Low-frequency neural tracking of natural speech envelope reflects evoked responses to acoustic edges, not oscillatory entrainment

Katsuaki Kojima*[1,2,3], Yulia Oganian*[2], Chang Cai[3], Anne Findlay[3], Edward Chang[2]⁑, Srikantan Nagarajan[3]⁑

1. Department of Pediatrics, University of California, San Francisco, Box 0734, 550 16th Street, 5th Floor, San Francisco, CA 94143, USA

2. Department of Neurological Surgery, University of California, San Francisco, 675 Nelson Rising Lane, San Francisco, CA 94158, USA

3. Department of Radiology, University of California, San Francisco, 513 Parnassus Avenue, S362, San Francisco, CA 94143-0628

*, ⁑ authors contributed equally

Corresponding authors:

Edward Chang: Edward.chang@ucsf.edu

Srikantan Nagarajan: Srikantan.Nagarajan@ucsf.edu

Abstract

The amplitude envelope of speech is crucial for accurate comprehension, and several studies have shown that the phase of neural activity in the theta-delta bands (1 - 10 Hz) tracks the phase of the speech amplitude envelope during listening, a process referred to as envelope tracking. However, the mechanisms underlying envelope tracking have been heavily debated. A dominant model posits that envelope tracking reflects continuous entrainment of endogenous low-frequency oscillations to the speech envelope. However, it has proven challenging to distinguish this from the alternative that envelope tracking reflects evoked responses to acoustic landmarks within the envelope. Here we recorded magnetoencephalography while participants listened to natural and slowed speech to test two critical predictions of the entrainment model: (1) that the frequency range of phase-locking reflects the stimulus speech rate and (2) that an entrained oscillator will resonate for multiple cycles after a landmark-driven phase reset. We found that peaks in the rate of envelope change, acoustic edges, induced evoked responses and theta phase locking. Crucially, the frequency range of this phase locking was independent of the speech rate and transient, in line with the evoked response account. Further comparisons between regular and slowed speech revealed that encoding of acoustic edge magnitudes was invariant to contextual speech rate, demonstrating that it was normalized for speech rate. Taken together, our results show that the evoked response model provides a better account of neural phase locking to the speech envelope than oscillatory entrainment.

Introduction

Speech comprehension is essential to human communication. A major computational step in neural processing of speech is tracking of its amplitude envelope, the overall intensity of speech across spectral bands. The speech envelope is dominated by fluctuations in the range of ~ 1–10 Hz, which cue the syllabic structure of speech, and the removal of which from the speech signal severely impairs intelligibility (Drullman et al., 1994a, 1994b). A multitude of studies have shown a consistent relationship between the phase of band-limited low-frequency neural activity measured in M/EEG over auditory cortical areas and the phase of the amplitude envelope of speech, a phenomenon widely known as envelope tracking (Ahissar et al., 2001; Kösem et al., 2018; Luo & Poeppel, 2007). Subsequent studies found that the strength of envelope tracking is correlated with speech intelligibility, suggesting that it could constitute an essential stage in speech comprehension (Abrams et al., 2008; Peelle et al., 2013). However, the neural computations underlying speech envelope tracking are controversially debated (Gwilliams, 2019; Obleser & Kayser, 2019; Zoefel, ten Oever, et al., 2018).

According to a dominant framework, envelope tracking reflects the neural entrainment (i.e., phase alignment) of endogenous neural oscillations in the delta-theta range (1-10 Hz) to fluctuations in the amplitude envelope of speech, with the primary entrained frequency reflecting the syllabic rate of the speech signal (Ding et al., 2015; Giraud & Poeppel, 2012; Zoefel, 2018). Functionally, oscillatory entrainment is thought to benefit speech processing due to its resonance properties (Haegens & Zion Golumbic, 2018; Helfrich et al., 2019), implementing automatic prediction of the timing of upcoming speech (e.g., the next syllable (Hovsepyan et al., 2020)). It is thought that in natural speech, phase resetting of entrained oscillations is driven by discrete acoustic landmark events in the speech signal. These could be amplitude peaks, marking vowel midpoints (Ghitza, 2013), or acoustic edges (i.e., peaks in the rate of acoustic amplitude change) that mark the perceived onset of vowels (Doelling et al., 2014; Hertrich et al., 2012), with recent intracranial evidence of acoustic edge encoding in speech cortical areas pointing to the latter (Oganian & Chang, 2019).

However, recent evidence has indicated that phase alignment of low-frequency neural activity in the presence of dynamic streams does not necessarily reflect oscillatory entrainment but could also stem from evoked responses to transient events (Breska & Deouell, 2017; Capilla et al., 2011). While this data has mostly come from the visual modality, it cannot be rejected that a similar mechanism accounts for phase locking to natural speech, which would have major implications for theories of speech perception. For instance, instead of oscillatory resonance, predictive processing of speech could rely on non-oscillatory temporal prediction mechanisms guided by statistical learning (Breska & Deouell, 2017; Friston et al., 2020; Sohoglu & Davis, 2016). Recent M/EEG studies of continuous speech characterized evoked response dynamics for different speech features, in particular the impulse response to the continuous

speech envelope (Daube et al., 2019; Di Liberto et al., 2015; Lalor & Foxe, 2010). This work provides a foundational framework for the study of M/EEG responses to natural speech. However, to date, the rapid dynamics of natural speech left it open in which parts of the envelope drive the neural response and whether they additionally entrain an oscillator (Ruhnau et al., 2020; Zoefel et al., 2019).

Here, we took an innovative approach to this problem by comparing speech at a regular speech rate and equally intelligible speech slowed to 1/3 of its original rate (Figure 1A), during magnetoencephalography (MEG) recordings. Slowing increased the time interval between consecutive acoustic landmarks in the speech envelope (Figure 1), hence allowing us to directly test two central predictions of the oscillatory entrainment theory against those of an evoked response mechanism. First, oscillatory entrainment predicts that slowing of the speech signal would shift the frequency of neural phase locking to peak at lower frequencies, due to the reduced syllabic rate. In contrast, an evoked response mechanism predicts similar frequency distributions of phase locking in regular and slow speech, as the shape and frequency content of the evoked response to landmarks should not be affected by slowing. Second, oscillatory entrainment predicts that phase coherence following a phase reset would last for several frequency cycles, as long as another phase-resetting event has not occurred (Figure 1B bottom right). In contrast, an evoked response mechanism predicts that phase alignment will only persist for the duration of the evoked response (Figure 1B, middle panel). In slow speech, the interval to the next landmark is extended, allowing to differentiate these patterns. Before examining these two predictions, we identified the phase-resetting landmark, hypothesizing it to be acoustic edges.

Our use of slow speech allowed us to investigate a central computational mechanism in speech perception that is thought to rely on temporal predictions mediated by oscillatory entrainment, namely speech rate normalization. Speech rate normalization allows listeners to adjust perceptual processes, in particular envelope tracking, to differences in speech rate. It has been proposed that speech rate normalization is implemented through shifts in the phase-locking frequency with changes in speech rate (Ahissar et al., 2001; Bosker & Kösem, 2017; Kösem et al., 2018; Nourski et al., 2009; Pefkou et al., 2017). The comparison of MEG responses to the regular and slowed speech in this study allowed us to examine how the strength and frequency range of phase resets at acoustic edges depend on the speech rate. In particular, as the envelope slope at acoustic edges scales down with slowing of speech (Figure 1B left vs. right), the comparison between slow and regular speech allowed testing whether phase locking to acoustic edges depends on amplitude slopes, and how this relationship is re-adjusted to the range of envelope slopes at each speech rate.

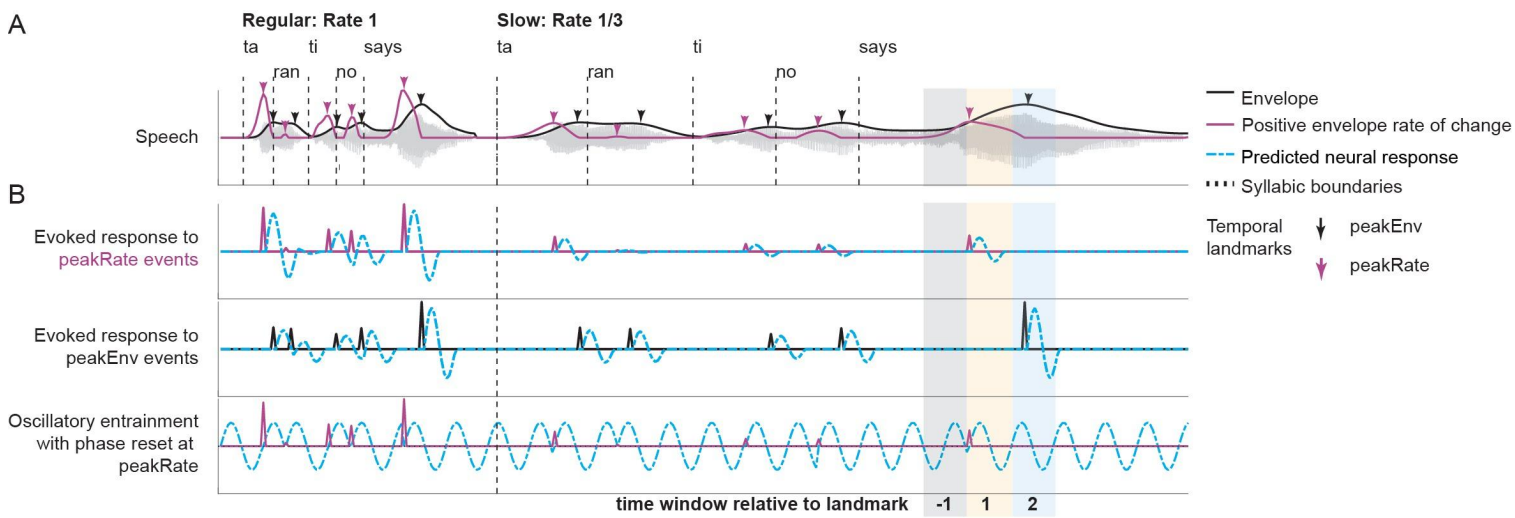

Figure 1. Stimuli and model predictions. A. The acoustic waveform of an example utterance ('Tarantino says…"), with syllable boundaries, amplitude envelope, and first temporal derivative of the envelope superimposed on it. The same utterance is shown at a regular rate (left) and slowed (right) speech rate. Arrows mark temporal landmark that could drive phase resetting (Black: local peaks in the envelope, peakEnv; Purple: acoustic edges, defined as local peaks in the rate of change of the envelope, peakRate). **B.** Schematic illustrations of the predicted neural response to the utterance in A using three different models. Top: Evoked response to peakRate events; Middle: peakEnv events; Bottom: Fixed-frequency oscillatory entrainment with phase reset. Colored bars mark the duration of a single oscillatory cycle at that frequency, relative to peakRate event occurrence for an example event.

Results

We recorded MEG while participants (n = 12) listened to ten minutes of continuous speech, once at the original rate, and once slowed to 1/3 of the original speech rate (Figure 1A). Slowing did not impair participants' comprehension behavior, as indicated by responses to 3-4 multiple choice comprehension questions after each story (chance: 50 %, regular speech: mean = 83%, SD = 13%; slow speech: mean = 90%, SD = 9.5%; $t(11) = -1.85$, $p = 0.09$; Figure 2A). Stimulus materials were selected from the Boston University Radio Speech Corpus, BURSC (Ostendorf, Price, & Shattuck-Hufnagel, 1995); see Tables S1 and S2 for full stimulus transcript and list of comprehension questions.

In a first step, we characterized the temporal dynamics of acoustic landmark events in our speech stimulus (Figure S1). At the regular speech rate, the average interval between landmarks (similar for peakRate and peakEnv) was 175ms, corresponding to an average landmark frequency in the theta band at 5.7 Hz (SD = 2.9 Hz, Figure 2D), in line with the typical power spectrum of the amplitude envelope in natural speech (Ding et al., 2017). When slowed to 1/3 of the original speech rate, the average frequency of landmarks was 1.9 Hz (SD = 1 Hz, similar for peakRate and peakEnv), shifting the peak in the envelope power spectrum to the delta band.

To focus our analyses on sensors that capture auditory sensory processing, we selected ten sensors with the largest M100 response to speech onsets after silence periods from each hemisphere (Figure 2B) for all further analyses. The M100 response showed the typical dipole pattern in each hemisphere (Chait et al., 2004).

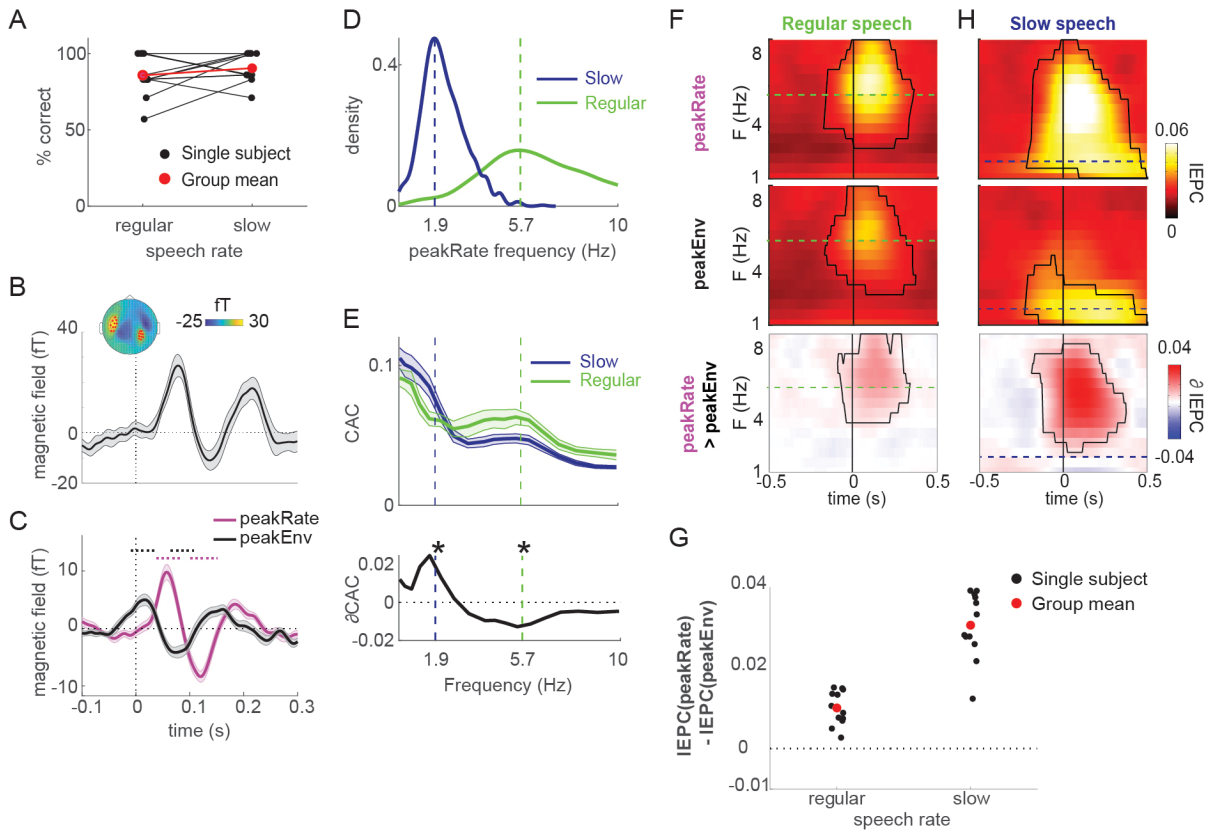

Figure 2. Phase locking during speech comprehension. A. Single-subject (black) and group-average (red) comprehension performance for regular and slow speech. **B.** Sensor selection was based on M100 response to utterance onsets. Top: Topographic map of a group-averaged M100 response with selected sensors marked in red. Bottom: Group-averaged evoked response across all 20 sensors included in the analysis. Error bars are ± 1 SEM across subjects. **C.** Group-averaged evoked response aligned to peakRate and peakEnv events. Dotted lines mark clusters with $p < 0.05$ with a cluster-based permutation test against 0. Error bars are ± 1 SEM across subjects. **D.** Frequency of peakRate/peakEnv events in slow speech (blue) and regular speech (green). Dashed vertical lines mark the average frequency of peakRate events in slow (blue, 1.9 Hz) and regular speech (green, 5.7 Hz). **E.** Cerebro-acoustic phase coherence (CAC) between MEG responses and speech envelope (upper panel), and the difference between slow and regular speech (ΔCAC, lower panel). Data were filtered in semi-logarithmically spaced bands between 0.3 and 10 Hz for this analysis. Dashed vertical lines mark the average frequency of peakRate events in each condition, as shown in D. * $p < 0.01$ in post-hoc t-tests with interaction $p < 0.01$. Error bars are ± 1 SEM

across subjects. **F.** Inter-event phase coherence (IEPC) with alignment to peakRate events (upper) and peakEnv events (middle) in regular speech. The dashed line indicates the average frequency of the peakRate events. Outlined area: $p < 0.01$ with a cluster-based permutation test. Bottom: Difference in IEPC between alignments to peakRate and peakEnv. Outlined area: $p < 0.01$ with a cluster-based permutation test. **G.** Theta-band IEPC difference between alignment to peakRate and peakEnv events in single subjects (black) and the group-average difference for each speech rate (red). **H.** Same as F for slow speech. All average responses are averaged across subjects and sensors.

Cerebro-acoustic phase coherence between speech envelope and MEG reflects speech rate

To confirm that cortical speech envelope tracking was present in our data (Peelle & Davis, 2012), we calculated the cerebro-acoustic phase coherence (CAC) between neural responses and the speech envelope in frequency bands between 0.3 and 10 Hz. CAC typically peaks at the frequency corresponding to the speech rate (Pefkou et al., 2017), which in our data corresponds to the frequency of peakRate and peakEnv in each rate condition (regular: 5.7 Hz, slow: 1.9 Hz). Indeed, speech rate had opposite effects on CAC in these two frequency bands (2-way repeated-measures ANOVA, interaction $F(1, 11) = 31.20$, $p < 0.001$, $\eta^2 = 0.30$, Figure 2E). At 5.7 Hz, CAC was higher for regular speech ($t(11) = 5.6$, $p < 0.001$, $\eta^2 = 0.42$), while at 1.9 Hz it was higher for slow speech ($t(11) = 3.4$, $p = 0.006$, $\eta^2 = 0.29$). Moreover, CAC was overall higher at lower frequencies ($F(1, 11) = 16.44$, $p < 0.001$, $\eta^2 = 0.39$), reflecting the 1/f shape of MEG frequency spectrum. No other frequency band showed a significant effect of speech rate on CAC (all Bonferroni-corrected $p > 0.05$). Overall, this result replicates previous findings of cortical speech envelope tracking in frequency bands corresponding to the modulation frequency of the speech envelope.

Acoustic edges drive evoked response and theta phase locking

Next, we asked which landmark in the speech envelope drives the phase locking to the envelope in regular speech. First, we tested whether averaging the time-domain neural data (band-pass filtered 1-40 Hz) with alignment to peakRate and peakEnv led to evoked responses and examined their characteristics. While peakEnv closely follows on peakRate in regular speech, the interval between them varies. Thus, aligning to the incorrect landmark should lead to (1) a reduced magnitude of the averaged evoked neural signal due to smearing, and (2) shifts in response onset times away from the acoustic event. We found that alignment to either peakRate or peakEnv led to a robust transient evoked response (Figure 2C). Crucially, the evoked response was of larger magnitude when neural data were aligned to peakRate events than to peakEnv events (peak magnitude: $t(11) = 5.9$, $p < 0.001$). Moreover, this response started after peakRate events, but before peakEnv events (response latency relative to the event for peakEnv: -12.5 ms;

peakRate: +50 ms, determined as the first significant time point in a cluster-based permutation test against 0). Together, these analyses indicate that peakRate events, not peakEnv events, drove the response.

We then turned to examine whether peakRate events also drove the low-frequency phase reset in envelope tracking. To test this, we computed the inter-event phase coherence (IEPC) with an alignment to peakRate and peakEnv. While CAC quantifies phase consistency across time, IEPC is calculated across single event occurrences (i.e., single trials) for each time point. IEPC thus enables tracking of the temporal dynamics of phase locking (Gross et al., 2013). Consistent with the time-domain analysis, we observed stronger IEPC in the theta band (4-8 Hz, Figure 2F) following peakRate events than with alignment to peakEnv events (Figure 2F, two-dimensional cluster-based permutation $p < 0.001$).

Taken together, the analysis of regular speech confirms that acoustic edges, and not envelope peaks, drive evoked responses to the speech envelope and trigger low-frequency phase resets, extending our previous work (Oganian & Chang, 2019).

Slowing speech rate does not eliminate theta phase locking

Next, we tested how slowing the speech signal would affect the frequency range of phase locking to peakRate events. In the regular speech condition, phase locking was observed in the theta range (as typically found (Peelle & Davis, 2012)), which is also the frequency of peakRate events. If this reflects the reset of an oscillator of matched frequency, phase locking in slowed speech should be observed at the frequency of peakRate events in slow speech (1.9 Hz). In contrast, if this reflects evoked responses, in slowed speech phase locking should still be observed in the theta range. In line with the latter, we found that, in slowed speech, IEPC to peakRate in the theta band remained pronounced (Figures 2H, two-dimensional cluster-based permutation against pre-event baseline $p < 0.001$), suggesting that this effect was independent of the stimulus speech rate, as predicted by the evoked response account. However, additionally, the IEPC increase in slow speech also extended into the delta band, to include the frequency of peakRate events in slow speech (1.9 Hz).

To better understand the source of phase locking to peakRate in theta and delta bands, we compared IEPC between alignments to peakRate and peakEnv. This comparison revealed significantly stronger IEPC for alignment to peakRate in the theta band, but no difference in IEPC in the delta band, around the frequency of events in slowed speech, which was increased for both alignments (Figure 2H, cluster-based permutation $p < 0.001$). This comparison suggested that the delta band component was not dependent on the correct alignment to the phase-resetting event. To further test whether the delta component can be accounted for by an evoked response component without any oscillatory entrainment, we simulated the IEPC pattern under an evoked response model (Figure S2, see supplementary materials for simulation details). If this low-frequency phase locking in slow speech is a by-product of evoked responses to

peakRate events, it should also emerge from an analysis of IEPC in an evoked response model. Indeed, in our model, delta-band IEPC increased under peakRate and peakEnv alignments, in addition to the theta-band IEPC time-locked to peakRate (Figure S2). Based on this result, we hypothesized that delta-IEPC in slowed speech reflects the frequency of evoked responses to peakRate events. To test for such adaptive frequency adjustment, we examined whether, within regular speech, the dominant IEPC frequency varied depending on the time interval between neighboring peakRate events. As predicted, we found a gradual shift in IEPC frequency ranges depending on the interval of peakRate events to their adjacent events. Our model simulation showed that this pattern is also expected under an evoked response mechanism (Figure S3).

Collectively, these findings disagree with an oscillatory entrainment account, instead suggesting that the low-frequency phase locking to the speech envelope in MEG is driven by evoked responses to peaks in the envelope rate of change (peakRate). This also shows that IEPC to peakRate events reflected two different sources: (1) local responses to individual peakRate events and (2) the rate of occurrence of responses to peakRate events. Furthermore, the shift in IEPC frequency bands with changes in speech rate may be the product of a time-frequency decomposition of a series of evoked responses, rather than a shift in the frequency of an entrained oscillator (Kösem et al., 2018).

Theta phase locking is limited to a single cycle after peakRate events.

A second prediction of the oscillatory entrainment model is that phase alignment would persist for multiple oscillatory cycles after a phase reset. In our data, this should be expressed as increased IEPC extending beyond a single oscillatory cycle after peakRate events (Haegens & Zion Golumbic, 2018; Helfrich et al., 2019; Meyer et al., 2019). In contrast, if phase locking results from evoked responses to peakRate events, the increase in IEPC should be limited to the duration of the evoked response (approximately 250 ms, see Figure 2). To test this, we focused on peakRate events in the slowed speech stimulus that were at least 500 ms apart from preceding and subsequent peakRate events (n = 456 events, 40% of all events). For these events, IEPC increases in the theta band (4-8 Hz, average cycle length 125 to 250 ms) would not be affected by preceding/subsequent peakRate events for at least two oscillatory cycles, allowing us to test for continuous entrainment (Figures 3A, B). Notably, such cycles were distributed throughout the slowed speech stimulus and not limited to sentence or phrase ends. We extracted IEPC for each frequency band, for a single theoretical cycle width preceding and up to three cycle widths after each peakRate event (defined based on the center frequency of each frequency band used in the analysis). Additionally, using a shuffling distribution, we calculated an IEPC baseline. We first tested whether IEPC in the cycle immediately following peakRate events was elevated as compared

to surrounding cycles. In a second step, we tested whether IEPC in the pre- and post-event cycles differed significantly from baseline.

For bands with two post-event cycles (4 bands in the range of 4 – 6Hz, see Figure 3), IEPC differed between cycles (repeated-measures ANOVA: main effect of cycle, F (2, 22) = 14.82, $p < 0.001$, $\eta^2 = 0.26$), and this modulation did not differ between bands (interaction: $p = 0.6$, n.s., main effect of band: Greenhouse-Geyser $p > 0.05$). Planned comparisons showed that IEPC was increased in the first post-event cycle relative to the pre-event cycle ($p < 0.001$) as well as relative to the 2nd post-event cycle ($p < 0.001$). In contrast, we observed no significant difference in IEPC between the pre-event and the second post-event cycles ($p = 0.3$, n.s.), and IEPC in both cycles did not differ from the permutation-based baseline (all Bonferroni-corrected p's > 0.05). Similarly, for bands with three post-event cycles (three bands in the range of 6-8 Hz), IEPC differed between cycles (F (3, 33) = 6.58, $p < 0.001$, $\eta^2 = 0.15$), being higher in the cycle immediately following peakRate events relative to the other cycles (planned contrast of 1st post-event cycle to all other cycles: $p = 0.002$). Here as well, we saw no significant IEPC difference between frequency bands (p = 0.15 n.s.), and the effect of the cycle did not differ significantly between bands ($p = 0.054$). Moreover, we found that IEPC did not differ between the permutation-based baseline and cycles that were not immediately following peakRate events (baseline vs. 1st post-event cycle: Bonferroni-corrected $p = 0.006$; baseline vs. all other cycles Bonferroni-corrected p's > 0.05, n.s.). Notably, this pattern was evident in single subjects (see Figure S3 for single-subject data).

Overall, the IEPC increase in the theta band extended for the duration of the evoked response, but not beyond. This analysis thus illustrates the transient nature of theta phase locking to peakRate events, which is more in line with an evoked response mechanism of speech envelope tracking than with an oscillatory entrainment model.

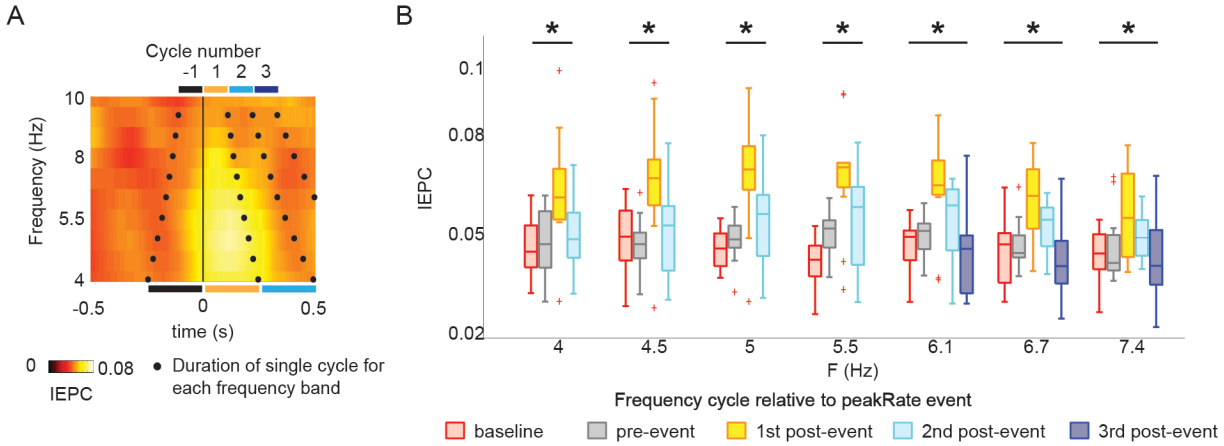

Figure 3. Theta phase locking is limited to the duration of the evoked response after peakRate events. A. IEPC across selected peakRate events that were at least 500 ms apart from preceding and subsequent events. Black dots mark the extent of one oscillatory cycle in each frequency band. **B**. Group

distribution IEPC in 3-4 cycles around peakRate events and baseline in each frequency band. * $p < 0.05$. Boxplots reflect median, 25th and 75th percentile, and range of values. Outliers are plotted individually.

Speech rate normalization of peakRate IEPC

The perceptual ability to adapt to variation in the speech signal resulting from changes in the speech rate, i.e., the number of syllables produced per second, is referred to as speech rate normalization. Changes in speech rate result in acoustic changes in the speech signal, including slower amplitude increases at acoustic edges, resulting in lower peakRate magnitudes (Figure 4A-C). We had previously found that responses to peakRate monotonically scale with peakRate magnitude being larger for faster changes in the speech amplitude (Oganian & Chang, 2019). Efficient envelope tracking across speech rates would thus require remapping of neural responses to peakRate magnitude, to account for this overall reduction. Here, we assessed the effect of speech rate on the magnitude of theta IEPC to peakRate events. In the slowed speech, stimuli peakRate magnitudes were 1/3 of those in regular speech (Figures 4A-C). If no normalization occurs, IEPC magnitudes in slow speech should reflect absolute peakRate values, resulting in an overall reduction in IEPC (Figure 4F, dark dots). In contrast, if theta IEPC to peakRate is invariant to speech rate, it should reflect peakRate values relative to the contextual speech rate, resulting in similar IEPC magnitudes in both speech rate conditions (Figure 4F, light dots).

An evaluation of IEPC after peakRate events, split by peakRate magnitude quantiles, showed comparable theta IEPC in both speech rate conditions (Figure 4D-E), such that average theta IEPC was higher for larger peakRate magnitudes across both rate conditions (main effect of peakRate quantile: b = 0.01, SD = 0.001, t = 1.4, χ^2= 55.0, $p = 10^{-13}$). Crucially, they did not differ between regular and slow speech (peakRate quantile X speech rate: b = 0.003, SD = 0.005, t = 0.6, n.s., Figure 4G), as expected in case of speech rate normalization (Figure 4F). Thus, the magnitude of phase reset induced by peakRate depended on its magnitude relative to the local speech rate context, allowing for the flexible encoding of peakRate information at different speech rates.

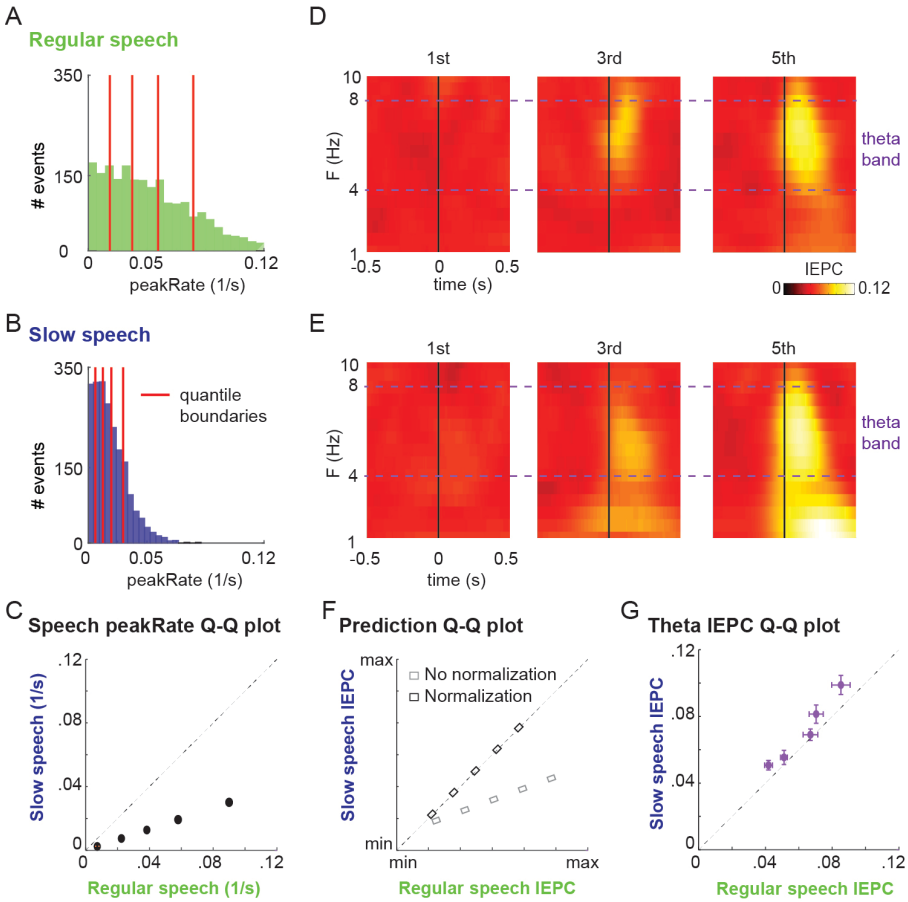

Figure 4. Normalization of peakRate IEPC for contextual speech rate. A. Histogram of peakRate magnitudes in regular speech, with quantile boundaries marked in red. **B.** Same as A for slow speech **C**. Quantile-Quantile plot of peakRate magnitudes in regular and slowed speech stimulus. peakRate values in slowed speech stimulus are 1/3 of peakRate values in regular speech stimulus. **D.** IEPC in 1st, 3rd, 5th peakRate magnitude quantile. Horizontal lines mark the theta frequency range (4-8Hz). **E**. Same as D for slow speech. **F.** Predicted quantile-quantile plots of theta IEPC in regular and slowed speech with (dark) or without (light) normalization. **G.** Quantile-quantile plot of theta-band IEPC (mean, error bars mark ± 1 SEM across subjects) in regular and slow speech. Theta IEPC quantile-quantile values are close to the diagonal, indicating similar magnitudes of theta IEPC in regular and slowed speech conditions.

Discussion

We evaluated local temporal dynamics in MEG neural responses to continuous speech under the predictions of an evoked response model and an oscillatory entrainment model. Confirming previous results, we found that acoustic edges (peakRate events), but not peaks in the envelope (peakEnv events), drove evoked responses and phase locking in sensors over auditory cortical areas (Hertrich et al., 2012;

Oganian & Chang, 2019). Phase locking to acoustic edges had two components: a local component reflective of the evoked response, and a global component, which captured the frequency of evoked responses in the stimulus. Both components were predicted by a simulation of an evoked response model of envelope tracking. The local evoked response component in the theta band was transient and limited to the duration of the evoked response after peakRate events, providing no evidence for sustained oscillatory phase locking, the hallmark marker of entrainment models (Helfrich et al., 2019; Peelle & Davis, 2012). Moreover, we found that the magnitude of the evoked phase reset to acoustic edges reflected the amplitude slope at the acoustic edge after normalization for speech rate, offering novel neural evidence for speech rate normalization. Overall, our findings confer with the predictions of the evoked response model and suggest that neural phase locking induced by evoked responses to acoustic edges is the primary source of speech envelope tracking in the theta-delta band in MEG.

Phase locking between low-frequency MEG components and the amplitude envelope of speech is well established (Ahissar et al., 2001; Luo & Poeppel, 2007). In natural speech, which does not show strong temporal rhythmicity, such phase locking must rely on repeated realignment of neural phase to the speech signal (Haegens & Zion Golumbic, 2018). It is, however, disputed which acoustic events in the speech signal induce these phase resets (Aubanel et al., 2016; Doelling et al., 2014; Ghitza, 2011; Peelle & Davis, 2012). Here, we show that acoustic edges drive phase reset of neural activity in low-frequency bands in MEG, in line with recent intracranial results (Oganian & Chang, 2019). This supports the central role of acoustic edges, which mark the onsets of stressed vowel nuclei and closely correspond to perceptual centers (P-centers) of syllables (Harsin, 1993; Scott, 1998) in speech comprehension, and establishes them as the basis for the cortical representation of the speech envelope.

Neural phase resetting may be due to either the superposition of evoked responses or the entrainment of endogenous oscillatory activity. A hallmark signature of oscillatory entrainment is the persistence of an oscillation after the phase-resetting stimulus (Breska & Deouell, 2017; Haegens & Zion Golumbic, 2018). To test for this, we focused on speech portions where the intervals between consecutive acoustic edges were long enough to detect reverberation in the theta band (4-8 Hz), the main frequency band that showed a phase reset to acoustic edges. We found that phase locking increased for the duration of the evoked response, but diminished right after to the baseline level, showing no evidence of reverberation and supporting the interpretation of theta phase locking as a transient evoked response. Moreover, this analysis showed no phase alignment in the cycle preceding peakRate events, providing no evidence for the anticipatory phase locking to the speech envelope in the theta band.

With the slowing of the speech signal, the phase reset after acoustic edges expanded to include not only the theta (4-8 Hz) but also the delta frequency range (1-4 Hz). While phase locking in the theta range was temporally aligned to peakRate events, phase locking in the delta range appeared more extended in

time. Using a simulation of an evoked response model, we found that such low-frequency phase locking does not necessarily result from oscillatory alignment. Instead, it can reflect the overall rate of occurrence of evoked responses, in contrast to theta-band phase locking, which reflects individual evoked responses. Therefore, crucially, our analyses show that both components are accounted for by an evoked response model. It should be noted that our results do not refute the possibility that an additional oscillatory component contributes to phase locking at the speech rate frequency in each condition. Rather, our results show that such a component is not necessary to produce phase-locking patterns as observed for natural speech in MEG, but that these can be explained more parsimoniously as reflecting evoked responses. Previous research has suggested distinct functional roles of delta and theta band entrainment in linguistic processing, with theta implicated in envelope tracking, and delta in the processing of phrasal boundaries (Donhauser & Baillet, 2020; Meyer et al., 2017). In the present study, however, both theta and delta effects reflect responses to the acoustic amplitude envelope of speech, as the effect in the delta arose with slowing of the speech stimulus.

Shifting of the entrained oscillatory frequency to match the overall speech rate of the input was previously proposed as a mechanism for speech rate normalization, a behavioral (Reinisch, 2016; Wade & Holt, 2005) and neural (Alexandrou et al., 2018b; Kösem et al., 2018) phenomenon in speech perception. While our data do not support this account, we do find that the magnitude of phase locking to acoustic edges was normalized to reflect the distribution of peakRate magnitudes at each speech rate. Namely, while phase locking to acoustic edges was stronger for sharper edges (Biermann & Heil, 2000; Oganian & Chang, 2019), it was comparable in regular and slowed speech, despite acoustic edges being flatter in slow speech. This result suggests that the cortical representation of acoustic edges reflects the magnitude of an edge relative to the contextual speech rate. Such shifting of the dynamic range for the representation of acoustic edge magnitude constitutes a flexible mechanism that maximizes the dynamic range for the representation of speech temporal dynamics (Diehl, Souther, & Convis, 1980; Hirataa & Lambacher, 2004). Further studies are necessary to elucidate the neural mechanisms underlying this process at the circuit level. One possibility is that the local distribution of peakRate values is continuously learned via dynamic adjustment of synaptic weights.

Our approach represents a methodological departure from previous investigations of speech envelope tracking. Namely, previous studies focused on cerebro-acoustic coherence (CAC), which reflects the consistency of phase differences between the neural signal and the acoustic stimulus across time (Peelle, Gross, & Davis, 2013). CAC is primarily sensitive to regularities across time, such as the rate of phase resets. In contrast, we used inter-event phase coherence (IEPC), which focuses on assessing temporally local similarities in neural phase across repeated occurrences of the same acoustic event (see (Gross et al., 2013) for IEPC to speech onsets). As IEPC is sensitive to local phase resets as well as to their rate of

occurrence, this approach revealed that both components are reflected in IEPC to acoustic edges. In regular speech, both components overlapped within the theta frequency range. In contrast, slowing of the speech signal separated them in an evoked response component in the theta band, and a slower event frequency component. The comparison of IEPC and CAC results suggests that the CAC analysis (Figure 2) was sensitive only to the latter component, whereas IEPC uncovered both.

Speech rate manipulations are frequently used to study neural tracking of speech, in particular, speech envelope tracking (Ahissar et al., 2001; Ghitza & Greenberg, 2009; Nourski et al., 2009; Pefkou et al., 2017). Previous studies used compressed speech, in which speech rate is increased, to study temporal boundaries on envelope tracking and intelligibility. In contrast, here we used slowed speech as we aimed to spread out distinct acoustic envelope features in time. It is essential to reconsider previous findings with compressed speech in light of the evoked response framework. For example, while envelope tracking and intelligibility deteriorate for speech rates higher than 8 Hz, insertion of brief silence periods in compressed speech, which returns the effective speech rate to below 8 Hz, improves intelligibility (Ghitza & Greenberg, 2009). While this result is typically interpreted as evidence for oscillatory envelope tracking in the theta range, an evoked response framework would argue that it is reflective of the minimal refractory period of neural populations that encode acoustic edges in speech.

Recent studies have demonstrated that neural responses to speech in M/EEG reflect not only the amplitude envelope but also phonetic and semantic information (Broderick et al., 2019; Di Liberto et al., 2015; O'Sullivan et al., 2015). These studies focused on impulse responses (evoked responses) to different speech features and their modulation by higher-order information content. We extend upon these studies by pinpointing the envelope feature that drives neural responses to the envelope and show that this feature, acoustic edges, does not entrain an oscillator beyond the impulse response. Future studies will be necessary to address whether the same is true for responses at the phonemic or phrasal levels. Furthermore, we found responses to speech onsets and amplitude modulations in ongoing speech on the same sensors over temporal cortical areas. This stands in contrast to intracranial results, which showed that distinct areas of the superior temporal gyrus represent speech onsets and the content of ongoing speech (Hamilton et al., 2018; Oganian & Chang, 2019). However, this is not surprising as we focus on an analysis of MEG data from sensors over the temporal cortex, reflecting the summation of neural sources across auditory cortical areas.

Natural speech does not have strong temporal rhythmicity (Alexandrou et al., 2018a). Our focus on envelope tracking for natural speech indicates that in this case, neural signatures of envelope tracking are well explained by an evoked response model without the need for an oscillatory component. These results seemingly contradict recent findings of predictive entrainment to music (Doelling et al., 2019). However, our study employed natural speech with considerable variability in inter-edge intervals, whereas Doelling

et al. used rhythmic musical stimuli. Critically, recent neuropsychological work dissociated neural mechanisms for prediction based on rhythmic streams from predictions in non-rhythmic streams (Breska & Ivry, 2018). This adds an important caveat to the current debate, suggesting that the results of Doelling et al. may perhaps not extend to natural speech with inherent temporal variability and reduced rhythmicity. The present study thus calls to reevaluate the role of oscillatory entrainment in natural speech comprehension. However, it does not preclude the possibility that the introduction of additional rhythmicity to speech, e.g., in poetry or song, recruits additional neural processes associated with the processing of rhythms. Such additional processes might support speech comprehension and could underlie some of the recent findings obtained with a rhythmic speech stimulus (Ding et al., 2015; ten Oever & Sack, 2015; Zoefel et al., 2019). Therefore, caution needs to be exercised when extending findings obtained with rhythmic stimuli (e.g., (Ding et al., 2015; Doelling et al., 2019; Zoefel, Archer-Boyd, et al., 2018)) to natural speech.

In summary, our results show that an evoked response model accounts for the main neural signatures of speech envelope tracking. The neural representation of acoustic edges serves as a cue to syllabic nucleus onsets and informs about speech rate via the inter-event interval. Moreover, the speech rate normalization of these responses renders this mechanism flexibly adaptable to changes in speech rate. To conclude, evoked responses to acoustic edges track the syllabic rate in speech and provide a flexible framework for temporal analysis and prediction during speech perception.

Methods

Participants

Twelve healthy, right-handed volunteers (six females; age range 22-44 years, median 25 years) participated in the study. All participants were native speakers of English. All participants provided informed written consent and received monetary compensation for their participation. The study was approved by the University of California, San Francisco Committee on Human Research. All participants gave informed written consent before experimental testing.

Speech stimulus

Task stimuli contained a selection of two 3-minute new excerpts from the Boston University Radio Speech Corpus (BURSC, Table S1) (Ostendorf et al., 1995), produced by a male and a female voice. The stimuli were split into 62 utterances of 2.2 –14.9 s, with silent periods of 500 – 1100 ms between utterances. Slowed speech stimuli consisted of the same utterances slowed to 1/3 of the original speech rate. Speech stimuli were slowed using the PSOLA (Pitch Synchronous Overlap and Add) algorithm, as

implemented in the software Praat (Boersma & Weenik, 2019), which slows down the temporal structure of the speech signal while keeping its spectral structure constant (Moulines & Charpentier, 1990).

Stimulus presentation

All stimuli were presented binaurally at a comfortable ambient loudness (~ 70 dB) through MEG compatible headphones using custom-written MATLAB R2012b scripts (Mathworks, https://www.mathworks.com). Speech stimuli were sampled at 16,000 Hz. Participants were asked to listen to the stimuli attentively and to keep their eyes closed during stimulus presentation.

Regular and slow speech were presented interleaved in 4 blocks, such that half the stimulus was heard first in the regular condition and half in the slowed condition. Comprehension was assessed with 3-4 unique comprehension questions after each block (Table S2). Comprehension was compared between regular and slow speech using a two-sided paired t-test.

Data analysis

All analyses were conducted in MATLAB R2019a (Mathworks, https://www.mathworks.com) using standard toolboxes and custom-written scripts. The FieldTrip toolbox (Oostenveld et al., 2011) was used for visualization of topography.

Acoustic feature extraction

We extracted the broad amplitude envelope of speech stimuli by low-pass filtering to 10 Hz and down-sampling to 100 Hz. We then calculated the derivative of the resulting loudness contours as a measure of the rate of change in the amplitude envelope. Finally, we extracted the sparse time series of local peaks in the amplitude envelope (peakEnv) and its derivative (peakRate). All features are depicted in Figure 1A, for an example stimulus excerpt. Overall the stimulus set contained 2106 peakRate and peakEnv events per condition.

Speech preprocessing

To assess cerebro-acoustic phase coherence between the speech envelope and MEG responses, the speech envelope was processed using the same procedure that was applied to the MEG responses: Down-sampling to 400 Hz and band-passed between 0.2 and 200 Hz using the second-order Butterworth filter. We then used the Hilbert transform to extract the phase in the delta-theta frequency range using eighteen band-pass filters (Gaussian filters, logarithmically increasing center frequencies between 0.2 – and 10 Hz with semi-logarithmically increasing bandwidth).

Neural data acquisition and preprocessing

MEG recordings were obtained with a 275-axial gradiometers whole-head MEG system (CTF, Coquitlam, British Columbia, Canada) at a sampling rate of 1,200 Hz. Three fiducial coils were placed on the nasion and left and right pre-auricular points to triangulate the position of the head relative to the MEG sensor array. The position of the patient's head in the device relative to the MEG sensors was determined using indicator coils before and after each recording interval to verify an adequate sampling of the entire field. The fiducial markers were later co-registered onto a structural magnetic resonance imaging scan to generate head shape (Teichmann et al., 2013).

Offline data preprocessing included (in this order) artifact rejection with dual signal subspace projection (DSSP) and down-sampling to 400 Hz. DSSP is a MEG interference rejection algorithm based on spatial and temporal subspace definition (Sekihara et al., 2016). Its performance has been recently validated using clinical data (Cai et al., 2019). To assess the broadband response, data were then band-pass filtered between 1 and 40 Hz using a second-order Butterworth filter. To evaluate the power and phase of each frequency component, data were band-pass filtering between 0.1 and 10 Hz using a second-order Butterworth filter. We extracted the phase and analytic amplitude in this frequency range using eighteen band-pass filters (Gaussian filters, logarithmically increasing center frequencies with semi-logarithmically increasing bandwidths) using the Hilbert transform.

Data segmentation

Evoked responses, band-passed power, and phase were segmented into symmetric epochs of 1000 ms duration around onsets of utterances, peakEnv events, and peakRate events, respectively. Segments containing sensor data above 1.5pT and visually identified artifacts (including muscle, eye blink, and motion) were flagged as bad events and removed from further processing (0.2 % of segments).

For analyses of evoked responses and narrow-band power, baseline correction (- 200 ms to 0 ms relative to event onset for the evoked response, -400 ms to -200 ms for narrow-band power) was applied before averaging across events. No baseline correction was applied to the instantaneous phase.

Sensor selection

To focus analyses on responses originating in temporal auditory areas, we selected ten sensors with maximal M100 responses from each hemisphere. The M100 magnitude was determined based on the time window of 60-100ms after speech onset.

Cerebro-acoustic phase coherence (CAC)

Phase locking between the speech envelope and MEG response was calculated across the entire duration of every utterance within each frequency band, using the Cerebro-acoustic phase coherence (CAC):

$$CAC(\varphi) = \frac{1}{N}\left|\sum_{t=1}^{T} \exp\left(i * (ph(\varphi, t) - phs(\varphi, t))\right)\right|$$

where φ is the center frequency of a frequency band, T is the number of time samples in an utterance, ph is the phase of the neural signal, and phs is the phase of the speech envelope in band φ at time t. To equate the number of time points entering the analysis for slow and regular speech, slow speech utterances were split into three equal parts before CAC calculation, and resultant CAC values were averaged. CAC was averaged across sensors for each hemisphere.

A priori, we hypothesized that CAC would differ between conditions in the frequency bands corresponding to the average frequency of peakRate events in each rate condition (regular: 5.7 Hz; slow: 1.9 Hz, Figure 2D). We tested this hypothesis using a 3-way repeated-measures ANOVA with factors frequency band (high/low), factor speech rate (slow/regular), and hemisphere (left/right).

To test for further differences in each frequency band, we assessed the effect of speech rate and hemisphere onto CAC using a two-way repeated-measures ANOVA with factor speech rate (slow/regular) and hemisphere (left/right). Significance in this analysis was Bonferroni-corrected for multiple comparisons across bands.

Inter-event phase coherence (IEPC)

From the phase time series for each frequency band, epochs with 1000 ms duration (-500 ms to + 500 ms relative to event onset) were extracted for peakRate events. Inter-event phase coherence (IEPC) (Lachaux, Rodriguez, Martinerie, & Varela, 1999) was computed for each condition, for segmentations around peakRate and peakEnv events. IEPC was calculated using the following formula:

$$IEPC(\varphi, t) = \frac{1}{N}\left|\sum_{k=1}^{N} \exp\left(i * ph_k(\varphi, t)\right)\right|$$

where N is the number of events, ph is the phase of the neural signal in trial k, for the frequency band φ and timepoint t. IEPC were averaged across subjects and selected sensors. The resulting time x frequency IEPC was compared with the pre-event baseline using 2-D cluster-based permutation t-tests (Maris & Oostenveld, 2007) with 3000 permutations, a peak threshold of $p < 0.05$, and a cluster threshold of $p < 0.01$. Baseline IEPC in each frequency band was calculated as the average IEPC between -400 ms

to -100 ms relative to event onset. Differences in IEPC between alignments to peakRate and peakEnv were assessed using 2-D cluster-based permutation *t*-tests.

Theta-band IEPC for peakRate events with large inter-event distance

To assess the temporal extent of theta-band IEPC between peakRate events, we analyzed IEPC following peakRate events with a temporal interval of more than 500 ms to adjacent peakRate events. For each frequency band between 4 and 8 Hz, we computed the average IEPC in time windows corresponding to a single oscillatory cycle at the band center frequency, for a single cycle before and up to 3 cycles post-event. Post-event, the maximal number of cycles within the 500 ms window was included for each band. Bands were split into two sets, according to the number of post-event cycles in the analysis (2 cycles: 4Hz, 4.5Hz, 5Hz, 5.5Hz; 3 cycles: 6.1Hz, 6.7Hz, 7.4Hz). The IEPC baseline was calculated using circular permutation of peakRate time series across sentences. In each frequency band, baseline IEPC was averaged across the time window of a single oscillatory cycle. For each frequency range, IEPC in neighboring oscillatory cycles was compared using 2-way repeated-measures ANOVAs with factors frequency band x cycle relative to the event. Then, IEPC in the pre- and end post-event cycles was compared with baseline using 2-way repeated-measures ANOVAs with factors frequency band x cycle time window relative to event time. Pairwise comparisons between cycles were Holm-corrected for multiple comparisons. Pairwise comparisons to the permutation-based baseline were Bonferroni-corrected for multiple comparisons across frequency bands. Note that IEPC in the cycle before peakRate events and the last cycle following the peakRate event are not necessarily identical: increased IEPC in the cycle before peakRate event would indicate predictive processing based on the preceding peakRate event (Doelling et al., 2019), whereas increased IEPC in the last extracted cycle would be an indicator of continuous entrainment, even if it does not capture the timing of the next event. Those two would be identical for an isochronous stimulus sequence but diverge otherwise, for example, in natural speech.

Effect of peakRate magnitude on IEPC

In each rate condition, peakRate events were split into five quantiles, and IEPC was separately calculated within each quantile. Then, we extracted the average IEPC in the theta band (4 – 8 Hz) across all the time points for one cycle of the given frequency band after the event. IEPC in each quantile was compared using 2-way ANOVA with factors quantile and speech rate (regular speech, slow speech).

Data and code availability

All custom-written analysis and stimulus presentation code are available upon request from the corresponding authors (SN and EC). All data are available upon request from the corresponding authors.

Author contributions: Y.O and E.F.C conceived the study, Y.O and K.K designed the experiments and analyzed the data. K.K., C.C., and A.F collected and preprocessed the data. K.K. and Y.O. wrote the manuscript, K.K., Y.O, E.F.C, and S.N revised the manuscript.

References

Abrams, D. A., Nicol, T., Zecker, S., & Kraus, N. (2008). Right-Hemisphere Auditory Cortex Is Dominant for Coding Syllable Patterns in Speech. *J Neurosci*, *28*(15), 3958–3965. https://doi.org/10.1523/JNEUROSCI.0187-08.2008

Ahissar, E., Nagarajan, S. S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(23), 13367–13372. https://doi.org/10.1073/pnas.201400998

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018a). Cortical entrainment: What we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience*, *0*(0), 1–13. https://doi.org/10.1080/23273798.2018.1518534

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018b). Cortical Tracking of Global and Local Variations of Speech Rhythm during Connected Natural Speech Perception. *Journal of Cognitive Neuroscience*, *26*(3), 1–16. https://doi.org/10.1162/jocn_a_01295

Aubanel, V., Davis, C., & Kim, J. (2016). Exploring the Role of Brain Oscillations in Speech Perception in Noise: Intelligibility of Isochronously Retimed Speech. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00430

Biermann, S., & Heil, P. (2000). Parallels Between Timing of Onset Responses of Single Neurons in Cat and of Evoked Magnetic Fields in Human Auditory Cortex. *Journal of Neurophysiology*, *84*(5), 2426–2439.

Boersma, P., & Weenik, D. (2019). Praat: Doing phonetics by computer [Computer program]. Version 6.1.08, retrieved 5 December 2019 from http://www.praat.org/. *Amsterdam: University of Amsterdam*.

Bosker, H. R., & Kösem, A. (2017). An Entrained Rhythm's Frequency, Not Phase, Influences Temporal Sampling of Speech. *Interspeech 2017*, 2416–2420. https://doi.org/10.21437/Interspeech.2017-73

Breska, A., & Deouell, L. Y. (2017). Neural mechanisms of rhythm-based temporal prediction: Delta phase-locking reflects temporal predictability but not rhythmic entrainment. *PLOS Biology*, *15*(2), e2001665. https://doi.org/10.1371/journal.pbio.2001665

Breska, A., & Ivry, R. B. (2018). Double dissociation of single-interval and rhythmic temporal prediction in cerebellar degeneration and Parkinson's disease. *Proceedings of the National Academy of Sciences*, *115*(48), 12283–12288. https://doi.org/10.1073/pnas.1810596115

Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic Context Enhances the Early Auditory Encoding of Natural Speech. *The Journal of Neuroscience*, 0584–19. https://doi.org/10.1523/JNEUROSCI.0584-19.2019

Cai, C., Xu, J., Velmurugan, J., Knowlton, R., Sekihara, K., Nagarajan, S. S., & Kirsch, H. (2019). Evaluation of a dual signal subspace projection algorithm in magnetoencephalographic recordings from patients with intractable epilepsy and vagus nerve stimulators. *Neuroimage*, *188*, 161–170. https://doi.org/10.1016/j.neuroimage.2018.11.025

Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., & Gross, J. (2011). Steady-State Visual Evoked Potentials Can Be Explained by Temporal Superposition of Transient Event-Related Responses. *PLoS ONE*, *6*(1), e14543. https://doi.org/10.1371/journal.pone.0014543

Chait, M., Simon, J. Z., & Poeppel, D. (2004). Auditory M50 and M100 responses to broadband noise: Functional implications: *NeuroReport*, *15*(16), 2455–2458. https://doi.org/10.1097/00001756-200411150-00004

Daube, C., Ince, R. A. A., & Gross, J. (2019). Simple Acoustic Features Can Explain Phoneme-Based Predictions of Cortical Responses to Speech. *Current Biology*, 1–14. https://doi.org/10.1016/j.cub.2019.04.067

Di Liberto, G. M., O'Sullivan, J. A., & Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, *25*(19), 2457–2465. https://doi.org/10.1016/j.cub.2015.08.030

Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2015). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*(1), 158–164. https://doi.org/10.1038/nn.4186

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, *85*(0 2), 761–768. https://doi.org/10.1016/j.neuroimage.2013.06.035

Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., & Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *Proceedings of the National Academy of Sciences*, *116*(20), 201816414. https://doi.org/10.1073/pnas.1816414116

Donhauser, P. W., & Baillet, S. (2020). Two Distinct Neural Timescales for Predictive Speech Processing. *Neuron*, *105*(2), 385-393.e9. https://doi.org/10.1016/j.neuron.2019.10.019

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America*, *95*(2), 1053–1064. https://doi.org/10.1121/1.408467

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, *95*(5), 2670–2680. https://doi.org/10.1121/1.409836

Friston, K. J., Sajid, N., Quiroga-Martinez, D. R., Parr, T., Price, C. J., & Holmes, E. (2020). Active listening. *Hearing Research*, 107998. https://doi.org/10.1016/j.heares.2020.107998

Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, *2*(JUN), 1–13. https://doi.org/10.3389/fpsyg.2011.00130

Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in Psychology*, *4*(MAR), 138. https://doi.org/10.3389/fpsyg.2013.00138

Ghitza, O., & Greenberg, S. (2009). On the Possible Role of Brain Rhythms in Speech Perception: Intelligibility of Time-Compressed Speech with Periodic and Aperiodic Insertions of Silence. *Phonetica*, *66*(1–2), 113–126. https://doi.org/10.1159/000208934

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech Rhythms and Multiplexed Oscillatory Sensory Coding in the Human Brain. *PLoS Biology*, *11*(12), e1001752. https://doi.org/10.1371/journal.pbio.1001752

Gwilliams, L. (2019). Hierarchical oscillators in speech comprehension: A commentary on Meyer Sun and Martin 2019. *Language, Cognition and Neuroscience*. https://doi.org/10.1080/23273798.2020.1740749

Haegens, S., & Zion Golumbic, E. (2018). Rhythmic facilitation of sensory processing: A critical review. *Neuroscience & Biobehavioral Reviews*, *86*, 150–165. https://doi.org/10.1016/j.neubiorev.2017.12.002

Hamilton, L. S., Edwards, E., & Chang, E. F. (2018). A Spatial Map of Onset and Sustained Responses to Speech in the Human Superior Temporal Gyrus. *Current Biology*, *28*(12), 1860-1871.e4. https://doi.org/10.1016/j.cub.2018.04.033

Harsin. (1993). Acoustics of perceptual centers in speech. *The Journal of the Acoustical Society of America*, *94*(3), 1864–1864. https://doi.org/10.1121/1.407646

Helfrich, R. F., Breska, A., & Knight, R. T. (2019). Neural entrainment and network resonance in support of top-down guided attention. *Current Opinion in Psychology*, *29*, 82–89. https://doi.org/10.1016/j.copsyc.2018.12.016

Hertrich, I., Dietrich, S., Trouvain, J., Moos, A., & Ackermann, H. (2012). Magnetic brain activity phase-locked to the envelope, the syllable onsets, and the fundamental frequency of a perceived speech signal. *Psychophysiology*, *49*(3), 322–334. https://doi.org/10.1111/j.1469-8986.2011.01314.x

Hovsepyan, S., Olasagasti, I., & Giraud, A.-L. (2020). Combining predictive coding and neural oscillations enables online syllable recognition in natural speech. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16956-5

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural Entrainment Determines the Words We Hear. *Current Biology*, *28*(18), 2867-2875.e3. https://doi.org/10.1016/j.cub.2018.07.023

Lalor, E. C., & Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, *31*(1), 189–193. https://doi.org/10.1111/j.1460-9568.2009.07055.x

Luo, H., & Poeppel, D. (2007). Phase Patterns of Neuronal Responses Reliably Discriminate Speech in Human Auditory Cortex. *Neuron*, *54*(6), 1001–1010. https://doi.org/10.1016/j.neuron.2007.06.004

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., & Friederici, A. D. (2017). Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations. *Cerebral Cortex*, *27*(9), 4293–4302. https://doi.org/10.1093/cercor/bhw228

Meyer, L., Sun, Y., & Martin, A. E. (2019). Synchronous, but not entrained: Exogenous and endogenous cortical rhythms of speech and language processing. *Language, Cognition and Neuroscience*, 1–11. https://doi.org/10.1080/23273798.2019.1693050

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5–6), 453–467. https://doi.org/10.1016/0167-6393(90)90021-Z

Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Howard, M. A., & Brugge, J. F. (2009). Temporal Envelope of Time-Compressed Speech Represented in the Human Auditory Cortex. *Journal of Neuroscience*, *29*(49), 15564–15574. https://doi.org/10.1523/JNEUROSCI.3065-09.2009

Obleser, J., & Kayser, C. (2019). Neural Entrainment and Attentional Selection in the Listening Brain. *Trends Cogn Sci*, *23*(11), 913–926. https://doi.org/10.1016/j.tics.2019.08.004

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science Advances*. https://doi.org/10.1101/388280

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Ostendorf, M., Price, P. J., & Shattuck-Hufnagel, S. (1995). The Boston University radio news corpus. *Linguistic Data Consortium*, 1–19.

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., Slaney, M., Shamma, S. A., & Lalor, E. C. (2015). Attentional Selection in a Cocktail Party Environment Can Be Decoded from Single-Trial EEG. *Cerebral Cortex*, *25*(7), 1697–1706. https://doi.org/10.1093/cercor/bht355

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*(SEP), 1–17. https://doi.org/10.3389/fpsyg.2012.00320

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378–1387. https://doi.org/10.1093/cercor/bhs118

Pefkou, M., Arnal, L. H., Fontolan, L., & Giraud, A.-L. (2017). θ-Band and β-Band Neural Activity Reflects Independent Syllable Tracking and Comprehension of Time-Compressed Speech. *The Journal of Neuroscience*, *37*(33), 7930–7938. https://doi.org/10.1523/JNEUROSCI.2882-16.2017

Reinisch, E. (2016). Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics*, *37*(6), 1397–1415. https://doi.org/10.1017/S0142716415000612

Ruhnau, P., Rufener, K. S., Heinze, H.-J., & Zaehle, T. (2020). Pulsed transcranial electric brain stimulation enhances speech comprehension. *Brain Stimulation*. https://doi.org/10.1016/j.brs.2020.07.011

Scott, S. K. (1998). The point of P-centres. *Psychological Research Psychologische Forschung*, *61*(1), 4–11. https://doi.org/10.1007/PL00008162

Sekihara, K., Kawabata, Y., Ushio, S., Sumiya, S., Kawabata, S., Adachi, Y., & Nagarajan, S. S. (2016). Dual signal subspace projection (DSSP): A novel algorithm for removing large interference in biomagnetic measurements. *J Neural Eng*, *13*(3), 036007. https://doi.org/10.1088/1741-2560/13/3/036007

Sohoglu, E., & Davis, M. H. (2016). Perceptual learning of degraded speech by minimizing prediction error. *Proceedings of the National Academy of Sciences*, *113*(12), E1747–E1756. https://doi.org/10.1073/pnas.1523266113

Teichmann, M., Kas, A., Boutet, C., Ferrieux, S., Nogues, M., Samri, D., Rogan, C., Dormont, D., Dubois, B., & Migliaccio, R. (2013). Deciphering logopenic primary progressive aphasia: A clinical, imaging and biomarker investigation. *Brain*, *136*(Pt 11), 3474–3488. https://doi.org/10.1093/brain/awt266

ten Oever, S., & Sack, A. T. (2015). Oscillatory phase shapes syllable perception. *Proceedings of the National Academy of Sciences*, *112*(52), 15833–15837. https://doi.org/10.1073/pnas.1517519112

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics*, *67*(6), 939–950. https://doi.org/10.3758/BF03193621

Zoefel, B. (2018). Speech Entrainment: Rhythmic Predictions Carried by Neural Oscillations. *Current Biology*, *28*(18), R1102–R1104. https://doi.org/10.1016/j.cub.2018.07.048

Zoefel, B., Allard, I., Anil, M., & Davis, M. H. (2019). Perception of Rhythmic Speech Is Modulated by Focal Bilateral Transcranial Alternating Current Stimulation. *Journal of Cognitive Neuroscience*, *32*(2), 226–240. https://doi.org/10.1162/jocn_a_01490

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech. *Current Biology*, *28*(3), 401-408.e5. https://doi.org/10.1016/j.cub.2017.11.071

Zoefel, B., ten Oever, S., & Sack, A. T. (2018). The Involvement of Endogenous Neural Oscillations in the Processing of Rhythmic Input: More Than a Regular Repetition of Evoked Neural Responses. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00095

Supplemental Online Materials

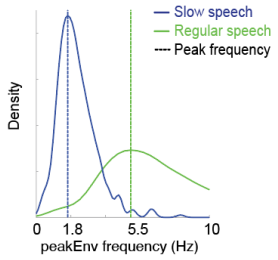

Figure S1. Frequency of peakEnv events. Frequency of peakEnv events in slow speech (blue) and regular speech (green). Dashed vertical lines mark average event frequency in slow (blue, 1.8 Hz) and regular speech (green, 5.5 Hz).

Evoked low-frequency power following peakRate events

Evoked increase in power is a marker of evoked neural responses and is used to distinguish between evoked responses and oscillatory activity. In addition to calculating the ERP to peakRate events, we thus also tested whether band-passed power would increase after peakRate events. However, we found no significant effects of peakRate on evoked power in theta or delta bands ($p > 0.05$, cluster-based permutation test, data not shown). Our hypothesis that this was due to higher susceptibility of power measures to noise was confirmed in a simulation of the evoked response model (see below).

Simulation of the evoked response model

The predictions of an oscillatory entrainment theory of speech envelope tracking have been described in detail in the literature (Myers, Lense, & Gordon, 2019; Obleser & Kayser, 2019), with the major prediction that phase locking continues after phase reset for multiple oscillatory cycles. In contrast, the phase-locking pattern to a continuous stimulus under an evoked-response model is less well described. We thus simulated the patterns of Inter-Event-Phase-Coherence under the evoked response model. To this end, we simulated the time series of neural responses to our speech stimuli. Specifically, we simulated neural responses for the 20 sensors used in all analyses, modeling the evoked response to peakRate events after the evoked response to utterance onsets. To create a simulated response time series, the evoked response was convolved with a discrete time series marking time and magnitude of peakRate events. Finally, autocorrelated noise was added to the response time series. We then analyzed it with the same pipeline as the experimental data (see Methods section of the main manuscript). Reported results are averaged across all 20 sensors simulations.

Simulation results

IEPC after peakRate event. First, we used the simulation to describe the pattern of IEPC expected with the alignment of neural data to peakRate and to peakEnv events, under the assumption that peakRate is the acoustic event that drives the phase reset. The simulation of an evoked response model reproduced the main patterns observed in our MEG data:

1) Stronger IEPC following peakRate than following peakEnv events in the theta band. This effect appeared in regular and in slow speech and was stronger in slow speech (Figure S2 C-D).

2) For slow speech, increase in IEPC around peakRate and peakEnv events in the delta band around the frequency of events in this condition (1.9 Hz, Figure S2D).

The simulation results showed that the observed pattern could be entirely explained by an evoked response model, including the increase in IEPC in the delta band. They also supported our interpretation that the IEPC increase was driven by peakRate and not by peakEnv events.

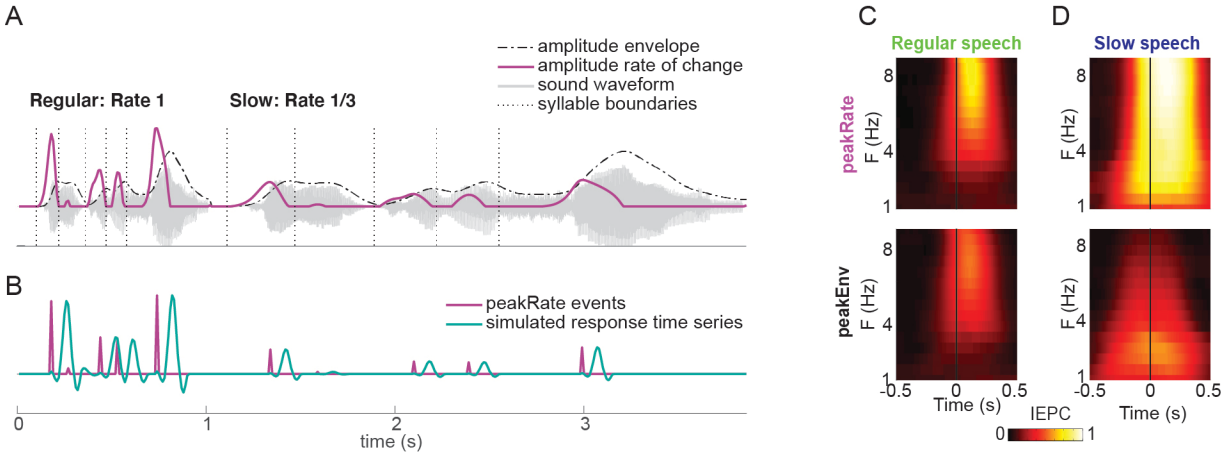

Figure S2. IEPC in a simulation of the evoked response model. A. Simulated neural time course for the example utterance in regular and slow speech, for the same excerpt, as shown in Figure 1A. The same utterance is depicted at a regular rate (left) and slow (right) rate. **B.** Simulated neural response to the example utterance at regular (left) and slow (right) speech rate. **C.** IEPC in regular speech, with alignment to peakRate events (upper) and peakEnv events (lower). **D.** IEPC in slow speech, as in C.

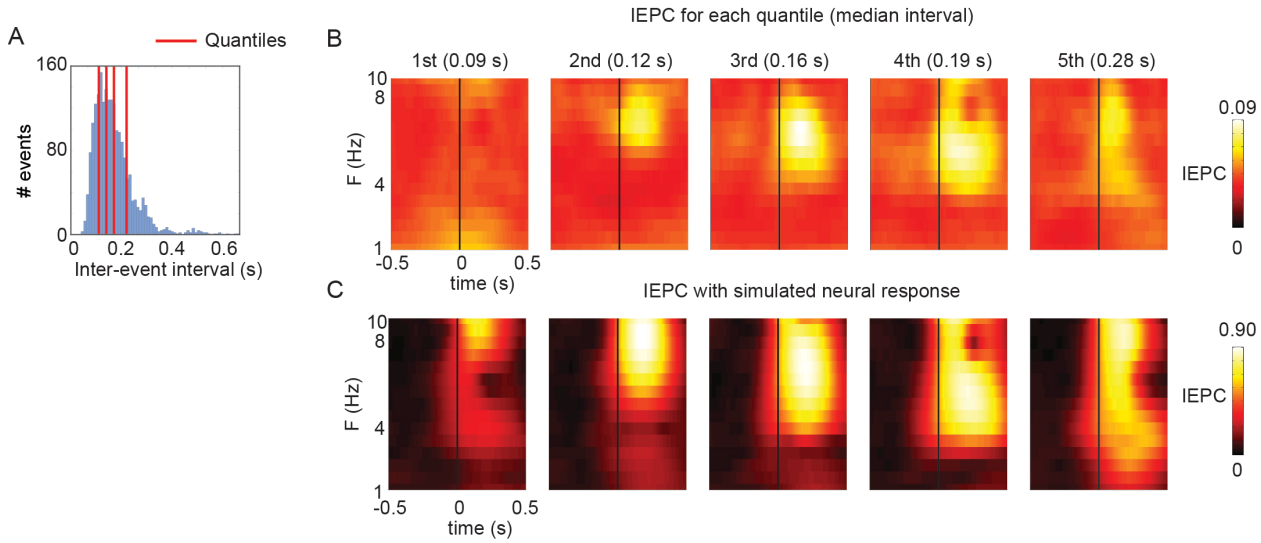

Figure S3. IEPC frequency range shifts with an increase in the inter-peakRate interval in regular speech. A. Histogram of inter-event interval in regular speech, with quantile boundaries marked in red. **B**. IEPC in each inter-event interval quantile. **C**. Same as B for the simulated neural response.

Effect of noise

Finally, as we did not observe an increase in power in theta or delta bands following peakRate events, we hypothesized that this might reflect the high susceptibility of power increases to noise. To assess the effect of noise onto power and phase measures, we tested the evoked response model at noise levels of 1 to 10 relative to response magnitude. We evaluated the effect of noise onto power and IEPC in the theta band (4-8Hz) in the window of a single cycle for a given frequency band after event onset. The effects of noise on power and IEPC were compared using two-sided paired *t*-tests at each noise level (n = 20 simulated responses), with Bonferroni correction for the number of comparisons.

We found continuously large effect sizes for IEPC even at high levels of noise, whereas the effect size for power deteriorated rapidly with the addition of noise.

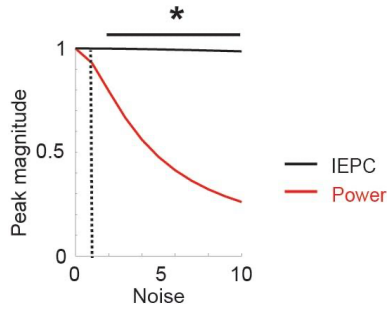

Figure S4. Effect of noise level on IEPC (black) and power (red) after peakRate events in theta band (4-8Hz) for regular speech. * $p < 0.01$

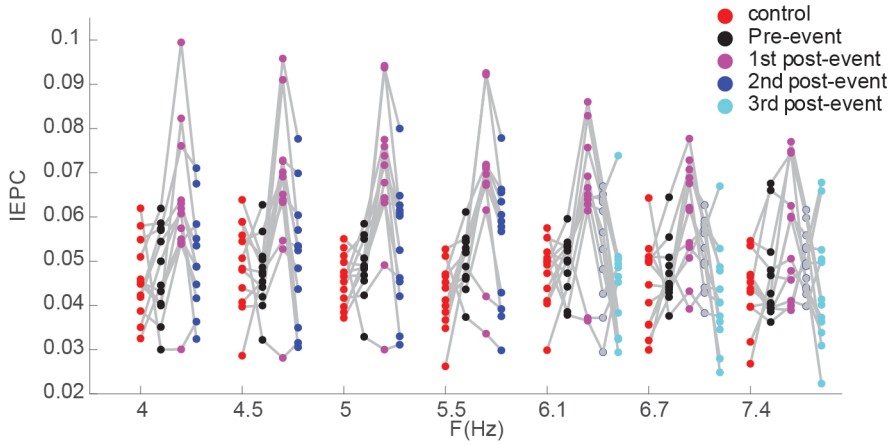

Figure S5. The duration of MEG phase locking is consistent across subjects. Single-subject data for Figure 3B. Average IEPC during baseline (red), pre-event (black), 1st post-event (pink), 2nd post-event (blue), and 3rd post-event cycle (light blue) in each frequency band. IEPC increased from pre-event to 1st post-event and decreased from 1st post-event to 2nd post-event across subjects.

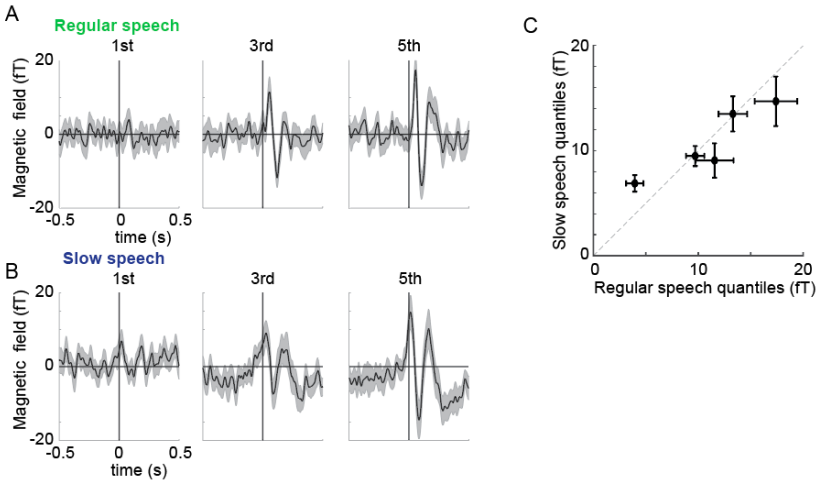

Figure S6. Normalization of peakRate evoked response for contextual speech rate.

A. Evoked response for peakRate events in 1st, 3rd, 5th peakRate magnitude quantile. Group-averaged ERPs, error bars are ± 1 SEM. **B**. Same as A for slow speech. **C.** Quantile-quantile plot of peak evoked response (mean (SEM)) in regular and slow speech. Peak evoked response quantile-quantile values are close to the diagonal, indicating the similar distribution of evoked response in regular and slow speech conditions.

1

2 **Table S1. Speech stimulus transcription**

Stimulus ID	Transcript
f1ajrlp1	Wanted: Chief Justice of the Massachusetts Supreme Court. In April, the S.J.C.'s current leader Edward Hennessy reaches the mandatory retirement age of seventy, and a successor is expected to be named in March. It may be the most important appointment Governor Michael Dukakis makes during the remainder of his administration and one of the toughest. As WBUR's Margo Melnicove reports, Hennessy will be a hard act to follow.
f1ajrlp2	In nineteen seventy-six, Democratic Governor Michael Dukakis fulfilled a campaign promise to de-politicize judicial appointments. He named Republican Edward Hennessy to head the State Supreme Judicial Court. For Hennessy, it was another step along a distinguished career that began as a trial lawyer and led to an appointment as associate Supreme Court Justice in nineteen seventy- one. That year Thomas Maffy, now president of the Massachusetts Bar Association, was Hennessy's law clerk.
f1ajrlp3	The author of more than eight hundred State Supreme Court opinions, Hennessy is widely respected for his legal scholarship and his administrative abilities. Admirers give Hennessy much of the credit for sweeping court reform that began a decade ago, and for last year's legislative approval of thirty-five new judgeships and three hundred million dollars to restore crumbling court houses. Despite the state's massive budget deficit, Hennessy recently urged colleagues in the bar association not to retreat from these hard won gains.
f1ajrlp4	Hennessy is the S.J.C.'s thirty-second chief justice. Holding the court system on the course he has set and plotting it's future agenda won't be an easy job for his successor.
f1ajrlp5	Attorney Haskel Kassler chairs the Judicial Nominating Council, eighteen attorneys and laypeople charged with screening applicants for vacancies on the bench. Usually the J.N.C. refers three nominees to the Governor. His top choice is rated by bar associations and grilled by the Governor's executive council. Kassler says, unlike the Federal Supreme Court, there's no litmus test on particular issues that Massachusetts high court nominees must pass.

f1ajrlp6	All but one of the Chief Justices since eighteen ninety- nine, when Oliver Wendell Holmes was appointed, came from the ranks of S.J.C. associate justices. If he sticks with tradition, Dukakis is likely to elevate one of his appointees to chief. That means Paul Leocos or Ruth Abrams, the only woman on the court. Another possible choice is Herbert Wilkins, a Governor Sargent appointee, and next to Hennessy, the court's most senior member. The other three associate justices were put on the bench by Governor Edward King. And many lawyers say that despite Dukakis' promise to keep the judiciary above the political fray, it's unlikely Dukakis will choose a King appointee to run the state's highest court. For WBUR, I'm Margo Melnicove.
m2brrlp1	Massachusetts may now have the toughest drunken driving law in the nation, thanks to the Safe Roads Act that became law this week. The legislation came complete with a message from the governor to those seeking a little too much comfort and joy this holiday season and beyond.
m2brrlp2	Under the new law the consequences of driving drunk will be swift and severe. Under the old, already tough law, first time offenders used to lose their license for thirty days. Well now they could lose it for as many as ninety. The prison sentence for a second time offender has now doubled. The law is also tougher for vehicular homicide, and it cracks down on minors who drink and drive. State trooper Joseph Holly says this tough new law will mean fewer drunken drivers.
m2brrlp3	Well that's what supporters of the Safe Roads Act are hoping anyway, but will it work? Massachusetts got tough on drunken drivers years ago. What's going to make the difference this time around? Ralph Henksen is chief of the social and behavioral science section at the Boston University school of public health. He says research into what deters the drunken driver shows that often those who drink and drive are risk takers. Even when they're sober they're more likely to run red lights and shun the seat belt. If they think they can drink and drive and get away with it, they'll try.
m2brrlp4	And in fact the state is not planning on putting more police on the road. State officials say there will be the same levels of enforcement. B.U.'s professor Henksen believes the key to the new law's success is its enforcement.
m2brrlp5	The state Safe Roads Act is only a few days old, but it's already facing a challenge. The man who directed the successful effort last election to get the mandatory seat belt law stripped from the law books is now vowing to try to repeal part of this law. Lynn activist Chip Ford has yet to give a name to his committee, but

	he's already set its goal: to repeal what is the most controversial clause in the Safe Roads Act, the per se clause, which allows someone to be punished for drinking and driving before they've even been tried in court. Under the per se part of this new law, if you're arrested, take a breath test, and your blood content level registers point one oh or higher, you can automatically lose your license for ninety days. And that's just at your arraignment. Then you still got to face DWI charges at a trial. State officials say about a quarter of all those arrested for drunken driving are repeat offenders. They say the per se clause will get those people off the road sooner. Chip Ford says it may indeed get them off the road, but it will do it at the expense of their civil rights.
m2brrlp6	Ford and other opponents of the per se clause say it relies exclusively on the breath test, and they say breathalizers are simply unreliable. John Tarantino is a Providence attorney who specializes in drunken driving cases.
m2brrlp7	Tarantino says because of people's physiological differences, the breathalizer test will always be inaccurate on about fourteen percent of the population. Those people will register as having more alcohol in their blood than they actually do. But state officials aren't convinced that's a problem. While they don't say breathalizers are one hundred percent accurate, they do say they are extremely reliable, and if they err at all it's on the conservative side, the side favorable to the driver. As for those who say that taking away someone's license based on the breathalizer results is tantamount to trial by machine, state officials say that simply isn't so. Losing one's license under the per se clause is not a criminal offense. You won't get a criminal record for it. It is a civil infraction and one that they believe will be a key deterrent to the drunken driver. If the per se clause is tested in court, the state says it will pass constitutional muster. And if opponents gather enough signatures to make it a ballot question in nineteen eighty-eight, state officials predict voters will keep it intact.

Table S2. Comprehension questions

Stimulus ID	Question	Answer
f1ajrlp1	What is the mandatory retirement age for chief justice?	**a: 70** b: 30
f1ajrlp2	What did Michael Dukakis promise in his campaign?	a: To legalize medical Marijuana, **b:**

		To de-politicize judicial appointments.
f1ajrlp3	How many laws were passed under Hennessy?	**a: more than 800** b: more than 8000
f1ajrlp4	How many chief justices were there before Hennessy?	**a: 31** b: 10
f1ajrlp5	How many nominees for the court are passed to the governor?	a: 5 **b: 3**
f1ajrlp6	How many chief Justices so far were not from the ranks of SJC associate justices?	a: 5 **b: 1**
m2brrlp1	What is the name of the new law in Massachusetts?	**a: Safe Roads Act**, b: Seatbelt act
m2brrlp2	For how long do first-time offenders lose their license under the new law?	a: 10 days, **b: 90 days**
m2brrlp3	Research shows that those that drink, and drive are	**a: risk-takers**; b: not aware of the risks
m2brrlp4	Are there more police on the road after the new law?	a: Yes, **b: No**
m2brrlp5	How many of those arrested for drunk driving are repeat offenders?	**a: a quarter** b: More than half
m2brrlp6	Is the breathalyzer test reliable?	a: yes, **b: no**
m2brrlp7	Does one get a criminal record for losing their driver's license?	a: yes, **b: no**

1