Hitting the right pitch: Cortical tracking of fundamental frequency changes across speech rates in auditory and sensorimotor regions

Yorguin-Jose Mantilla-Ramos[1,2,] Ana-Sofía Hincapié-Casas[1], Annalisa Pascarella[3], Tarek Lajnef[1], Richard M. Leahy[4], Emily Coffey[5] Karim Jerbi[1,6*] and Véronique Boulenger[7*]

[1] Cognitive and Computational Neuroscience Laboratory (CoCo Lab), Psychology Department, University of Montreal, Montreal, Montréal, QC H3C 3J7, Canada

[2] Neuropsychology and Behavior Group (GRUNECO), Medical School, Universidad de Antioquia, Antioquia, 050010 Colombia

[3] Institute for Applied Mathematics Mauro Picone, National Research Council, 00185 Rome, Italy

[4] Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, 90089 California, USA

[5] Department of Psychology, Concordia University, Montreal, QC, H3G 1M8 Canada

[6] Mila (Quebec AI Institute), QC, H2S 3H1 Canada

[7] CNRS, Université Lyon 2, Laboratoire Dynamique du Langage, 69007 Lyon, France

[*] These authors equally contributed to this work.

Corresponding authors: yorguin.jose.mantilla.ramos@umontreal.ca and veronique.boulenger@cnrs.fr

Abstract

Neuronal entrainment to speech properties is essential for language processing, particularly through oscillatory tracking of slower rhythms corresponding to the syllabic rate. However, it remains less explored whether brain rhythms also synchronize with higher-frequency speech components, particularly the fundamental frequency (F0) or pitch. We used magnetoencephalography (MEG) to investigate cortical tracking of F0 while participants listened to sentences produced at natural normal and fast rates, but also to time-compressed speech. We examined how pitch changes accompanying natural increases in speech rate modulate brain-to-speech coupling and compared this with artificially accelerated speech, where F0 is unchanged. We also explored whether this coupling is confined to the auditory cortex or involves a broader cortical network. We computed whole-brain cortico-acoustic coupling between the MEG source time-series and the speech signal, alongside spectral power modulations in frequency bands centered on the mean F0 of the speech material. We found significant brain-to-F0 coupling in the right auditory, inferior parietal, insular, and pre- and postcentral regions across all speech rate conditions. Importantly, the peak neuro-acoustic coupling frequency adjusted to reflect the F0 increase due to natural rate acceleration. Interestingly, we found significant brain-speech coupling around F0 not only in the primary auditory cortex but also in a postcentral somatosensory region likely corresponding to the ventral larynx area. These findings provide new insights into frequency-specific cortical tracking of F0 during the perception of naturally-produced speech at varying rates and suggest the involvement of an auditory-somato-motor network that may facilitate the recognition of articulatory features during speech perception.

Significance statement

Low-frequency neural oscillations play a key role in the parsing of speech into syllabic and phrasal units. Little is known however about the cortical tracking of higher-frequencies in speech such as the fundamental frequency (F0), resulting from vocal fold vibration during production and conveying phonemic and prosodic cues for comprehension. Using magnetoencephalography (MEG), we show that neural oscillations synchronize with speech F0 but most importantly, that they adapt their coupling frequency to match F0 variations accompanying natural speech rate changes. This coupling occurs beyond the right primary auditory cortex to include regions involved in auditory-somato-motor integration. These findings suggest a dynamic interplay between auditory and motor systems in the encoding of pitch and articulatory information during speech perception.

Keywords

Speech perception, Neural Oscillations, Fundamental frequency, Pitch, Cortico-acoustic coupling, Frequency-Following Response (FFR), Magnetoencephalography (MEG)

Introduction

Speech periodicity is a temporal cue of chief importance for speech perception and communication as it conveys segmental (voicing, manner of articulation) and suprasegmental (intonation, stress) information (Rosen, 1992). It results from the quasi-periodic vibration of the vocal folds during production and includes a representation of the speaker's fundamental frequency (F0), which is the prime acoustic correlate of voice pitch. Many studies on pitch encoding measured the Frequency-Following Response (FFR), a sustained oscillatory response generally recorded by averaging responses to repeated syllables or tones that mimics stimulus periodicity and reflects precise auditory temporal processing (Russo et al., 2004; Kraus and Nicol, 2005; Aiken and Picton, 2008; Coffey et al., 2019). Although the FFR recorded in EEG at the scalp is believed to originate primarily from the brainstem (Chandrasekaran and Kraus, 2010), contributions from the thalamus and auditory cortex were also reported, with cortical activity peaking at the stimulus F0 (Coffey et al., 2016, 2017; Bidelman, 2018; Hartmann and Weisz, 2019; Ross et al., 2020; Bidelman and Momtaz, 2021; Gnanateja et al., 2021; Guo et al., 2021). Brain tracking of F0 in continuous sound streams has been given less attention (but see Puschmann et al., 2019; Greenlaw et al., 2020), and studies available so far focusing on neural encoding of the F0 in naturalistic speech showed auditory cortical responses at the F0 of isolated sentences or stories (Hertrich et al., 2012; Kulasingham et al., 2020; Van Canneyt et al., 2021).

Here, we provide novel contribution to this line of research by assessing whether neural oscillations, in the auditory cortex but also in other cortical regions, track the F0 of natural speech rate variations. Besides spectro-temporal modifications (Berry, 2011), naturally speaking faster induces changes in the phonetic realization of F0 (Caspers and van Heuven, 1993; Fougeron and Jun, 1998). We aimed to characterize how F0 representation changes as a function of speech rate in natural speech. Following earlier work that low-frequency neural

oscillations (e.g., 2-9 Hz) adjust their coupling frequency to match the faster syllable rate (e.g., Pefkou et al., 2017; Hincapié-Casas et al., 2021a), we asked whether such frequency-shift also occurs for higher-frequency oscillations to track the F0 changes accompanying speech rate increase. Using MEG, we examined brain coupling to the F0 of sentences naturally produced at a normal or fast rate, as well as of artificially time-compressed sentences. Given that time-compression reduces signal duration while keeping its spectral and pitch content intact (Janse, 2004), time-compressed speech has the same F0 as normal rate speech despite an increased syllabic rate. By contrast, F0 is higher in naturally accelerated speech compared to both normal rate and time-compressed speech. Using Minimum Norm Estimation source reconstruction methods, we compared cortico-acoustic coupling and power modulations during sentence perception in the three speech rate conditions. In keeping with previous studies (Keitel et al., 2018; Hincapié-Casas et al., 2021b), we analysed coupling in two frequency ranges centred on the mean F0 of our sentence material (77-84 Hz for normal rate/time-compressed sentences and 90-97 Hz for natural fast speech). In addition, a whole-brain analysis allowed to investigate whether tracking of sentences' F0 only occurs in the auditory cortex or whether it involves a more distributed network, noting that previous work used region-of-interest based analyses (e.g., Coffey et al., 2016; Kulasingham et al., 2020). We predicted that if neural oscillations phase-lock to F0 to encode speech periodic information, cortico-acoustic coupling should be found in frequency ranges that are specific to the sentences' fundamental frequency. Neural tracking should thus occur in a higher frequency band for natural fast speech compared to normal rate speech, mirroring the increase in speech F0. On the contrary, no change in coupling frequency should be observed for time-compressed sentences. Such findings may bring evidence that by adjusting to the rate of vocal fold vibration, cortical oscillations contribute to the encoding of pitch and of phonemic information conveyed by the periodicity of the speech signal.

Materials and Methods

Participants

Twenty-four (14 females) participants aged between 18 and 45 years old (mean = 23 years old) volunteered for the study. All were monolingual French native speakers and right-handed (mean score at the Edinburgh handedness inventory = 94; Oldfield, 1971). None of them reported hearing or language problems nor history of neurological or psychiatric disorders. The protocol conformed to the Declaration of Helsinki and was approved by the local ethical committee (Comité de Protection des Personnes Lyon Sud-Est II; ID RCB: 2012-A00857-36). Written informed consent was obtained for all participants before the experiment, and they were compensated for their participation.

Stimuli

A set of 288 meaningful sentences (7-9 words) with the same syntactic structure (e.g., "Le guide organise une excursion dans le désert" / *The guide organizes a trip to the desert*) served as stimuli. A French native professional theatre actor, who was able to produce sentences at the required fast rate while remaining intelligible, recorded the stimuli in a sound-attenuated booth using ROC*me*! software (Ferragne et al., 2012). Each sentence was recorded twice (44.1 kHz, mono, 16 bits), first at a normal and then at a fast rate. The mean syllable rate of the sentences (i.e. number of produced syllables divided by sentence duration, as calculated with Praat, Boersma & Weenik, 2012) was 6.76 syllables/s (SD 0.57) for the normal rate condition, while it increased to 9.15 syllables/s (SD 0.60) in the natural fast condition. Time compression, which is a re-synthesis process only affecting the temporal structure of the stimuli, was then applied to normal rate sentences by digitally shortening them with a PSOLA algorithm (Pitch Synchronous Overlap and Add; Moulines & Charpentier, 1990). Compression rates were

computed for each sentence, namely each time-compressed sentence was matched in terms of syllable rate to its natural fast equivalent. Therefore, time-compressed sentences had a mean syllable rate of 9.15 syllables/s (SD 0.60) as well. For the total of 864 sound files (288×3 speech rate variants), we applied an 80 Hz high-pass filter and smoothed sentences' amplitude envelope initially and finally. The intensity of the sound files was finally peak normalized.

Computation of the power spectrum in the 64-128 Hz range for sentences in the three speech rate conditions using Python showed peak frequencies at 81.6 Hz (SD 3.2) and 81 Hz (SD 3.3) for the normal rate and time-compressed conditions respectively, and at 91.9 Hz (SD 9.8) for the fast rate condition (see Figure 1). These peak frequencies are consistent with their computation with Praat.

As control conditions, we generated amplitude-modulated noise stimuli (i.e., Gaussian white noise with no linguistic content) for each of the 288 normal rate and 288 fast rate sentences, using sentences' envelopes. We also added forty-eight filler sentences (different from the experimental stimuli but following the same syntactic structure) in which beep-sounds were inserted at different possible time-points (i.e. 200, 600 or 800 ms from onset) in the sentences for attentional control (see Experimental procedure). The whole set of stimuli was split into two experimental lists, each including 288 sentences (96 for each speech rate condition), 192 amplitude-modulated noise stimuli (96 modulated with the normal rate envelope and 96 with the fast rate envelope) and 48 filler sentences, for a total of 528 stimuli per list. Each stimulus appeared in each rate condition across all participants but only once per list.

Experimental procedure

While comfortably sitting in a sound-attenuated, magnetically-shielded recording room with a screen in front of them, participants were presented binaurally with stimuli through air-conducting tubes with foam insert earphones (Etymotic ER2 and ER3). Prior to the MEG

recording, their auditory detection thresholds were determined for each ear with a 1-minute pure tone. The level was then adjusted in order to present stimuli at 50 dB Sensation Level (SL) with a central position (stereo) with respect to the participant's head.

Each trial of the experiment ran as follows: after a fixation cross displayed at the centre of the screen for 1500 ms, an auditory stimulus (sentence or amplitude-modulated noise) was presented, with the fixation cross remaining on the screen. Participants were instructed to attentively listen to all stimuli while looking at the fixation cross and to press a button (response button Neuroscan, Pantev) with their left index finger as quickly as possible whenever they detected beep-sounds in filler sentences. The next trial started after an inter-trial interval (grey screen) of 1250 ms. Participants detected 100% of the filler trials, which were excluded from subsequent analysis. Each participant listened to the stimuli from one of the two experimental lists. The total of 528 stimuli were pseudo-randomly divided into 8 blocks (66 stimuli each) separated by short breaks to allow the participants to rest. Participants were familiarized with the stimuli and task in a training phase (5 sentences different from the experimental stimuli). To maintain their attention throughout the experiment, they were informed that questions on the heard sentences would be asked at the end of the experiment. More precisely, they had to indicate, by ticking a yes/no box on a paper sheet, whether 8 proposed sentences (3 of which were not experimental stimuli) were presented during the MEG experiment. The experiment was run with Presentation software (Neurobehavioral Systems).

MEG data acquisition and analysis

We recorded neuromagnetic brain activity of the 24 participants using a 275-channel whole-head MEG system (CTF OMEGA 275, Canada) at 1200 Hz sampling rate. Head position within the MEG helmet was monitored with three fiducial coils (nasion, left and right pre-auricular points) placed on the participant's head. Horizontal and vertical eye movements were also

recorded with four electrooculography (EOG) electrodes. We recorded reference head position before each of the 8 experimental blocks and tracked head movements throughout the experiment using continuous head position identification (HPI).

Preprocessing of the MEG data was done with custom written Matlab scripts (Mathworks Inc., MA, USA) and the Fieldtrip toolbox (Oostenveld et al., 2011). Subsequent coupling analyses were carried out in Python using custom scripts and the MNE-Python library (Gramfort et al., 2013). In the following we describe the processing of 1) speech recordings, 2) MEG data, 3) Magnetic Resonance Imaging (MRI) data, and the analyses of 4) source localization, power and coupling, and 5) coupling maps and statistics.

1) Audio Recordings (reference signals)

To examine the neural tracking of speech F0, we used the speech signal as the reference. Inspection of the F0 distributions and power spectral densities revealed some spectral (including F0s) overlap between sentences in the different speech rate conditions (see Figure 1 A and C). To test our hypothesis that brain oscillatory activity adjusts to F0 increase with speech rate acceleration, we optimized the distinction between conditions in terms of F0s by selecting a proportion of sentences in the natural normal rate and time-compressed conditions with the lowest F0s (71% and 77% of the sentences on average per participant, respectively). Similarly, we selected sentences in the natural fast rate condition with the highest F0s (~62% on average per participant). To do this, we collected sentences from the normal rate/time-compressed and fast rate conditions starting from opposite sides of the spectrum, thus maximizing the spectral distance between these two subsets while keeping an acceptable number of sentences in each condition (see Table 1). This resulted in normal rate and time-compressed sentences with the lowest F0s (i.e. below 83 Hz) and a F0 peak at 80 Hz, and natural fast rate sentences with the highest F0s (i.e. above 87 Hz) peaking at 93 Hz (see Figure 1 B and D). We analyzed speech-

brain coupling in two frequency ranges centered on the F0 peak values of the two subsets of sentences (with 7Hz bandwidth), namely 77-84 Hz and 90-97 Hz. For this coupling analysis, we selected speech segments between 200 ms and 1000 ms post-stimulus onset, as the shortest sentences in the natural fast and time-compressed conditions were ~1 s long (this time-window for speech segments equals the time-window of interest of the functional MEG data, see next section).

Condition	Lowest % (and number) of sentences kept across participants	Mean % (and number) of sentences kept across participants	Highest % (and number) of sentences kept across participants
Normal rate	66.67% (37)	70.95% (53.04)	79.63% (64)
Time-Compressed	71.25% (42)	77.41% (58.39)	83.13% (69)
Fast rate	55.88% (35)	62.10% (46.65)	70.27% (57)

Table 1: Mean percentage and percentage range of included sentences across participants by speech rate condition, after selecting sentences to reduce F0 overlap between subsets of sentences in the three conditions. For each MEG recording (corresponding to a participant and a condition), sentences were selected to reduce the F0 overlap between the Fast rate condition and the other two conditions (Normal rate and Time-Compressed). After this procedure, the number and percentage of included sentences within that MEG recording were calculated. Subsequently, the inclusion values were characterized by mean and range (lowest, highest) across participants within a condition. None of the participants was excluded after this selection procedure. The minimum % of included sentences is ~56% for one participant in the fast rate condition. Nevertheless, for this condition, the overall mean inclusion percentage is ~62% (with a maximum of ~70%). All conditions have around 3% standard deviation of the inclusion percentage. The values in brackets represent the number of sentences corresponding to the

given percentage. For the range, these are integer numbers as they represent a specific participant. For the mean, fractional values appear as these represent the average value of included sentences across participants.

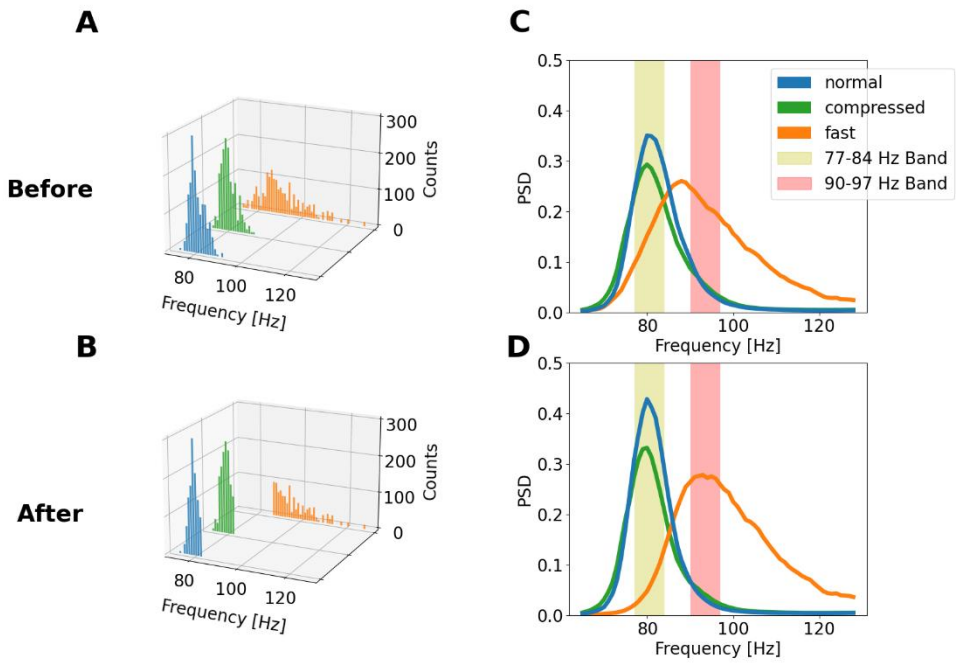

Figure 1: Optimizing sentence selection (normal rate vs. fast rate) in order to minimize F0 overlap across speech conditions. Sentences with the lowest F0s (<83 Hz) in the normal rate/time-compressed conditions and sentences with the highest F0s (>87 Hz) in the fast rate condition were selected. In this way, the spectral distance between the two subsets of stimuli was maximized. Sentences were chosen in such a way that at least 60% of them were retained in each condition, irrespective of which participant they correspond to. This resulted in a selection of, on average, 71% of the normal rate sentences, 77% of the time-compressed sentences, and 62% of the fast rate sentences (see Table 1 for the percentage of sentence selection across participants). (A) shows the original distribution of F0s which exhibits the overlap of F0 across the sets of stimuli (i.e. before the selection procedure). (B) displays the distribution of F0s after the selection procedure was applied. Note the absence of F0 overlap between the two subsets. (C) Mean spectral density of the audio spectrum per condition prior to reducing the overlap (corresponding to panel A). The highest portion of the curves for normal rate and time-compressed sentences overlaps with the lowest portion and almost the peak of the curve for fast rate sentences. (D)

Same as (C) but showing the mean power spectral density after the stimulus selection procedure. Notice the right shift of the peak of audio spectra for fast rate sentences and the reduced overlap of its lowest portion with the highest portion of normal rate and time-compressed curves. The two bands of interest that we used for the coupling analysis (77-84 Hz and 90-97 Hz) are shaded, respectively, in gold and red in panels (C) and (D).

2) MEG data

The MEG continuous data were segmented into 3 s periods (from 1 s before to 2 s after stimulus onset). A standard semi-automatic procedure available in the Fieldtrip toolbox was used to reject data segments contaminated by eye blinks, heartbeat and muscle artefacts. To remove the power line artefact and its harmonics notch-filtering was applied at 50, 100 and 150 Hz. We then detected and rejected EOG, jumps and muscle artifacts and visually double-checked the trials. Finally, an Independent Component Analysis (ICA) was used to correct for ECG artifacts. For each trial, the time interval of interest was set between 200 ms and 1000 ms after stimulus onset, while the baseline extended from 1000 ms to 200 ms before onset. The cleaned and segmented data was then imported to Python to perform the coupling analysis.

3) MRI data

T1-weighted structural MRIs (MRI 1.5 T, Siemens AvantoFit) were acquired for 23 of the 24 participants after the MEG study. Alignment of each MRI to the MEG coordinate system was performed with the localization coils marked on the MRIs and the interactive alignment interface available in MNE. Each MRI was segmented to compute each participant's head model using the single-shell as volume conductor model (Nolte, 2003). To carry out group analyses in a common space, we first spatially normalized a template anatomical grid from the Montreal Neurological Institute (MNI) common coordinate system. MRI grids of each participant were then aligned to this normalized template (this common grid has 5124 vertices).

We finally computed the cortical source space modelling for each participant using the individual MRI volume, but with the surface grid based on the MNI template. This allowed transforming the results to the same brain template for all participants for easier comparison.

4) Source localization, power and cortico-acoustic coupling estimation

We estimated source power and cortico-acoustic coupling given source reconstructions calculated using Minimum Norm Estimation (Dale and Sereno, 1993; Hämäläinen and Ilmoniemi, 1994; Gramfort et al., 2014). Before computing the coupling, we resampled the speech signal to 1200 Hz to match the sampling frequency of the MEG data. In this way, we could suitably assess the neural tracking of F0 by computing the coupling between auditory signals and cortical activity using Phase Lag Index and coherence measures. The Phase Lag Index provides information regarding the synchronization of two signals with respect to the sign of their phase difference (Stam et al., 2007). It has the advantage of being insensitive to the amplitude of the signals and moreover, neglecting zero-phase synchronizations that commonly arise because of volume conduction. It can be mathematically defined as:

$$PLI(r_{ef}, r_c, f) = |E[sign(Im[C_s(r_{ef}, r_c, f)])]|$$

where $\boldsymbol{r_{ef}}$ is the reference signal, namely the acoustic signal here; $\boldsymbol{r_c}$ is the brain signal at each vertex of the anatomical grid; $\boldsymbol{f}$ is the frequency bin and $\boldsymbol{Cs}$ is the Cross Spectral Density (CSD) matrix. The operator E[x] refers to the average over trials, |x| is the absolute value, sign(x) returns the sign of the value, and Im[x] gives the imaginary part of a complex number.

Coherence refers to the linear correlation between two signals as a function of frequency. It can be mathematically defined as:

$$Coh(r_{ef}, r_c, f) = \frac{|E[Cs(r_{ef}, r_c, f)]|}{\sqrt{E[Cs(r_{ef}, r_{ef}, f)]E[Cs(r_c, r_c, f)]}}$$

where the notation is the same as the one used in the previous equation.

We computed PLI and coherence using the MNE-Connectivity package within the MNE-Python ecosystem (Gramfort et al., 2013). Specifically, the CSD matrix was computed using the multitaper Fast Fourier Transform (FFT) from 64 to 128 Hz. Cortico-acoustic coupling was calculated for the active period (200 to 1000 ms), the baseline (-1000 to -200 ms), and for shuffled data, the last two serving as control conditions for statistical analyses for power and coupling, respectively. In the case of surrogate data, we randomly shuffled the samples of the audio signals before computing cortico-acoustic coupling (i.e. the structure of the audio signal was lost and consequently did not match its MEG counterpart for the coupling computation). We repeated this procedure 100 times and averaged coupling across iterations for each participant, and each experimental condition and frequency bin. For both active, baseline and shuffled data, the couplings (and power) were averaged across the frequency bins inside each band of interest.

5) Cortico-acoustic coupling maps, power maps and statistical analyses

The average coupling and power maps used for visualization in the Results section were obtained by first applying the Fisher transform of the values. Then we computed the difference between the active period and a set of control data (shuffled audio for coupling, and baseline for power) for each participant and condition. In the case of power, the relative difference was used. The values were then averaged across participants for each experimental condition and subsequently mapped to the brain plots. Group statistical analysis was conducted on the raw coupling and power values (without the Fisher transform) for the 23 participants with individual MRIs. For PLI and power, we applied cluster-based permutation tests with 1024 permutations (Maris and Oostenveld, 2007). For coherence, we performed an FDR corrected t-test. The tests

are one-tailed for cortico-acoustic coupling analysis and two-tailed for power analysis. All tests use a statistical significance threshold of 0.001. We used the Desikan-Killiany atlas to identify the regions detected as significant in the statistical analysis (Desikan et al., 2006). The top 10 regions with the highest mean t-value based on the PLI statistical analysis for each frequency range, condition, and hemisphere can be found in Table S1.

To inspect more closely the changes in coupling frequency with F0 variation, we additionally computed - for each of the three speech rate conditions - the coupling spectra during the encoding period, both for PLI and coherence, across a wide range of frequencies (64-128 Hz) and averaged through participants. This was done on all voxels and then averaged into the regions of the Desikan-Killiany atlas (aparc parcellation; Desikan et al., 2006). We simultaneously explored two other parcellations: aparc_sub (corresponding to subdivisions of the regions of the aparc parcellation; Khan et al., 2018) and PALS-B12 (which delineates Brodmann areas; van Essen, 2005). Then, the top 8 regions in the right hemisphere with the highest PLI peak independent of the speech rate conditions were identified for each of the three parcellations. These regions are detailed in Figures S1, S2 and S3. We further compared the top region of each parcellation with the frontal pole, selected as the control region since it was not expected to show any F0 response (see Figure 4; see also Coffey et al., 2016). Additionally, to assess the potential effects of power modulations, we computed the power spectra of the encoding brain signals at these locations in a similar manner. To better compare the coupling spectra to the oscillatory peaks in the brain power, we removed the aperiodic component using the FOOOF (specparam) methodology (Donoghue et al., 2020). The details of the model that fits each of the spectra involved in Figure 4 can be seen in supplementary Figures S4.

Results

PLI analyses revealed significant phase-locking of neural oscillations to speech fundamental frequency, with a shift in coupling frequency that mirrored the changes in F0 following natural syllable rate acceleration (Figure 2). Cortico-acoustic coupling was found at 77-84 Hz for normal rate sentences mainly in bilateral Heschl's, superior temporal, supramarginal and pre-/postcentral gyri as well as in the insula (see Table 1 for more details). For the natural fast condition, PLI increased in the higher 90-97 Hz range in a right-lateralized network mostly including the same regions, namely Heschl's, superior temporal, supramarginal and postcentral gyri (note from Figure 2A that coupling, albeit not significant, was also observed in the left hemisphere). As previously outlined, time-compressed sentences had the same F0 as normal rate sentences. This was well reflected in the brain by a significant PLI increase for this condition in the lower, normal rate pitch range only (77-84 Hz), in a similar network as the one found for normal rate speech. No significant phase-coupling was on the contrary observed for naturally accelerated sentences in this frequency range, stressing the specificity of the effect. Additional coherence analyses corroborated this pattern of results (see Figure 3). Indeed, coherence significantly increased bilaterally in the lower frequency range for normal rate and time-compressed sentences, but in the higher range for natural fast sentences. As expected, for the amplitude-modulated noise conditions, neither PLI nor coherence measures showed any significant coupling increase in any of the two frequency bands of interest (Figures 2 and 3).

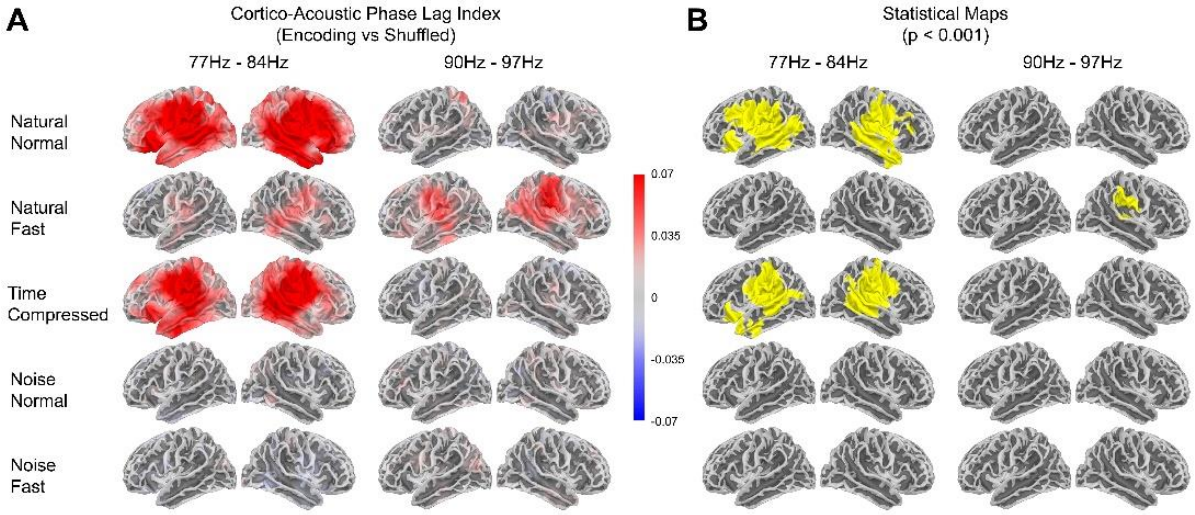

Figure 2. Neural tracking of fundamental frequency in the frequency ranges encompassing the F0s of normal rate and time-compressed sentences (77-84 Hz) and of natural fast rate sentences (90-97 Hz). (A) Phase Lag Index (PLI) maps for coupling between the acoustic signal and brain oscillatory activity during stimulus encoding (active period), as compared to surrogate (i.e. shuffled) data. The values represent the difference in Fisher z-transformed PLI between the compared data. (B) Cluster-based statistical maps (corrected, $\alpha = .001$, one-tailed) are shown for the different speech and noise conditions. Noise Normal and Noise Fast correspond to noise stimuli modulated with the amplitude envelope of normal rate and fast rate sentences, respectively.

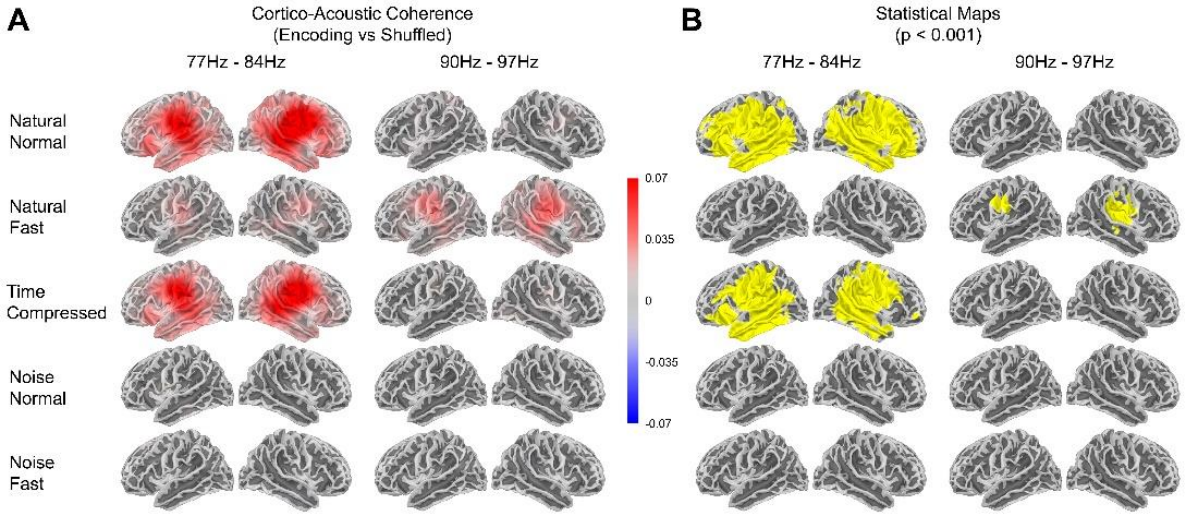

Figure 3: Cortico-acoustic coherence in the frequency ranges corresponding to the F0s in the normal rate and time-compressed conditions (77-84 Hz) and the natural fast rate condition (90-97 Hz). (A) Coherence maps for coupling between the acoustic signal and brain oscillatory activity during stimulus encoding (active period), as compared to surrogate (i.e. shuffled) data. The values represent the difference in Fisher z-transformed coherence between the compared data. (B) Statistical maps (FDR statistics, corrected, $\alpha = .001$) are shown for the different speech and noise conditions. Noise stimuli modulated with the amplitude envelope of normal rate and fast rate sentences are referred to by Noise Normal and Noise Fast, respectively.

To more thoroughly assess the specific change in coupling frequency with F0 variation, we computed the coupling spectra for PLI and coherence across the three speech rate conditions, averaged over all participants and for each voxel. We then calculated the mean spectra separately over regions delimited by three different brain parcellations (see Methods). We identified the top regions in each parcellation with the highest spectral PLI peak cross all speech condition. For this analysis, we focused on the right hemisphere as it is where our PLI and coherence analyses above showed significant effects for all speech conditions. As displayed in Figure 4, this highlighted the Heschl's gyrus (aparc parcellation), the first subzone of the postcentral gyrus (aparc_sub parcellation), and Brodmann area 43 corresponding to the subcentral area (PALS-B12 parcellation). Analyses confirmed that for both PLI and coherence,

the coupling spectra in these regions peaked at the sentences' F0, that is at ~80 Hz for normal rate and time-compressed speech, with a shift at ~90 Hz for naturally-accelerated speech (Figure 4 A, B, C for PLI and E, F, G for coherence). These peaks were not found when running the same analysis in the frontal pole which served as a control region (Figure 4 D, H). Note from Supplementary Figures S1, S2 and S3 that the three identified auditory, inferior postcentral and subcentral regions were found among the top 8 regions showing the highest PLI and coherence peaks at sentences' F0 irrespective of the parcellation used. Note also that while power modulations might explain coherence peaks (Figure 4 E-M, F-N and G-O), this cannot be the case for the PLI curves as their calculation relies on phase measures rather than on the amplitude spectrum.

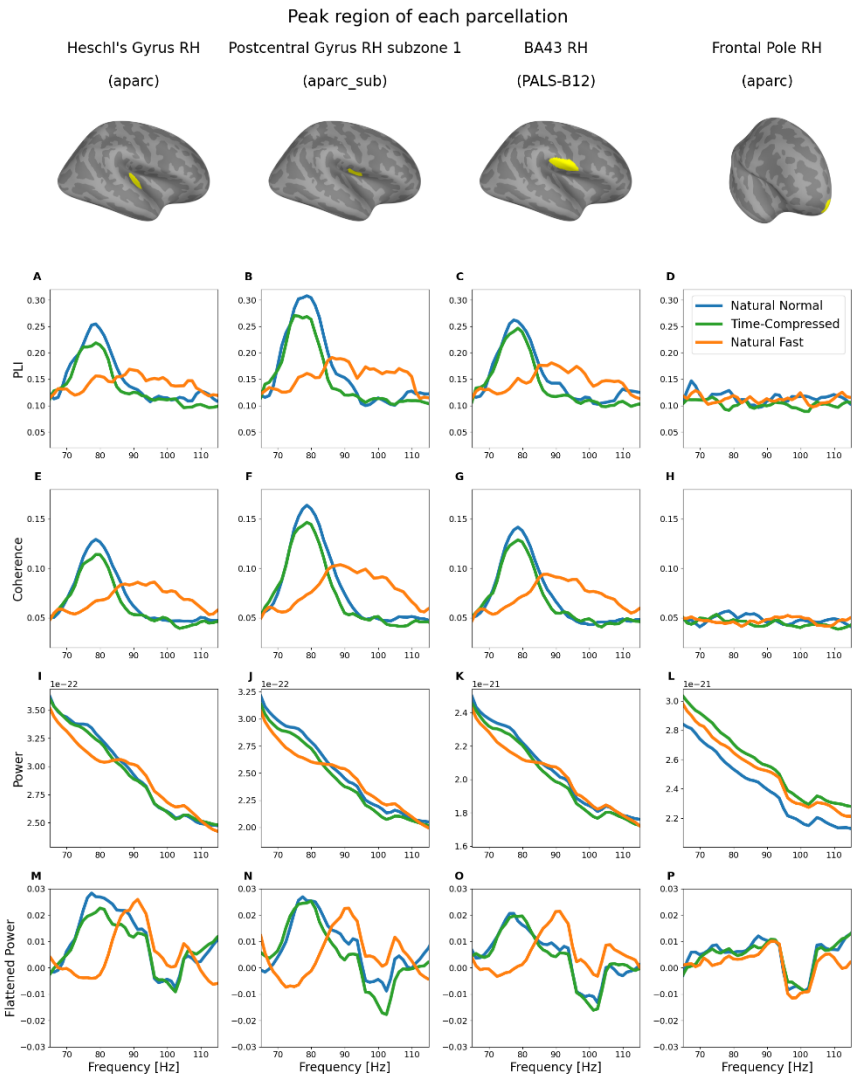

Figure 4. Coupling and power spectra for the three speech conditions in the top regions of the right hemisphere for each brain parcellation. The regions shown (one per column) correspond to the regions in each parcellation (respectively, aparc, aparc_sub and PALS-B12) with the highest peak of Phase Lag Index (PLI) coupling, which remains the same when evaluating Coherence. The curves shown correspond to the average over subjects and voxels inside each region of PLI, coherence, power and flattened power, respectively. The coupling spectra were calculated between the acoustic signal and brain oscillatory activity during stimulus encoding (active period). For comparison, the same results are provided for the frontal pole as a control region. (A) to (D) show the PLI and (E) to (H) the Coherence coupling spectra, respectively, for natural normal rate (in blue), time-compressed (in green), and natural fast rate speech (in orange). Note that Supplementary Figures S1, S2 and S3 detail the top 8 peak PLI regions for each of the three parcellations. To evaluate the possible effects of power modulations on the coupling, (I) to (L) display the brain activity's power spectrum during the encoding period and (M) to (P) show the same spectra but after correcting for the 1/f-like slope (using FOOOF; Donoghue et al., 2020). The specific models used for the aperiodic fits of these curves can be seen in Supplementary Figures S4.

We finally examined whether tracking of speech F0 is also reflected in the amplitude of the cortical responses by analyzing source power modulations in each experimental condition and frequency band of interest. To this end, we contrasted power during the stimulus encoding period with power measured during the pre-stimulus baseline, using cluster-based statistical correction. Results showed a specific increase of power as a function of sentences' F0 (Figure 5). In the normal and time-compressed conditions with F0 peaking around 80 Hz, brain activity significantly increased in the 77-84 Hz range in the right (and most of the time left for time-compressed speech) Heschl's, superior temporal, supramarginal, insular and postcentral gyri. For natural fast rate sentences, the power increase was found at 90-97 Hz in a similar right-lateralized network also encompassing the precentral gyrus. No significant increase of power was found for amplitude-modulated noise in any of the frequency ranges.

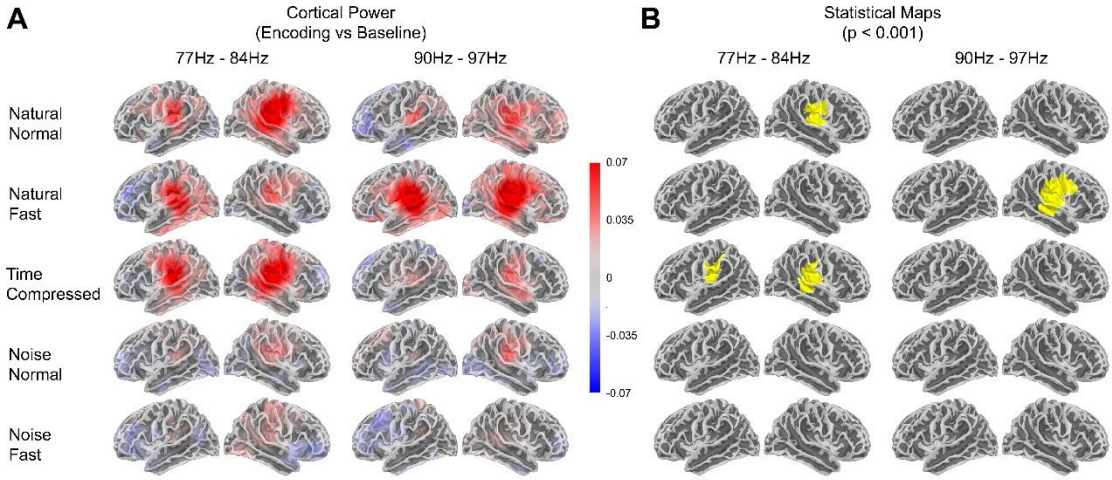

Figure 5. **Power modulations in the frequency bands matching the F0s of natural normal rate and time-compressed sentences (77-84 Hz) and of natural fast rate sentences (90-97 Hz).** (A) Power maps during the active period (i.e. stimulus encoding), as compared to the pre-stimulus baseline. The values represent the relative difference in Fisher z-transformed power between the compared data. (B) cluster-based statistical maps (corrected, α = .001, two tailed) are reported for the different speech and noise conditions.

Discussion

Our results revealed neural tracking of speech fundamental frequency, with a shift in coupling frequency and power increase that echoed the F0 increase with natural speech rate acceleration. Whereas phase-coupling (PLI), coherence and amplitude were enhanced at 77-84 Hz for normal rate and time-compressed speech (F0 peak at 80 Hz), coupling shifted to 90-97 Hz for naturally produced fast speech (F0 peak at 93 Hz). Crucially, this coupling was not restricted to the auditory cortex but encompassed a network distributed over sensorimotor regions. These findings provide novel evidence that high-frequency cortical oscillations track and adapt to the fundamental frequency of speech for natural speech rate variations.

Coupling and power modulations to F0 were mainly found in the right or bilateral superior temporal (including the primary auditory cortex), supramarginal, pre-/postcentral gyri and insula, in line with previous work on pitch processing and voicing perception (Johnsrude et al.,

2000; Zatorre and Belin, 2001; Meyer et al., 2002; Hyde et al., 2008; Tamura et al., 2022). FFR studies also reported responses in the right or bilateral auditory cortex to syllables or continuous speech (Coffey et al., 2016; Hartmann and Weisz, 2019; Ross et al., 2020; Guo et al., 2021). Importantly, our thorough whole-brain analysis using three different cortical parcellations highlighted the right Heschl's gyrus, ventral postcentral gyrus and subcentral area (BA 43, at the inferior edge of the postcentral gyrus) as showing maximal coupling to sentences' F0. Note the close spatial proximity of the latter two regions (14.7 mm based on a computation of distance between their centroids) which may correspond to the same somatosensory area. Each of these three regions furthermore showed up among the top eight regions with strong neuro-acoustic coupling across the three parcellations, indicating the reliability of the results. Other right-lateralized regions such as the insula and the supramarginal and precentral gyri also consistently synchronized their oscillatory activity to speech F0, irrespective of the parcellation used. Interestingly, these regions are part of a dorsal network underlying speech auditory-motor integration (Hickok and Poeppel, 2007; Rauschecker and Scott, 2009). The somatosensory cortex has also been advocated to contribute to speech perception by closely interacting with auditory regions (Franken et al., 2022). In this regard, the subcentral gyrus (BA 43) identified in our study, owning both sensory and motor properties (Cogan et al., 2014; Belyk and Brown, 2017), is seen as the ventral larynx motor area activated for vocalization production and pitch control (Grabski et al., 2012; Conant et al., 2014; Dichter et al., 2018; Eichert et al., 2020, 2021). High-gamma activity in this region increases before and during phoneme articulation (Bouchard et al., 2013) and for voiced consonant production (Toyoda et al., 2014), suggesting its role in the control of vocal fold movements. This area furthermore increases its connectivity with the auditory cortex during mere voicing perception (Tamura et al., 2022), and it can decode articulatory features of heard phonemes (Arsenault and Buchsbaum, 2015; Correia et al., 2015; Archila-Meléndez et al., 2018). Our results concur with a functional role of this ventral

somatosensory area in speech perception as it is one of the regions that best synchronizes its high-frequency oscillatory activity to speech F0.

A dorsal larynx motor area, located in the precentral gyrus, is also thought to be involved in voicing and pitch production and perception (Brown et al., 2008; D'Ausilio et al., 2011; Simonyan, 2014; Dichter et al., 2018; Eichert et al., 2020; Liang et al., 2023; Lu et al., 2023). In fact, and although this should be further tested, the precentral region coupling to speech F0 in our study may correspond to this dorsal motor region. This could reflect some simulation of the laryngeal gestures underlying pitch control during speech perception (Sammler et al., 2015). We also found reliable tracking of F0 in the supramarginal gyrus and insula. This agrees with the inferior parietal cortex acting as a sensorimotor interface to coordinate speech perception and production (Rauschecker and Scott, 2009; Rauschecker, 2011; Shum et al., 2011) as well as with its role in phonological processing (Hartwigsen et al., 2010). The insula is also assumed to contribute to auditory-motor speech integration (Mutschler et al., 2009), as it is engaged in speech production and the control of laryngeal muscles and F0 modulations (Dronkers, 1996; Ackermann and Riecker, 2010; Howell et al., 2012), but also in prosodic perception and auditory temporal processing (Bamiou et al., 2003, 2006; Tregellas et al., 2006; Oh et al., 2014). In keeping with our work, high-gamma activity in the posterior insula was shown to track the F0 of a repetitive vowel similarly as in Heschl's gyrus (Zhang et al., 2019). The insula moreover shares connections with auditory (Ghaziri et al., 2017) and motor regions, including the laryngeal motor cortex (Simonyan et al., 2009; Simonyan and Horwitz, 2011). Altogether, our findings putatively suggest that besides the primary auditory cortex targeted by previous work (Coffey et al., 2016, 2017, 2021; Kulasingham et al., 2020), a dorsal sensorimotor cortical network involved in speech auditory-motor integration tracks the periodicity of vocal fold vibration in perceived speech.

Whether and how our findings relate to the cortical FFR to repetitive syllables (Coffey et al., 2016, 2017; Bidelman, 2018; Hartmann and Weisz, 2019; Ross et al., 2020) is not straightforward. The FFR is averaged over several hundreds of the same stimulus, whereas we assessed cortico-acoustic coupling and power modulations to a set of varying sentences. Still, the FFR is thought to indicate phase-locked neural activity to the stimulus F0. The different measures may therefore reflect at least some common processes. As previously mentioned, a FFR source in the right posterior auditory cortex has been described and would be involved in pitch computation (Coffey et al., 2016, 2017). Among the few studies that investigated F0 tracking in sentences or continuous speech, some focused on the brainstem (Forte et al., 2017; Maddox and Lee, 2018; Etard et al., 2019), while others examined high-gamma auditory cortex activity in light of the cortical FFR (Puschmann et al., 2019; Kulasingham et al., 2020; Guo et al., 2021; Van Canneyt et al., 2021; Schüller et al., 2023). To the best of our knowledge, all of these FFR-F0 studies, whether conducted on repetitive syllables or continuous speech, assessed brain responses in the auditory cortex only. We are aware of only one stereotactic EEG / MEG study showing FFR in a temporo-parieto-frontal cortical network, however this was in response to musical tones (Lerousseau et al., 2021). Crucially, by analyzing both coupling (using two measures, PLI and coherence) and power modulations at high frequencies, we bring new support for brain tracking of F0 in natural speech that occurs well beyond the auditory cortex, in regions that are involved in auditory-somato-motor integration.

Although the increase in coherence we found might be explained, at least partly, by power increases at the same frequencies, this cannot hold for the phase-lag index which is insensitive to amplitude modulations. Cortical coupling to speech F0 therefore most probably captures the periodic nature of the phenomenon. Coffey and collaborators (Coffey et al., 2021) likewise proposed that their auditory cortex FFR entailed oscillatory entrainment since the brain

response outlasted syllable offset (see also Lerousseau at al., 2021). Along this view, oscillatory coupling to F0 in our study could reflect a predictive coding mechanism comparing speech input periodicity with internal temporal expectations. Such an interpretation, although still speculative, would support the idea that high-gamma activity mediates pitch-level prediction errors and their feedforward propagation up to higher integrative areas in order to update predictions (Brodski et al., 2015; Lewis and Bastiaansen, 2015; van Pelt et al., 2016). F0 tracking in (pre)motor regions furthermore agrees with their known role in predictive coding, despite this is usually observed in lower (delta and beta) frequencies (Park et al., 2015; Keitel et al., 2017; Morillon and Baillet, 2017). Further cross-frequency coupling and functional connectivity analyses between auditory and motor regions are needed to fully substantiate the predictive coding interpretation and further delineate the role of speech F0 tracking at the cortical level. As speech periodicity conveys segmental information, phase-locking to F0 may contribute to phoneme identification by encoding the articulatory feature of voicing or by extracting vowels' formant structure. Another, not exclusive possibility is that it could serve more general communicative purposes, to track the relevant targets in an auditory scene (Hertrich et al., 2012) and decrease sensitivity to environmental noise or competing speakers (Coffey et al., 2021).

As in previous work in this domain, several critical aspects need to cautiously be taken into account when interpreting our speech-brain oscillatory coupling. Distinguishing true neuronal entrainment and oscillations from a succession of neural representations of events that happen to have some periodicity is a topic of current debate (Duecker et al., 2024). Moreover, the reliability of the widespread network reported here would benefit from further investigation to rule out possible effects of spurious activity arising from the smoothing inherent to source reconstruction.

In conclusion, our study suggests that cortical oscillations exhibit synchronization with the fundamental frequency (F0) of speech but also appear to adaptively adjust their coupling frequency to match variations in F0. Using magnetoencephalography (MEG), we identified significant brain-to-speech F0 coupling across a network extending beyond the auditory cortex to include the right auditory, inferior parietal, insular, and pre- and postcentral regions, implicating a dynamic auditory-somato-motor network in the encoding of speech. These findings advance our understanding of how high-frequency cortical oscillations contribute to language processing and hint at a complex interplay between auditory and motor systems in speech perception, and pave the way for future work exploring how these mechanisms function under different auditory and communicative conditions.

Conflict of interest statement: The authors declare no competing financial interests.

Acknowledgements: This research was supported by the French National Research Agency (ANR-11 JSH2 005 1; PI: V.B) and the Laboratory of Excellence ASLAN (ANR-10-LABX-0081) of Université de Lyon within the program "Investissements d'Avenir" (ANR-11-IDEX-0007; V.B) of the French government operated by the ANR. Y-J.M-R was supported by Mitacs through the Mitacs Globalink program. A-S.H-C was supported in part by a scholarship from Comisión Nacional de Ciencia y Tecnología de Chile (CONICYT, Colfuturo, Colombia) and by the French ANR project ANR-DEFIS n°09-EMER-002 CoAdapt. This research was also undertaken thanks to funding from the Canada Research Chairs program and a Discovery Grant (RGPIN-2015-04854) awarded by the Natural Sciences and Engineering Research Council (NSERC) of Canada to K.J (the NSERC Discovery Grant also provided support to A-S.H-C). We would like to thank Damien Gouy for recording the stimuli, Emmanuel Ferragne for his help with Praat software, and Mélanie Canault for discussions on articulatory phonetics.

References

Ackermann H, Riecker A (2010) The contribution(s) of the insula to speech production: a review of the clinical and functional imaging literature. Brain Struct Funct 214:419–433.

Aiken SJ, Picton TW (2008) Envelope and spectral frequency-following responses to vowel sounds. Hear Res 245:35–47.

Archila-Meléndez ME, Valente G, Correia JM, Rouhl RPW, Kranen-Mastenbroek VH van, Jansma BM (2018) Sensorimotor Representation of Speech Perception. Cross-Decoding of Place of Articulation Features during Selective Attention to Syllables in 7T fMRI. eNeuro 5 Available at: https://www.eneuro.org/content/5/2/eneuro.0252-17.2018 [Accessed January 17, 2022].

Arsenault JS, Buchsbaum BR (2015) Distributed Neural Representations of Phonological Features during Speech Perception. J Neurosci 35:634–642.

Bamiou D-E, Musiek FE, Luxon LM (2003) The insula (Island of Reil) and its role in auditory processing: Literature review. Brain Res Rev 42:143–154.

Bamiou D-E, Musiek FE, Stow I, Stevens J, Cipolotti L, Brown MM, Luxon LM (2006) Auditory temporal processing deficits in patients with insular stroke. Neurology 67:614–619.

Belyk M, Brown S (2017) The origins of the vocal brain in humans. Neurosci Biobehav Rev 77:177–193.

Berry JJ (2011) Speaking Rate Effects on Normal Aspects of Articulation: Outcomes and Issues. Perspect Speech Sci Orofac Disord 21:15–26.

Bidelman GM (2018) Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. NeuroImage 175:56–69.

Bidelman GM, Momtaz S (2021) Subcortical rather than cortical sources of the frequency-following response (FFR) relate to speech-in-noise perception in normal-hearing listeners. Neurosci Lett 746:135664.

Boersma P, Weenik D (2012) Doing phonetics by computer [Computer program]. Available at: http://www.praat.org/.

Bouchard KE, Mesgarani N, Johnson K, Chang EF (2013) Functional Organization of Human Sensorimotor Cortex for Speech Articulation. Nature 495:327–332.

Brodski A, Paasch G-F, Helbling S, Wibral M (2015) The Faces of Predictive Coding. J Neurosci 35:8997–9006.

Brown S, Ngan E, Liotti M (2008) A Larynx Area in the Human Motor Cortex. Cereb Cortex 18:837–845.

Caspers J, van Heuven VJ (1993) Effects of Time Pressure on the Phonetic Realization of the Dutch Accent-Lending Pitch Rise and Fall. Phonetica 50:161–171.

Chandrasekaran B, Kraus N (2010) The scalp-recorded brainstem response to speech: Neural origins and plasticity. Psychophysiology 47:236–246.

Coffey EBJ, Arseneau-Bruneau I, Zhang X, Baillet S, Zatorre RJ (2021) Oscillatory entrainment of the Frequency Following Response in auditory cortical and subcortical structures. J Neurosci

Available at: https://www.jneurosci.org/content/early/2021/03/16/JNEUROSCI.2313-20.2021 [Accessed March 23, 2021].

Coffey EBJ, Herholz SC, Chepesiuk AMP, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. Nat Commun 7:11070.

Coffey EBJ, Musacchia G, Zatorre RJ (2017) Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. J Neurosci 37:830–838.

Coffey EBJ, Nicol T, White-Schwoch T, Chandrasekaran B, Krizman J, Skoe E, Zatorre RJ, Kraus N (2019) Evolving perspectives on the sources of the frequency-following response. Nat Commun 10:1–10.

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

Conant D, Bouchard KE, Chang EF (2014) Speech Map in the Human Ventral Sensory-Motor Cortex. Curr Opin Neurobiol 0:63–67.

Correia JM, Jansma BMB, Bonte M (2015) Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions. J Neurosci 35:15015–15025.

Dale AM, Sereno MI (1993) Improved Localizadon of Cortical Activity by Combining EEG and MEG with MRI Cortical Surface Reconstruction: A Linear Approach. J Cogn Neurosci 5:162–176.

D'Ausilio A, Bufalari I, Salmas P, Busan P, Fadiga L (2011) Vocal pitch discrimination in the motor system. Brain Lang 118:9–14.

Desikan RS, Ségonne F, Fischl B, Quinn BT, Dickerson BC, Blacker D, Buckner RL, Dale AM, Maguire RP, Hyman BT, Albert MS, Killiany RJ (2006) An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31:968–980.

Dichter BK, Breshears JD, Leonard MK, Chang EF (2018) The Control of Vocal Pitch in Human Laryngeal Motor Cortex. Cell 174:21-31.e9.

Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, Noto T, Lara AH, Wallis JD, Knight RT, Shestyuk A, Voytek B (2020) Parameterizing neural power spectra into periodic and aperiodic components. Nat Neurosci 23:1655–1665.

Dronkers NF (1996) A new brain region for coordinating speech articulation. Nature 384:159–161.

Duecker K, Doelling KB, Breska A, Coffey EBJ, Sivarao DV, Zoefel B (2024) Challenges and Approaches in the Study of Neural Entrainment. J Neurosci Off J Soc Neurosci 44:e1234242024.

Eichert N, Papp D, Mars RB, Watkins KE (2020) Mapping Human Laryngeal Motor Cortex during Vocalization. Cereb Cortex 30:6254–6269.

Eichert N, Watkins KE, Mars RB, Petrides M (2021) Morphological and functional variability in central and subcentral motor cortex of the human brain. Brain Struct Funct 226:263–279.

Etard O, Kegler M, Braiman C, Forte AE, Reichenbach T (2019) Decoding of selective attention to continuous speech from the human auditory brainstem response. NeuroImage 200:1–11.

Ferragne E, Flavier S, Fressard C (2012) Recording of Oral Corpora Made Easy (ROCme): logiciel d'enregistrement de corpus.

Forte AE, Etard O, Reichenbach T (2017) The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention Shinn-Cunningham BG, ed. eLife 6:e27203.

Fougeron C, Jun S-A (1998) Rate effects on French intonation: prosodic organization and phonetic realization. J Phon 26:45–69.

Franken MK, Liu BC, Ostry DJ (2022) Towards a somatosensory theory of speech perception. J Neurophysiol 128:1683–1695.

Ghaziri J, Tucholka A, Girard G, Houde J-C, Boucher O, Gilbert G, Descoteaux M, Lippé S, Rainville P, Nguyen DK (2017) The Corticocortical Structural Connectivity of the Human Insula. Cereb Cortex 27:1216–1228.

Gnanateja GN, Rupp K, Llanos F, Remick M, Pernia M, Sadagopan S, Teichert T, Abel TJ, Chandrasekaran B (2021) Frequency-Following Responses to Speech Sounds Are Highly Conserved across Species and Contain Cortical Contributions. eNeuro 8:ENEURO.0451-21.2021.

Grabski K, Lamalle L, Vilain C, Schwartz J-L, Vallée N, Tropres I, Baciu M, Bas J-FL, Sato M (2012) Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. Hum Brain Mapp 33:2306–2321.

Gramfort A, Luessi M, Larson E, Engemann D, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hämäläinen M (2013) MEG and EEG data analysis with MNE-Python. Front Neurosci 7 Available at: https://www.frontiersin.org/articles/10.3389/fnins.2013.00267 [Accessed May 25, 2023].

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and EEG data. NeuroImage 86:446–460.

Greenlaw KM, Puschmann S, Coffey EBJ (2020) Decoding of Envelope vs. Fundamental Frequency During Complex Auditory Stream Segregation. Neurobiol Lang 1:268–287.

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

Guo N, Si X, Zhang Y, Ding Y, Zhou W, Zhang D, Hong B (2021) Speech frequency-following response in human auditory cortex is more than a simple tracking. NeuroImage 226:117545.

Hämäläinen MS, Ilmoniemi RJ (1994) Interpreting magnetic fields of the brain: minimum norm estimates. Med Biol Eng Comput 32:35–42.

Hartmann T, Weisz N (2019) Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. NeuroImage 203:116185.

Hartwigsen G, Baumgaertner A, Price CJ, Koehnke M, Ulmer S, Siebner HR (2010) Phonological decisions require both the left and right supramarginal gyri. Proc Natl Acad Sci U S A 107:16494–16499.

Hertrich I, Dietrich S, Trouvain J, Moos A, Ackermann H (2012) Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. Psychophysiology 49:322–334.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hincapié-Casas AS, Lajnef T, Pascarella A, Guiraud H, Laaksonen H, Bayle D, Jerbi K, Boulenger V (2021a) Neural oscillations track natural but not artificial fast speech: Novel insights from speech-brain coupling using MEG. bioRxiv:2020.10.20.344895.

Hincapié-Casas AS, Lajnef T, Pascarella A, Guiraud-Vinatea H, Laaksonen H, Bayle D, Jerbi K, Boulenger V (2021b) Neural oscillations track natural but not artificial fast speech: Novel insights from speech-brain coupling using MEG. NeuroImage 244:118577.

Howell P, Jiang J, Peng D, Lu C (2012) Neural control of fundamental frequency rise and fall in Mandarin tones. Brain Lang 121:35–46.

Hyde KL, Peretz I, Zatorre RJ (2008) Evidence for the role of the right auditory cortex in fine pitch resolution. Neuropsychologia 46:632–639.

Janse E (2004) Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. Speech Commun 42:155–173.

Johnsrude IS, Penhune VB, Zatorre RJ (2000) Functional specificity in the right human auditory cortex for perceiving pitch direction. Brain 123:155–163.

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

Keitel A, Ince RAA, Gross J, Kayser C (2017) Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. NeuroImage 147:32–42.

Kraus N, Nicol T (2005) Brainstem origins for cortical ‘what’ and ‘where’ pathways in the auditory system. Trends Neurosci 28:176–181.

Kulasingham JP, Brodbeck C, Presacco A, Kuchinsky SE, Anderson S, Simon JZ (2020) High gamma cortical processing of continuous speech in younger and older listeners. NeuroImage 222:117291.

Lerousseau JP, Trébuchon A, Morillon B, Schön D (2021) Frequency Selectivity of Persistent Cortical Oscillatory Responses to Auditory Rhythmic Stimulation. J Neurosci 41:7991–8006.

Lewis AG, Bastiaansen M (2015) A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. Cortex 68:155–168.

Liang B, Li Y, Zhao W, Du Y (2023) Bilateral human laryngeal motor cortex in perceptual decision of lexical tone and voicing of consonant. Nat Commun 14:4710.

Lu J, Li Y, Zhao Z, Liu Y, Zhu Y, Mao Y, Wu J, Chang EF (2023) Neural control of lexical tone production in human laryngeal motor cortex. Nat Commun 14:6917.

Maddox RK, Lee AKC (2018) Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. eNeuro 5 Available at: https://www.eneuro.org/content/5/1/ENEURO.0441-17.2018 [Accessed April 12, 2021].

Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Methods 164:177–190.

Meyer M, Alter K, Friederici AD, Lohmann G, von Cramon DY (2002) FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. Hum Brain Mapp 17:73–88.

Morillon B, Baillet S (2017) Motor origin of temporal predictions in auditory attention. Proc Natl Acad Sci U S A 114:E8913–E8921.

Moulines E, Charpentier F (1990) Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun 9:453–467.

Mutschler I, Wieckhorst B, Kowalevski S, Derix J, Wentlandt J, Schulze-Bonhage A, Ball T (2009) Functional organization of the human anterior insular cortex. Neurosci Lett 457:66–70.

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.

Oh A, Duerden EG, Pang EW (2014) The role of the insula in speech and language processing. Brain Lang 135:96–103.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci 2011.

Park H, Ince RAA, Schyns PG, 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.

Pefkou M, Arnal LH, Fontolan L, Giraud A-L (2017) Theta- and beta-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. J Neurosci:2882–16.

Puschmann S, Baillet S, Zatorre RJ (2019) Musicians at the Cocktail Party: Neural Substrates of Musical Training During Selective Listening in Multispeaker Situations. Cereb Cortex 29:3253–3265.

Rauschecker JP (2011) An expanded role for the dorsal auditory pathway in sensorimotor control and integration. Hear Res 271:16–25.

Rauschecker JP, Scott SK (2009) Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Rosen S (1992) Temporal information in speech: Acoustic, auditory and linguistic aspects. Philos Trans R Soc B Biol Sci 336:367–373.

Ross B, Tremblay KL, Alain C (2020) Simultaneous EEG and MEG recordings reveal vocal pitch elicited cortical gamma oscillations in young and older adults. NeuroImage 204:116253.

Russo N, Nicol T, Musacchia G, Kraus N (2004) Brainstem responses to speech syllables. Clin Neurophysiol 115:2021–2030.

Sammler D, Grosbras M-H, Anwander A, Bestelmeyer PEG, Belin P (2015) Dorsal and Ventral Pathways for Prosody. Curr Biol 25:3079–3085.

Schüller A, Schilling A, Krauss P, Rampp S, Reichenbach T (2023) Attentional Modulation of the Cortical Contribution to the Frequency-Following Response Evoked by Continuous Speech. J Neurosci 43:7429–7440.

Shum M, Shiller DM, Baum SR, Gracco VL (2011) Sensorimotor integration for speech motor learning involves the inferior parietal cortex. Eur J Neurosci 34:1817–1822.

Simonyan K (2014) The laryngeal motor cortex: its organization and connectivity. Curr Opin Neurobiol 28:15–21.

Simonyan K, Horwitz B (2011) Laryngeal Motor Cortex and Control of Speech in Humans. Neurosci Rev J Bringing Neurobiol Neurol Psychiatry 17:197–208.

Simonyan K, Ostuni J, Ludlow CL, Horwitz B (2009) Functional But Not Structural Networks of the Human Laryngeal Motor Cortex Show Left Hemispheric Lateralization during Syllable But Not Breathing Production. J Neurosci 29:14912–14923.

Stam CJ, Nolte G, Daffertshofer A (2007) Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. Hum Brain Mapp 28:1178–1193.

Tamura S, Hirose N, Mitsudo T, Hoaki N, Nakamura I, Onitsuka T, Hirano Y (2022) Multi-modal imaging of the auditory-larynx motor network for voicing perception. NeuroImage 251:118981.

Toyoda G, Brown EC, Matsuzaki N, Kojima K, Nishida M, Asano E (2014) Electrocorticographic correlates of overt articulation of 44 English phonemes: Intracranial recording in children with focal epilepsy. Clin Neurophysiol 125:1129–1137.

Tregellas JR, Davalos DB, Rojas DC (2006) Effect of task difficulty on the functional anatomy of temporal processing. NeuroImage 32:307–315.

Van Canneyt J, Wouters J, Francart T (2021) Neural tracking of the fundamental frequency of the voice: The effect of voice characteristics. Eur J Neurosci 53:3640–3653.

van Pelt S, Heil L, Kwisthout J, Ondobaka S, van Rooij I, Bekkering H (2016) Beta- and gamma-band activity reflect predictive coding in the processing of causal events. Soc Cogn Affect Neurosci 11:973–980.

Zatorre RJ, Belin P (2001) Spectral and Temporal Processing in Human Auditory Cortex. Cereb Cortex 11:946–953.

Zhang Y, Zhou W, Wang S, Zhou Q, Wang H, Zhang B, Huang J, Hong B, Wang X (2019) The Roles of Subdivisions of Human Insula in Emotion Perception and Auditory Processing. Cereb Cortex 29:517–528.

Supplementary Material

Condition	77-84 Hz								90-97 Hz							
	Left Hemisphere				Right Hemisphere				Left Hemisphere				Right Hemisphere			
	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count
Normal	Transverse Temporal Gyrus	5.95	0.42	17	Posterior Cingulate Cortex	5.05	0.98	41	None				None			
	Superior Temporal Gyrus	4.95	0.83	79	Transverse Temporal Gyrus	4.79	0.35	10								
	Banks of the Superior Temporal Sulcus	4.85	0.58	35	Insular Cortex	4.56	0.67	52								
	Postcentral Gyrus	4.75	0.66	72	Paracentral Lobule	4.53	0.62	30								
	Insular Cortex	4.72	0.88	62	Supramarginal Gyrus	4.49	0.63	45								
	Inferior Parietal Lobule	4.71	0.83	33	Postcentral Gyrus	4.33	0.45	73								
	Caudal Middle Frontal Gyrus	4.40	0.45	18	Superior Temporal Gyrus	4.08	0.39	93								
	Supramarginal Gyrus	4.29	0.61	82	Precentral Gyrus	4.02	0.41	114								
	Precentral Gyrus	4.28	0.51	75	Banks of the Superior Temporal Sulcus	3.86	0.22	30								
	Middle Temporal Gyrus	4.18	0.45	36	Inferior Parietal Lobule	3.79	0.17	5								
Fast	Left Hemisphere				Right Hemisphere				Left Hemisphere				Right Hemisphere			
	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count
	None				None				None				Supramarginal Cortex	4.08	0.40	64
													Transverse Temporal Cortex	3.91	0.28	6
													Postcentral Cortex	3.91	0.30	21
													Superior Temporal Cortex	3.89	0.24	13
Time Compressed	Left Hemisphere				Right Hemisphere				Left Hemisphere				Right Hemisphere			
	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count	Region	Mean t-value	SD	Voxel Count
	Transverse Temporal Gyrus	4.91	0.51	17	Transverse Temporal Gyrus	5.11	0.43	10	None				None			
	Supramarginal Gyrus	4.77	0.60	98	Posterior Cingulate Cortex	4.95	0.86	38								
	Postcentral Gyrus	4.70	0.65	101	Supramarginal Gyrus	4.95	0.87	104								
	Insular Cortex	4.62	0.66	42	Postcentral Gyrus	4.71	0.66	89								
	Posterior Cingulate Cortex	4.60	0.52	34	Insular Cortex	4.61	0.72	45								
	Superior Temporal Gyrus	4.34	0.52	72	Superior Temporal Gyrus	4.51	0.50	59								
	Precentral Gyrus	4.15	0.46	42	Paracentral Lobule	4.25	0.50	33								
	Temporal Pole	4.14	0.34	9	Banks of the Superior Temporal Sulcus	4.22	0.30	36								
	Banks of the Superior Temporal Sulcus	4.12	0.26	29	Caudal Anterior Cingulate Cortex	4.14	0.21	9								
	Lateral Orbitofrontal Cortex	4.11	0.40	40	Precentral Gyrus	4.12	0.44	106								

Table S1: Top 10 statistically significant regions from the cluster-based analysis of the PLI for each condition, frequency range and hemisphere. This represents a more tabular view of Figure 2. For each condition (Normal, Fast, Time-Compressed), frequency range (77-84 Hz and 90-97Hz) and hemisphere (left and right), the statistically significant regions with the highest mean t-value (averaged over the voxels of the region) are shown. Each region is also described by its number of voxels (voxel count) and the standard deviation of the t-values (SD).

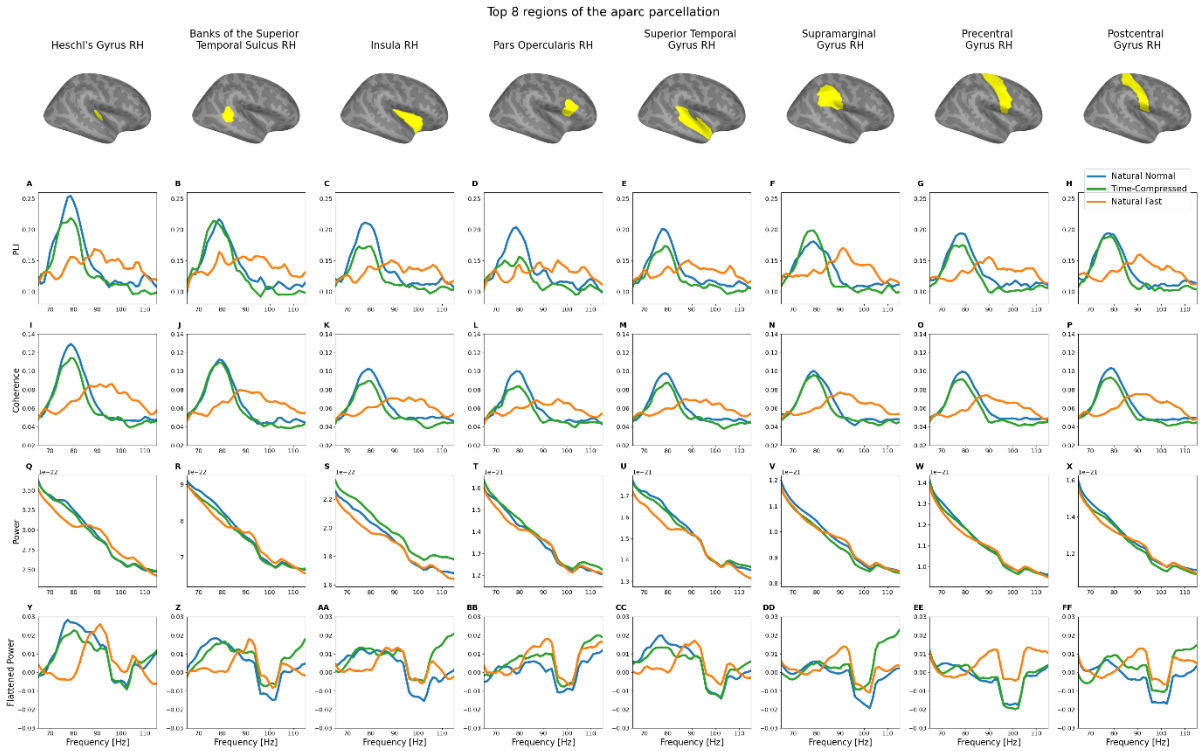

Figure S1: Top 8 regions of the right hemisphere of the aparc parcellation showing the highest spectral PLI peak. The regions shown (one per column) correspond to the 8 regions of the aparc parcellation with the highest peak PLI coupling (from left to right). The curves correspond to the average over subjects and voxels inside each region of PLI, coherence, power and flattened power respectively. The coupling spectra were calculated between the acoustic signal and brain oscillatory activity during stimulus encoding (active period). (A to H) show the Phase Lag Index (PLI) and (I to P) the Coherence coupling spectra, respectively, for natural normal rate (in blue), time-compressed (in green), and natural fast rate speech (in orange). To evaluate the possible effects of power modulations on the coupling, (Q) to (X) display the brain activity's power spectrum during the encoding period, and (Y) to (FF) show their flattened curve using FOOOF.

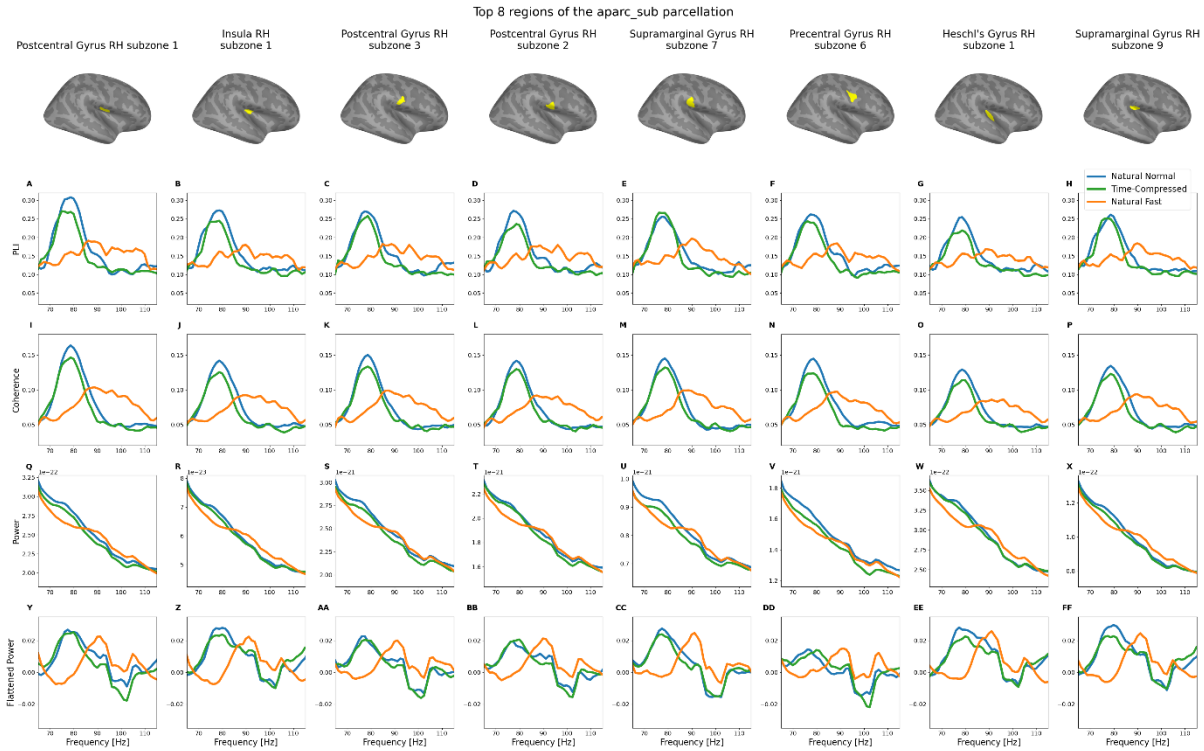

Figure S2: Top 8 regions of the right hemisphere of the aparc_sub parcellation. The regions shown (one in each column) correspond to the 8 regions of the aparc_sub parcellation with the highest peak PLI coupling (from left to right). The curves correspond to the average over subjects and voxels inside each region of PLI, coherence, power and flattened power respectively. The coupling spectra were calculated between the acoustic signal and brain oscillatory activity during stimulus encoding (active period). (A) to (H) show the Phase Lag Index (PLI) and (I) to (P) the Coherence coupling spectra, respectively, for natural normal rate (in blue), time-compressed (in green), and natural fast rate speech (in orange). To evaluate the possible effects of power modulations on the coupling, (Q) to (X) display the brain activity's power spectrum during the encoding period, and (Y) to (FF) show their flattened curve using FOOOF.

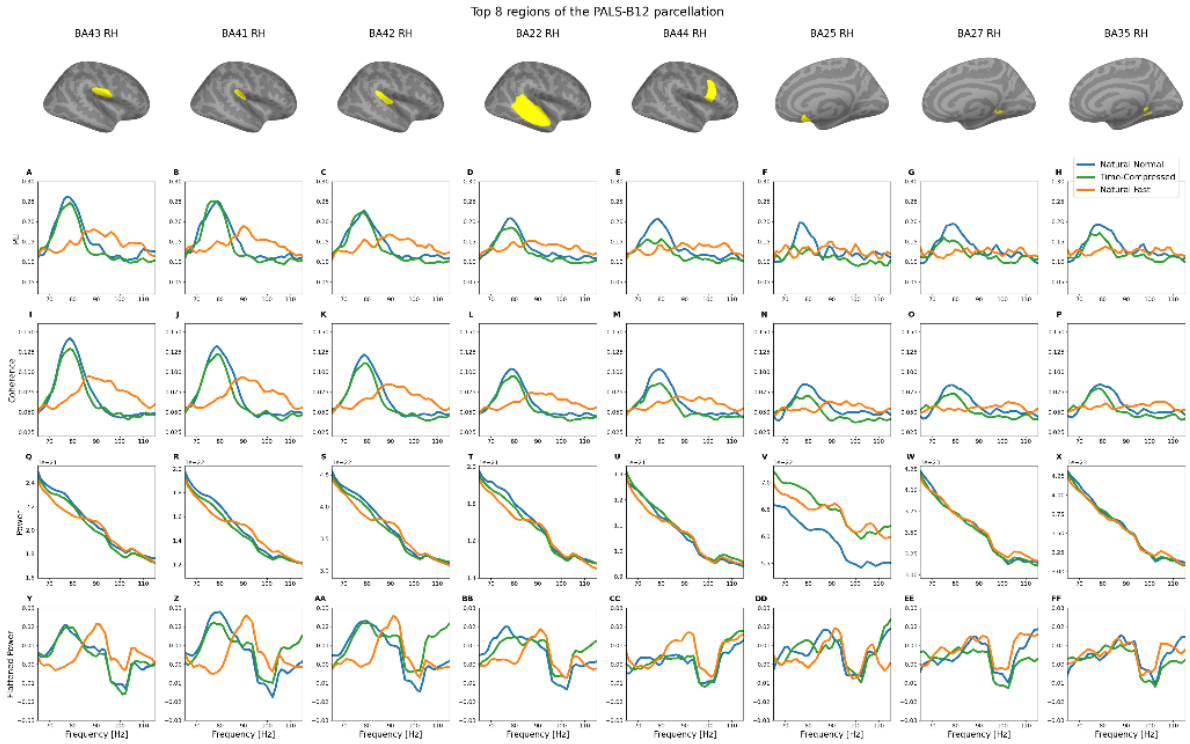

Figure S3: Top 8 regions of the right hemisphere of the PALS-B12 parcellation. The regions shown (one in each column) correspond to the 8 regions of the PALS-B12 parcellation with the highest peak PLI coupling (from left to right). The curves shown correspond to the average over subjects and voxels inside each region of PLI, coherence, power and flattened power respectively. The coupling spectra were calculated between the acoustic signal and brain oscillatory activity during stimulus encoding (active period). (A) to (H) show the Phase Lag Index (PLI) and (I) to (P) the Coherence coupling spectra, respectively, for natural normal rate (in blue), time-compressed (in green), and natural fast rate speech (in orange). To evaluate the possible effects of power modulations on the coupling, (Q) to (X) display the brain activity's power spectrum during the encoding period, and (Y) to (FF) show their flattened curve using FOOOF.

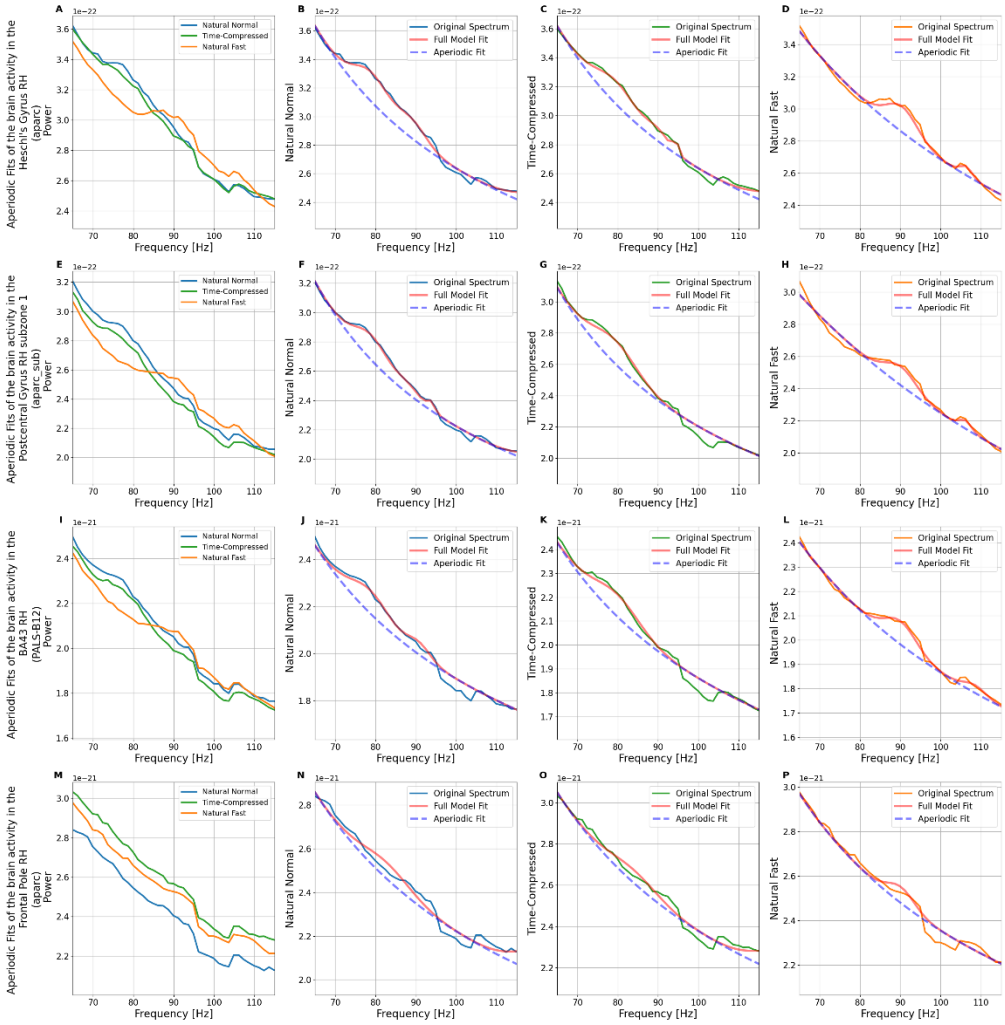

Figure S4: Aperiodic fits of the mean brain activity power spectra in each region. (A), (B), (C) , (D) correspond to the right Heschl's gyrus (from aparc parcellation)**.** (A) Power spectra for the three speech conditions (same as Figure 2I; normal rate in blue, time-compressed in green and fast rate in orange). (B), (C) and (D) show the fitting using the FOOOF procedure (Donoghue et al., 2020) for normal rate, time-compressed and fast rate conditions, respectively. On each, the purple dashed line represents the aperiodic component, the red line the full model fit (the sum of the periodic and aperiodic components), and the third line shows the original spectra of the respective speech condition following the same color code as in (A). Similarly, the plots are repeated for the other regions. The first subzone of the right Postcentral Gyrus (from aparc_sub parcellation) corresponds to (E), (F), (G), (H). The right Brodmann Area 43 (from PALS-B12 parcellation) corresponds to (I), (J), (K), (L). The right Frontal Pole is used as a control region (from aparc parcellation); it corresponds to (M), (N), (O), (P).