PLOS BIOLOGY

SHORT REPORTS

Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features

Anne Keitel[1]*, **Joachim Gross**[1,2], **Christoph Kayser**[1,3]

1 Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom, **2** Institute for Biomagnetism and Biosignalanalysis, University of Münster, Münster, Germany, **3** Cognitive Neuroscience, Bielefeld University, Bielefeld, Germany

* anne.keitel@glasgow.ac.uk

Citation: Keitel A, Gross J, Kayser C (2018) Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. PLoS Biol 16(3): e2004473. https://doi.org/10.1371/journal.pbio.2004473

Academic Editor: Jennifer Bizley, University College London, United Kingdom of Great Britain and Northern Ireland

Received: October 9, 2017

Accepted: February 21, 2018

Published: March 12, 2018

Data Availability Statement: All relevant data underlying the results presented in the manuscript are available from the Dryad database (doi:10.5061/dryad.1qq7050).

Funding: BBSRC http://www.bbsrc.ac.uk/ (grant number BB/L027534/1) to CK and JG. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. European Research Council https://erc.europa.eu/ (grant number 646657) to CK. The funder had no role in study design, data collection

Abstract

During online speech processing, our brain tracks the acoustic fluctuations in speech at different timescales. Previous research has focused on generic timescales (for example, delta or theta bands) that are assumed to map onto linguistic features such as prosody or syllables. However, given the high intersubject variability in speaking patterns, such a generic association between the timescales of brain activity and speech properties can be ambiguous. Here, we analyse speech tracking in source-localised magnetoencephalographic data by directly focusing on timescales extracted from statistical regularities in our speech material. This revealed widespread significant tracking at the timescales of phrases (0.6–1.3 Hz), words (1.8–3 Hz), syllables (2.8–4.8 Hz), and phonemes (8–12.4 Hz). Importantly, when examining its perceptual relevance, we found stronger tracking for correctly comprehended trials in the left premotor (PM) cortex at the phrasal scale as well as in left middle temporal cortex at the word scale. Control analyses using generic bands confirmed that these effects were specific to the speech regularities in our stimuli. Furthermore, we found that the phase at the phrasal timescale coupled to power at beta frequency (13–30 Hz) in motor areas. This cross-frequency coupling presumably reflects top-down temporal prediction in ongoing speech perception. Together, our results reveal specific functional and perceptually relevant roles of distinct tracking and cross-frequency processes along the auditory–motor pathway.

Author summary

How we comprehend speech—and how the brain encodes information from a continuous speech stream—is of interest for neuroscience, linguistics, and research on language disorders. Previous work that examined dynamic brain activity has addressed the issue of comprehension only indirectly, by contrasting intelligible speech with unintelligible speech or baseline activity. Recent work, however, suggests that brain areas can show similar stimulus-driven activity but differently contribute to perception or comprehension. To directly address the perceptual relevance of dynamic brain activity for speech encoding, we used a straightforward, single-trial comprehension measure. Furthermore, previous work has

and analysis, decision to publish, or preparation of the manuscript. Wellcome Trust https://wellcome.ac.uk/ (grant number 098433) to JG. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: BA, Brodmann area; c_v, coefficient of variation; FDR, false discovery rate; HG, Heschl gyrus; LCMV, linear constraint minimum variance; MEG, magnetoencephalography; MI, mutual information; MRI, magnetic resonance image; MTG, middle temporal gyrus; n.s., not significant; PAC, phase-amplitude coupling; PM, premotor; T_{sum}, summed *t* values.

been vague regarding the analysed timescales. We therefore base our analysis directly on the timescales of phrases, words, syllables, and phonemes of our speech stimuli. By incorporating these two conceptual innovations, we demonstrate that different areas of the brain track acoustic information at the time-scales of words and phrases. Moreover, our results suggest that the motor cortex uses a cross-frequency coupling mechanism to predict the timing of phrases in ongoing speech. Our findings suggest spatially and temporally distinct brain mechanisms that directly shape our comprehension.

Speech consists of hierarchically organised linguistic segments such as phrases, words, syllables, and phonemes [1,2]. To comprehend speech, a listener needs to parse the continuous stream into these segments [3]. One mechanism that has been proposed to fulfil such a role is the tracking of speech information in dynamic brain activity (often termed speech-to-brain entrainment) [4,5]. Such tracking is observed as a precise alignment of rhythmic brain activity to the temporal characteristics of speech. Previous studies commonly focused on brain activity at the timescales of traditional delta (1–4 Hz) and theta (4–8 Hz) bands [2,6,7,8]. These have been suggested to reflect the neural processing of prosodic and syllabic speech features [e.g., 3]. However, such a general association between timescales in brain activity and speech properties is difficult [9]. First, there are large interindividual differences in speaking rate and use of prosody. For example, the syllabic rate (sometimes interchangeably used with speech modulation rate) has been found within a range of 2 to 20 Hz across studies [10,11,12,13]. This indicates that the syllabic rate does not always fall into the often used theta frequency band. Second, the association between specific timescales and linguistic or phonological features remains ambiguous because multiple properties have been associated with delta (stress, intonation, phrase structure) [14,15,16] and theta bands (syllables, jaw movements) [17,18]. In the present study, we therefore take a different approach and first extract linguistic timescales from the speech corpus, based on stimulus-specific regularities. We then use these to investigate the neural mechanisms underlying speech encoding. We believe that any observed neural or perceptual effects at these linguistically specific timescales allow a more straightforward and mechanistically more specific interpretation than effects at generic timescales.

The functional interpretation of speech-to-brain entrainment is further hampered by a lack of a systematic assessment of where (in the brain) and at which timescale it is relevant for a perceptual outcome. By combining neural recordings with a behavioural measure, it has recently been shown that much distributed neural activity represents processes contributing to sensory encoding and perception, and only some focal activity is causally related to perceptual decisions [19,20]. Brain areas can therefore have similar stimulus-driven responses but different causal relationships with perceptual decisions [20]. By contrast, most studies attempting to link comprehension and speech tracking have done so only indirectly (i.e., without probing perceptual decisions) (but see [21]). These studies typically vary speech intelligibility by using background noise [22,23], noise vocoding [7,24], reversed speech [2,25,26,27], or by shuffling syllables [28]. Thereby, they have revealed a link between neural delta and theta band tracking and intelligibility during listening to continuous speech [2,5,21]. Interestingly, some of these studies have demonstrated widespread low-frequency tracking across the brain, going beyond auditory cortex and encompassing prefrontal and motor areas [2,22]. However, because many studies contrast brain activity during actual speech encoding with a surrogate dataset reflecting no speech–brain relation, these only demonstrate the existence of a speech-to-brain encoding process but not its functional or perceptual relevance. This notion is also supported by the

finding that auditory entrainment can be observed for subthreshold stimuli [29], suggesting that there is no necessary perceptual consequence of this tracking process. Therefore, it remains uncertain whether and where the tracking of speech by dynamic brain activity is perceptually relevant for comprehension at the single-trial level.

In the present study, participants performed a comprehension task on natural, structurally identical sentences embedded in noise. Sentences were semantically unpredictable, but meaningful. Using magnetoencephalography (MEG), we analysed speech tracking (quantified by mutual information [MI]) in source-localised data at the single-trial level. We then directly tested the perceptual relevance of speech tracking at timescales mapping onto linguistic categories such as phrasal elements, words, syllables, and phonemes.

Results

Behavioural results

Participants listened to single sentences and indicated after each trial which (out of 4) target words occurred in the sentence. Participants reported the correct target word, on average, in 69.7 ± 7.1% (mean ± SD) of trials, with chance level being 25%. Performance of individual participants ranged from 56.1% to 81.1%, allowing a comparison between correct and incorrect trials for each participant.

Overall speech tracking

The MI between the source-localised, Hilbert-transformed MEG time series and the Hilbert-transformed speech envelope was computed within 4 frequency bands. These reflected the rates of phrases (0.6–1.3 Hz), words (1.8–3 Hz), syllables (2.8–4.8 Hz), and phonemes (8–12.4 Hz) in the stimulus corpus (for an example sentence, see Fig 1). The boundaries of each band were defined based on the slowest and fastest event per linguistic category across sentences (see Materials and methods). When compared to surrogate data, speech MI was significant in all analysed frequency bands (Fig 2A; phrases: $T_{sum}(19) = 32{,}262.9$, $p < .001$; words: $T_{sum}(19) = 22{,}2243.8$, $p < .001$; syllables: $T_{sum}(19) = 13{,}428.6$, $p < .001$; phonemes: $T_{sum}(19) = 1{,}294.0$, $p = .018$). As in previous studies, MI was strongest in early auditory areas [2,22,30] and decreased with increasing frequency [6,22]. Tracking of phrases, words, and syllables was reflected in a bilateral cluster, whereas phoneme tracking was only significant in the right hemisphere. These results confirm the previously reported existence of speech encoding in rhythmic brain activity versus a null hypothesis of no encoding but do not speak on the perceptual relevance.

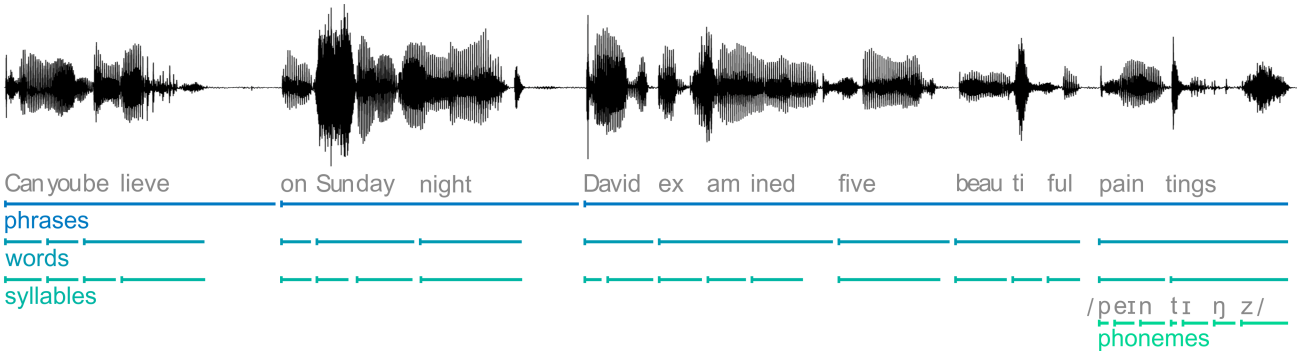

Fig 1. Example sentence from the stimulus material. Shown is the acoustic waveform (black line) as well as its segmentation into phrases, words, syllables, and phonemes (last word only). Example sentence deposited in the Dryad repository: https://doi.org/10.5061/dryad.1qq7050 [31].

https://doi.org/10.1371/journal.pbio.2004473.g001

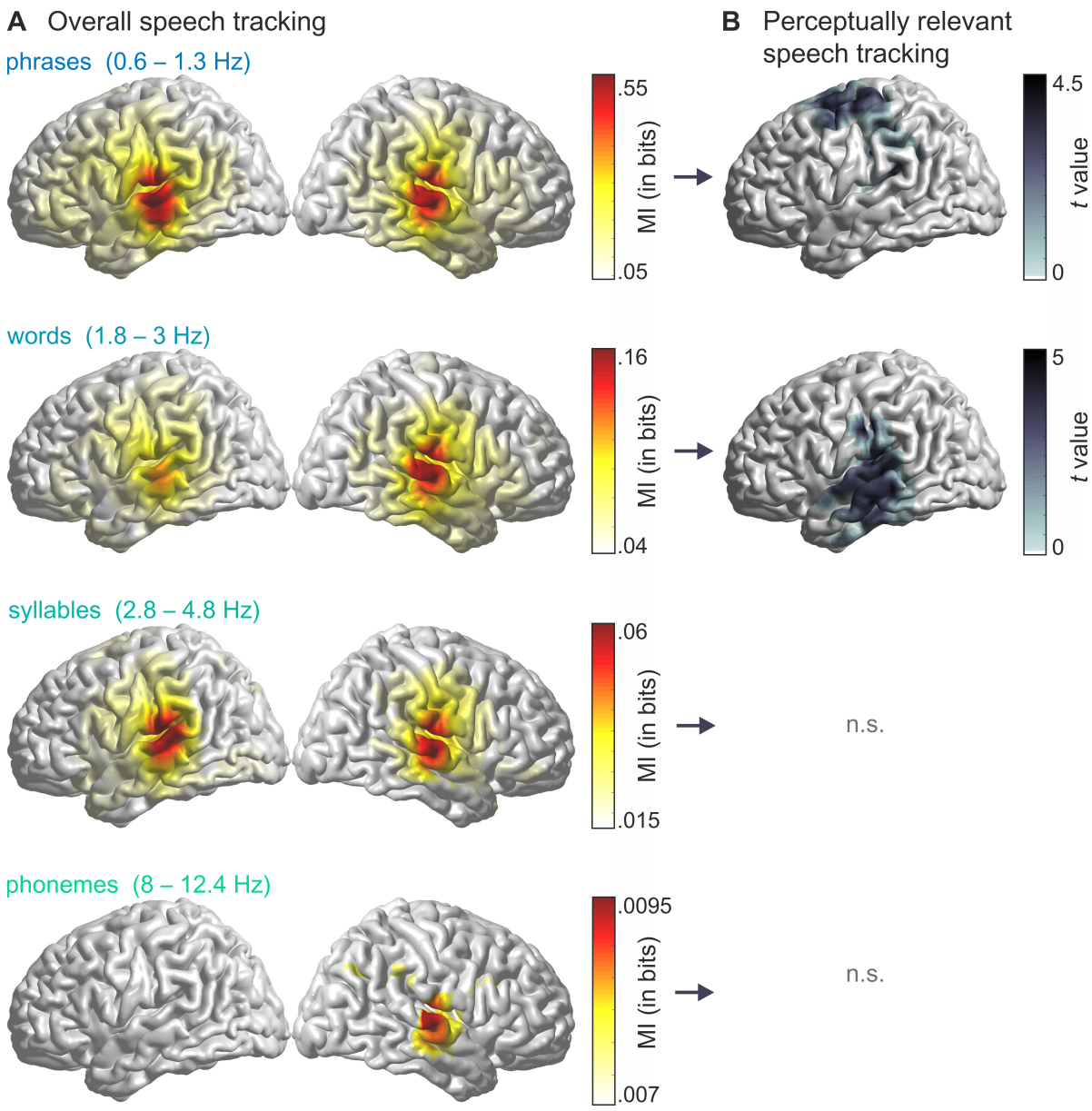

Fig 2. Overall and perceptually relevant speech tracking in the 4 stimulus-tailored frequency bands. (A) Significant areas for comparison between true MI values and surrogate data (*t* test, cluster-corrected). The used frequency bands map onto timescales for phrases (0.6–1.3 Hz), words (1.8–3 Hz), syllables (2.8–4.8 Hz), and phonemes (8–12.4 Hz). (B) Clusters where speech tracking MI was larger for correctly comprehended compared to incorrectly comprehended trials (*t* test, cluster-corrected). Data deposited in the Dryad repository: https://doi.org/10.5061/dryad.1qq7050 [31]. MI, mutual information; n.s., not significant.

https://doi.org/10.1371/journal.pbio.2004473.g002

Perceptually relevant speech tracking

To localise cortical regions where entrainment was functionally relevant for comprehension, we statistically compared MI between correct and incorrect trials within each band (hereafter called 'perceptually relevant'). This yielded significant left-hemispheric clusters in 2 frequency bands (Fig 2B). For the phrasal timescale (0.6–1.3 Hz), MI was larger for correctly versus incorrectly comprehended trials in a cluster comprising left pre- and postcentral regions, supramarginal gyrus, and Heschl gyrus (HG; $T_{sum}(19) = 568.00$, $p_{cluster} = .045$; 205 grid points).

The effect peaked in the left premotor (PM) cortex (left Brodmann area [BA] 6). For the word timescale (1.8–3 Hz), MI was larger in a cluster comprising left superior, middle, and inferior temporal gyrus as well as supramarginal gyrus and HG ($T_{sum}(19) = 739.59$, $p_{cluster} = .018$; 263 grid points). The effect peaked in the left middle temporal gyrus (MTG; left BA 21). There was a small overlap of the clusters for phrasal and word timescales, peaking in the left HG (left BA 41, see Fig 3A).

We performed several further analyses to determine whether these effects were specific to those timescales extracted from the stimulus corpus. First, we performed posthoc *t* tests at the peak grid points of each cluster to see whether phrasal and word effects were indeed significant only for the respective timescales (Fig 3B). As expected, MI differed between correct and incorrect trials at the phrasal timescale in left PM cortex and HG ($t(19) = 4.90$, $p_{FDR} < .001$ and $t(19) = 2.53$, $p_{FDR} = .031$, respectively). Likewise, MI differed at the word timescale in the left MTG and HG ($t(19) = 5.22$, $p_{FDR} < .001$ and $t(19) = 3.48$, $p_{FDR} = .005$, respectively). MI neither differed at the phrasal scale in MTG ($t(19) = -1.78$, $p_{FDR} = .11$) nor at the word scale in PM cortex ($t(19) = -0.57$, $p_{FDR} = .58$). We also compared correct and incorrect trials at the same peak grid points for syllable and phoneme timescales, although the whole-brain analysis did not indicate that effects were perceptually relevant for the task at these timescales. This was to make sure that effects had not been overlooked due to corrections for multiple comparisons. None of the comparisons were significant (all $p_{FDR} > .56$, see S1 Fig), indicating that none of the peak grid points in PM, HG, or MTG showed perceptual relevance at the faster scales of syllables or phonemes.

Second, we compared brain-wide MI values between correctly and incorrectly comprehended trials in 7 generic, 2 Hz–wide frequency bands (from 0–8 Hz, in 1-Hz steps) to confirm that the above intelligibility-related effects were indeed specific to frequency bands matched to the specific temporal structure of the speech material. This is an important contrast because most previous studies used generic bands with a predefined fixed frequency spacing. Perceptually relevant effects were found in only 2 bands (S2A Fig). For the 1–3 Hz band, which largely overlaps with the word scale, MI was larger for comprehended than uncomprehended trials in a cluster centred around auditory cortex ($T_{sum}(19) = 1{,}078.85$, $p_{cluster} = .030$), confirming the relevance of auditory regions for word-level encoding. For the 2–4 Hz band, which spans the scale of words and syllables, MI was only marginally enhanced for comprehended sentences in a cluster covering middle and inferior temporal cortex ($T_{sum}(19) = 751.93$, $p_{cluster} = .046$).

Third, using posthoc statistics, we also verified that the MI at the previously identified peak grid points (see Fig 3A for peaks) differed between correct and incorrect trials only at those timescales that matched the stimulus-specific bands (S2B Fig). The motor cortex was not found to be perceptually relevant in any of the probed generic bands, which suggests that exact frequency boundaries are necessary to detect phrase tracking in motor areas.

Phase-amplitude coupling in motor cortex

Rhythmic brain activity represents neuronal excitability changes [32,33]. In auditory areas, this mechanism has been suggested to reflect a segmentation of the incoming sensory stream [34,35]. But what is the role of slow excitability changes in the motor system? The motor system plays a role in the temporal prediction of rhythms and beats [36,37,38,39]. Previous studies have suggested that these predictions rely on the coupling of delta phase to rhythmic activity in the beta band [36,40,41]. We therefore hypothesised that speech entrainment at the phrasal scale—and its perceptual relevance—is directly linked to phase-amplitude coupling (PAC) with motor cortical beta activity. Indeed, we found that the coupling of the phase of phrasal-scale activity (0.6–1.3 Hz) to beta power (13–30 Hz) was significantly stronger for

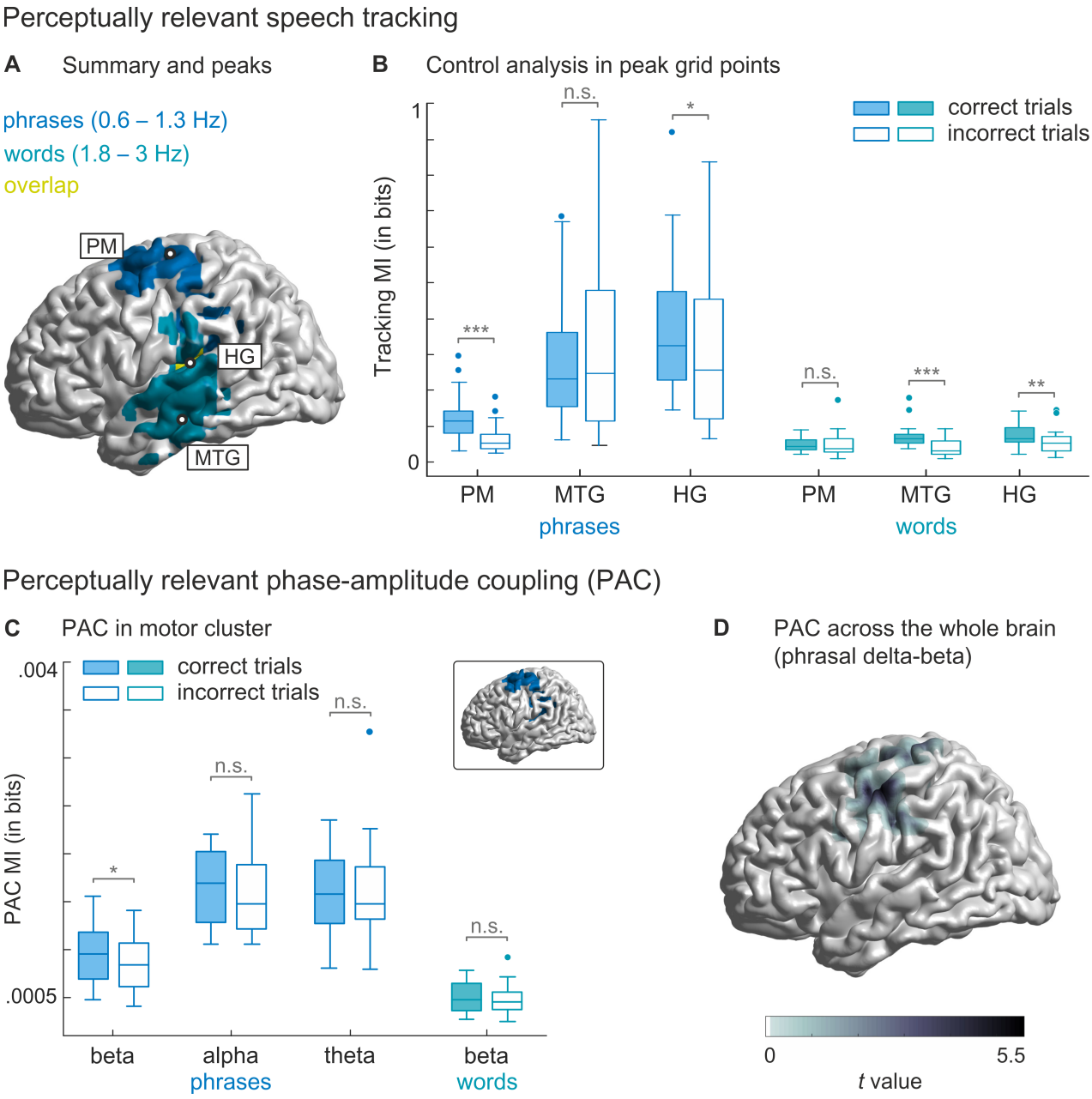

Fig 3. Summary of perceptually relevant effects and control analyses. (A) Brain regions with perceptually relevant speech tracking. For phrases (blue), the effect was strongest in the left PM area. For words (turquoise), it was strongest in the left MTG. Areas that were perceptually relevant for both phrases and words (yellow) include left HG and supramarginal gyrus. Peak grid points are denoted with circles. (B) Comparison of tracking MI values at peak grid points in PM cortex, MTG, and HG for phrase and word scales. Boxes denote interquartile range with median line; error bars show minimum and maximum, excluding outliers. (C) MI between the phase at the phrasal timescale (0.6–1.3 Hz) and power in beta, alpha, and theta bands (blue plots), as well as the phase at the word timescale (1.8–3 Hz) and beta power (turquoise plot) for correctly and incorrectly comprehended trials, averaged across all grid points in the motor cluster (cluster shown in inlet). Only PAC between the phase at the phrasal timescale and beta power was perceptually relevant. (D) Whole-brain analysis across all 12,337 grid points confirming that PAC between phrasal phase and beta power is indeed confined to motor regions. Coloured area denotes the cluster where PAC between phrasal phase and beta power was larger for correctly comprehended than uncomprehended trials. Significance in panel B and C is denoted with: *** = $p < .001$, ** = $p < .01$, * = $p < .05$. Data deposited in the Dryad repository: https://doi.org/10.5061/dryad.1qq7050 [31]. HG, Heschl gyrus; MI, mutual information; MTG, middle temporal gyrus; n.s., not significant; PAC, phase-amplitude coupling; PM, premotor.

https://doi.org/10.1371/journal.pbio.2004473.g003

correctly versus incorrectly comprehended trials in our motor cluster ($t(19) = 2.96$, $p_{FDR} = .032$; Fig 3C). In contrast, there was no such cross-frequency coupling for the phase of word-scale activity relative to beta power ($t(19) = 1.14$, $p_{FDR} = .356$) or of phrasal phase to either alpha ($t(19) = 1.38$, $p_{FDR} = .356$) or theta power ($t(19) = -0.38$, $p_{FDR} = .708$).

To confirm that the PAC between phrasal phase and beta power is confined to motor regions, we performed a further whole-brain analysis, comparing PAC between correct and incorrect trials. This analysis yielded 1 cluster in which PAC was larger for comprehended than uncomprehended trials ($T_{sum} = 203.94$, $p_{cluster} = .030$, one-sided; Fig 3D). The cluster included left pre- and postcentral regions. Therefore, perceptually relevant PAC was confined to the left motor system, overlapping with the speech tracking effect in left motor areas.

Phrasal tracking and PAC in additional dataset

The sentence structure in the present study was relatively rigid and predictable, which could have emphasised effects at the phrasal timescale. We therefore tested the presence of the phrasal tracking effect and the PAC between phrasal phase and beta power in the motor cortex in an additional dataset in which the sentence and phrase structure were more variable. Here, participants listened to a natural 7-min narration while their MEG was recorded [2,42]. The phrasal rate of the narration ranged between 0.1 Hz and 1.5 Hz. We specifically tested phrase tracking and PAC in the motor cluster, defined by results in the main dataset above, by contrasting the actual MI with surrogate data. Phrase tracking was larger in actual data ($t(22) = 6.22$, $p_{FDR} < .001$), as was PAC between phrasal phase and beta power ($t(22) = 4.52$, $p_{FDR} < .001$; Fig 4). These results suggest that the found mechanisms in the present study also exist in an unrelated dataset with highly variable sentence structure.

Discussion

By focusing on stimulus-specific timescales and measuring comprehension on individual trials, we show that distinct neural and linguistic timescales relate to speech encoding along the auditory–motor pathway. Our findings provide specific functional roles of speech tracking at two timescales within the delta band, one relevant in motor areas—accompanied by

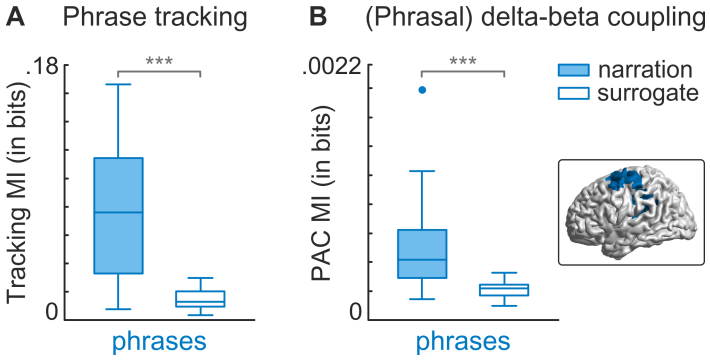

Fig 4. Phrase tracking and PAC in motor cortex in an additional dataset. MI values for phrase tracking and PAC were compared with surrogate data. Surrogate data were created by reversing the time series for speech and computing MI between forward brain time series and reversed speech time series. Surrogate data represent values that would be expected by chance. (A) Speech tracking at the phrasal time-scale (0.1–1.5 Hz) for real and surrogate data in the motor cortex (see inlet for analysed area). Phrase tracking was significantly larger in real data than in surrogate data. (B) MI between the phase at the phrasal timescale (0.1–1.5 Hz) and power in the beta band (13–30 Hz). PAC was significantly larger in real data than in surrogate data. Significance is denoted with: *** = $p < .001$. Data deposited in the Dryad repository: https://doi.org/10.5061/dryad.1qq7050 [31]. MI, mutual information; PAC, phase-amplitude coupling.

https://doi.org/10.1371/journal.pbio.2004473.g004

modulations in delta–beta coupled oscillations—and one relevant in temporal areas. Previous studies typically show that speech tracking peaks in early auditory areas, independent of the analysed frequency band [2,22,30]. Although a topographical distinction between hierarchical time-scales, as shown in the present data, has been hinted at before [1], this has not been straightforwardly demonstrated.

The motor system predicts phrasal timing using beta oscillations

The motor system plays a causal role in speech perception [43,44,45]. Previous studies have attributed functions for simulating speech production [46,47] or sensorimotor speech processing [48] to the motor system. Furthermore, the PM cortex and the motor system generally have been associated with generating temporal predictions [49,50,51,52,53] and the processing of rhythms and beats [37,38,39]. In the present study, we increase the knowledge about its role for natural speech processing by uncovering two specific neural mechanisms. The first mechanism is a perceptually relevant speech tracking specifically at the phrasal timescale, peaking in the left PM cortex. Notably, the timing of phrasal elements in the used stimulus corpus was relatively predictable because all sentences followed the same structure. The phrasal structure was also defined by prominent pauses between phrasal elements (evident in the clear peak in the frequency spectrum, S3A Fig). On the other hand, words (and therefore syllables and phonemes) were not semantically—or temporally—predictable due to the recombination of words across sentences. The motor system likely exploited the temporally predictive phrasal information for parsing and segmenting the sentences, thus facilitating comprehension by providing a temporal prediction of when the relevant target word was likely to occur. Our results therefore suggest that perceptually relevant speech entrainment emerges not only at the time-scale of the directly task-relevant feature (here words) but also at those time-scales that can be exploited to better detect or encode this feature. We confirmed this motor mechanism in a second dataset, which featured a less stereotyped phrasal structure. Yet it is possible that this mechanism is not specific to the phrasal structure per se. Instead, it could be that the motor system would exploit any temporal regularities [38], regardless of their linguistic or metalinguistic relevance. Future research is required to directly compare acoustic and linguistic regularities and their relevance for speech tracking.

It has been suggested that delta entrainment to speech in the left hemisphere reflects a motor-driven top-down modulation [42,54]. These top-down modulations have been associated with beta oscillations [50,55,56,57], which are prevalent in the motor system [58]. Beta power in the motor system has also been related to speech comprehension [57]. The finding that the temporal prediction of tone sequences is mediated by prestimulus delta–beta coupled oscillations further supports this hypothesis ([36], see also [40,41]). Here, we show—to our knowledge, for the first time—that such a cross-frequency mechanism also operates during the encoding and perception of continuous speech. This coupling is (i) specific to phrasal delta phase (0.6–1.3 Hz) and beta power (13–30 Hz) and (ii) only perceptually relevant in left motor areas. Furthermore, in an additional dataset, we show that this phrasal delta–beta coupling is also present during the processing of a natural, spontaneous narration. Based on the above-mentioned findings [36,40,41], we speculate that this cross-frequency motor coupling reflects top-down temporal prediction, which is relevant both for the perception of simple sounds [36] and speech.

Word segmentation in the temporal cortex

Speech tracking at the word scale was perceptually relevant across the entire midtemporal gyrus, peaking in MTG and including superior and inferior temporal gyrus as well as inferior

supramarginal gyrus. Previous MEG studies that have localised tracking processes typically show that this peaks in early auditory areas independent of the frequency band (for example, when contrasted with the null hypothesis of no speech encoding) [2,22,30]. Only by using a direct comprehension measure can we show that perceptually relevant word segmentation peaks in the left MTG. The MTG is associated with lexical semantic processes [59,60] and is one endpoint of the ventral auditory pathway, mapping sound to meaning [61]. It is plausible that stronger speech tracking, and therefore better word-scale segmentation in these regions, is directly linked to comprehension performance. The result that the effect at the word scale extends dorsally to supramarginal gyrus seems to contradict models of a ventral focus of word comprehension. However, it is consistent with the notion of a dorsal lexicon, thought to store articulatorily organised word form representations [62].

Specificity of linguistic timescales

An analysis of 2 Hz–wide generic bands showed that (i) the activity in the motor system was not predictive for comprehension in any generic band; (ii) the 1–3 Hz band, which largely overlaps with the word scale, yielded a similar pattern as the word-specific timescale; and (iii) the 2–4 Hz band also overlapped with the effect at the word timescale, albeit only minimally significantly (S2 Fig). These results suggest that perceptually relevant speech tracking in the motor system is specific to the phrasal timescale in the stimulus material. In temporal regions, perceptually relevant tracking was found in the delta band (above 1 Hz and below 4 Hz), independent of the specific boundaries of the used bands (although 0–2 Hz did not yield a significant effect). This suggests that speech tracking in temporal areas emerges at more widespread timescales, perhaps because word length is more variable than phrasal length in the present stimulus material. Analyses of the coefficient of variation (c_v) supported this interpretation: when compared with phrases ($c_v = 0.27$), words varied in length almost twice as much ($c_v = 0.48$).

We chose to base the timescales on linguistic categories of phrases, words, syllables, and phonemes. This is the most pragmatic approach because the language system ultimately has to parse the speech stream into these segments. However, one could argue that these linguistic categories overlap with other metalinguistic elements that also follow temporal modulations below 4 Hz such as prosodic features [16]. The most relevant prosodic features for speech segmentation are pauses, stress, and intonation [14,63]. The phrases in the stimulus material are defined by pauses, and therefore phrasal timescale and timing of pauses can be considered one and the same. The interaction between linguistic categories and lexical stress is more complex. If we consider every third syllable as stressed [64], one can derive a 'stress timescale' of 0.9 to 1.6 Hz, which partly overlaps with the phrasal timescale (0.6–1.3 Hz). The role of stress is manifold in speech (disambiguation of phonemically identical words, highlighting the meaning of words, metrical stress), and we cannot rule out that stressed syllables are reflected in neural activity. However, the segmentation into phrases does not typically have stressed syllables as boundaries because this would often yield nonsense phrases. Therefore, although stress is important and useful in speech comprehension, focusing on the phrasal timescale (as opposed to the 'stress timescale') is a direct way to address phrase segmentation. Fluctuations in pitch, or intonation, also occur in the delta band (see S3B Fig for spectral analysis of pitch, or its acoustic correlate the fundamental frequency). Pitch fluctuations can signal phrasal boundaries [65], and an overlap with the phrasal timescale is therefore not surprising. Because the auditory system is able to track pitch fluctuations [9] and fundamental frequency and intensity are related, we cannot completely disentangle pitch tracking from envelope tracking. But language comprehension requires the grouping of words into phrases [15], and focusing the

analysis on the phrasal timescale is the most direct way of analysing phrasal processing. Future research needs to address the question of how much phrasal segmentation relies on the acoustic envelope, pitch fluctuations, or both. Taken together, linguistic and metalinguistic events in natural speech have a tendency to co-occur [66], and their interaction is complex. However, for natural speech processing, the division into linguistic categories, as done in the present study, seems the most pragmatic and ecologically valid solution to gain specificity about speech comprehension effects.

Finally, in the present study, the average speech rate was approximately 130 words per minute. In two other studies that reported speech rate, it was considerably higher, at approximately 160 words per minute [22] and approximately 210 words per minute [25]. The rate of syllables is typically associated with frequencies between 4 and 8 Hz [3,67]. In the present study, it was 2.8 to 4.8 Hz, and in another study, it was even lower at 2 to 4 Hz [4]. These differences demonstrate that, even in experimental contexts, speech rates can deviate from the assumed standard. Furthermore, a recent study has shown that the auditory system is not limited by traditionally imposed frequency bands [57]. It therefore is highly beneficial to calculate stimulus-specific speech regularities for speech tracking analyses instead of applying generic frequency bands (cf. [68] for visual modality).

Not all speech tracking processes are perceptually relevant

Our results regarding overall speech tracking (compared with chance) replicate previous reports of widespread speech-to-brain entrainment at multiple timescales [2,22]. However, in our data, only speech tracking in specific bands within the delta frequency range differed between trials with correct and incorrect comprehension and was therefore likely relevant for the perceptual outcome. One interpretation of why only the slow timescales were directly perceptually relevant is that the comprehension task focused on words, thus stressing the word timescale. Furthermore, the stereotyped phrasal structure of the sentence provided a temporal structure on which the emergence of the target word could be expected. Therefore, participants may have relied on the encoding of the phrasal structure to exploit its regularity, thereby stressing the phrasal timescale. However, it has been suggested that speech tracking in the delta and theta bands index different functional roles for speech perception [69], such that theta tracking reflects the analysis of acoustic features and delta tracking reflects linguistic representations. In line with the this is the notion that only speech-to-brain entrainment in the delta band reflects active speech-specific processing, as opposed to a passive, low-level synchronisation to acoustic properties at other timescales [30]. Therefore, these and our findings tentatively support the conclusion that only speech tracking in the delta band might indicate a speech-specific, perceptually relevant process during continuous speech processing.

Recent findings also highlight the distinction between widely distributed versus focal (but perceptually relevant) auditory encoding [19,20] that could contribute to this pattern of results. In these accounts, perceptual choices are determined by the efficient readout of a restricted neural area, whereas widespread neural activity represents collateral processes of sensory processing. Therefore, such distributed processes could also explain the widespread speech tracking at all timescales we found when compared to chance level. In the present data, speech tracking at the syllabic and phonetic scales did not differ between trials with correct and incorrect comprehension. But for the participants to comprehend target words correctly, at least some syllables must have been encoded phonetically. It could be that the use of a noisy background prevented the robust encoding of individual syllables or phonemes, thus reducing the tracking at these timescales or reducing the statistical power in detecting between-trial differences. Furthermore, as mentioned above, the use of a word-related task could have highlighted

effects at the word level and obscured effects at faster timescales. Additional work is required to understand whether speech tracking at the syllabic and phonetic timescales is indeed a robust marker of the actual neural encoding of these features or whether only speech tracking at timescales below the syllabic rate directly indexes functionally and perceptually relevant processes. Furthermore, the left-lateralised perceptually relevant speech tracking at slow timescales stands in contrast with bilateral overall speech tracking at these scales (Fig 2). This supports the notion that 'early' acoustic processes are bilateral, whereas 'higher-order' speech comprehension is left-lateralised [70].

Materials and methods

Ethics statement

All participants provided written informed consent prior to testing and received monetary compensation of £10 per h. The experiment was approved by a local ethics committee (College of Science and Engineering, University of Glasgow, application number 300140078) and conducted in compliance with the Declaration of Helsinki.

Participants and data acquisition

Following previous sample sizes of MEG studies that used MI to study speech tracking [2,22], as well as previous recommendations [71,72,73], 20 healthy, native British participants took part in the study (9 female, age 23.6 ± 5.8 years [mean ± SD], age range: 18 to 39 years). All participants were right-handed [Edinburgh Handedness Inventory; 74], had normal hearing [Quick Hearing Check; 75], and normal or corrected-to-normal vision. Furthermore, participants had no self-reported current or previous neurological or language disorders.

MEG was recorded with a 248-magnetometer, whole-head MEG system (MAGNES 3600 WH, 4-D Neuroimaging, San Diego, CA) at a sampling rate of 1 KHz. Head positions were measured at the beginning and end of each run, using 5 coils placed on the participants' heads. Coil positions were codigitised with head shape (FASTRAK, Polhemus Inc., Colchester, VT). Participants sat upright and fixated at a fixation point projected centrally on screen with a DLP projector. Sounds were transmitted binaurally through plastic earpieces, and 3.7 m–long plastic tubes connected to a sound pressure transducer. Stimulus presentation was controlled with Psychophysics toolbox [76] for MATLAB (The MathWorks, Inc., Natick, MA).

Stimuli

The stimulus material consisted of 2 structurally equivalent sets of 90 sentences (180 unique sentences in total) that were spoken by a trained, male, native British actor. The speaker was instructed to speak clearly and naturally. Sentences were constructed to be meaningful but unpredictable. Each sentence consisted of the same basic elements and therefore had the same structure. For example, the sentence 'Did you notice, on Sunday night, Graham offered ten fantastic books' consists of a 'filler' phrase ('Did you notice'), a time phrase ('on Sunday night'), a name, a verb, a numeral, an adjective, and a noun. There were 18 possible names, verbs, numerals, adjectives, and nouns that were each repeated 10 times. Sentence elements were randomly combined within a set of 90 sentences. To measure comprehension, a target word was included that was either the adjective in 1 set of sentences ('fantastic' in the above example or 'beautiful' in Fig 1) or the number in the other set (for example, 'twenty-one'). The duration of sentences ranged from 4.2 s to 6.5 s (5.4 ± 0.4 s [mean ± SD]). Sentences were presented at a sampling rate of 22,050 Hz.

During the experiment, speech stimuli were embedded in noise. The noise consisted of ecologically valid environmental sounds (traffic, car horns, people talking), combined into a

mixture of 50 different background noises. The individual noise level for each participant was determined with a staircase procedure that was designed to yield a performance of around 70% correct. For the staircase procedure, only the 18 possible target words were used instead of whole sentences. Participants were presented with single target words embedded in noise and subsequently saw 2 alternatives on screen. They indicated by button press which word they had heard. Depending on whether their choice was correct or incorrect, the noise level was increased or decreased (one-up-three-down procedure) until a reliable level was reached. The average signal-to-noise ratio across participants was approximately −6 dB.

Experimental design

The 180 sentences were presented in 4 blocks with 45 sentences each. In each block, participants either indicated the comprehended adjective or the comprehended number, resulting in 2 'adjective blocks' and 2 'number blocks'. The order of sentences and blocks was randomised for each participant. The first trial of each block was a 'dummy' trial that was discarded for subsequent analysis; this trial was repeated at the end of the block. After each sentence, participants were prompted with 4 target words (either adjectives or numbers) on the screen. They then had to indicate which one they heard by pressing 1 of 4 buttons on a button box. After 2 s, the next trial started automatically. Each block lasted approximately 10 min, and participants could rest in between blocks. The session, including instructions, questionnaires, preparation, staircase procedure, and 4 blocks, took approximately 3 to 3.5 hours.

Speech preprocessing

For each sentence, we computed the wideband speech envelope at a sampling rate of 150 Hz following procedures of previous studies [2,6,12,77]. Acoustic waveforms were first filtered into 8 frequency bands (between 100 and 8,000 Hz; third-order Butterworth filter; forward and reverse) that were equidistant on the cochlear frequency map [77]. From these 8 individual bands, the wideband speech envelope was extracted by averaging the magnitude of the Hilbert transformed signals from each band.

To define the timescales on which to probe speech encoding, we evaluated the rates of phrases, words, syllables, and phonemes in the stimulus material. For this, the duration between onsets of linguistic categories (i.e., phrases, words, and phonemes) was calculated. The exact onset timing was extracted from the speech signals using Penn Phonetics Lab Forced Aligner (P2FA; [78]). Phrases were defined as the first 2 clauses in each sentence (for example, 'I have heard' and 'on Tuesday night'). These phrases had distinct pauses (see Fig 1 for an example sentence) that determined the rhythm of the sentence (also visible in the frequency spectrum S3A Fig). The syllable rate is generally difficult to assess [18,79]. Here, we chose to count the actually produced syllables for each sentence. Finally, timescales were converted to frequencies, and the specific frequency bands for each category were then defined as the minimum and maximum frequencies across all 180 sentences. This led to the following bands: 0.6–1.3 Hz (phrases), 1.8–3.0 Hz (words), 2.8–4.8 Hz (syllables), and 8–12.4 Hz (phonemes). Mean and standard deviations for linguistic categories were as follows: 1.0 ± 0.1 Hz for phrases, 2.4 ± 0.3 Hz for words, 3.8 ± 0.4 Hz for syllables, and 10.4 ± 0.8 Hz for phonemes. Furthermore, the fundamental frequency for each sentence was extracted using Praat [80]. This was used to determine the frequency spectrum of the pitch fluctuations (see S3B Fig).

MEG preprocessing

Preprocessing of MEG data was carried out in MATLAB (The MathWorks, Inc., Natick, MA) using the Fieldtrip toolbox [81]. The 4 experimental blocks were preprocessed separately.

Single trials were extracted from continuous data starting 2 s before sound onset and until 10 s after sound onset. MEG data were denoised using the reference signal. Known faulty channels ($N = 7$) were removed before further preprocessing. Trials with SQUID jumps (3.5% of trials) were detected and removed using Fieldtrip procedures with a cutoff z-value of 30. Before further artifact rejection, data were filtered between 0.2 and 150 Hz (fourth-order Butterworth filters, forward and reverse) and down-sampled to 300 Hz. Data were visually inspected to find noisy channels (4.37 ± 3.38 on average across blocks and participants) and trials (0.66 ± 1.03 on average across blocks and participants). Finally, heart and eye-movement artifacts were removed by performing an independent component analysis with 30 principal components. Data were further down-sampled to 150 Hz to match the sampling rate of the speech signal.

Source localisation

Source localisation was performed using Fieldtrip, SPM8, and the Freesurfer toolbox. We acquired T1-weighted structural magnetic resonance images (MRIs) for each participant. These were coregistered to the MEG coordinate system using a semiautomatic procedure [2,6]. MRIs were then segmented and linearly normalised to a template brain (MNI space). A volume conduction model was constructed using a single-shell model [82]. We projected sensor-level waveforms into source space using frequency-specific linear constraint minimum variance (LCMV) beamformers [83] with a regularisation parameter of 7% and optimal dipole orientation (singular value decomposition method). Grid points had a spacing of 6 mm, resulting in 12,337 points covering the whole brain.

Analysis of speech tracking in brain activity

We quantified the statistical dependency between the speech envelope and the source-localised MEG data using MI [2,6,34,84]. The speech envelopes, as well as MEG data, were filtered in the 4 frequency bands reflecting the rates of each linguistic category using third-order (for delta and theta bands) forward and reverse Butterworth filters. Within these bands, we computed the Hilbert transform and used real and imaginary parts for further analysis. Both parts were normalised separately and combined as a two-dimensional variable for the MI calculation [84]. To take into account the stimulus–brain lag, we computed MI at 5 different lags (from 60 to 140 ms in 20-ms steps) and summed the MI values across lags. This procedure prevents spurious results that can occur when using a single lag. First, we calculated the overall MI for each source grid point. For a robust computation of MI values, we concatenated MEG and speech data from all trials. The resulting MI values were compared with surrogate data to determine their statistical significance. Surrogate data were created by randomly shuffling trials 50 times and averaging surrogate MI values across iterations. This repetition was necessary because all sentences followed the same structure and their envelope was often comparable, especially when filtered at low frequencies. We used a dependent t test for statistical comparison for each grid point and corrected for multiple comparisons with cluster-based permutation. Specifically, we used Monte-Carlo randomisation with 1,000 permutations and a critical t value of 2.1, which represents the critical value of the Student t distribution for 20 participants and a two-tailed probability of $p = .05$. The significance level for accepting clusters was 5%. We report summed t values (T_{sum}) as indicator of effect size.

For the analysis of perceptual relevance, we compared MI between trials in which participants responded correctly and incorrectly. Because the number of trials differed between these samples (on average, approximately 70% correct and 30% incorrect), we performed the calculations based on 80% of the minimally available number of trials. This way, the number of compared correct and incorrect trials was equal. However, because this included only a small

part of all available trials, we repeated the analysis 20 times with a random selection of trials to yield representative values. The resulting MI values were averaged. Again, trials were concatenated to yield robust MI values. MI values between correct and incorrect trials were compared using the same method and parameters as for the comparison between overall MI and surrogate MI.

To examine the specificity of the effects, we compared MI between correct and incorrect trials for all peak grid points in both frequency bands (i.e., phrasal and word timescales). Peak grid points were those with the largest *t* values in each cluster and the largest summed *t* values for the overlap of grid points. This led to 12 comparisons (3 peak grid points × 4 frequency bands). MI values were compared using dependent sample *t* tests, corrected for multiple comparisons using the FDR method [85].

PAC

To examine the hypothesis that beta power is coupled with delta phase in the motor cluster and that this is perceptually relevant, we quantified PAC using the MI between beta power and delta phase. Phase and power were derived from Hilbert-transformed time series and filtered in the phrasal (0.6–1.3 Hz) and beta band (13–30 Hz). Phase was expressed as a unit magnitude complex number. To get an equal number of trials for correct and incorrect trials, we again took 80% of trials of the smaller sample, concatenated trials, and repeated the calculation 50 times. This was done for all grid points within the motor cluster ($N = 205$) and then averaged across grid points and iterations. PAC was compared between correct and incorrect trials across participants using a dependent sample *t* test.

We performed 3 control analyses within the motor cluster to address the frequency specificity of the effect. First, we analysed PAC between phrasal phase (0.6–1.3 Hz) and alpha power (8–12 Hz) as well as theta power (4–8 Hz). Second, we analysed PAC between the word phase (1.8–3 Hz) and beta power. All *p*-values were corrected for multiple comparisons using the FDR method [85].

To address the spatial specificity of the delta–beta PAC, we also performed a whole-brain analysis. Based on the results in the motor cluster, we hypothesised that PAC should be larger in correct than incorrect trials. PAC between phrasal delta phase (0.6–1.3 Hz) and beta power (13–30 Hz) was compared between correct and incorrect trials, again equalling sample sizes by using 80% of the minimally available number of trials and repeating the analysis 20 times. PAC MI was averaged across all iterations and then compared between correct and incorrect trials across participants using a dependent sample *t* test for each grid point. To correct for multiple comparisons, we used the same parameters for cluster correction as in all previous analyses except that the significance level to choose significant clusters was one-sided, due to the clear hypothesis.

Analysis of a previously published dataset

We analysed an additional and previously published dataset to confirm the present effects in the motor cortex. For this, we used data from 23 participants [2,42] who passively listened to a 7-min continuous natural narration. Preprocessing and analysis were identical to the procedures of the main data. We compared (i) phrasal tracking and (ii) PAC between phrasal phase and beta power with surrogate data in the motor cortex. The phrasal rate of the speech stimulus was 0.5 ± 0.26 Hz (mean ± SD) and ranged between 0.1 Hz and 1.5 Hz. Surrogate data were created by reversing the time series for speech and computing MI between forward brain time series and reversed speech time series. This represents values that would be expected by chance [2]. Values were computed for all grid points in the motor cluster and then spatially averaged.

Actual MI values and surrogate data were compared using a dependent *t* test, and *p*-values for both tests were FDR corrected.

Data were deposited in the Dryad repository (https://doi.org/10.5061/dryad.1qq7050) [31].

Supporting information

S1 Fig. Comparison of MI values at peak grid points in PM gyrus, MTG, and HG for syllable and phoneme scales. Boxes denote interquartile range with median line; error bars show minimum and maximum, excluding outliers. None of the comparisons reached significance (all $p_{FDR} > .56$, $p_{uncorrected} > .20$). Data deposited in the Dryad repository: https://doi.org/10.5061/dryad.1qq7050 [31]. FDR, false discovery rate; HG, Heshl gyrus; MI, mutual information; MTG, middle temporal gyrus; PM, premotor.
(TIF)

S2 Fig. Analyses in generic 2 Hz–wide bands. Seven overlapping frequency bands were analysed (from 0–8 Hz, in 2 Hz–wide bands, in 1-Hz steps). The first 3 of these bands are displayed here. (A) Perceptually relevant tracking (larger MI for correctly comprehended than incorrectly comprehended trials) was found at the 1–3 Hz scale ($T_{sum}(19) = 1{,}078.85$, $p_{cluster} = .030$) and at the 2–4 Hz scale ($T_{sum}(19) = 751.93$, $p_{cluster} = .046$). Effects in all other bands were $p > .11$. (B) Analysis of the peak grid points that showed the strongest effect in stimulus-specific bands. Larger MI for correctly than incorrectly comprehended trials is found in HG in the generic 1–3 Hz band ($t(19) = 4.54$, $p_{FDR} = .002$) and in MTG in the 2–4 Hz band ($t(19) = 3.38$, $p_{FDR} = .014$). All other comparisons are $p > .08$. The peak grid point in the PM cortex does not show a comprehension modulation in any of the generic bands. Data deposited in the Dryad repository: https://doi.org/10.5061/dryad.1qq7050 [31]. FDR, false discovery rate; HG, Heschl gyrus; MTG, middle temporal gyrus; PM, premotor; T_{sum}, summed *t* values.
(TIF)

S3 Fig. Power spectral density estimates of speech envelope and pitch. Welch's periodograms are shown for speech envelopes (A) and fundamental frequency (F0-contours/pitch) (B) of all 180 stimulus sentences (thin gray lines) and their average (thick black line), for frequencies between 0.1 and 12 Hz (in 0.1-Hz steps). For envelope spectra, visible peaks that correspond to rates used in the analysis are marked with arrows (i.e., for phrases, words, and—less pronounced—syllables). Data deposited in the Dryad repository: https://doi.org/10.5061/dryad.1qq7050 [31].
(TIF)

Acknowledgments

We thank Christian Keitel and Edwin Robertson for valuable comments on earlier versions of this manuscript, and Christoph Daube for providing the stimulus onsets in the seven-minute narration.

Author Contributions

Conceptualization: Anne Keitel, Christoph Kayser.

Formal analysis: Anne Keitel, Christoph Kayser.

Funding acquisition: Joachim Gross, Christoph Kayser.

Investigation: Anne Keitel.

Methodology: Anne Keitel, Christoph Kayser.

Project administration: Christoph Kayser.

Supervision: Joachim Gross, Christoph Kayser.

Validation: Anne Keitel.

Visualization: Anne Keitel.

Writing – original draft: Anne Keitel.

Writing – review & editing: Anne Keitel, Joachim Gross, Christoph Kayser.

References

1. Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016) Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 19: 158–164. https://doi.org/10.1038/nn.4186 PMID: 26642090
2. Gross J, Hoogenboom N, Thut G, Schyns P, Panzeri S, et al. (2013) Speech rhythms and multiplexed oscillatory sensory coding in the human brain. PLoS Biol 11(12): e1001752. https://doi.org/10.1371/journal.pbio.1001752 PMID: 24391472
3. Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: emerging computational principles and operations. Nat Neurosci 15: 511–517. https://doi.org/10.1038/nn.3063 PMID: 22426255
4. Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85 Pt 2: 761–768.
5. Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8: 311. https://doi.org/10.3389/fnhum.2014.00311 PMID: 24904354
6. Keitel A, Ince RA, Gross J, Kayser C (2017) Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks. Neuroimage 147: 32–42. https://doi.org/10.1016/j.neuroimage.2016.11.062 PMID: 27903440
7. Ding N, Chatterjee M, Simon JZ (2014) Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. Neuroimage 88C: 41–46.
8. Luo H, Liu Z, Poeppel D (2010) Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. PLoS Biol 8(8): e1000445. https://doi.org/10.1371/journal.pbio.1000445 PMID: 20711473
9. Obleser J, Herrmann B, Henry MJ (2012) Neural Oscillations in Speech: Don't be Enslaved by the Envelope. Front Hum Neurosci 6: 250. https://doi.org/10.3389/fnhum.2012.00250 PMID: 22969717
10. Greenberg S, Carvey H, Hitchcock L, Chang SY (2003) Temporal properties of spontaneous speech—a syllable-centric perspective. Journal of Phonetics 31: 465–485.
11. Pellegrino F, Coupe C, Marsico E (2011) A cross-language perspective on speech information rate. Language 87: 539–558.
12. Chandrasekaran C, Trubanova A, Stillittano S, Caplier A, Ghazanfar AA (2009) The natural statistics of audiovisual speech. PLoS Comput Biol 5(7): e1000436. https://doi.org/10.1371/journal.pcbi.1000436 PMID: 19609344
13. Ding N, Patel AD, Chen L, Butler H, Luo C, et al. (2017) Temporal modulations in speech and music. Neurosci Biobehav Rev 81: 181–187. https://doi.org/10.1016/j.neubiorev.2017.02.011 PMID: 28212857
14. Ghitza O (2017) Acoustic-driven delta rhythms as prosodic markers. Lang Cogn Neurosci 32: 545–561.
15. Meyer L, Henry MJ, Gaston P, Schmuck N, Friederici AD (2016) Linguistic bias modulates interpretation of speech via neural delta-band oscillations. Cereb Cortex.
16. Goswami U, Leong V (2013) Speech rhythm and temporal structure: Converging perspectives? Laboratory Phonology 4: 67–92.
17. Ghitza O (2013) The theta-syllable: a unit of speech information defined by cortical function. Front Psychol 4: 138. https://doi.org/10.3389/fpsyg.2013.00138 PMID: 23519170
18. Cummins F (2012) Oscillators and syllables: a cautionary note. Front Psychol 3: 364. https://doi.org/10.3389/fpsyg.2012.00364 PMID: 23060833
19. Bouton S, Chambon V, Tyrand R, Guggisberg AG, Seeck M, et al. (2018) Focal versus distributed temporal cortex activity for speech sound category assignment. Proc Natl Acad Sci U S A.

20. Tsunada J, Liu AS, Gold JI, Cohen YE (2016) Causal contribution of primate auditory cortex to auditory perceptual decision-making. Nat Neurosci 19: 135–142. https://doi.org/10.1038/nn.4195 PMID: 26656644

21. Ahissar E, Nagarajan S, Ahissar M, Protopapas A, Mahncke H, et al. (2001) Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. Proc Natl Acad Sci U S A 98: 13367–13372. https://doi.org/10.1073/pnas.201400998 PMID: 11698688

22. Giordano BL, Ince RAA, Gross J, Schyns PG, Panzeri S, et al. (2017) Contributions of local speech encoding and functional connectivity to audio-visual speech perception. Elife 6.

23. Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33: 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013 PMID: 23536086

24. Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23: 1378–1387. https://doi.org/10.1093/cercor/bhs118 PMID: 22610394

25. Di Liberto GM, O'Sullivan JA, Lalor EC (2015) Low-Frequency Cortical Entrainment to Speech Reflects Phoneme-Level Processing. Curr Biol 25: 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030 PMID: 26412129

26. Peña M, Melloni L (2012) Brain oscillations during spoken sentence processing. Journal of Cognitive Neuroscience 24: 1149–1164. https://doi.org/10.1162/jocn_a_00144 PMID: 21981666

27. Zoefel B, VanRullen R (2015) Selective perceptual phase entrainment to speech rhythm in the absence of spectral energy fluctuations. Journal of Neuroscience 35: 1954–1964. https://doi.org/10.1523/JNEUROSCI.3484-14.2015 PMID: 25653354

28. Ding N, Melloni L, Yang A, Wang Y, Zhang W, et al. (2017) Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (EEG). Frontiers in Human Neuroscience 11.

29. Ten Oever S, Schroeder CE, Poeppel D, van Atteveldt N, Mehta AD, et al. (2017) Low-frequency cortical oscillations entrain to subthreshold rhythmic auditory stimuli. J Neurosci 37: 4903–4912. https://doi.org/10.1523/JNEUROSCI.3658-16.2017 PMID: 28411273

30. Molinaro N, Lizarazu M (2018) Delta (but not theta)-band cortical entrainment involves speech-specific processing. European Journal of Neuroscience. [Cited 18 January 2018]. Available from: https://doi.org/10.1111/ejn.13811 PMID: 29283465

31. Keitel A, Gross J, Kayser C (2018) Data from: Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. Dryad Digital Repository. [Cited 2 March 2018]. Available from: https://doi.org/10.5061/dryad.1qq7050.

32. Wang XJ (2010) Neurophysiological and computational principles of cortical rhythms in cognition. Physiol Rev 90: 1195–1268. https://doi.org/10.1152/physrev.00035.2008 PMID: 20664082

33. Kayser C, Wilson C, Safaai H, Sakata S, Panzeri S (2015) Rhythmic auditory cortex activity at multiple timescales shapes stimulus-response gain and background firing. J Neurosci 35: 7750–7762. https://doi.org/10.1523/JNEUROSCI.0268-15.2015 PMID: 25995464

34. Kayser SJ, Ince RA, Gross J, Kayser C (2015) Irregular Speech Rate Dissociates Auditory Cortical Entrainment, Evoked Responses, and Frontal Alpha. J Neurosci 35: 14691–14701. https://doi.org/10.1523/JNEUROSCI.2243-15.2015 PMID: 26538641

35. Lakatos P, Shah AS, Knuth KH, Ulbert I, Karmos G, et al. (2005) An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. J Neurophysiol 94: 1904–1911. https://doi.org/10.1152/jn.00263.2005 PMID: 15901760

36. Arnal LH, Doelling KB, Poeppel D (2015) Delta-beta coupled oscillations underlie temporal prediction accuracy. Cereb Cortex 25: 3077–3085. https://doi.org/10.1093/cercor/bhu103 PMID: 24846147

37. Bengtsson SL, Ullen F, Ehrsson HH, Hashimoto T, Kito T, et al. (2009) Listening to rhythms activates motor and premotor cortices. Cortex 45: 62–71. https://doi.org/10.1016/j.cortex.2008.07.002 PMID: 19041965

38. Grahn JA, Brett M (2007) Rhythm and beat perception in motor areas of the brain. J Cogn Neurosci 19: 893–906. https://doi.org/10.1162/jocn.2007.19.5.893 PMID: 17488212

39. Morillon B, Schroeder CE, Wyart V (2014) Motor contributions to the temporal precision of auditory attention. Nat Commun 5: 5255. https://doi.org/10.1038/ncomms6255 PMID: 25314898

40. Cravo AM, Rohenkohl G, Wyart V, Nobre AC (2011) Endogenous modulation of low frequency oscillations by temporal expectations. J Neurophysiol 106: 2964–2972. https://doi.org/10.1152/jn.00157.2011 PMID: 21900508

41. Saleh M, Reimer J, Penn R, Ojakangas CL, Hatsopoulos NG (2010) Fast and slow oscillations in human primary motor cortex predict oncoming behaviorally relevant cues. Neuron 65: 461–471. https://doi.org/10.1016/j.neuron.2010.02.001 PMID: 20188651

42. Park H, Ince RA, 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. https://doi.org/10.1016/j.cub.2015.04.049 PMID: 26028433

43. Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7: 701–702. https://doi.org/10.1038/nn1263 PMID: 15184903

44. Watkins KE, Strafella AP, Paus T (2003) Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 41: 989–994. PMID: 12667534

45. Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci 15: 399–402. PMID: 11849307

46. Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M (2007) The essential role of premotor cortex in speech perception. Curr Biol 17: 1692–1696. https://doi.org/10.1016/j.cub.2007.08.064 PMID: 17900904

47. Iacoboni M (2008) The role of premotor cortex in speech perception: Evidence from fmri and rtms. Journal of Physiology-Paris 102: 31–34.

48. Scott SK, McGettigan C, Eisner F (2009) A little more conversation, a little less action—candidate roles for the motor cortex in speech perception. Nature Reviews Neuroscience 10: 295–302. https://doi.org/10.1038/nrn2603 PMID: 19277052

49. Schubotz RI (2007) Prediction of external events with our motor system: towards a new framework. Trends Cogn Sci 11: 211–218. https://doi.org/10.1016/j.tics.2007.02.006 PMID: 17383218

50. Arnal LH, Giraud AL (2012) Cortical oscillations and sensory predictions. Trends Cogn Sci 16: 390–398. https://doi.org/10.1016/j.tics.2012.05.003 PMID: 22682813

51. Schroeder CE, Wilson DA, Radman T, Scharfman H, Lakatos P (2010) Dynamics of Active Sensing and perceptual selection. Curr Opin Neurobiol 20: 172–176. https://doi.org/10.1016/j.conb.2010.02.010 PMID: 20307966

52. Grahn JA, Rowe JB (2013) Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. Cereb Cortex 23: 913–921. https://doi.org/10.1093/cercor/bhs083 PMID: 22499797

53. Breska A, Deouell LY (2017) Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. PLoS Biol 15(2): e2001665. https://doi.org/10.1371/journal.pbio.2001665 PMID: 28187128

54. Morillon B, Hackett TA, Kajikawa Y, Schroeder CE (2015) Predictive motor control of sensory dynamics in auditory active sensing. Curr Opin Neurobiol 31: 230–238. https://doi.org/10.1016/j.conb.2014.12.005 PMID: 25594376

55. Engel AK, Fries P (2010) Beta-band oscillations—signalling the status quo? Curr Opin Neurobiol 20: 156–165. https://doi.org/10.1016/j.conb.2010.02.015 PMID: 20359884

56. Morillon B, Baillet S (2017) Motor origin of temporal predictions in auditory attention. Proceedings of the National Academy of Sciences: 201705373.

57. Pefkou M, Arnal LH, Fontolan L, Giraud AL (2017) theta-band and beta-band neural activity reflects independent syllable tracking and comprehension of time-compressed speech. J Neurosci 37: 7930–7938. https://doi.org/10.1523/JNEUROSCI.2882-16.2017 PMID: 28729443

58. Keitel A, Gross J (2016) Individual human brain areas can be identified from their characteristic spectral activation fingerprints. PLoS Biol 14(6): e1002498. https://doi.org/10.1371/journal.pbio.1002498 PMID: 27355236

59. Friederici AD (2012) The cortical language circuit: from auditory perception to sentence comprehension. Trends Cogn Sci 16: 262–268. https://doi.org/10.1016/j.tics.2012.04.001 PMID: 22516238

60. Leonard MK, Chang EF (2014) Dynamic speech representations in the human temporal lobe. Trends in Cognitive Sciences 18: 472–479. https://doi.org/10.1016/j.tics.2014.05.001 PMID: 24906217

61. Saur D, Kreher BW, Schnell S, Kummerer D, Kellmeyer P, et al. (2008) Ventral and dorsal pathways for language. Proc Natl Acad Sci U S A 105: 18035–18040. https://doi.org/10.1073/pnas.0805234105 PMID: 19004769

62. Gow DW (2012) The cortical organization of lexical knowledge: a dual lexicon model of spoken language processing. Brain Lang 121: 273–288. https://doi.org/10.1016/j.bandl.2012.03.005 PMID: 22498237

63. Vaissière J (1983) Language-independent prosodic features. Prosody: Models and measurements: Springer. pp. 53–66.

64. Kochanski G, Grabe E, Coleman J, Rosner B (2005) Loudness predicts prominence: fundamental frequency lends little. J Acoust Soc Am 118: 1038–1054. PMID: 16158659

65. Vaissière J (2005) Perception of intonation. The handbook of speech perception: 236–263.

66. Lehiste I, Lass NJ (1976) Suprasegmental features of speech. Contemporary issues in experimental phonetics 225: 239.

67. Ghitza O (2012) On the role of theta-driven syllabic parsing in decoding speech: intelligibility of speech with a manipulated modulation spectrum. Front Psychol 3: 238. https://doi.org/10.3389/fpsyg.2012.00238 PMID: 22811672

68. Keitel C, Thut G, Gross J (2017) Visual cortex responses reflect temporal structure of continuous quasi-rhythmic sensory stimulation. Neuroimage 146: 58–70. https://doi.org/10.1016/j.neuroimage.2016.11.043 PMID: 27867090

69. Kösem A, Van Wassenhove V (2017) Distinct contributions of low-and high-frequency neural oscillations to speech comprehension. Language, Cognition and Neuroscience 32: 536–544.

70. Peelle JE, Davis MH (2012) Neural Oscillations Carry Speech Rhythm through to Comprehension. Front Psychol 3: 320. https://doi.org/10.3389/fpsyg.2012.00320 PMID: 22973251

71. Simmons JP, Nelson LD, Simonsohn U (2011) False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. Psychological science 22: 1359–1366. https://doi.org/10.1177/0956797611417632 PMID: 22006061

72. Bieniek MM, Bennett PJ, Sekuler AB, Rousselet GA (2016) A robust and representative lower bound on object processing speed in humans. European Journal of Neuroscience 44: 1804–1814. https://doi.org/10.1111/ejn.13100 PMID: 26469359

73. Poldrack RA, Baker CI, Durnez J, Gorgolewski KJ, Matthews PM, et al. (2017) Scanning the horizon: towards transparent and reproducible neuroimaging research. Nature Reviews Neuroscience 18: 115–126. https://doi.org/10.1038/nrn.2016.167 PMID: 28053326

74. Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9: 97–113. PMID: 5146491

75. Koike KJ, Hurst MK, Wetmore SJ (1994) Correlation between the American-Academy-of-Otolaryngology-Head-and-Neck-Surgery 5-minute hearing test and standard audiological data. Otolaryngology-Head and Neck Surgery 111: 625–632. https://doi.org/10.1177/019459989411100514 PMID: 7970802

76. Brainard DH (1997) The Psychophysics Toolbox. Spat Vis 10: 433–436. PMID: 9176952

77. Smith ZM, Delgutte B, Oxenham AJ (2002) Chimaeric sounds reveal dichotomies in auditory perception. Nature 416: 87–90. https://doi.org/10.1038/416087a PMID: 11882898

78. Yuan J, Liberman M (2008) Speaker identification on the SCOTUS corpus. Journal of the Acoustical Society of America 123: 3878.

79. Strauss A, Schwartz JL (2017) The syllable in the light of motor skills and neural oscillations. Language Cognition and Neuroscience 32: 562–569.

80. Boersma P (2001) Praat, a system for doing phonetics by computer. Glot International 5: 341–345.

81. 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: 156869. https://doi.org/10.1155/2011/156869 PMID: 21253357

82. 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. PMID: 14680264

83. Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. IEEE Trans Biomed Eng 44: 867–880. https://doi.org/10.1109/10.623056 PMID: 9282479

84. Ince RA, Giordano BL, Kayser C, Rousselet GA, Gross J, et al. (2017) A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. Hum Brain Mapp 38: 1541–1573. https://doi.org/10.1002/hbm.23471 PMID: 27860095

85. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological 57: 289–300.