Hitting the right pitch: Cortical tracking of speech fundamental frequency in auditory and somatomotor regions

Yorguin-Jose Mantilla-Ramos[1,2,6] 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 contributed equally.

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

Abstract

Low-frequency neural oscillations contribute to the parsing of continuous speech into linguistic units. Little is known however on the coupling of brain rhythms to higher-frequencies in speech such as fundamental frequency (F0) or pitch. Using magnetoencephalography, we investigated whole-brain cortical tracking of F0 while participants listened to sentences produced at normal rate or fast rate, where pitch naturally increases, and to artificially accelerated sentences, where F0 remains unchanged. Our results revealed significant brain-to-F0 coupling across all speech rates not only in right auditory but also in right parietal, insular, and pre- and postcentral regions, likely including the ventral larynx area. Importantly, the cortico-acoustic coupling peak frequency was higher for natural fast speech to reflect the corresponding F0 increase compared to normal rate and time-compressed speech. These findings demonstrate the engagement of an auditory-somato-motor network in F0 tracking, supporting its role in facilitating phonemic processing during the perception of naturally-produced speech.

Introduction

By tracking speech temporal structure, neural oscillations contribute to the parsing of the acoustic stream into hierarchically-organized linguistic constituents[1–4]. Substantial evidence from electro/magnetoencephalography (E/MEG) revealed phase-coupling of delta- (~1-3 Hz) and theta-band (~4-8 Hz) auditory cortical activity to the slow modulations of speech amplitude envelope, which mainly reflect the syllable rate[5–9]. Speech envelope has also been shown to modulate low (35-45 Hz) and high gamma (> 50 Hz) amplitude as well as theta phase/low gamma amplitude coupling[6,10–14].

Besides syllabic structure, speech periodicity is another temporal cue of chief importance for speech perception and communication as it conveys segmental (voicing, manner of articulation) and suprasegmental (intonation, stress) information[15]. 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. Only limited evidence exists regarding cortical tracking of F0 in naturally-produced speech. 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[16–19]. Although the FFR recorded in EEG at the scalp is believed to originate primarily from the brainstem[20], contributions from the thalamus and auditory cortex were also reported, with cortical activity peaking at the stimulus F0[21–28]. Brain coupling to F0 in continuous sound streams has been given less attention. The studies available so far focusing on neural encoding of the F0 in naturalistic speech showed responses in the auditory cortex at the F0 of isolated sentences or stories[29–32]. For instance, using a linear modelling approach in EEG, Van Canneyt and colleagues[33] reported that the F0 of stories could be reconstructed from brain activity likely arising from the brainstem and the primary auditory cortex.

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, couple their oscillatory activity with the F0 of natural speech produced at different rates. Besides spectro-temporal modifications (e.g., shortening of acoustic cues, enhanced coarticulation)[34], naturally speaking faster induces changes in the phonetic realization of F0. Acoustic studies in fact reported a shift in the frequency of the pitch range with increasing speech rate, namely decreased or increased pitch depending on speakers and languages[35,36]. We aimed to characterize how cortical 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[37,38], 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 accelerated (i.e. time-compressed) sentences. Given that time-compression reduces signal duration while keeping its spectral and pitch content intact[39], time-compressed speech has the same F0 as normal rate speech despite a faster syllabic rate. By contrast, F0 inherently increases in naturally-produced 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[7,37], 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 coupling to sentences' F0 only occurs in the auditory cortex or whether it involves a more distributed cortical network, noting that previous work used region-of-interest based analyses, mostly in the primary auditory

116 cortex[e.g.23,32]. We are aware of only one stereotactic EEG / MEG study showing FFR in a
117 temporo-parieto-frontal cortical network in response to musical tones[40].
118 Our results reveal significant cortico-acoustic coupling in frequency ranges that are specific to
119 the sentences' fundamental frequency. Coupling indeed occurred in the higher frequency band
120 (90-97 Hz) for naturally-produced fast speech compared to normal rate speech, mirroring the
121 natural increase in speech F0. No change in coupling frequency (77-84 Hz) was on the contrary
122 observed for time-compressed sentences which have similar F0 to normal rate sentences.
123 Crucially, brain-to-F0 coupling was found in a right-lateralized network encompassing
124 temporal, parietal, precentral and postcentral cortices, known to contribute to speech perception
125 as well as to production. Such findings bring evidence that by adjusting to the rate of vocal fold
126 vibration, neural oscillations in sensorimotor regions contribute to the encoding of pitch and of
127 phonemic information conveyed by the periodicity of the speech signal.
128
129 **Results**
130 Twenty-four French native participants listened to natural normal rate (mean = 6.76 $\pm$ 0.57
131 syllables/s), natural fast rate (9.15 $\pm$ 0.60 syllables/s) and time-compressed sentences (9.15 $\pm$
132 0.60 syllables/s) while their MEG brain activity was recorded. The mean F0 of the natural
133 normal rate and time-compressed sentences was 81.3 Hz (SD 3.3) as calculated from the peaks
134 of the power spectrum of the sentences using Python. It increased to 91.9 Hz (SD 9.8) for
135 naturally accelerated sentences. Amplitude-modulated noise stimuli, generated with the
136 envelopes of normal rate and fast rate stimuli, and which lack pitch and verbal information,
137 were included as control conditions. We assessed neural tracking of changes in fundamental
138 frequency by computing cortico-acoustic coupling between the acoustic signals and source-
139 localized MEG time-series using phase-lag index (PLI) and, as a complementary approach,
140 coherence measures (see Methods for more details about the selected coupling metrics).

Because sentences' F0 in the different speech rate conditions overlapped to some extent, we analysed speech-brain coupling in subsets of sentences for which the F0s of the normal and time-compressed conditions were most distant from the F0s of the fast condition (see Methods and Supplementary Figure S1 and Table S1). The F0 in these subsets peaked at 80 Hz for natural normal rate and time-compressed sentences, and at 93 Hz for natural fast rate sentences. We therefore conducted the coupling analysis in two frequency bands of interest centred on these values, respectively, 77-84 Hz and 90-97 Hz. For statistical analysis, PLI and coherence measures for actual stimulus encoding were compared with measures obtained on surrogate data using a shuffling procedure and correction for multiple comparisons using, respectively, cluster-based correction and FDR (see Methods).

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 1). 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 Supplementary Table S2 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 1A 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 2). 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 cortico-acoustic coupling increase in any of the two frequency bands of interest (Figures 1 and 2).

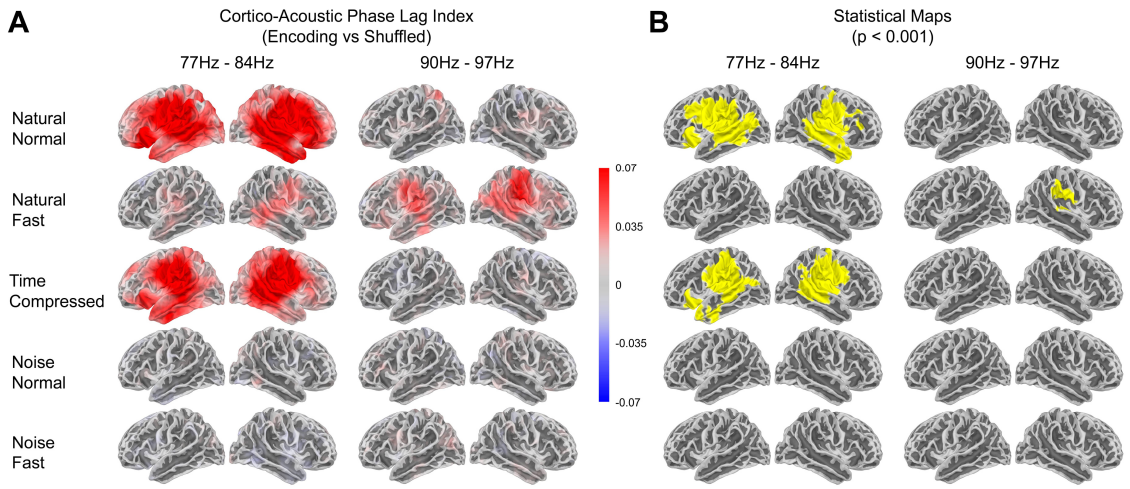


Figure 1. Cortical coupling to fundamental frequency in the frequency ranges encompassing the F0s of normal rate and time-compressed sentences (77-84 Hz) and of natural fast rate sentences (90-97 Hz). (A) Phase Lag Index (PLI) maps for coupling between the acoustic signal and brain oscillatory activity 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 white noise stimuli modulated with the amplitude envelope of normal rate and fast rate sentences, respectively.

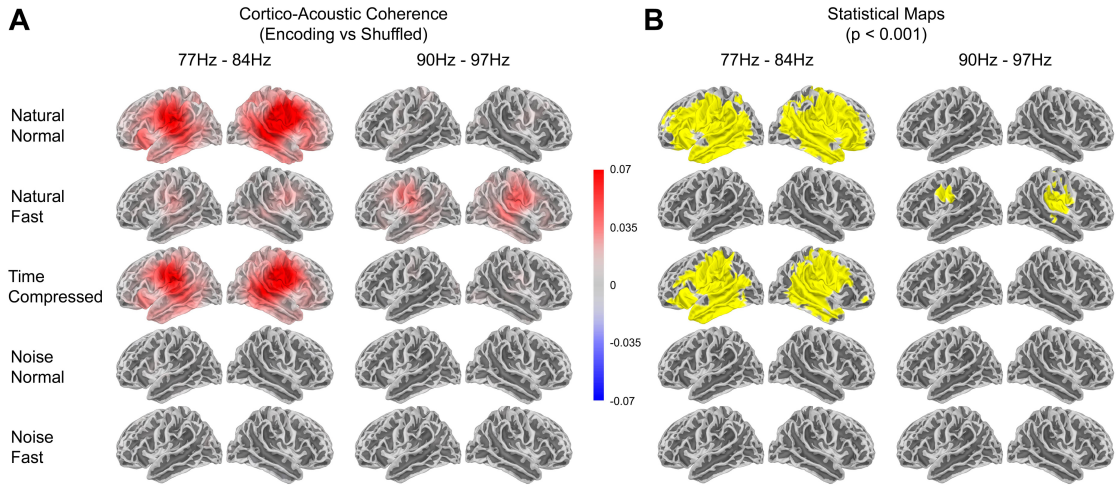


Figure 2: Cortico-acoustic coherence in the frequency ranges corresponding to the F0s in the normal rate and time-compressed conditions (77-84 Hz) and the natural fast rate condition (90-97 Hz). (A) Coherence maps for coupling between the acoustic signal and brain oscillatory activity during stimulus 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 across all speech conditions. 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 3, 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 3 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 3 D, H). Note from Supplementary Figures S2, S3 and S4 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 3 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.

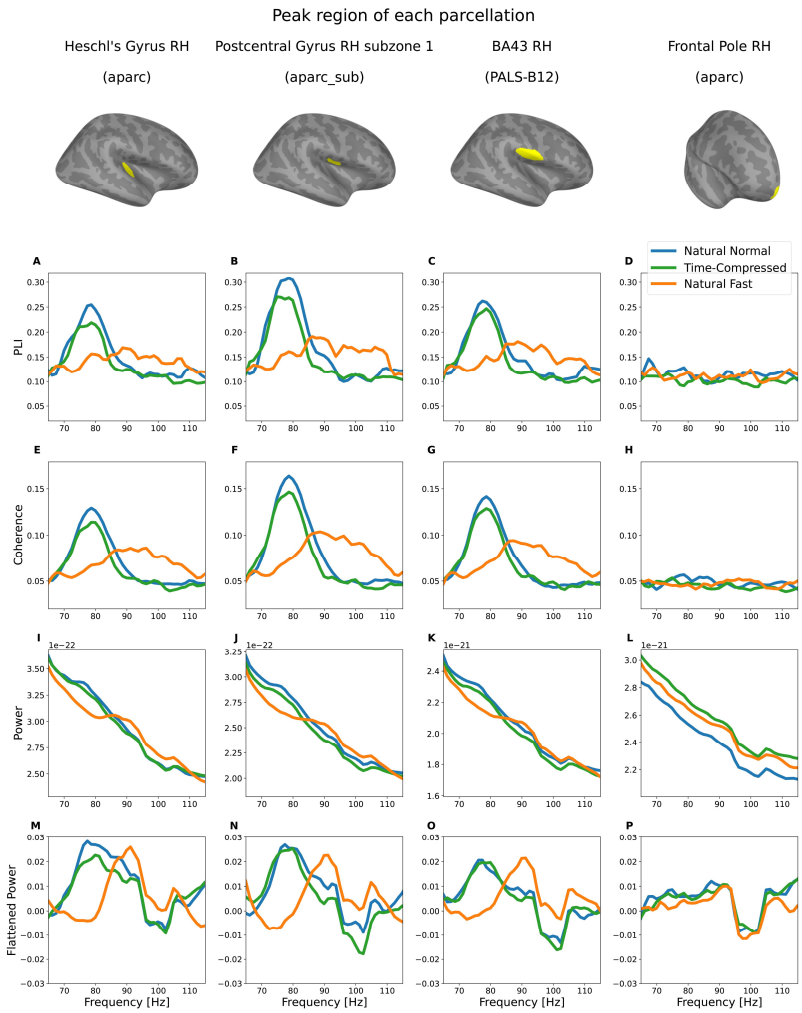


Figure 3. Coupling and power spectra for the three speech conditions in the top regions of the right hemisphere for each brain parcellation. The regions shown (one per column) correspond to the regions in each parcellation (respectively, aparc, aparc_sub and PALS-B12) with the highest peak of Phase Lag 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 S2, S3 and S4 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[41]). The specific models used for the aperiodic fits of these curves can be seen in Supplementary Figure S5.

We finally examined whether brain coupling to 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 4). 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 stimuli in any of the frequency ranges.

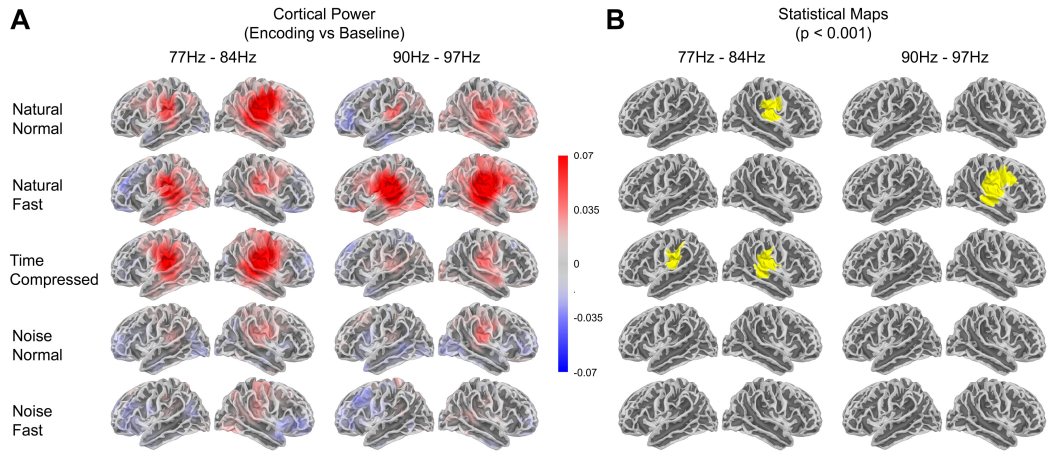


Figure 4. **Power modulations in the frequency bands matching the F0s of natural normal rate and time-compressed sentences (77-84 Hz) and of natural fast rate sentences (90-97 Hz).** (A) Power maps during the active period (i.e. stimulus encoding), as compared to the pre-stimulus baseline. The

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 brain coupling to speech fundamental frequency, with a shift in coupling frequency and power increase that echoed the natural 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. Overall, these findings therefore provide novel evidence that high-frequency cortical oscillations couple with the fundamental frequency of speech in case of 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[42–46]. FFR studies also reported responses in the right or bilateral auditory cortex to syllables or continuous speech[23,26–28]. The lack of coupling to amplitude-modulated noise further suggests that phase-locking does not reflect responses to the temporal envelope of sentences but to their high-frequency carrier information. 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 cortico-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 coupled their oscillatory activity with speech F0, irrespective of the parcellation used. Interestingly, these regions are part of a dorsal network underlying speech auditory-motor integration[47,48]. The somatosensory cortex has also been advocated to contribute to speech perception by closely interacting with auditory regions[49]. In this regard, the subcentral gyrus (BA 43) identified in our study, owning both sensory and motor properties[50,51], is seen as the ventral larynx motor area activated for the production of vocalizations and pitch control[52–56]. High-gamma activity in this region has been reported to increase before and during phoneme articulation[57] and for voiced consonant production[58], suggesting its role in the control of vocal fold movements. This area furthermore increases its connectivity with the auditory cortex during mere voicing perception[45], and it can decode articulatory features of heard phonemes[59–61]. Our results concur with a functional role of this ventral somatosensory area in speech perception as it is one of the regions that best couple its high-frequency oscillatory activity with 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[53,54,62–66]. 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[67]. We also found reliable coupling to 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[48,68,69] as well as with its role in phonological processing[70]. The insula is also assumed to contribute to auditory-motor speech integration[71]. It is indeed engaged in speech production and the control of laryngeal muscles and of F0 modulations[72–74], but also in prosodic perception and auditory temporal processing[75–78]. 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[79]. The insula moreover shares connections with auditory[80] and motor regions, including the laryngeal motor cortex[81,82]. Altogether, our findings thus putatively suggest that besides the primary auditory cortex targeted by previous work[23,24,32,83], a dorsal sensorimotor cortical network involved in speech auditory-motor integration sync to the periodicity of vocal fold vibration in perceived speech.

Whether and how our findings relate to the cortical FFR to repetitive syllables[21,23,24,27,28] is not straightforward. One difference is that 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[23,24]. Among the few studies that investigated F0 coupling to sentences or continuous speech, some focused on the brainstem[84–86], while others examined high-gamma auditory cortex activity in light of the cortical FFR[26,30,32,33,87]. 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. Crucially, by analyzing both coupling (using two measures, PLI and coherence) and power modulations at high frequencies, we bring new support for brain coupling to F0 in natural speech that occurs well beyond the auditory cortex, in regions that are involved in auditory-somato-motor integration (in agreement with [40] for tones).

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

319 to amplitude modulations. Cortical coupling to speech F0 may therefore capture the periodic
320 nature of the phenomenon. Coffey and collaborators[83] likewise proposed that their auditory
321 cortex FFR entailed oscillatory entrainment since the brain response outlasted syllable offset
322 (see also [40]). Along with this view, oscillatory coupling to F0 in our study could reflect a
323 predictive coding mechanism comparing speech input periodicity with internal temporal
324 expectations. Such an interpretation, although still speculative, would support the idea that
325 high-gamma activity mediates pitch-level prediction errors and their feedforward propagation
326 up to higher integrative areas in order to update predictions[88–90]. Cortico-acoustic coupling to
327 F0 in (pre)motor regions furthermore agrees with their known role in predictive coding, despite
328 this is usually observed in lower (delta and beta) frequencies[91–93]. Further cross-frequency
329 coupling and functional connectivity analyses between auditory and motor regions are needed
330 to fully substantiate the predictive coding interpretation and further delineate the role of speech
331 F0 coupling at the cortical level. As speech periodicity conveys segmental information, phase-
332 locking to F0 might contribute to phoneme identification by encoding the articulatory feature
333 of voicing or by extracting vowels' formant structure. Another, not exclusive possibility is that
334 it could serve more general communicative purposes, to track the relevant targets in an auditory
335 scene[29] and decrease sensitivity to environmental noise or competing speakers[83].
336 One last question relates to the upper limit of F0 cortical tracking. In our study and previous
337 work[23,27,83], phase-locking was examined for typical male voices with F0 not exceeding 100-
338 110 Hz. It has been proposed that the auditory cortex is not able, or hardly tracks frequencies
339 beyond 100 Hz[20], for pure tones[94,95] or for syllables produced by males[21] or females[33,87]. Some
340 intracranial recordings still showed phase-locked responses up to 200 Hz[19,26,96,97]. In the present
341 study, cortical tracking of natural fast speech was significant but less widespread than for
342 normal rate and time-compressed speech, although the three conditions highlighted a
343 comparable cortical network. This may relate to the general decrease in cortical frequency

representation towards higher frequencies and/or the greater variance in F0 for rapidly-produced sentences. Part of the F0 distribution in this condition indeed lied beyond 110 Hz (up to ~120 Hz) even after the selection procedure (Figure S1, panels B and D), which may have decreased the tracking capacity of cortical oscillations. Speech F0 therefore appears as an important parameter for the emergence of cortical tracking of speech periodicity, since phase-locking seems to decline with increasing F0. Alternatively, greater variability in natural fast speech F0 may have reduced statistical power, explaining the globally weaker effects in this condition. Further examining F0 tracking in different voices and under various speech rate or noisy conditions, will help better understanding this cortical contribution.

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[98]. 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 high-frequency cortical oscillations adaptively couple with the fundamental frequency (F0) of speech to match variations in F0 that inherently accompany natural speech rate acceleration. 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, involving 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.

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[99] = 94). 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 male 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 [100]. 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[101], was 6.76 syllables/s (SD 0.57) for the natural 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[102]). 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 across 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 S5 in the supplementary material). These peak frequencies are consistent with the computation of the median pitch values with Praat.

To assess whether brain-to-F0 coupling occurs with the high-frequency carrier information in speech and not with the modulations of the amplitude envelope, we generated amplitude-modulated noise stimuli) as control conditions. A total of 576 control stimuli were created by modulating Gaussian white noise (lacking linguistic content) with the amplitude envelope of each of the 288 natural normal rate and 288 natural fast rate sentences. 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. These filler sentences were used to encourage participants to pay attention to the auditory stimuli during the experiment (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

418 normal rate envelope and 96 with the fast rate envelope) and 48 filler sentences, for a total of
419 528 stimuli per list and per participant. Each stimulus appeared in each rate condition across all
420 participants but only once per list.
421
422 ***Experimental procedure***
423 While comfortably sitting in a sound-attenuated, magnetically-shielded recording room with a
424 screen in front of them, participants were presented binaurally with stimuli through air-
425 conducting tubes with foam insert earphones (Etymotic ER2 and ER3). Prior to the MEG
426 recording, their auditory detection thresholds were determined for each ear with a pure tone.
427 The level was then adjusted in order to present stimuli at 50 dB Sensation Level (SL) with a
428 central position (stereo) with respect to the participant's head.
429 Each trial of the experiment ran as follows: after a fixation cross displayed at the centre of the
430 screen for 1500 ms, an auditory stimulus (sentence or amplitude-modulated noise) was
431 presented, with the fixation cross remaining on the screen. Participants were instructed to
432 attentively listen to all stimuli while looking at the fixation cross and to press a button (response
433 button Neuroscan, Pantev) with their left index finger as quickly as possible whenever they
434 detected beep-sounds in (filler) sentences. The next trial started after an inter-trial interval (grey
435 screen) of 1250 ms. Participants detected 100% of the filler trials, which were excluded from
436 subsequent analysis. Each participant listened to the stimuli from one of the two experimental
437 lists. The total of 528 stimuli were pseudo-randomly divided into 8 blocks (66 stimuli each)
438 separated by short breaks to allow the participants to rest. Participants were familiarized with
439 the stimuli and task in a training phase (5 sentences different from the experimental stimuli).
440 To further maintain their attention throughout the experiment, they were informed that
441 questions on the heard sentences would be asked at the end of the experiment. More precisely,
442 they had to indicate, by ticking a yes/no box on a paper sheet, whether eight proposed sentences

443 (three of which were not experimental stimuli) were presented during the MEG experiment.
444 The experiment was run with Presentation software (Neurobehavioral Systems).
445
446 ***MEG data acquisition and analysis***
447 We recorded neuromagnetic brain activity of the 24 participants using a 275-channel whole-
448 head MEG system (CTF OMEGA 275, Canada) at 1200 Hz sampling rate. Head position within
449 the MEG helmet was monitored with three fiducial coils (nasion, left and right pre-auricular
450 points) placed on the participant's head. Horizontal and vertical eye movements were also
451 recorded with four electrooculography (EOG) electrodes. We recorded reference head position
452 before each of the 8 experimental blocks and tracked head movements throughout the
453 experiment using continuous head position identification (HPI).
454 Preprocessing of the MEG data was done with custom written Matlab scripts (Mathworks Inc.,
455 MA, USA) and the Fieldtrip toolbox[103]. Subsequent coupling analyses were carried out in
456 Python using custom scripts and the MNE-Python library[104]. In the following we describe the
457 processing of 1) speech recordings, 2) MEG data, 3) Magnetic Resonance Imaging (MRI) data,
458 and the analyses of 4) source localization, power and coupling, and 5) coupling maps and
459 statistics.
460
461 *Audio Recordings (reference signals)*
462 To examine the neural tracking of speech F0, we used the speech signal as the reference.
463 Inspection of the F0 distributions and power spectral densities revealed some spectral (including
464 F0s) overlap between sentences in the different speech rate conditions (see Supplementary
465 Figure S1 A and C). To test our hypothesis that brain oscillatory activity adjusts to F0 increase
466 with speech rate acceleration, we optimized the distinction between conditions in terms of F0s
467 by selecting a proportion of sentences in the natural normal rate and time-compressed

conditions with the lowest mean 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 mean 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 Supplementary Table 12). 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 Supplementary Figure S1 B and D). We analyzed brain-to-speech coupling in two frequency ranges centered on the F0 peak values (across sentences) 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).

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.

493

494 *MRI data*

495 T1-weighted structural MRIs (MRI 1.5 T, Siemens AvantoFit) were acquired for 23 of the 24
496 participants after the MEG study. Alignment of each MRI to the MEG coordinate system was
497 performed with the localization coils marked on the MRIs and the interactive alignment
498 interface available in MNE. Each MRI was segmented to compute each participant's head
499 model using the single-shell as volume conductor model[105]. To carry out group analyses in a
500 common space, we first spatially normalized a template anatomical grid from the Montreal
501 Neurological Institute (MNI) common coordinate system. MRI grids of each participant were
502 then aligned to this normalized template (this common grid has 5124 vertices). We finally
503 computed the cortical source space modelling for each participant using the individual MRI
504 volume, but with the surface grid based on the MNI template. This allowed transforming the
505 results to the same brain template for all participants for easier comparison.

506

507

508 *Source localization, power and cortico-acoustic coupling estimation*

509 We estimated source power and cortico-acoustic coupling given source reconstructions
510 calculated using Minimum Norm Estimation[106–108]. Before computing the coupling, we
511 resampled the speech signal to 1200 Hz to match the sampling frequency of the MEG data. In
512 this way, we could suitably assess the neural tracking of F0 by computing the coupling between
513 auditory signals and cortical activity using Phase Lag Index and coherence measures. The Phase
514 Lag Index provides information regarding the coupling of two signals with respect to the sign
515 of their phase difference[110]. It has the advantage of being insensitive to the amplitude of the
516 signals and moreover, neglecting zero-phase coupling that commonly arise because of volume
517 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[104]. 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.

Cortico-acoustic coupling maps, power maps and statistical analyses

542 The average coupling and power maps used for visualization in the Results section were
543 obtained by first applying the Fisher transform of the values. Then we computed the difference
544 between the active period and a set of control data (shuffled audio for coupling, and baseline
545 for power) for each participant and condition. In the case of power, the relative difference was
546 used. The values were then averaged across participants for each experimental condition and
547 subsequently mapped to the brain plots. Group statistical analysis was conducted on the raw
548 coupling and power values (without the Fisher transform) for the 23 participants with individual
549 MRIs. For PLI and power, we applied cluster-based permutation tests with 1024 permutations
550 [111]. For coherence, we performed an FDR corrected t-test. The tests are one-tailed for cortico-
551 acoustic coupling analysis and two-tailed for power analysis. All tests use a statistical
552 significance threshold of 0.001. We used the Desikan-Killiany atlas to identify the regions
553 detected as significant in the statistical analysis[112]. The top 10 regions with the highest mean t-
554 value based on the PLI statistical analysis for each frequency range, condition, and hemisphere
555 can be found in Supplementary Table S2.
556
557 To inspect more closely the changes in coupling frequency with F0 variation, we additionally
558 computed - for each of the three speech rate conditions - the coupling spectra during the
559 encoding period, both for PLI and coherence, across a wide range of frequencies (64-128 Hz)
560 and averaged through participants. This was done on all voxels and then averaged into the
561 regions of the Desikan-Killiany atlas[112] (aparc parcellation). We simultaneously explored two
562 other parcellations: aparc_sub[113] (corresponding to subdivisions of the regions of the aparc
563 parcellation) and PALS-B12[114] (which delineates Brodmann areas). Then, the top 8 regions in
564 the right hemisphere with the highest PLI peak independent of the speech rate conditions were
565 identified for each of the three parcellations. These regions are detailed in Figures S2, S3 and
566 S4. 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 3; see also[23]). 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[41]. The details of the model that fits each of the spectra involved in Figure 3 can be seen in Supplementary Figure S4.

Data availability: The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Ding, N. & Simon, J. Z. Cortical entrainment to continuous speech: functional roles and interpretations. *Front. Hum. Neurosci.* **8**, (2014).
2. Ghitza, O. Linking Speech Perception and Neurophysiology: Speech Decoding Guided by Cascaded Oscillators Locked to the Input Rhythm. *Front. Psychol.* **2**, (2011).
3. Giraud, A.-L. & Poeppel, D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat. Neurosci.* **15**, 511–517 (2012).
4. Peelle, J. E. & Davis, M. H. Neural Oscillations Carry Speech Rhythm through to Comprehension. *Front. Psychol.* **3**, (2012).
5. Ding, N., Chatterjee, M. & Simon, J. Z. Robust Cortical Entrainment to the Speech Envelope Relies on the Spectro-temporal Fine Structure. *NeuroImage* **88**, 41–46 (2014).
6. Gross, J. *et al.* Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biol.* **11**, e1001752 (2013).
7. Keitel, A., Gross, J. & Kayser, C. Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLOS Biol.* **16**, e2004473 (2018).
8. Kösem, A., Basirat, A., Azizi, L. & Wassenhove, V. van. High-frequency neural activity predicts word parsing in ambiguous speech streams. *J. Neurophysiol.* **116**, 2497–2512 (2016).

9. Peelle, J. E., Gross, J. & Davis, M. H. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* **23**, 1378–1387 (2013).

10. Kubanek, J., Brunner, P., Gunduz, A., Poeppel, D. & Schalk, G. The Tracking of Speech Envelope in the Human Cortex. *PLOS ONE* **8**, e53398 (2013).

11. Lizarazu, M., Lallier, M. & Molinaro, N. Phase–amplitude coupling between theta and gamma oscillations adapts to speech rate. *Ann. N. Y. Acad. Sci.* **1453**, 140–152 (2019).

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

13. Nourski, K. V. *et al.* Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. *J. Neurosci.* **29**, 15564–15574 (2009).

14. Zion Golumbic, E. M. *et al.* Mechanisms Underlying Selective Neuronal Tracking of Attended Speech at a 'Cocktail Party'. *Neuron* **77**, 980–991 (2013).

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

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

17. Kraus, N. & Nicol, T. Brainstem origins for cortical 'what' and 'where' pathways in the auditory system. *Trends Neurosci.* **28**, 176–181 (2005).

18. Kraus, N., Anderson, S., White-Schwoch, T., Fay, R. & Popper, A. The Frequency-Following Response: A Window into Human Communication. in *The frequency-following response* 1–15 (New York: Springer, 2017). doi:10.1007/978-3-319-47944-6.

19. Coffey, E. B. J. *et al.* Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* **10**, 1–10 (2019).

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

21. Bidelman, G. M. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage* **175**, 56–69 (2018).

22. Bidelman, G. M. & Momtaz, S. 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 (2021).

23. Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S. & Zatorre, R. J. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* **7**, 11070 (2016).

24. Coffey, E. B. J., Musacchia, G. & Zatorre, R. J. Cortical Correlates of the Auditory Frequency-Following and Onset Responses: EEG and fMRI Evidence. *J. Neurosci.* **37**, 830–838 (2017).

25. Gnanateja, G. N. *et al.* Frequency-Following Responses to Speech Sounds Are Highly Conserved across Species and Contain Cortical Contributions. *eNeuro* **8**, ENEURO.0451-21.2021 (2021).

26. Guo, N. *et al.* Speech frequency-following response in human auditory cortex is more than a simple tracking. *NeuroImage* **226**, 117545 (2021).

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

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

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

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

31. Greenlaw, K. M., Puschmann, S. & Coffey, E. B. J. Decoding of Envelope vs. Fundamental Frequency During Complex Auditory Stream Segregation. *Neurobiol. Lang.* **1**, 268–287 (2020).

32. Kulasingham, J. P. *et al.* High gamma cortical processing of continuous speech in younger and older listeners. *NeuroImage* **222**, 117291 (2020).

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

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

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

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

37. Hincapié-Casas, A. S. *et al.* Neural oscillations track natural but not artificial fast speech: Novel insights from speech-brain coupling using MEG. *bioRxiv* 2020.10.20.344895 (2021) doi:10.1101/2020.10.20.344895.

38. Pefkou, M., Arnal, L. H., Fontolan, L. & Giraud, A.-L. Theta- and beta-band neural activity reflect independent syllable tracking and comprehension of time-compressed speech. *J. Neurosci.* 2882–16 (2017) doi:10.1523/JNEUROSCI.2882-16.2017.

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

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

41. Donoghue, T. *et al.* Parameterizing neural power spectra into periodic and aperiodic components. *Nat. Neurosci.* **23**, 1655–1665 (2020).

42. Hyde, K. L., Peretz, I. & Zatorre, R. J. Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia* **46**, 632–639 (2008).

43. Johnsrude, I. S., Penhune, V. B. & Zatorre, R. J. Functional specificity in the right human auditory cortex for perceiving pitch direction. *Brain* **123**, 155–163 (2000).

44. Meyer, M., Alter, K., Friederici, A. D., Lohmann, G. & von Cramon, D. Y. FMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Hum. Brain Mapp.* **17**, 73–88 (2002).

45. Tamura, S. *et al.* Multi-modal imaging of the auditory-larynx motor network for voicing perception. *NeuroImage* **251**, 118981 (2022).

677 46. Zatorre, R. J. & Belin, P. Spectral and Temporal Processing in Human Auditory Cortex. *Cereb.*
678 *Cortex* **11**, 946–953 (2001).
679 47. Hickok, G. & Poeppel, D. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8**,
680 393–402 (2007).
681 48. Rauschecker, J. P. & Scott, S. K. Maps and streams in the auditory cortex: nonhuman primates
682 illuminate human speech processing. *Nat. Neurosci.* **12**, 718–724 (2009).
683 49. Franken, M. K., Liu, B. C. & Ostry, D. J. Towards a somatosensory theory of speech perception.
684 *J. Neurophysiol.* **128**, 1683–1695 (2022).
685 50. Belyk, M. & Brown, S. The origins of the vocal brain in humans. *Neurosci. Biobehav. Rev.* **77**,
686 177–193 (2017).
687 51. Cogan, G. B. *et al.* Sensory–motor transformations for speech occur bilaterally. *Nature* **507**, 94–
688 98 (2014).
689 52. Conant, D., Bouchard, K. E. & Chang, E. F. Speech Map in the Human Ventral Sensory-Motor
690 Cortex. *Curr. Opin. Neurobiol.* **0**, 63–67 (2014).
691 53. Dichter, B. K., Breshears, J. D., Leonard, M. K. & Chang, E. F. The Control of Vocal Pitch in
692 Human Laryngeal Motor Cortex. *Cell* **174**, 21-31.e9 (2018).
693 54. Eichert, N., Papp, D., Mars, R. B. & Watkins, K. E. Mapping Human Laryngeal Motor Cortex
694 during Vocalization. *Cereb. Cortex* **30**, 6254–6269 (2020).
695 55. Eichert, N., Watkins, K. E., Mars, R. B. & Petrides, M. Morphological and functional variability
696 in central and subcentral motor cortex of the human brain. *Brain Struct. Funct.* **226**, 263–279
697 (2021).
698 56. Grabski, K. *et al.* Functional MRI assessment of orofacial articulators: Neural correlates of lip,
699 jaw, larynx, and tongue movements. *Hum. Brain Mapp.* **33**, 2306–2321 (2012).
700 57. Bouchard, K. E., Mesgarani, N., Johnson, K. & Chang, E. F. Functional Organization of Human
701 Sensorimotor Cortex for Speech Articulation. *Nature* **495**, 327–332 (2013).
702 58. Toyoda, G. *et al.* Electrocorticographic correlates of overt articulation of 44 English phonemes:
703 Intracranial recording in children with focal epilepsy. *Clin. Neurophysiol.* **125**, 1129–1137 (2014).

704 59. Archila-Meléndez, M. E. *et al.* Sensorimotor Representation of Speech Perception. Cross-
705 Decoding of Place of Articulation Features during Selective Attention to Syllables in 7T fMRI.
706 *eNeuro* **5**, (2018).
707 60. Arsenault, J. S. & Buchsbaum, B. R. Distributed Neural Representations of Phonological Features
708 during Speech Perception. *J. Neurosci.* **35**, 634–642 (2015).
709 61. Correia, J. M., Jansma, B. M. B. & Bonte, M. Decoding Articulatory Features from fMRI
710 Responses in Dorsal Speech Regions. *J. Neurosci.* **35**, 15015–15025 (2015).
711 62. Brown, S., Ngan, E. & Liotti, M. A Larynx Area in the Human Motor Cortex. *Cereb. Cortex* **18**,
712 837–845 (2008).
713 63. D'Ausilio, A., Bufalari, I., Salmas, P., Busan, P. & Fadiga, L. Vocal pitch discrimination in the
714 motor system. *Brain Lang.* **118**, 9–14 (2011).
715 64. Liang, B., Li, Y., Zhao, W. & Du, Y. Bilateral human laryngeal motor cortex in perceptual decision
716 of lexical tone and voicing of consonant. *Nat. Commun.* **14**, 4710 (2023).
717 65. Lu, J. *et al.* Neural control of lexical tone production in human laryngeal motor cortex. *Nat.*
718 *Commun.* **14**, 6917 (2023).
719 66. Simonyan, K. The laryngeal motor cortex: its organization and connectivity. *Curr. Opin.*
720 *Neurobiol.* **28**, 15–21 (2014).
721 67. Sammler, D., Grosbras, M.-H., Anwander, A., Bestelmeyer, P. E. G. & Belin, P. Dorsal and
722 Ventral Pathways for Prosody. *Curr. Biol.* **25**, 3079–3085 (2015).
723 68. Rauschecker, J. P. An expanded role for the dorsal auditory pathway in sensorimotor control and
724 integration. *Hear. Res.* **271**, 16–25 (2011).
725 69. Shum, M., Shiller, D. M., Baum, S. R. & Gracco, V. L. Sensorimotor integration for speech motor
726 learning involves the inferior parietal cortex. *Eur. J. Neurosci.* **34**, 1817–1822 (2011).
727 70. Hartwigsen, G. *et al.* Phonological decisions require both the left and right supramarginal gyri.
728 *Proc. Natl. Acad. Sci. U. S. A.* **107**, 16494–16499 (2010).
729 71. Mutschler, I. *et al.* Functional organization of the human anterior insular cortex. *Neurosci. Lett.*
730 **457**, 66–70 (2009).

731 72. Ackermann, H. & Riecker, A. The contribution(s) of the insula to speech production: a review of
732 the clinical and functional imaging literature. *Brain Struct. Funct.* **214**, 419–433 (2010).
733 73. Dronkers, N. F. A new brain region for coordinating speech articulation. *Nature* **384**, 159–161
734 (1996).
735 74. Howell, P., Jiang, J., Peng, D. & Lu, C. Neural control of fundamental frequency rise and fall in
736 Mandarin tones. *Brain Lang.* **121**, 35–46 (2012).
737 75. Bamiou, D.-E., Musiek, F. E. & Luxon, L. M. The insula (Island of Reil) and its role in auditory
738 processing: Literature review. *Brain Res. Rev.* **42**, 143–154 (2003).
739 76. Bamiou, D.-E. *et al.* Auditory temporal processing deficits in patients with insular stroke.
740 *Neurology* **67**, 614–619 (2006).
741 77. Oh, A., Duerden, E. G. & Pang, E. W. The role of the insula in speech and language processing.
742 *Brain Lang.* **135**, 96–103 (2014).
743 78. Tregellas, J. R., Davalos, D. B. & Rojas, D. C. Effect of task difficulty on the functional anatomy
744 of temporal processing. *NeuroImage* **32**, 307–315 (2006).
745 79. Zhang, Y. *et al.* The Roles of Subdivisions of Human Insula in Emotion Perception and Auditory
746 Processing. *Cereb. Cortex* **29**, 517–528 (2019).
747 80. Ghaziri, J. *et al.* The Corticocortical Structural Connectivity of the Human Insula. *Cereb. Cortex*
748 **27**, 1216–1228 (2017).
749 81. Simonyan, K., Ostuni, J., Ludlow, C. L. & Horwitz, B. Functional But Not Structural Networks
750 of the Human Laryngeal Motor Cortex Show Left Hemispheric Lateralization during Syllable But
751 Not Breathing Production. *J. Neurosci.* **29**, 14912–14923 (2009).
752 82. Simonyan, K. & Horwitz, B. Laryngeal Motor Cortex and Control of Speech in Humans. *Neurosci.*
753 *Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **17**, 197–208 (2011).
754 83. Coffey, E. B. J., Arseneau-Bruneau, I., Zhang, X., Baillet, S. & Zatorre, R. J. Oscillatory
755 entrainment of the Frequency Following Response in auditory cortical and subcortical structures.
756 *J. Neurosci.* (2021) doi:10.1523/JNEUROSCI.2313-20.2021.
757 84. Etard, O., Kegler, M., Braiman, C., Forte, A. E. & Reichenbach, T. Decoding of selective attention
758 to continuous speech from the human auditory brainstem response. *NeuroImage* **200**, 1–11 (2019).

85. Forte, A. E., Etard, O. & Reichenbach, T. The human auditory brainstem response to running speech reveals a subcortical mechanism for selective attention. *eLife* **6**, e27203 (2017).

86. Maddox, R. K. & Lee, A. K. C. Auditory Brainstem Responses to Continuous Natural Speech in Human Listeners. *eNeuro* **5**, (2018).

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

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

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

90. van Pelt, S. *et al.* Beta- and gamma-band activity reflect predictive coding in the processing of causal events. *Soc. Cogn. Affect. Neurosci.* **11**, 973–980 (2016).

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

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

93. Park, H., Ince, R. A. A., Schyns, P. G., Thut, G. & Gross, J. Frontal Top-Down Signals Increase Coupling of Auditory Low-Frequency Oscillations to Continuous Speech in Human Listeners. *Curr. Biol.* **25**, 1649–1653 (2015).

94. Gorina-Careta, N., Kurkela, J. L. O., Hämäläinen, J., Astikainen, P. & Escera, C. Neural generators of the frequency-following response elicited to stimuli of low and high frequency: A magnetoencephalographic (MEG) study. *NeuroImage* **231**, 117866 (2021).

95. Holmes, E., Purcell, D. W., Carlyon, R. P., Gockel, H. E. & Johnsrude, I. S. Attentional Modulation of Envelope-Following Responses at Lower (93–109 Hz) but Not Higher (217–233 Hz) Modulation Rates. *J. Assoc. Res. Otolaryngol.* **19**, 83–97 (2018).

96. Brugge, J. F. *et al.* Coding of Repetitive Transients by Auditory Cortex on Heschl's Gyrus. *J. Neurophysiol.* **102**, 2358–2374 (2009).

97. Nourski, K. V. *et al.* Sound identification in human auditory cortex: Differential contribution of local field potentials and high gamma power as revealed by direct intracranial recordings. *Brain Lang.* **148**, 37–50 (2015).

98. Duecker, K. *et al.* Challenges and Approaches in the Study of Neural Entrainment. *J. Neurosci. Off. J. Soc. Neurosci.* **44**, e1234242024 (2024).

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

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

101. Boersma, P. & Weenik, D. *Doing Phonetics by Computer [Computer Program]*. (2012).

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

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

104. Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, (2013).

105. Nolte, G. 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**, 3637–52 (2003).

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

107. Gramfort, A. *et al.* MNE software for processing MEG and EEG data. *NeuroImage* **86**, 446–460 (2014).

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

109. Gross, J. *et al.* Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 694–9 (2001).

110. Stam, C. J., Nolte, G. & Daffertshofer, A. Phase lag index: Assessment of functional connectivity from multichannel EEG and MEG with diminished bias from common sources. *Hum. Brain Mapp.* **28**, 1178–1193 (2007).

111. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–90 (2007).

112. Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).

113. Khan, S. *et al.* Maturation trajectories of cortical resting-state networks depend on the mediating frequency band. *NeuroImage* **174**, 57–68 (2018).

114. Van Essen, D. C. A Population-Average, Landmark- and Surface-based (PALS) atlas of human cerebral cortex. *NeuroImage* **28**, 635–662 (2005).

Acknowledgements: We thank Damien Gouy for recording the stimuli, Emmanuel Ferragne for his help with Praat software, and Mélanie Canault for discussions on articulatory phonetics. V.B. discloses support for the research of this work from the French National Research Agency ANR (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). Y-J.M-R discloses support from 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 DEFIS program (ANR-09-EMER-002). K.J. discloses support for the research of this work 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 (the NSERC Discovery Grant also provided support to A-S.H-C).

Author contributions: Yorguin-Jose Mantilla-Ramos: Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. Ana-Sofía Hincapié-Casas: Data curation, Formal analysis, Software. Annalisa Pascarella: Methodology. Tarek Lajnef: Methodology. Richard M. Leahy: Methodology. Emily Coffey: Writing – Review and editing.

Karim Jerbi: Conceptualization, Methodology, Software, Validation, Resources, Writing – original draft, Supervision, Project administration, Funding acquisition. Véronique Boulenger: Conceptualization, Methodology, Validation, Investigation, Resources, Writing – original draft, Supervision, Project administration, Funding acquisition.

Declaration of competing interests: The authors declare no competing interests.

Supplementary Material

Content:

Table S1-S2

Figures S1-S5

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 S1: 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 were 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.

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 S2: 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 1. 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).

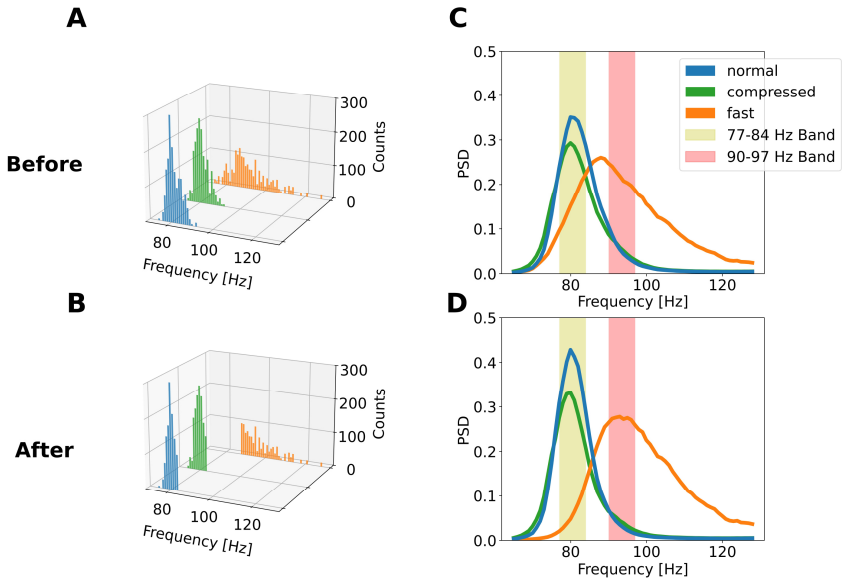


Figure S1: Optimizing sentence selection (normal rate vs. fast rate) in order to minimize F0 overlap across speech conditions. Sentences with the lowest F0s (<83 Hz) in the normal rate/time-compressed conditions and sentences with the highest F0s (>87 Hz) in the fast rate condition were selected. In 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 S2 in the supplementary material 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).

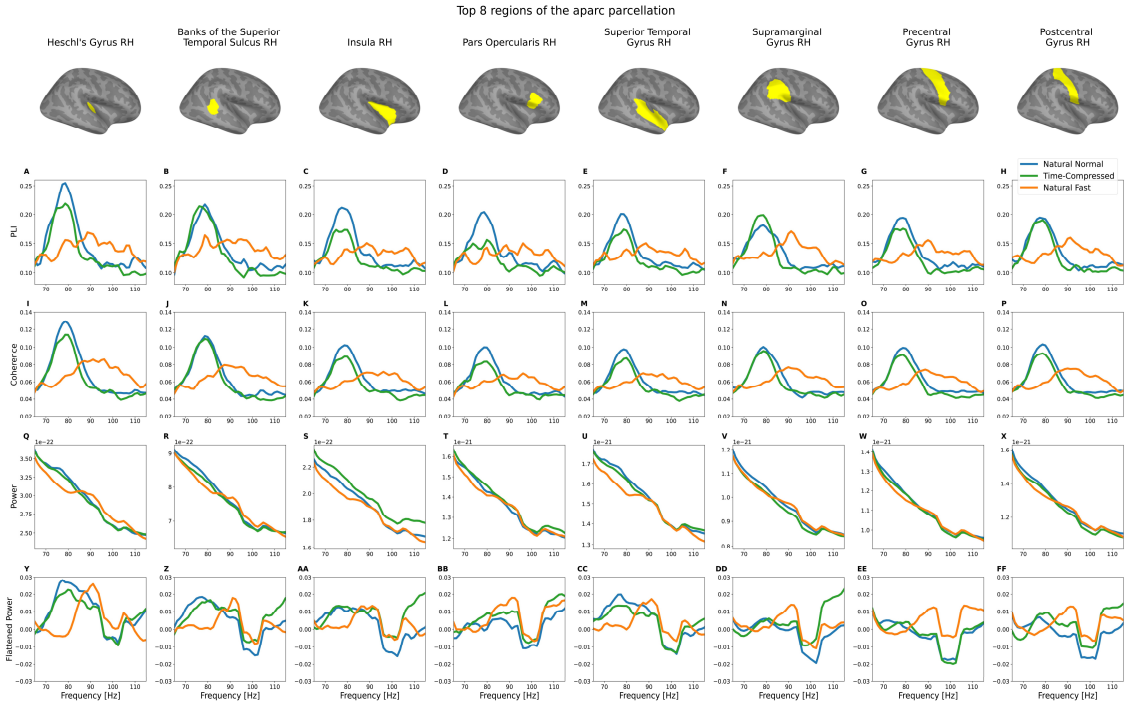
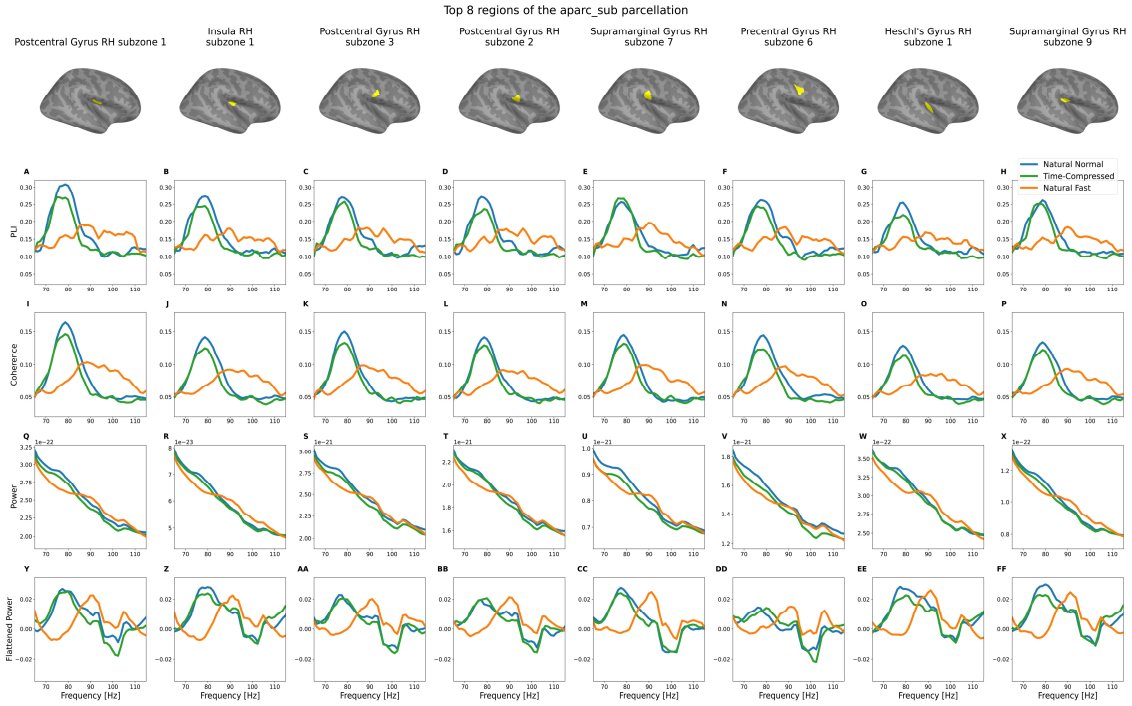


Figure S2: Top 8 regions of the right hemisphere of the aparc parcellation showing the highest spectral PLI peak. The regions shown (one per column) correspond to the 8 regions of the aparc parcellation with the highest peak PLI coupling (from left to right). The curves correspond to the average over 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.



927

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

938

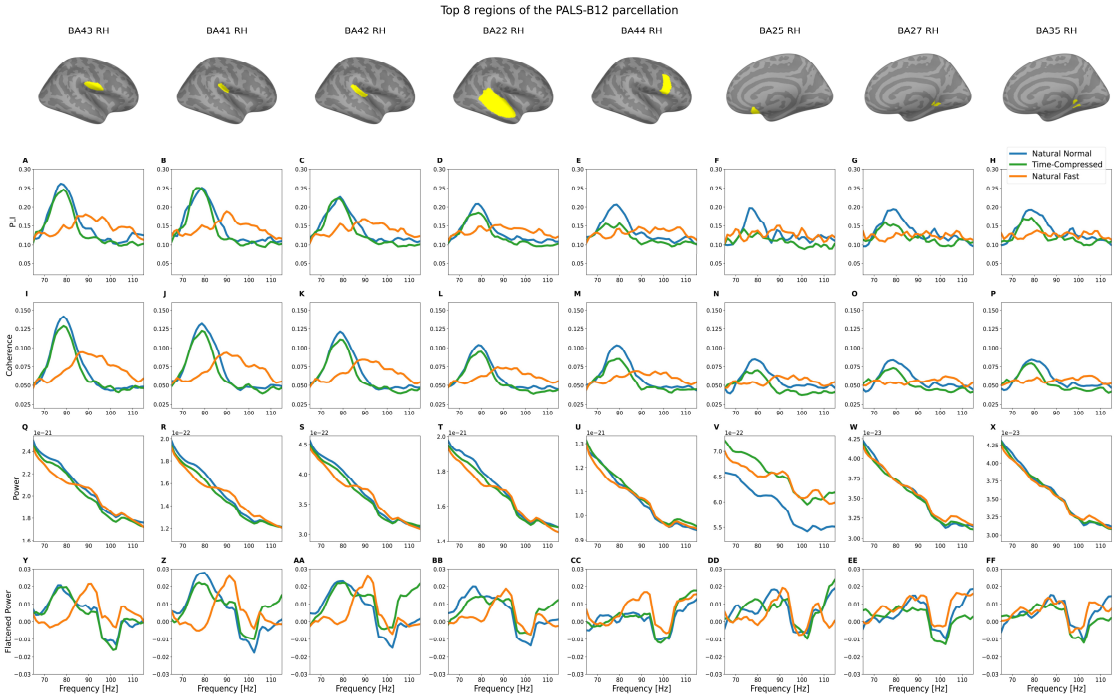


Figure S4: Top 8 regions of the right hemisphere of the PALS-B12 parcellation. The regions shown (one in each column) correspond to the 8 regions of the PALS-B12 parcellation with the highest peak PLI coupling (from left to right). The curves shown correspond to the average over subjects and voxels inside 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.

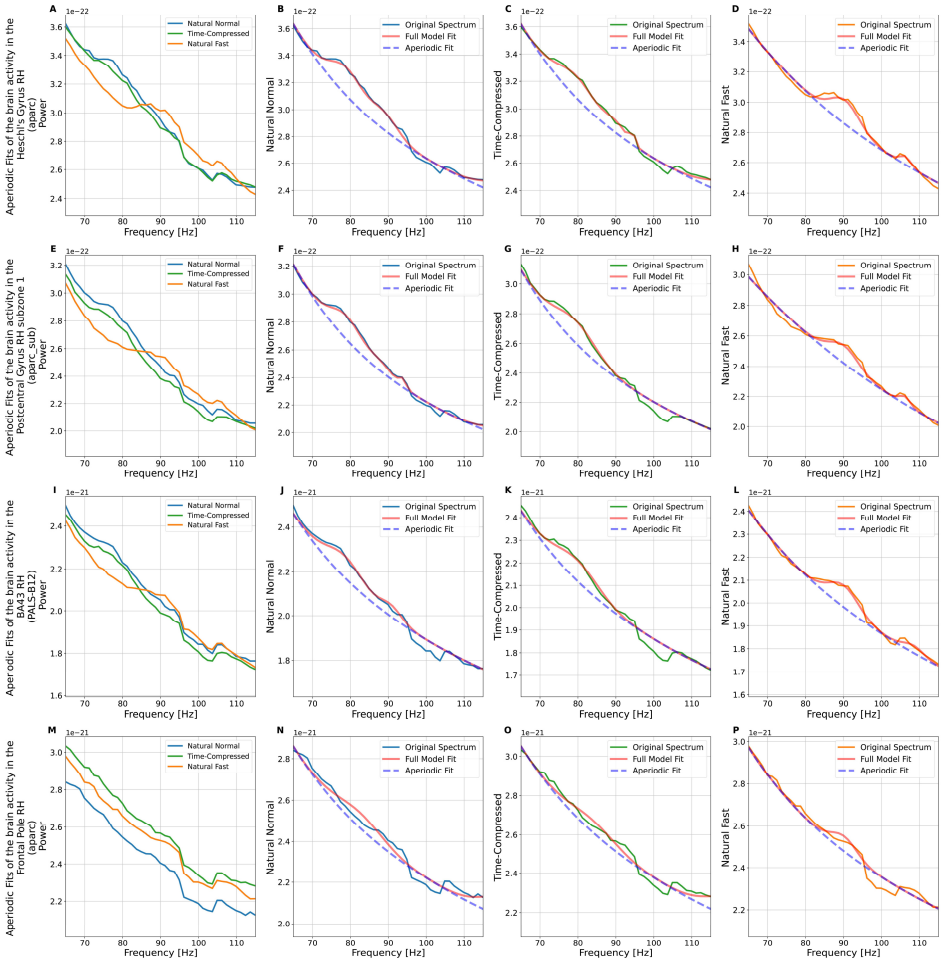


Figure S5: Aperiodic fits of the mean brain activity power spectra in each region. (A), (B), (C), (D) correspond to the right Heschl's gyrus (from aparc parcellation)**.** (A) Power spectra for the three speech conditions (same as Figure 3I; normal rate in blue, time-compressed in green and fast rate 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).