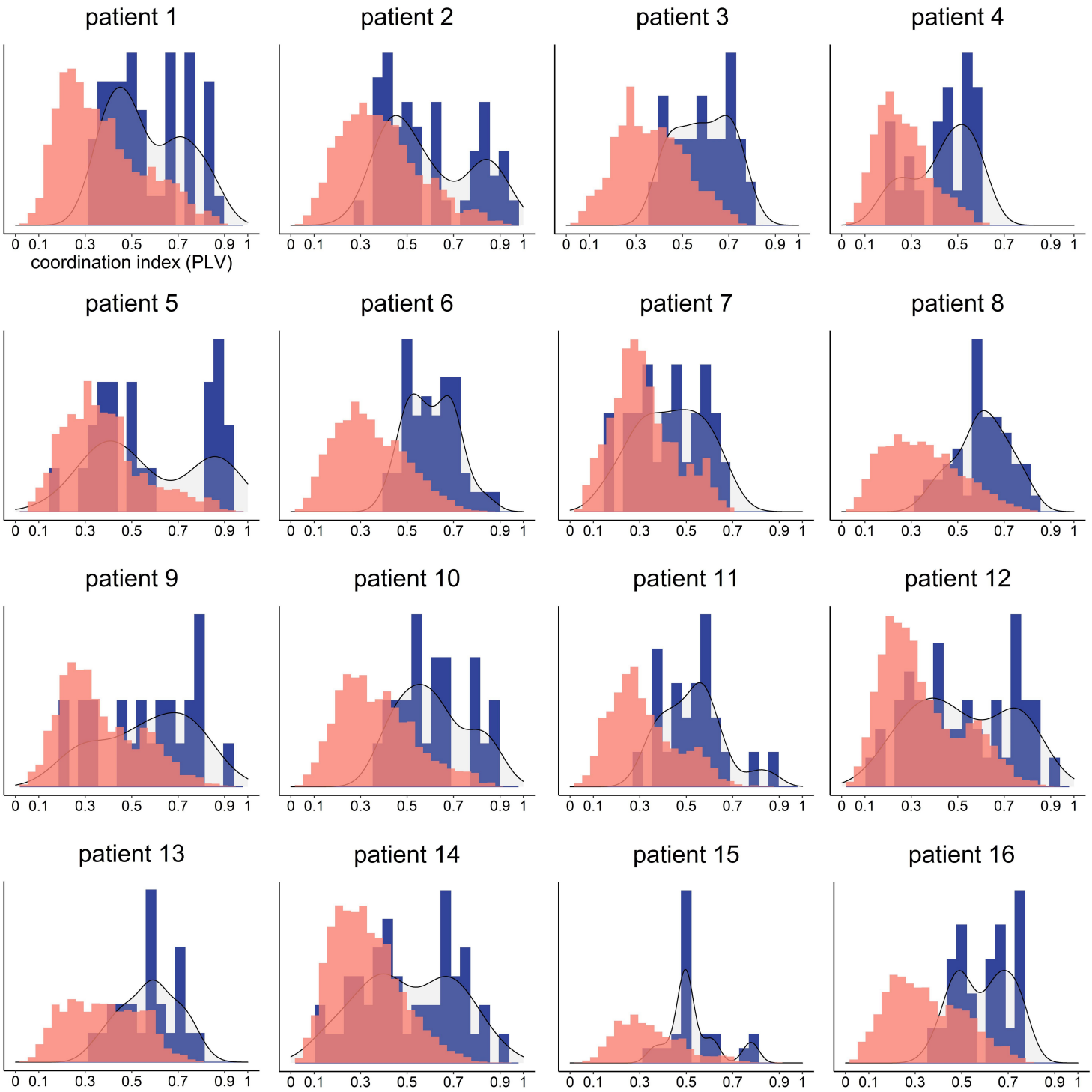


Figure S1 – Distribution of verbal coordination index across patients. *For each of the sixteen patients, this figure depicts the histogram of the coordination index for all trials (in blue) as well as the null distribution (random phase shift) computed using 500 permutations per trial (in red).*

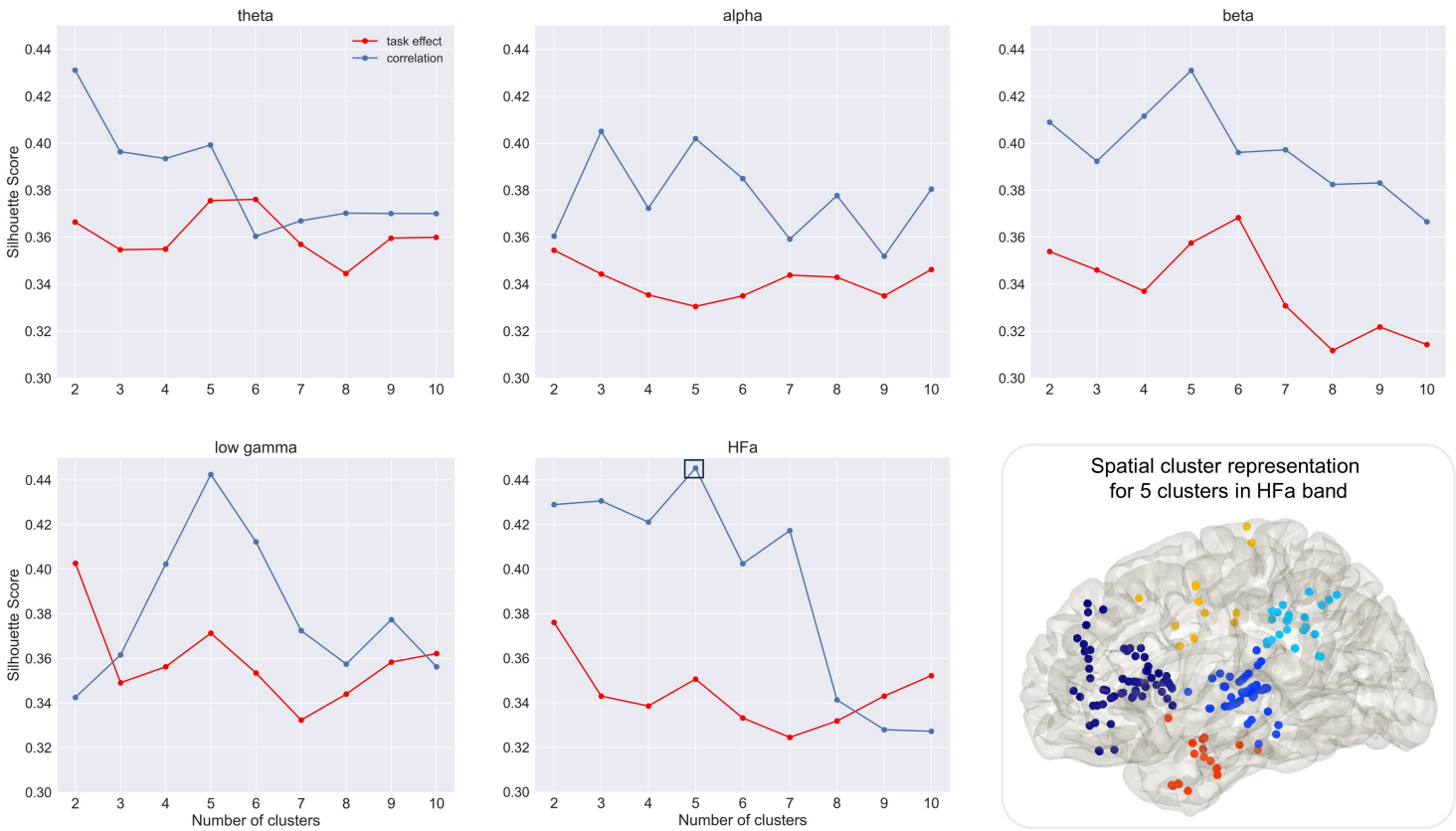


Figure S2 – Cluster analysis (silhouette score). *Illustration of the mean silhouette scores according to the number of clusters for global activity (in red) and behaviour-related activity (correlation between power changes and coordination index, in blue).* The highest silhouette score was obtained for five clusters in the HFa range for behaviour-related activity (framed in full black square). *Bottom right : spatial cluster representation for the highest mean silhouette score value in HFa range.*

References

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. Nature neuroscience, 22(4), 627-632.

Basilakos, A., Smith, K. G., Fillmore, P., Fridriksson, J., & Fedorenko, E. (2018). Functional characterization of the human speech articulation network. Cerebral Cortex, 28(5), 1816-1830.

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. Research on Language and Social Interaction, 50(1), 71-89.

Bögels, S., Magyari, L., & Levinson, S. C. (2015). Neural signatures of response planning occur midway through an incoming question in conversation. Scientific reports, 5(1), 12881.

Bradshaw, A. R., & McGettigan, C. (2021). Convergence in voice fundamental frequency during synchronous speech. PloS one, 16(10), e0258747.

Bradshaw, A. R., Lametti, D. R., Shiller, D. M., Jasmin, K., Huang, R., & McGettigan, C. (2023). Speech motor adaptation during synchronous and metronome-timed speech. Journal of Experimental Psychology: General. to articulate the transition.

Castellucci, G. A., Kovach, C. K., Howard III, M. A., Greenlee, J. D., & Long, M. A. (2022). A speech planning network for interactive language use. *Nature*, *602*(7895), 117-122.

Catani, M., & Ffytche, D. H. (2005). The rises and falls of disconnection syndromes. *Brain*, *128*(10), 2224-2239.

Chang, E. F., Edwards, E., Nagarajan, S. S., Fogelson, N., Dalal, S. S., Canolty, R. T., ... & Knight, R. T. (2011). Cortical spatio-temporal dynamics underlying phonological target detection in humans. Journal of cognitive neuroscience, 23(6), 1437-1446.

Combrisson, E., Nest, T., Brovelli, A., Ince, R. A., Soto, J. L., Guillot, A., & Jerbi, K. (2020). Tensorpac: An open-source Python toolbox for tensor-based phase-amplitude coupling measurement in electrophysiological brain signals. PLoS computational biology, 16(10), e1008302.

Cope, T. E., Sohoglu, E., Sedley, W., Patterson, K., Jones, P. S., Wiggins, J., ... & Rowe, J. B. (2017). Evidence for causal top-down frontal contributions to predictive processes in speech perception. *Nature communications*, *8*(1), 2154.

Corps, R. E., Gambi, C., & Pickering, M. J. (2018). Coordinating utterances during turn-taking: The role of prediction, response preparation, and articulation. Discourse processes, 55(2), 230-240.

Cummins, F. (2002). On synchronous speech. Acoustics Research Letters Online, 3(1), 7-11.

Cummins, F. (2003). Practice and performance in speech produced synchronously. Journal of Phonetics, 31(2), 139-148.

Cummins, F. (2009). Rhythm as entrainment: The case of synchronous speech. Journal of Phonetics, 37(1), 16-28.

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, *53*(1), 1-15.

Dixon, S. (2005, September). Live tracking of musical performances using on-line time warping. In *Proceedings of the 8th International Conference on Digital Audio Effects* (Vol. 92, p. 97).

Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., ... & Jiang, T. (2016). The human brainnetome atlas: a new brain atlas based on connectional architecture. *Cerebral cortex*, *26*(8), 3508-3526.

Ferpozzi, V., Fornia, L., Montagna, M., Siodambro, C., Castellano, A., Borroni, P., ... & Cerri, G. (2018). Broca's area as a pre-articulatory phonetic encoder: gating the motor program. *Frontiers in human neuroscience*, *12*, 64.

Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., & Crone, N. E. (2015). Redefining the role of Broca's area in speech. Proceedings of the National Academy of Sciences, 112(9), 2871-2875.

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.

Floegel, M., Kasper, J., Perrier, P., & Kell, C. A. (2023). How the conception of control influences our understanding of actions. *Nature Reviews Neuroscience*, *24*(5), 313-329.

Friston, K. J., Parr, T., Yufik, Y., Sajid, N., Price, C. J., & Holmes, E. (2020). Generative models, linguistic communication and active inference. *Neuroscience & Biobehavioral Reviews*, *118*, 42-64.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. Topics in Cognitive Science, 1(2), 292-304.

Groppe, D. M., Bickel, S., Dykstra, A. R., Wang, X., Mégevand, P., Mercier, M. R., ... & Honey, C. J. (2017). iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. *Journal of neuroscience methods*, *281*, 40-48.

Hashimoto, Y., & Sakai, K. L. (2003). Brain activations during conscious self-monitoring of speech production with delayed auditory feedback: An fMRI study. Human brain mapping, 20(1), 22-28.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: computational basis and neural organization. Neuron, 69(3), 407-422.

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., ... & Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. Neuroreport, 8(9), 2379-2382.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. Frontiers in human neuroscience, 5, 82.

Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A. L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. *elife*, *4*, e06213.

Jasmin, K. M., McGettigan, C., Agnew, Z. K., Lavan, N., Josephs, O., Cummins, F., & Scott, S. K. (2016). Cohesion and joint speech: right hemisphere contributions to synchronized vocal production. *Journal of Neuroscience*, *36*(17), 4669-4680.

Karlin, R., Naber, C., & Parrell, B. (2021). Auditory feedback is used for adaptation and compensation in speech timing. Journal of Speech, Language, and Hearing Research, 64(9), 3361-3381.

Kuramoto, Y. (1975). Self-entrainment of a population of coupled non-linear oscillators. In *International Symposium on Mathematical Problems in Theoretical Physics: January 23–29, 1975, Kyoto University, Kyoto/Japan* (pp. 420-422). Springer Berlin Heidelberg.

Kurteff, G. L., Lester-Smith, R. A., Martinez, A., Currens, N., Holder, J., Villarreal, C., ... & Hamilton, L. S. (2023). Speaker-induced suppression in EEG during a naturalistic reading and listening task. Journal of Cognitive Neuroscience, 35(10), 1538-1556.

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, *8*(4), 194-208.

Lancia, L., Chaminade, T., Nguyen, N., & Prevot, L. (2017, August). Studying the link between inter-speaker coordination and speech imitation through human-machine interactions. In Interspeech 2017 (p 859-863).

Leonard, M. K., Gwilliams, L., Sellers, K. K., Chung, J. E., Xu, D., Mischler, G., ... & Chang, E. F. (2024). Large-scale single-neuron speech sound encoding across the depth of human cortex. *Nature*, *626*(7999), 593-602.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. Trends in cognitive sciences, 20(1), 6-14.

Lizarazu, M., Lallier, M., & Molinaro, N. (2019). Phase− amplitude coupling between theta and gamma oscillations adapts to speech rate. *Annals of the new York Academy of Sciences*, *1453*(1), 140-152.

Long, M. A., Katlowitz, K. A., Svirsky, M. A., Clary, R. C., Byun, T. M., Majaj, N., ... & Greenlee, J. D. (2016). Functional segregation of cortical regions underlying speech timing and articulation. Neuron, 89(6), 1187-1193.

Magyari, L., Bastiaansen, M. C., De Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. Journal of Cognitive Neuroscience, 26(11), 2530-2539.

Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in cognitive science*, *1*(2), 320-339.

Mercier, M. R., Bickel, S., Megevand, P., Groppe, D. M., Schroeder, C. E., Mehta, A. D., & Lado, F. A. (2017). Evaluation of cortical local field potential diffusion in stereotactic electro-encephalography recordings: a glimpse on white matter signal. *Neuroimage*, *147*, 219-232.

Mercier, M. R., Dubarry, A. S., Tadel, F., Avanzini, P., Axmacher, N., Cellier, D., ... & Oostenveld, R. (2022). Advances in human intracranial electroencephalography research, guidelines and good practices. *Neuroimage*, *260*, 119438.

Mesgarani, N., & Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. Nature, 485(7397), 233-236.

Morillon, B., Lehongre, K., Frackowiak, R. S., Ducorps, A., Kleinschmidt, A., Poeppel, D., & Giraud, A. L. (2010). Neurophysiological origin of human brain asymmetry for speech and language. *Proceedings of the National Academy of Sciences*, *107*(43), 18688-18693.

Morillon, B., Liégeois-Chauvel, C., Bénar, C. G., & Giraud, A. L. (2012). Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Frontiers in psychology*, *3*, 24108.

Mukherjee, S., Badino, L., Hilt, P. M., Tomassini, A., Inuggi, A., Fadiga, L., ... & d'Ausilio, A. (2019). The neural oscillatory markers of phonetic convergence during verbal interaction. Human brain mapping, 40(1), 187-201.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. Journal of Neuroscience, 33(41), 16110-16116.

Nourski, K. V. (2017). Auditory processing in the human cortex: An intracranial electrophysiology perspective. Laryngoscope investigative otolaryngology, 2(4), 147-156.

Nourski, K. V., Steinschneider, M., Oya, H., Kawasaki, H., & Howard III, M. A. (2015). Modulation of response patterns in human auditory cortex during a target detection task: an intracranial electrophysiology study. International Journal of Psychophysiology, 95(2), 191-201.

Oestreich, L. K., Whitford, T. J., & Garrido, M. I. (2018). Prediction of speech sounds is facilitated by a functional fronto-temporal network. Frontiers in Neural Circuits, 12, 43.

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science advances*, *5*(11), eaay6279.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*, 1-9.

Ozker, M., Doyle, W., Devinsky, O., & Flinker, A. (2022). A cortical network processes auditory error signals during human speech production to maintain fluency. PLoS Biology, 20(2), e3001493.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (Eds.). (2011). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

Petkov, C. I., Kang, X., Alho, K., Bertrand, O., Yund, E. W., & Woods, D. L. (2004). Attentional modulation of human auditory cortex. Nature neuroscience, 7(6), 658-663.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. Psychological bulletin, 144(10), 1002.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. Behavioral and brain sciences, 27(2), 169-190.

Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature reviews neuroscience*, *21*(6), 322-334.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, *20*, 53-65.

Schuhmann, T., Schiller, N. O., Goebel, R., & Sack, A. T. (2009). The temporal characteristics of functional activation in Broca's area during overt picture naming. *cortex*, *45*(9), 1111-1116.

Shahapure, K. R., & Nicholas, C. (2020, October). Cluster quality analysis using silhouette score. In *2020 IEEE 7th international conference on data science and advanced analytics (DSAA)* (pp. 747-748). IEEE.

Steinschneider, M., Nourski, K. V., Rhone, A. E., Kawasaki, H., Oya, H., & Howard III, M. A. (2014). Differential activation of human core, non-core and auditory-related cortex during speech categorization tasks as revealed by intracranial recordings. Frontiers in neuroscience, 8, 240.

Stuart, A., Kalinowski, J., Rastatter, M. P., & Lynch, K. (2002). Effect of delayed auditory feedback on normal speakers at two speech rates. The Journal of the Acoustical Society of America, 111(5), 2237-2241.

Takaso, H., Eisner, F., Wise, R. J., & Scott, S. K. (2010). The effect of delayed auditory feedback on activity in the temporal lobe while speaking: a positron emission tomography study.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. Language and cognitive processes, 26(7), 952-981.

Villalon, S. M., Paz, R., Roehri, N., Lagarde, S., Pizzo, F., Colombet, B., ... & Bénar, C. G. (2018). EpiTools, A software suite for presurgical brain mapping in epilepsy: Intracerebral EEG. *Journal of neuroscience methods*, *303*, 7-15.

Yamamoto, K., & Kawabata, H. (2014). Adaptation to delayed auditory feedback induces the temporal recalibration effect in both speech perception and production. Experimental brain research, 232, 3707-3718.