Low-frequency neural tracking of speech envelope reflects evoked responses to acoustic edges

Katsuaki Kojima*[1], 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. Envelope tracking may reflect either continuous entrainment of endogenous low-frequency oscillations to the speech envelope or the combination of a series of evoked responses to acoustic landmarks within the envelope. To distinguish between these two accounts for envelope tracking, and to identify the acoustic features driving it, we recorded magnetoencephalography (MEG) while participants listened to natural and slowed speech. First, we found that acoustic edges in speech amplitude drove evoked responses and induced phase-resetting in the theta-delta band, supporting the evoked response account. In line with this, phase locking in the theta band was transient and independent of the modulation rate of the speech envelope but matched the temporal extent of the evoked response. Further analysis showed that the magnitude of theta phase-locking reflected the slope of amplitude increases. However, although amplitude increases were more gradual in slowed speech, the magnitude of phase-locking did not differ between slow and regular speech, reflecting a normalization for speech rate. Taken together, our results are in line with predictions of the evoked response account of speech envelope tracking and support acoustic edge detection as a flexible mechanism for tracking of temporal dynamics in speech.

Introduction

Speech comprehension is essential to human communication. A central source of information in speech is its amplitude envelope, which captures prominent fluctuations in the overall intensity of speech across spectral bands (Figure 1A). It is dominated by fluctuations in the range of ~ 1 – 10 Hz, the removal of which from the speech signal severely impairs intelligibility (Drullman et al., 1994a, 1994b). A multitude of studies has shown a consistent relationship between the phase of band-limited 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 reflects a necessary stage in speech comprehension (Abrams et al., 2008; Peelle et al., 2013).

While the phenomenon of cortical speech envelope tracking is well-documented and robust, the underlying neural computations are actively debated (Zoefel, Ten Oever, & Sack, 2018). According to one framework, the oscillatory entrainment theory of speech comprehension (Giraud & Poeppel, 2012), 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, reflected in increased cerebro-acoustic coherence (CAC) in the relevant frequencies (Ding et al., 2015; Zoefel, 2018). An alternative framework, referred to here as evoked response theory, suggests that neural tracking of the speech envelope may reflect the combination of evoked responses to discrete acoustic landmark events in the speech signal, such as amplitude peaks or acoustic edges (i.e., peaks in the rate of acoustic amplitude change (Doelling et al., 2014; Oganian & Chang, 2019), without contributions by endogenous oscillatory activity. These two theories are not necessarily mutually exclusive. For instance, an acoustic landmark may trigger an evoked response as well as an oscillatory phase reset, and entrainment may arise from phase adjustment driven by an acoustic feature in the speech envelope (Doelling et al., 2014; Ghitza, 2011). However, their distinct contributions to envelope tracking remain under debate (Gwilliams, 2019; Obleser & Kayser, 2019) . These two competing theories have major implications for mechanistic theories of speech comprehension, as they make opposing predictions regarding two aspects of neural responses to continuous speech.

Here, using magnetoencephalography (MEG) recordings, while participants listened to continuous natural speech, and equally intelligible speech slowed to 1/3 of its original rate (Figure 1A), we test opposing predictions of these two theories. First, whether an evoked response to acoustic features accompanies the phase-reset is left unspecified under the oscillatory entrainment theory, whereas the evoked response theory predicts the existence of an evoked response to a phase-resetting acoustic feature, in addition to a temporally transient phase-reset. Second, the oscillatory entrainment theory predicts that

any phase-reset should last for several oscillatory cycles, as long as another phase-resetting event has not occurred, functioning as a predictive mechanism in continuous speech processing (Rimmele et al., 2018). In contrast, the evoked response theory predicts that phase alignment reflects the evoked response, and thus the period of high phase consistency will only extend for the duration of the evoked response. Therefore, a critical test between distinguishing the two alternatives requires, first, identifying the critical phase-resetting landmark, and then examining the temporal extent of phase-locking in the absence of exogenous phase-resets.

We build upon recent intracranial electrocorticography (ECoG) recordings that showed that local neural populations on the superior temporal gyrus (STG) selectively respond to acoustic edges in the speech envelope and that envelope tracking by broadband high gamma amplitude was explained by transient power increases following acoustic edges (Oganian & Chang, 2019). Results from this study suggested that onset edges may be the primary acoustic feature driving speech envelope encoding. However, as this study focused only on high gamma power in local neuronal populations, the temporal extent of low-frequency envelope phase-locking and whether it is driven by acoustic edges in M/EEG, which reflects summed neural activity across many more neural sources, remains unknown.

First, we test the hypothesis that increases in low-frequency phase-locking in the neural signal are driven by peaks in the rate of amplitude change (i.e., acoustic edges, which we refer to as peakRate events, Figure 1A purple arrows and Figure 1B top panel). We also examine an alternative hypothesis that MEG phase-locking may reflect evoked responses to local amplitude peaks in the acoustic envelope (peakEnv, Figure 1A black arrows, Figure 1B middle panel). The use of slow speech allowed us to temporally decorrelate different acoustic landmarks in time (Figure 1 A, B, right side). Therefore, specifically in slow speech, we predicted that phase-locking and evoked responses across occurrences of a landmark event would be larger with alignment to the landmark that drives envelope tracking (Figure 1 right column). Second, we test whether the temporal dynamics of envelope tracking by low-frequency phase would be accounted for predictions of the oscillatory entrainment theory or the evoked response theory. Examination of slowed speech, where the time windows between consecutive acoustic landmarks were sufficiently long, would reveal if multiple cycles of phase-locking after a phase reset are observed, as predicted by the oscillatory entrainment theory (Figure 1B bottom right). Finally, to further characterize the adjustment of speech envelope tracking to different speech rates, we investigated how MEG responses to the speech envelope change with the slowing of speech. In particular, as the slope at acoustic edges scales down with slowing of speech (Figure 1B left vs. right), the comparison between slow and regular speech allowed to test for speech rate normalization of responses to acoustic edges. If acoustic edges are an efficient cue to speech perception across speech rates, flexible envelope tracking would require a re-adjustment of neural responses to the range of acoustic edge slopes.

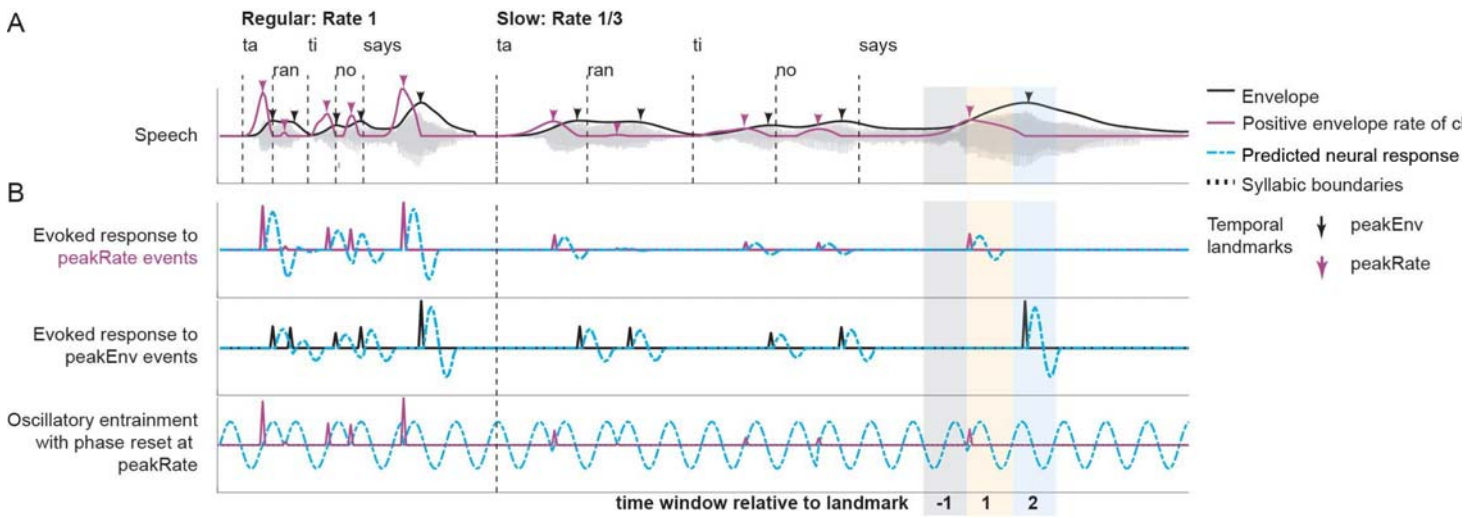

Figure 1. Stimuli and model predictions.

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. Black arrows mark local peaks in the envelope (peakEnv). Purple arrows mark acoustic edges, defined as local peaks in envelope rate of change (peakRate). **B.** Toy 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 selected from the Boston University Radio Speech Corpus (BURSC) (Ostendorf, Price, & Shattuck-Hufnagel, 1995) at the original speech rate and slowed to 1/3 of the original speech rate (see Figure 1A for stimulus excerpt). To characterize the temporal dynamics of the amplitude envelope of our speech stimulus, we calculated the frequency of occurrence of peakRate (Figure 2D) and peakEnv events (Figure S1). At the regular speech rate, the average frequency of peakRate and peakEnv landmark events across sentences was in the theta band at 5.7 Hz (SD = 2.9 Hz, Figure 2D), reflective of 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 peakRate and peakEnv landmarks was 1.9 Hz (SD = 1 Hz), shifting the peak in the envelope power spectrum to the delta band. 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; see Table S1, S2 for full stimulus transcript and list of comprehension questions).

To focus our analyses on sensors with auditory responses, 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).

Cerebro-acoustic phase coherence between speech envelope and MEG responses to speech

To confirm that cortical speech envelope tracking was present in our data, as expected from previous reports (Peelle & Davis, 2012), we calculated the narrow-band cerebro-acoustic phase coherence (CAC) between neural responses and the speech envelope in semi-logarithmically spaced bands between 0.3 and 10 Hz. Previous work found that CAC peaks at the frequency corresponding to the speech rate (Kösem et al., 2018). As the frequency of peakRate and peakEnv reflects the average speech rate in each condition, we predicted that CAC would peak at the frequency of these events in each condition (regular: 5.7 Hz, slow: 1.9 Hz). We tested this prediction with a 2-way repeated-measures ANOVA with factors frequency band (high, 5.7 Hz, and low, 1.9 Hz) X speech rate (regular, slow). CAC was differently affected by speech rate at low and high frequencies (interaction $F(1, 11) = 31.20$, $p < 0.001$, $\eta^2 = 0.30$). In line with our expectations, this interaction reflected a significantly higher CAC for regular than for slow speech in the band around 5.7 Hz ($t(11) = 5.6$, $p < 0.001$, $\eta^2 = 0.42$), and the opposite pattern in the low frequency band around 1.9 Hz ($t(11) = 3.4$, $p = 0.006$, $\eta^2 = 0.29$, Figure 2E). Moreover, CAC was overall higher at lower frequencies ($F(1, 11) = 16.44$, $p < 0.001$, $\eta^2 = 0.39$), probably reflecting the 1/f shape of MEG frequency spectrum. Bonferroni-corrected analyses of the effect of rate in single frequency bands showed no other significant effects. Overall, this result replicates previous findings of cortical speech envelope tracking in frequency bands corresponding to the modulation frequency of the speech envelope.

Evoked response to peakRate events

Next, we aimed to understand which of the two landmarks, peakRate or peakEnv, drives this effect. As a first step, we tested whether an evoked response to peakRate or peakEnv events would be evident in neural activity during listening to natural speech. We calculated the average evoked response (1-40Hz) across all mid-sentence occurrences of peakRate and peakEnv. We observed a transient evoked response of larger magnitude (two-sided paired t-test $t(11) = 5.9$, $p < .001$; Figure 2C) when neural data were aligned to peakRate events than to peakEnv events. Moreover, with alignment to peakEnv events only, the evoked response diverged from zero before the event time, suggesting that not peakEnv but an earlier event drove the response (first time point of significance: -12.5 ms for peakEnv, +50 ms for peakRate, based on cluster-based permutation test against 0). This showed that similar to local neural processing in

STG (Oganian & Chang, 2019), peakRate events, but not peakEnv events, induce evoked responses in MEG.

Low-frequency phase-locking to landmark events

Having established that the evoked response is locked to peakRate events, we hypothesized that it is also driving 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, by time point. IEPC thus enables tracking of the temporal dynamics of phase-locking. Using this metric, previous work found that speech onsets trigger increased phase-locking in the theta band (Gross et al., 2013). Here, this analysis aimed to test whether the increase in CAC across time is due to increased phase-locking following peakRate events. We predicted that IEPC would be stronger with an alignment to the landmark driving the neural response than with any other alternative alignment. In particular, we expected this difference to be most pronounced in the slow speech condition where peakRate and peakEnv events are more spread out in time.

In regular speech, we observed stronger IEPC in the theta band (4-8 Hz, Figure 2F) following peakRate events than with alignment to peakEnv events (Figure 2G, two-dimensional cluster-based permutation $p < 0.001$). In slow speech, IEPC to peakRate in the theta band remained pronounced and significantly stronger than IEPC to peakEnv, suggesting that this effect is independent of speech rate (Figures 2H, I). However, in this slowed condition, IEPC increase also extended into the delta band, to the frequency of peakRate events in slow speech, around 1.9 Hz. Directly comparing IEPC between alignments to peakRate and peakEnv revealed significantly stronger IEPC for alignment to peakRate across a broad frequency range in theta and delta bands, but no difference around the frequency of events in the slow condition (Figure 2I, cluster-based permutation $p < 0.001$).

Overall, these results suggest that theta band IEPC reflects local responses to peakRate, unaffected by speech rate. In contrast, with slowing, the second component in IEPC at the frequency of peakRate events became apparent (1.9 Hz, Figure 2H). This component was independent of event alignment. We thus hypothesized that this effect reflected the frequency of occurrence of peakRate events (and thus the frequency of evoked responses to peakRate events). To test whether such an effect can emerge after the band-pass filtering of a time series of evoked responses to single events (an evoked response model), we simulated the IEPC pattern under an evoked response model (Figure S2, see supplementary materials for simulation details). If our hypothesis was correct, model IEPC to our speech stimulus should follow the same pattern as the neural data, including similar levels of IEPC around 1.9 Hz in slow speech with alignment to peakRate and peakEnv.

The model simulation for slow speech indeed confirmed all predictions, showing an increase in theta-band IEPC for peakRate event and a delta-band increase in IEPC for peakRate and peakEnv alignments. We also hypothesized that if the frequency range of peak IEPC increase reflects the frequency of peakRate events, the IEPC frequency range in regular speech should vary depending on the time interval between neighboring peakRate events. This hypothesis was confirmed both by an analysis of the data and of the evoked response model simulation (Figure S3).

Collectively, this shows that evoked responses to peaks in the envelope rate of change (peakRate) were driving the low-frequency phase-reset in MEG during speech perception. Moreover, it showed that IEPC to peakRate events reflected local responses to individual peakRate events as well as their frequency of occurrence in the stimulus.

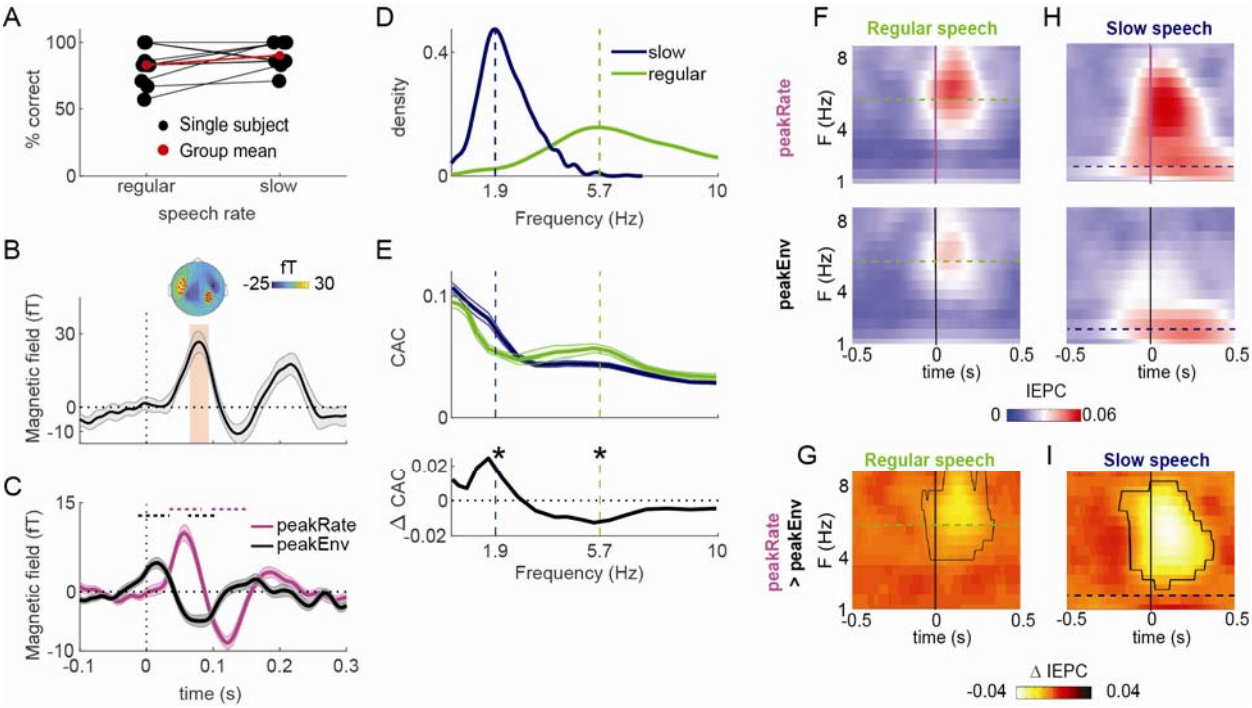

Figure 2. Phase locking during speech comprehension.

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-avereged evoked response aligned to peakRate and peakEnv events. Dotted line mark clusters with $p < 0.05$ with 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). Dashed vertical lines mark the average frequency of peakRate events in each condition, as shown in D. * $p < 0.01$ with 3-way ANOVA for speech rate x frequency band x hemisphere. Error bars are ± 1 SEM. across subjects. **F.** Inter-event phase coherence (IEPC) with alignment to peakRate events (upper) and peakEnv events (lower) in regular speech. The dashed line indicates the average frequency of the peakRate events. **G.** IEPC difference between alignment to peakRate and peakEnv events in regular speech. Outlined area: $p < 0.01$ with cluster-based permutation test. **H.** Same as F for slow speech. **I.** Same as G for slow speech. All panels depict responses averaged across subjects and sensors, shaded error areas (B, C, E) indicate standard error of the mean.

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

The above analyses indicated that acoustic edges (peakRate events) elicited an evoked response and phase-locking in the theta-band across speech rates. We aimed to further test whether the temporal dynamics of this phase-locking were more consistent with an evoked response or oscillatory entrainment. Qualitatively, if envelope tracking reflects evoked responses to peakRate events, the increase in IEPC should extend for the duration of the evoked response (250 ms). In contrast, if envelope tracking reflects an oscillatory phase-reset, we expected the increase in IEPC to be evident in the oscillatory cycles after the end of the evoked response. To test this, we focused on peakRate events that were at least 500 ms apart from neighboring peakRate events (n = 456). For these events, IEPC increases in the theta band (in which the average cycle length ranges between 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).

We extracted IEPC for each frequency band, for a single cycle preceding and up to three cycles after each peakRate event. As the continuous nature of the speech stimulus precluded us from calculating the baseline for IEPC in the absence of stimulation, we focused on the dynamics of IEPC across consecutive oscillatory cycles in each frequency band. IEPC was analyzed in repeated-measures ANOVAs with factors frequency band X cycle number relative to the event. Frequency bands with two post-event cycles and frequency bands with three post-event cycles were analyzed in two separate ANOVAs.

For bands with two post-event cycles (4 bands in the range of 4 – 6Hz, see Figure 3), the evoked response was contained within a single oscillatory cycle. Most importantly, we found a main effect of cycle (F (2, 22) = 16.4, $p < .001$, $\eta^2 = 0.27$), that did not differ between bands (interaction $p > .05$, n.s., main effect of band $p_{GG} > .05$ after Greenhouse-Geyser correction for violation of sphericity). Planned comparisons showed that IEPC increased from pre-event to the first post-event cycle ($p < .001$), but

returned to the pre-event level by the second oscillatory cycle (pre-event to 2st post-event: $p > .05$, n.s.; 1st post-event to 2st post-event: $p < .001$).

For bands with three post-event cycles (3 bands in the range of 6-8 Hz, see Figure 3), the evoked response was contained within 1-2 oscillatory cycles. While IEPC magnitudes differed between frequency bands ($F(2, 22) = 5.99$, $p_{GG} = .008$, $\eta^2 = 0.02$). Importantly, here as well, the main effect of cycle ($F(3, 33) = 8.98$, $p < .001$, $\eta^2 = 0.27$) did not differ between bands (interaction $p = .63$ n.s.). Planned pairwise comparisons between neighboring cycles indicated that IEPC was significantly higher in the first post-event cycle than in the pre-event ($p = .003$) and the subsequent post-event cycles ($p < .01$). In contrast, IEPC did not differ between the pre-event and the 2nd or 3rd post-event cycles ($p > .05$) and decreased between the second and third post-event cycle ($p = .04$). Notably, this pattern was evident and consistent within single subjects (see Figure S3 for single subject data).

Overall, across the entire theta-band, the increase in IEPC continued 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 appears more in line with an evoked response model of speech envelope tracking than with an oscillatory entrainment model.

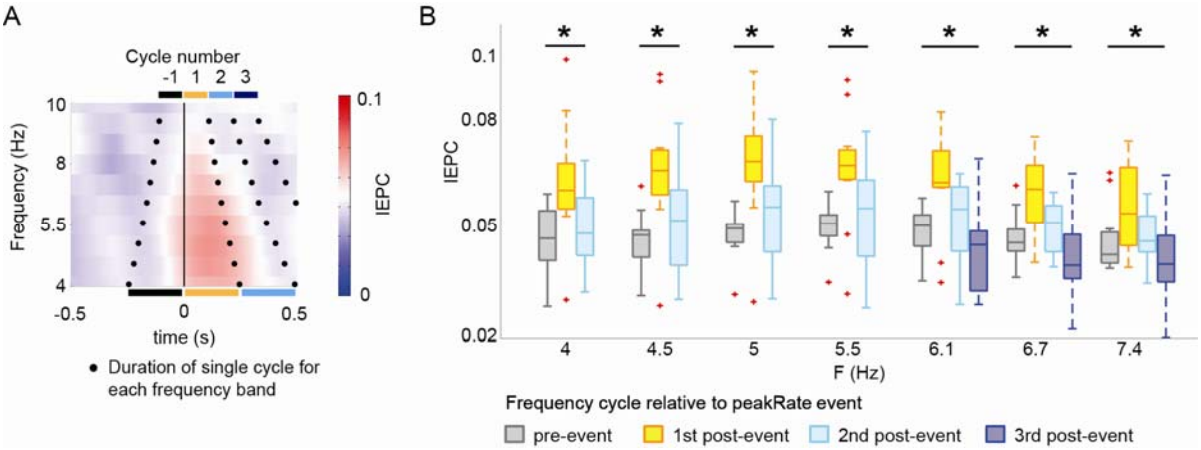

Figure 3. Theta phase-locking is limited to the duration of the evoked response after peakRate events.

A. IEPC across selected peakRate events that were at least .5 sec apart from preceding and subsequent events. Black dots mark the duration of a single oscillatory cycle in each frequency band. **B**. Group distribution IEPC in 3-4 cycles around peakRate events in each frequency band. * $p < 0.05$. Boxplots reflect median, 25th and 75th percentile, and range of values. Outliers (> 2.7 SD) are plotted individually.

Speech rate normalization of peakRate IEPC.

Finally, we assessed the effect of speech rate on the magnitude of theta IEPC to peakRate events. We had previously found that responses to peakRate monotonically scale with peakRate magnitude (Oganian

& Chang, 2019). As the rate of amplitude change is reduced with slowing, efficient envelope tracking across speech rates would require remapping of neural responses to peakRate events to account for this shift. If the representation of peakRate is normalized for contextual speech rate, we expect similar response magnitudes across speech rate conditions. In contrast, if no normalization occurs, IEPC magnitudes in slow speech should correspond to peakRate values, which are one-third of those in regular speech (Figures 4 A, B). To examine this, we split peakRate magnitudes into five quantiles for each rate condition. Figures 4A and B show the distribution of peakRate magnitudes in regular and slow speech, and Figure 4E (orange quantile plot) illustrates the reduction in the rate of amplitude change with slowing.

We calculated IEPC separately for each of the five peakRate quantiles in each rate condition (Figures 4C, D). We then evaluated the effects of quantile and speech rate on theta-band IEPC using a linear mixed-effects model with a maximal random effect structure. IEPC magnitudes increased for larger peakRate magnitudes (b = .01, std = 0.001, t = 1.4, χ^2= 55.0, p = 10^{-13}) in both conditions, replicating our previous results. Crucially, they did not differ between regular and slow speech (b = .003, SD = 0.005, t = .6, n.s., Figure 4E, purple quantile-quantile plot not significantly different from diagonal).

This analysis showed that 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.

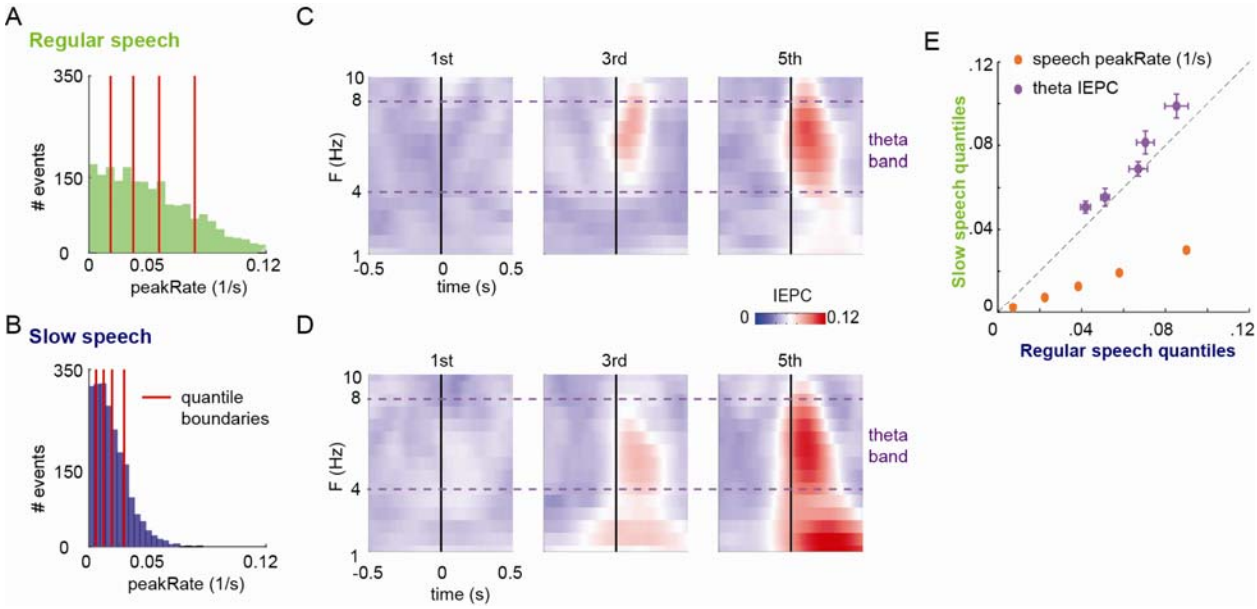

Figure 4. Normalization of peakRate IEPC for contextual speech rate.

A. Histogram of peakRate magnitudes in regular speech, with quantile boundaries marked in red. **B.** Same as A for slow speech **C**. IEPC in 1st, 3rd, 5th peakRate magnitude quantile. Horizontal lines mark theta frequency range (4-8Hz). **D**. Same as C for slow speech. **E.** Quantile-quantile plot of median peakRate magnitude (orange) and theta-band IEPC (purple, mean (error bars mark +- 1SEM across subjects)) in

regular and slow speech (orange). Theta IEPC quantile-quantile values are close to the diagonal, indicating the similar distribution of theta IEPC in regular and slow speech conditions. This is contrasted with the location of peakRate magnitude quantile-quantile values below the diagonal, reflecting smaller peakRate magnitudes in slow than in regular speech.

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. As predicted, we find that acoustic edges (peakRate events), but not to peaks in the envelope (peakEnv events), drive evoked responses on sensors over auditory cortical areas. Second, neural phase-locking induced by these evoked responses seems to be the primary source of speech envelope tracking in the theta-delta band (1 – 10 Hz) in MEG. Phase locking to acoustic edges had two components: a local component reflective of the evoked response and a global component, capturing the frequency of evoked responses and shifting with slowing to lower frequencies. Both components were predicted by a simulation of an evoked response model of envelope tracking. The 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). Finally, we find that the magnitude of evoked phase-reset to acoustic edges reflects the amplitude slope at the acoustic edge and is normalized for speech rate. Overall, our findings confer with the predictions of the evoked response model and suggest that low-frequency tracking of the speech envelope reflects the occurrence and frequency of evoked responses to acoustic edges.

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 has varying temporal dynamics, such phase locking requires a 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). Using recent intracranial recordings, it was recently discovered that local neural populations in the human speech cortex on the superior temporal gyrus respond specifically to acoustic edges in the envelope of continuous speech (Oganian & Chang, 2019). Here, we extend these results to MEG, showing that acoustic edges drive evoked responses and phase resets in low-frequency bands. These repeated phase-resets emerge as the basis of cortical envelope tracking. Taken together, our results establish acoustic edges as the basis for the cortical representation of the speech envelope.

A hallmark signature of oscillatory entrainment is the persistence of an oscillation after the phase-resetting stimulus (Breska & Deouell, 2017; Haegens & Zion Golumbic, 2018). Our results show that in speech, acoustic edges induce a phase reset in the theta band (4 – 8Hz). To test for ongoing oscillatory phase-locking following these phase resets, we focused on speech portions where the intervals between consecutive peakRate events were at least 500 ms long, corresponding to at least two oscillatory cycles in the 4-8 Hz range. This analysis showed no evidence for continuous theta-phase-locking beyond the duration of the evoked response. Instead, the increase in phase-locking was transient and continued only for one to two oscillatory cycles, after which it returned to the pre-event level, supporting the interpretation of theta phase locking as a transient evoked response.

With the slowing of the speech signal, the phase-reset after acoustic edges covered the entire delta-theta frequency range (1 -8 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 this such low-frequency phase-locking can reflect the overall frequency of evoked responses to acoustic edges, in contrast to the theta-band phase-locking that was due to the local evoked response. Crucially, our analyses show that both components can be accounted for by evoked responses. However, our results cannot refute the possibility that an additional oscillatory component contributes to phase-locking at the speech rate frequency in each condition. Rather our result show that such a component is not necessary to produce phase locking patterns as observed for speech in MEG. We also note that 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 delta arose with slowing of the speech stimulus.

Phase locking to acoustic edges was stronger for sharper edges (i.e., larger peakRate magnitudes), in line with prior results (Biermann & Heil, 2000; Oganian & Chang, 2019). Additionally, we found that the magnitude of phase-locking was normalized for the overall speech rate, a phenomenon previously described for speech perception at the behavioral (Reinisch, 2016; Wade & Holt, 2005) and neural levels (Alexandrou et al., 2018b; Kösem et al., 2018). Namely, although onset edges were flatter in slow speech, they induced comparable levels of phase-locking as in regular speech. This result suggests that cortical representation of and responses to acoustic edges reflect the relative magnitude of an edge relative to the contextual speech rate. Such shifting of the dynamic range for the representation of acoustic edge

magnitude can be a flexible mechanism for the precise representation of speech temporal dynamics (Diehl, Souther, & Convis, 1980; Hirataa & Lambacher, 2004). Further studies are necessary to elucidate the neural mechanisms underlying such speech rate normalization 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 frequency 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). Thus, IEPC is sensitive to long-range temporal regularities as well as local phase resets. IEPC revealed that phase locking to the speech envelope reflects both a local evoked response to acoustic edges as well as the frequency of occurrence of these events. 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. Overall, our results suggest that the latter component is the primary source of envelope tracking in CAC, but also that both can result from evoked responses, as this pattern was fully reproduced in our simulation of an evoked response model.

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 our previous work in ECoG, where we found 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, it 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. Future studies may be able to distinguish between onset responses and responses to ongoing amplitude changes using MEG.

Natural speech has varying temporal dynamics and no strong regularity (Alexandrou et al., 2018a). Our focus on envelope tracking for natural speech demonstrates that in this case, the neural signatures of envelope tracking are well explained by an evoked response model without an oscillatory component. This result 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 (Breska & Ivry, 2018). 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 to natural speech.

In summary, our results suggest that MEG speech envelope tracking relies on the detection of acoustic edges in the amplitude envelope of speech. Together with previous findings of acoustic edge encoding in the superior temporal gyrus, our findings suggest that acoustic edge detection is the main mechanism driving cortical speech envelope representations. The neural representation of these events 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 local and global changes in speech rate. In summary, an evoked response model of speech envelope tracking accounts for the main neural signatures of speech envelope tracking and provides a flexible temporal framework for speech comprehension.

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 assess 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 response, 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 the evoked response and narrow-band power analysis, 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 prior to averaging across events. No baseline correction was applied to the instantaneous phase.

Sensor selection

We selected sensors based on the magnitude of the group-averaged M100 waveform, 60-100 ms after the onset of utterance. To focus all further analyses on responses originating in temporal auditory areas, we selected ten sensors with maximal M100 responses from each hemisphere.

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 in band, and *phs* is the phase of 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 computed 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 between those computed with peakRate events and peakEnv events in regular speech. Differences in IEPC between alignments to peakRate and peakEnv were assessed using 2-D cluster-based permutation t-tests (Maris & Oostenveld, 2007) on the resulting maps, with 3000 permutations, a peak threshold of $p < 0.05$ and a cluster threshold of $p < 0.05$.

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). For each group, IEPC in neighboring oscillatory cycles was compared using 2-way repeated-measures ANOVAs with factors frequency band x cycle relative to the event. Pairwise comparisons between cycles were further assessed using paired comparisons with a Holm correction for multiple comparisons.

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.*

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

Rimmele, J. M., Morillon, B., Poeppel, D., & Arnal, L. H. (2018). Proactive Sensing of Periodic and Aperiodic Auditory Patterns. *Trends in Cognitive Sciences*, *22*(10), 870–882. https://doi.org/10.1016/j.tics.2018.08.003

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

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