Cognition and Behavior

θ-Band Cortical Tracking of the Speech Envelope Shows the Linear Phase Property

Jiajie Zou,[1,2] Chuan Xu,[3] Cheng Luo,[1,2] Peiqing Jin,[1,2] Jiaxin Gao,[1] Jingqi Li,[4] Jian Gao,[4] Nai Ding,[1,2] and Benyan Luo[3]

https://doi.org/10.1523/ENEURO.0058-21.2021

[1]Key Laboratory for Biomedical Engineering of Ministry of Education, College of Biomedical Engineering and Instrument Sciences, Zhejiang University, Hangzhou 310027, China, [2]Research Center for Advanced Artificial Intelligence Theory, Zhejiang Lab, Hangzhou 311121, China, [3]Department of Neurology, First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, China, and [4]Department of Rehabilitation, Hangzhou Mingzhou Brain Rehabilitation Hospital, Hangzhou 311215, China

Abstract

When listening to speech, low-frequency cortical activity tracks the speech envelope. It remains controversial, however, whether such envelope-tracking neural activity reflects entrainment of neural oscillations or superposition of transient responses evoked by sound features. Recently, it is suggested that the phase of envelope-tracking activity can potentially distinguish entrained oscillations and evoked responses. Here, we analyze the phase of envelope-tracking in humans during passive listening, and observe that the phase lag between cortical activity and speech envelope tends to change linearly across frequency in the θ band (4–8 Hz), suggesting that the θ-band envelope-tracking activity can be readily modeled by evoked responses.

Key words: neural entrainment; phase resetting; EEG

Significance Statement

During speech listening, cortical activity tracks the speech envelope, which is a critical cue for speech recognition. It is debated, however, what is the neural mechanism generating the envelope-tracking responses. Previous work has shown that δ-band envelope tracking responses recorded during music listening cannot be explained by a simple linear-system model. Here, however, we demonstrate that θ-band envelope tracking responses recorded during speech listening shows the linear phase property, which can be well explained by a linear-system model.

Introduction

The speech envelope, i.e., temporal modulations below 20 Hz, is critical for speech recognition (Drullman et al., 1994; Shannon et al., 1995; Shamma, 2001; Elliott and Theunissen, 2009; Ding et al., 2017), and large-scale cortical activity measured by MEG and EEG can track the speech envelope (Lalor et al., 2009; Ding and Simon, 2012b; Wang et al., 2012; Peelle et al., 2013; Doelling et al., 2014; Harding et al., 2019). Since the slow temporal modulations in speech are highly related to the ~5-Hz syllabic rhythm in speech, it has been hypothesized that θ-band neural synchronization to temporal modulations in speech provides a plausible mechanism to segment continuous speech into the perceptual units of syllables (Giraud and Poeppel, 2012; Poeppel and Assaneo, 2020).

Although low-frequency neural synchronization to slow temporal modulations has been extensively studied and is hypothesized to play a critical role in auditory perception, there is considerable debate about how it is generated (Ding and Simon, 2013; Doelling et al., 2014; Ding et al., 2016b; Haegens and Zion Golumbic, 2018; Zoefel et al., 2018; Alexandrou et al., 2020). On the one hand, it has been hypothesized that the low-frequency neural response to

Received February 9, 2021; accepted July 29, 2021; First published August 11, 2021.

The authors declare no competing financial interests.

Author contributions: J.Z., N.D., and B.L. designed research; J.Z., C.X., Jiax. Gao, J.L., and Jian Gao performed research; J.Z., C.X., C.L., P.J., Jiax. Gao, J.L., Jian Gao, N.D., and B.L. analyzed data; J.Z., C.X., C.L., P.J., N.D., and B.L. wrote the paper.

speech is generated by resetting the phase of intrinsic neural oscillations (Kayser et al., 2008; Lakatos et al., 2008, 2009; Schroeder et al., 2008; Kayser, 2009). On the other hand, it has been hypothesized that it is a sequence of transient responses evoked by sound features in speech (Lalor et al., 2009; Ding and Simon, 2012a,b). Distinguishing these two hypotheses, however, turns out to be extremely hard. For example, early studies have shown that the phase but not power of θ-band cortical activity is synchronized to speech (Luo and Poeppel, 2007; Howard and Poeppel, 2010), which supports the phase resetting hypothesis. It has been argued, however, that the same phenomenon can be observed for evoked responses, attributable to the different statistical sensitivity of response phase and power (Ding and Simon, 2013; Shah et al., 2004; Yeung et al., 2004). Furthermore, later studies observe consistent power and phase changes in the θ band during speech listening (Howard and Poeppel, 2012).

The phase resetting hypothesis and the evoked response hypothesis motivate different computational models for the neural responses to speech. Based on the evoked response hypothesis, the speech response can be simulated based on a linear time-invariant system, in which the phase lag between stimulus and response dramatically varies across frequency. A recent study, however, shows that neural synchronization to music violates the phase lag property predicted by the evoked response model, when listeners perform a pitch judgment task (Doelling et al., 2019). Instead, the response phase is more consistent with the prediction of a nonlinear oscillator model. This result suggests that cortical synchronization to music is potentially generated by more complicated nonlinear mechanisms than superposition of evoked responses. Since the evoked response model and the nonlinear oscillator model in Doelling et al. (2019) are computationally explicit, here we focus on these two models and test which model can better describe the neural response to speech.

The study by Doelling et al. (2019) questions the validity of using evoked response models to analyze neural activity synchronized to sound rhythms, since such models fail to predict the neural response phase during music listening. It remains possible, however, that the neural encoding scheme depends on the properties of sound. For example, the nonlinear oscillator models may be more appropriate for music, which is highly rhythmic, while the evoked response models may be sufficient to model the response to less rhythmic sound such as speech. It is also possible that the neural encoding scheme depends on the modulation frequency and appears to be different for music, which contains strong temporal modulations below 2 Hz, and speech, which contains strong temporal modulations around 5 Hz (Ding et al., 2017). Finally, it is also possible that active listening engages phase resetting mechanisms more than passive listening. Therefore, the primary goal of the current study is to quantify the phase lag property of the cortical response to speech during passive listening, and test whether it is more consistent with the prediction of the evoked response model or the nonlinear oscillator model in Doelling et al. (2019).

Materials and Methods

Participants

This study involved 15 healthy individuals (five males; 54.6 ± 10.12 years), who were right-handed with no history of neurologic diseases. Written informed consent was provided by participants.

Stimuli

Natural speech included two chapters from a novel, *The Supernova Era* by Cixin Liu (chapter 16, Fun country, and chapter 18, Sweet dream period). The story was narrated in Mandarin by a female speaker and digitized at 48 kHz. The speech was clear and highly intelligible. The two chapters were 34 min and 25 min in duration, respectively. Recordings of the two chapters were concatenated.

Procedures

All participants listened to speech while EEG responses were recorded. Speech was presented binaurally through headphones at a comfortable sound level. The experiment was separated into 2 d. On each day of the experiment, the spoken narrative was presented once. The 59-min speech stimulus was presented twice and therefore the total speech stimulus was almost 2 h, which was longer than the stimulus duration in most studies. The purpose of having the long stimulus was to reliably estimate the response phase. No other tasks were given, and therefore the participants listened passively.

EEG recording and signal preprocessing

EEG signals were recorded using a 64-electrodes BrainCap (Brain Products GmbH) in the international 10–20 system, and one of the 64 electrodes was placed under the right eye to record electrooculogram (EOG). EEG signals were referenced online to FCz, but were referenced offline to a common average reference (Poulsen et al., 2007). The EEG signals were filtered online with a 50-Hz notch filter to remove line noise (12th order zero-phase Butterworth filter), a low-pass antialiasing filter (70-Hz cutoff, eighth order zero-phase Butterworth filter), and a high-pass filter to prevent slow drifts (0.3-Hz cutoff,

This work was supported by the National Natural Science Foundation of China Grant 31771248 (to N.D.), the National Natural Science Foundation of China Grant 81870817(to B.L.), the National Key R&D Program of China Grant 2019YFC0118200 (to N.D.), the Major Scientific Research Project of Zhejiang Lab Grant 2019KB0AC02 (to N.D.), the Zhejiang Provincial Natural Science Foundation of China Grant LY20C090008 (to P.J.), and the Ministry of Education Frontier Science Center for Brain Science and Brain-Machine Integration, Zhejiang University.

Acknowledgements: We thank Yuhan Lu for thoughtful comments on a previous version of this manuscript.

Correspondence should be addressed to Benyan Luo at luobenyan@zju.edu.cn or Nai Ding at ding_nai@zju.edu.cn.

https://doi.org/10.1523/ENEURO.0058-21.2021

eighth order zero-phase Butterworth filter). The signals were sampled at 1 kHz. The EEG signal was processed following the procedure in Zou et al. (2019). All preprocessing and analysis in this study were conducted in the MATLAB software (The MathWorks).

EEG recordings were low-pass filtered below 50 Hz with a zero-phase anti-aliasing FIR filter (implemented using a 200-ms Kaiser window) and down-sampled to 100 Hz. EOG artifacts were regressed out based on the least-squares method. Similar to previous studies (Ding and Simon, 2012a,b), the speech response was averaged over the two representations on two recording days to increase the signal-to-noise ratio.

The envelopes of stimuli reflected how sound intensity fluctuated over time and were extracted by applying full-wave rectification to the stimulus. Similar to the preprocessing of EEG recordings, the envelopes were further low-pass filtered below 50 Hz with a zero-phase anti-aliasing FIR filter (implemented using a 200-ms Kaiser window) and down-sampled to 100 Hz.

Phase coherence analysis

To characterize the stimulus-response phase lag, the stimulus and response were both transformed into the frequency domain. Specifically, the acoustic envelope and EEG response were segmented into non-overlapping 2-s time bins, and all segments were converted into the frequency domain using the fast Fourier transform (FFT) algorithm. The response phase and stimulus phase were denoted as α_{ft} and β_{ft}, respectively, for frequency bin f and time bin t, and the stimulus-response phase lag was calculated as $\theta_{ft} = \alpha_{ft} - \beta_{ft}$. The coherence of the phase lag across time bins, also known as the cerebro-acoustic phase coherence (Peelle et al., 2013), was calculated using the following equation:

$$C(f) = \frac{(\sum_{t=1}^{T} cos(\theta_{ft}))^2 + (\sum_{t=1}^{T} sin(\theta_{ft}))^2}{T},$$

where $C(f)$ was the phase coherence in frequency bin f, and T is the total number of time bins. The phase coherence was independently calculated for each electrode and then averaged using the arithmetic mean. The phase coherence is in the range of 0–1, and higher phase coherence indicated that the response phase was more precisely synchronized to the stimulus phase.

In the response topography analysis, we considered a signed phase coherence. Specifically, we chose channel Fz as a reference. For each electrode, if the phase difference between this electrode and electrode Fz was larger than 90°, the phase coherence was negated. Otherwise, the phase coherence was kept positive. The signed phase coherence could illustrate the phase relationship between electrodes on top of showing the phase coherence. Since the phase coherence was strongest in central-frontal electrodes, fourteen centro-frontal electrodes, i.e., Fz, F1, F2, F3, F4, FC1, FC2, FC3, FC4, Cz, C1, C2, C3, and C4, were used to characterize the phase-frequency relationship.

Phase-frequency relationship

The stimulus-response phase lag at frequency f, i.e., θ_f, was calculated by averaging θ_{ft} over electrodes and all 2-s time bins using the circular average (Fisher, 1993). The group delay is defined based on the first-order derivative of the stimulus-response phase lag over frequency, i.e., $d(f) = (\theta(f) - \theta(f + \Delta f))/2\pi\Delta f$, which reflects how quickly a change in the stimulus is reflected in the response (Oppenheim et al., 1997). To calculate the group delay, we unwrapped the phase lag, fitted the phase lag with a straight line, and divided the slope of the straight line by 2π.

To evaluate the linearity of the phase-frequency curve, we defined a linearity measure as $L(f) = 1/|\theta(f) + \theta(f + 2\Delta f) - 2\theta(f + \Delta f)|$. This measure was the reciprocal of the absolute value of the second-order derivative of the phase-frequency curve and different electrodes was pooled with averaging. If phase lag linearly changed with frequency, the second-order derivative was 0 and the linearity measure was positive infinity. A large $d_2(f)$ indicated a roughly linear phase-frequency curve.

Since the phase-frequency curve was approximately linear in the θ band, we fitted the actual phase-frequency curve in this frequency range using a linear function: $\theta_L(f) = kf + b$, $4 \leq f < 8$. The slope parameter k and the intercept parameter b were fitted using the least-squares method, and the slope parameter k denoted the mean group delay between 4 and 8 Hz.

Statistics

Phase coherence

To evaluate whether the phase coherence at a frequency was significantly higher than chance, we estimated the chance-level phase coherence with a permutation strategy (Peelle et al., 2013; Harding et al., 2019). After the speech envelope and EEG response were segmented into 2-s time bins, we shuffled all time bins for the speech envelope so that the envelope and response were randomly paired. We calculated the phase coherence for the phase lag between the EEG response and randomly paired speech envelope. This procedure was repeated 5000 times, creating 5000 chance-level phase coherence. We averaged the phase coherence value over electrodes and participants, for both the actual phase coherence and the 5000 chance-level phase coherence. The significance level of the phase coherence at a frequency was $(N + 1)/5001$, if it was lower than N out of the 5000 chance-level coherence at that frequency (one-sided comparison).

Linearity

The chance-level linearity measure of the phase-frequency curve was estimated using a similar procedure. The linearity measure was significantly larger than chance with the significance level being $(N + 1)/5001$, if it was smaller than N of the 5000 chance-level values (one-sided comparison).

For the comparisons of the linearity measure of different frequency bands, statistical tests were performed using bias-corrected and accelerated bootstrap (Efron and

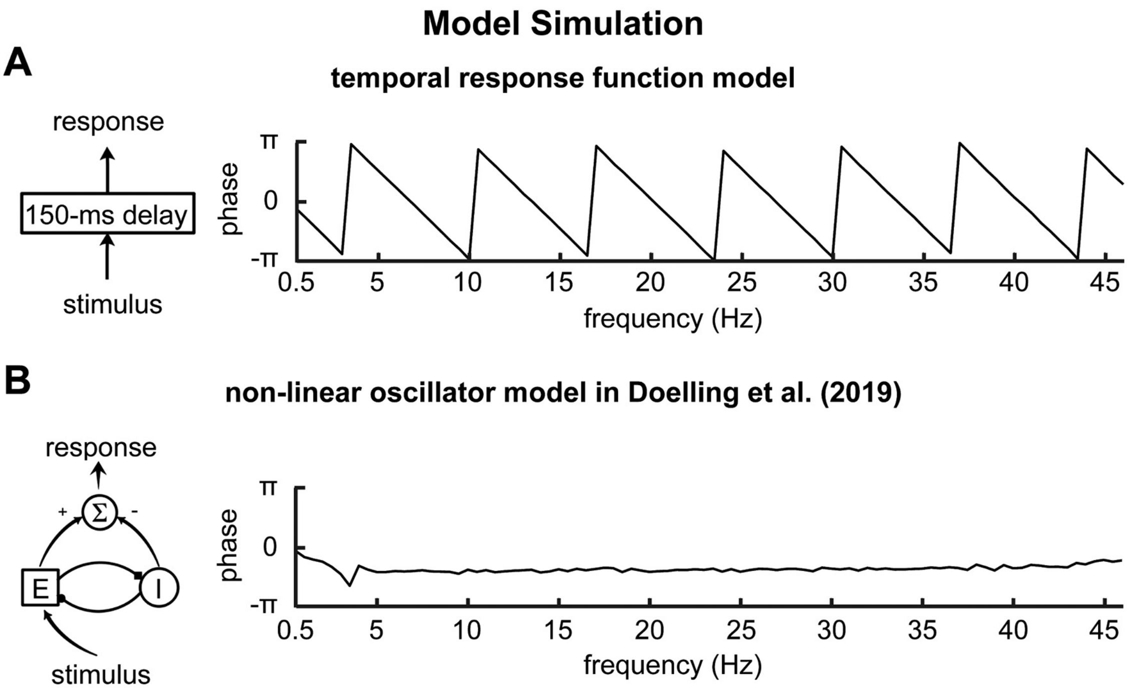

Figure 1. Simulated phase-frequency curve. The curve shows the stimulus-response phase lag as a function of response frequency. Panels ***A***, ***B*** separately show the results simulated based on the evoked response model and the nonlinear oscillator model proposed in Doelling et al. (2019). These two models are based on the evoked response hypothesis and the phase resetting hypothesis, respectively. In the current evoked response model, the response evoked by a unit change in stimulus is a delayed impulse and the function of the model is to delay the stimulus by 150 ms. Such a model predicts that the stimulus-response phase lag changes linearly across frequency. Consequently, the phase-frequency curve appears to have a sawtooth shape. The nonlinear oscillator model, in contrast, predicts that the stimulus-response phase lag only changes in a very limited range across frequency.

Tibshirani, 1994). In the bootstrap procedure, the differences between two frequency bands were resampled with replacement 5000 times. Each time the data sampled were averaged across participants, therefore a total of 5000 mean values were produced. If N out of the 5000 mean values were greater (or smaller) than 0, the significance level was $2(N + 1)/5001$ (two-sided comparison).

Correlation between phase coherence and phase linearity

For significance test of correlation between phase coherence and phase linearity, we used two-tailed Student's *t* test. When multiple comparisons were performed, the *p* value was further adjusted using the false discovery rate (FDR) correction (Amini and Hochberg, 1995).

Model simulation

We simulated the neural response to speech using two models. One was the evoked response model, in which the simulated neural response was simply the speech envelope convolving the response evoked by a unit change in the speech envelope. This model was formulated as follows:

$$r(t) = \int_0^T h(\tau)A(t-\tau)d\tau,$$

where $r(t)$ and $A(t)$ were the simulated neural response and the speech envelope, respectively. The $h(t)$ described the neural response evoked by a unit change in the stimulus. In the illustration in Figure 1*A*, $h(t)$ was a unit impulse function with 150-ms response latency, i.e., $h(t) = \delta(t - 150\,\text{ms})$, and in this case $r(t) = A(t - 150\,\text{ms})$. The simulation results would not be affected as long as $h(t)$ had a symmetric waveform centered at 150 ms.

The other model was the nonlinear oscillator model proposed by Doelling et al. (2019). The oscillator model was formulated as follows:

$$\tau\frac{dI(t)}{dt} = -I(t) + S(\rho_I + bE(t) - dI(t)),$$

where $A(t)$ was the speech envelope. $E(t)$ and $I(t)$ simulated the responses from an excitatory and an inhibitory neural population, respectively. The output of this model was the difference between excitatory and inhibitory populations, i.e., $E(t) - I(t)$. S denoted the sigmoid function. All the parameters were the same as in Doelling et al. (2019), i.e., $a = b = c = 10$, $d = -2$, $\rho_E = 2.3$, $\rho_I = -3.2$, $\kappa = 1.5$, and $\tau = 66$ ms. Interpretations of the parameters could be found in Doelling et al. (2019). The model was simulated using the ode3 method in MATLAB Simulink (The MathWorks) and the time step was 1 ms.

In the simulations, the input was the envelope of the entire 59-min speech stimulus, and the input-output phase lag was calculated the same way it was calculated in the EEG analysis.

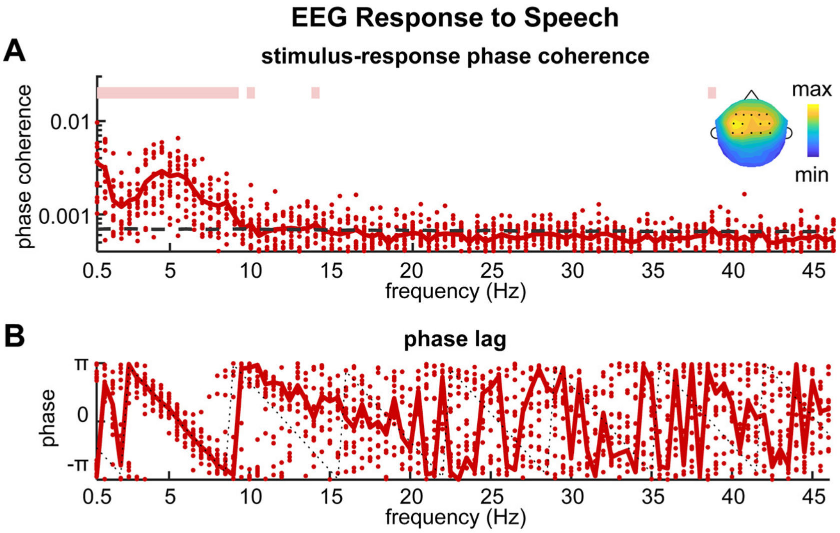

Figure 2. Phase analysis of the envelope-tracking response. ***A***, The phase coherence spectrum shows how precisely the response phase is phase locked to the stimulus. The dashed black line shows the 99% confidence interval of the chance-level phase coherence. The pink line on top denotes the frequency bins in which phase coherence is significantly higher than chance ($p < 0.01$, permutation test, FDR corrected). The topography shows the signed phase coherence averaged between 0.5 and 9 Hz. The dark dots denote the 14 centro-frontal electrodes selected for the further phase analysis. ***B***, The phase-frequency curve. The phase lag appears to linearly decrease over frequency in the frequency band where the phase coherence was higher than chance. The black dotted lines are fitted based on the phase lag in the θ band. Each red dot denotes a participant.

Results

We first analyzed in which frequency bands and EEG electrodes reliable cortical synchronization to speech was observed. The coherence of the stimulus-response phase lag was separately calculated in each frequency bin. Significant phase coherence was most reliably observed below 9 Hz. The topography of the low-frequency neural responses (<9 Hz) showed a

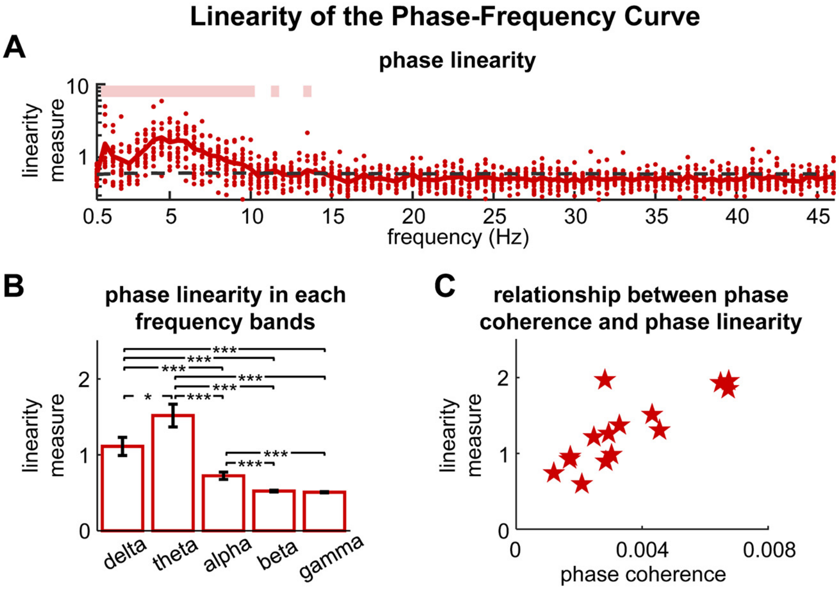

Figure 3. Linearity of the phase-frequency curve. ***A***, Phase linearity as a function of frequency. The dashed black line shows the 99% confidence interval of the chance-level phase linearity. The pink lines on top denotes the frequency bins in which the phase linearity is significantly higher than chance ($p < 0.01$, permutation test, FDR corrected). Each red dot denotes a participant. ***B***, The comparison of phase linearity across frequency bands. Error bars represent 1 standard error of the mean across participants. Significant differences between frequency bands are indicated by stars; *$p < 0.05$, ***$p < 0.001$ (bootstrap, FDR corrected). ***C***, The relationship between phase coherence and phase linearity. The phase coherence and the phase linearity are both averaged within the θ band. Each red marker denotes a participant. Participants with higher phase coherence generally show better linearity ($R = 0.805$, $p = 3 \times 10^{-4}$, two-tailed Student's t test).

centro-frontal distribution (Fig. 2*A*, upper right corner).

We next analyzed how the stimulus-response phase lag varied across frequency. The phase lag appeared to change linearly over frequency in the frequency range where the phase coherence was higher than chance (Fig. 2*B*). We then evaluated the linearity of the phase-frequency curve (see Materials and Methods). As shown in Figure 3*A*, the linearity measure was significantly higher than chance in the low-frequency bands ($p < 0.01$, permutation test, FDR corrected) and peaked in the θ band. We compared the averaged phase linearity across δ (1–4 Hz), θ, α (8–13 Hz), β (13–30 Hz), and γ (30–45 Hz) bands. As shown in Figure 3*B*, the phase linearity was significantly higher in the θ band than other frequency bands (θ band vs δ band: $p = 0.043$; θ band vs α, β, or γ band: $p = 4 \times 10^{-4}$, bootstrap, FDR corrected). In order to estimate the group delay in the θ band, we used a straight line to fit the linear trend, which was shown by the dotted gray line in Figure 2*B*. The mean group delay in the θ band, i.e., slope of the linear fit in Figure 2*B*, was 156 ± 50 ms.

Additionally, we investigated whether the phase-frequency curve tended to be more linear for participants who showed higher phase coherence. As the θ band shows significant and highest phase linearity (Fig. 3*A*,*B*), we compared the averaged phase linearity and the average phase coherence in the θ band (Fig. 3*C*). The phase linearity and the phase coherence were significantly correlated at the individual level ($R = 0.805$, $p = 3 \times 10^{-4}$, two-tailed Student's *t* test).

Discussion

The current study investigates the phase property of the EEG response to speech during passive listening. It is shown that the stimulus-response phase lag is approximately a linear function of frequency in the θ band. This linear phase property can be easily explained by the evoked response model and therefore does not require more sophisticated nonlinear oscillator models.

Based on systems theory (Oppenheim et al., 1997), if the stimulus-response phase lag changes linearly across frequency, it indicates that the evoked response has a finite duration and has a symmetric waveform centered at the group delay (Fig. 1*A* for a delay system, for example). The current results suggest that the EEG response resembles the speech envelope but delayed, which supports the evoked response hypothesis. It is possible that cortical activity tracks the speech envelope or related features (Ding and Simon, 2014), and it is also possible that discrete acoustic landmarks that are extracted using nonlinear mechanisms drive the evoked responses (Doelling et al., 2014).

The group delay observed here is around 150 ms. Above 8 Hz, neural activity is not precisely synchronized to the speech envelope. Below 4 Hz, the phase linearity was weaker, suggesting more complex generation mechanisms. A previous MEG study finds similar group delay for the response to amplitude modulated tones: the group delay is 131 and 147 ms in the left and right hemispheres respectively, in the frequency range between 1.5 and 8 Hz (Wang et al., 2012). The 150- to 200-ms group delay also corresponds to the latency of the N1 and P2 responses in the temporal response function derived from the envelope tracking response (Aiken and Picton, 2008; Lalor et al., 2009; Ding and Simon, 2012a; Horton et al., 2013).

The current study finds that the stimulus-response phase lag changes approximately linearly across frequency (Figs. 2*B*, 3*A*), and participants who have higher stimulus-response phase coherence generally showed the better phase linearity (Fig. 3*B*). A previous MEG study, however, has shown that the stimulus-response phase lag cannot be explained by simple evoked responses (Doelling et al., 2019) but is more consistent with the prediction of a nonlinear oscillator model (illustrated in Fig. 1*B*). These two studies, however, focus on the neural responses in different frequency bands and during different tasks. The study by Doelling et al. (2019) analyzes cortical activity synchronized to auditory rhythms at 0.5, 0.7, 1, 1.5, 5, and 8 Hz. Four of the 6 frequencies considered in the study are below the θ band, and the current study also finds that the stimulus-response relationship is complicated below the θ band. Therefore, the results from these two studies do not conflict but reveal different neural mechanisms in different frequency ranges.

During speech processing, the neural response below the θ band can encode higher-level linguistic structures, e.g., phrases and sentences, on top of slow acoustic modulations, even if these linguistic structures are mentally constructed based on syntactic rules instead of prosodic information (Ding et al., 2016a). These results suggest that multiple factors could drive very low-frequency neural synchronization to speech. The analysis in this study characterizes neural synchronization to the speech envelope and cannot capture purely syntactic-driven response components. In other words, the neural response shown here is the response to acoustic modulations in speech, instead of the response to linguistic structures. The slow acoustic modulations below the θ band, however, could serve as a prosodic cue for mental construction of phrasal-level linguistic structures (Frazier et al., 2006). It is possible that distinct mechanisms are employed to encode syllabic-level and higher-level speech information: a roughly linear code is employed to encode syllabic-level speech features while more complex neural mechanisms are employed to prosodic features, which allows for frequent interactions with the syntactic and semantic processing systems.

In sum, by analyzing the stimulus-response phase lag, we show that the speech response in the θ band was approximately a delayed version of the speech envelope in the same frequency range. A time-delay system can be readily implemented using a linear time-invariant system, which is consistent with the evoked response hypothesis. Future studies, however, are needed to study whether the response phase property is modulated by attention and whether similar results could be obtained when listening to other sound.

References

Aiken SJ, Picton TW (2008) Human cortical responses to the speech envelope. Ear Hear 29:139–157.

Alexandrou AM, Saarinen T, Kujala J, Salmelin R (2020) Cortical entrainment: what we can learn from studying naturalistic speech perception. Lang Cogn Neurosci 35:681–693.

Amini Y, Hochberg Y (1995) A practical and powerful approach to multiple testing. J Roy Stat Soc B 57:289–300.

Ding N, Simon JZ (2012a) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci USA 109:11854–11859.

Ding N, Simon JZ (2012b) Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. J Neurophysiol 107:78–89.

Ding N, Simon JZ (2013) Power and phase properties of oscillatory neural responses in the presence of background activity. J Comput Neurosci 34:337–343.

Ding N, Simon JZ (2014) Cortical entrainment to continuous speech: functional roles and interpretations. Front Hum Neurosci 8:311.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D (2016a) Cortical tracking of hierarchical linguistic structures in connected speech. Nat Neurosci 19:158–164.

Ding N, Simon JZ, Shamma SA, David SV (2016b) Encoding of natural sounds by variance of the cortical local field potential. J Neurophysiol 115:2389–2398.

Ding N, Patel AD, Chen L, Butler H, Luo C, Poeppel D (2017) Temporal modulations in speech and music. Neurosci Biobehav Rev 81:181–187.

Doelling KB, Arnal LH, Ghitza O, Poeppel D (2014) Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. Neuroimage 85:761–768.

Doelling KB, Assaneo MF, Bevilacqua D, Pesaran B, Poeppel D (2019) An oscillator model better predicts cortical entrainment to music. Proc Natl Acad Sci USA 116:10113–10121.

Drullman R, Festen JM, Plomp R (1994) Effect of temporal envelope smearing on speech reception. J Acoust Soc Am 95:1053–1064.

Efron B, Tibshirani RJ (1994) An introduction to the bootstrap. London: Chapman and Hall/CRC.

Elliott TM, Theunissen FE (2009) The modulation transfer function for speech intelligibility. PLoS Comput Biol 5:e1000302.

Fisher NI (1993) Statistical analysis of circular data. Cambridge: Cambridge University Press.

Frazier L, Carlson K, Clifton C Jr (2006) Prosodic phrasing is central to language comprehension. Trends Cogn Sci 10:244–249.

Giraud AL, Poeppel D (2012) Speech perception from a neurophysiological perspective. In: The human auditory cortex. New York: Springer.

Haegens S, Zion Golumbic E (2018) Rhythmic facilitation of sensory processing: a critical review. Neurosci Biobehav Rev 86:150–165.

Harding EE, Sammler D, Henry MJ, Large EW, Kotz SA (2019) Cortical tracking of rhythm in music and speech. Neuroimage 185:96–101.

Horton C, D'Zmura M, Srinivasan R (2013) Suppression of competing speech through entrainment of cortical oscillations. J Neurophysiol 109:3082–3093.

Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. J Neurophysiol 104:2500–2511.

Howard MF, Poeppel D (2012) The neuromagnetic response to spoken sentences: co-modulation of theta band amplitude and phase. Neuroimage 60:2118–2127.

Kayser C (2009) Phase resetting as a mechanism for supramodal attentional control. Neuron 64:300–302.

Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. Cereb Cortex 18:1560–1574.

Lakatos P, Karmos G, Mehta AD, Ulbert I, Schroeder CE (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320:110–113.

Lakatos P, O'Connell MN, Barczak A, Mills A, Javitt DC, Schroeder CE (2009) The leading sense: supramodal control of neurophysiological context by attention. Neuron 64:419–430.

Lalor EC, Power AJ, Reilly RB, Foxe JJ (2009) Resolving precise temporal processing properties of the auditory system using continuous stimuli. J Neurophysiol 102:349–359.

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. Neuron 54:1001–1010.

Oppenheim AV, Willsky AS, Nawab SH (1997) Signals and systems. Hoboken: Prentice Hall.

Peelle JE, Gross J, Davis MH (2013) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. Cereb Cortex 23:1378–1387.

Poeppel D, Assaneo MF (2020) Speech rhythms and their neural foundations. Nat Rev Neurosci 21:322–334.

Poulsen C, Picton TW, Paus T (2007) Age-related changes in transient and oscillatory brain responses to auditory stimulation in healthy adults 19-45 years old. Cereb Cortex 17:1454–1467.

Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A (2008) Neuronal oscillations and visual amplification of speech. Trends Cogn Sci 12:106–113.

Shah AS, Bressler SL, Knuth KH, Ding M, Mehta AD, Ulbert I, Schroeder CE (2004) Neural dynamics and the fundamental mechanisms of event-related brain potentials. Cereb Cortex 14:476–483.

Shamma S (2001) On the role of space and time in auditory processing. Trends Cogn Sci 5:340–348.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. Science 270:303–304.

Wang Y, Ding N, Ahmar N, Xiang J, Poeppel D, Simon JZ (2012) Sensitivity to temporal modulation rate and spectral bandwidth in the human auditory system: MEG evidence. J Neurophysiol 107:2033–2041.

Yeung N, Bogacz R, Holroyd CB, Cohen JD (2004) Detection of synchronized oscillations in the electroencephalogram: an evaluation of methods. Psychophysiology 41:822–832.

Zoefel B, Ten Oever S, Sack AT (2018) The involvement of endogenous neural oscillations in the processing of rhythmic input: more than a regular repetition of evoked neural responses. Front Neurosci 12:95.

Zou J, Feng J, Xu T, Jin P, Luo C, Zhang J, Pan X, Chen F, Zheng J, Ding N (2019) Auditory and language contributions to neural encoding of speech features in noisy environments. Neuroimage 192:66–75.