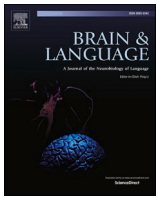
Brain & Language 256 (2024) 105463

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l



Cross-linguistic and acoustic-driven effects on multiscale neural synchrony to stress rhythms

Deling He [a,b], Eugene H. Buder [a,b], Gavin M. Bidelman [c,d,e,*]

[a] School of Communication Sciences & Disorders, University of Memphis, Memphis, TN, USA
[b] Institute for Intelligent Systems, University of Memphis, Memphis, TN, USA
[c] Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, USA
[d] Program in Neuroscience, Indiana University, Bloomington, IN, USA
[e] Cognitive Science Program, Indiana University, Bloomington, IN, USA

ARTICLE INFO

Keywords:
Brain oscillation
Phase locking
Delta-theta coupling
Cortical tracking
Delta band
Stress foot
Syllable rhythm
Hierarchical coherence

ABSTRACT

We investigated how neural oscillations code the hierarchical nature of stress rhythms in speech and how stress processing varies with language experience. By measuring phase synchrony of multilevel EEG-acoustic tracking and intra-brain cross-frequency coupling, we show the encoding of stress involves different neural signatures (delta rhythms = stress foot rate; theta rhythms = syllable rate), is stronger for amplitude vs. duration stress cues, and induces nested delta-theta coherence mirroring the stress-syllable hierarchy in speech. Only native English, but not Mandarin, speakers exhibited enhanced neural entrainment at central stress (2 Hz) and syllable (4 Hz) rates intrinsic to natural English. English individuals with superior cortical-stress tracking capabilities also displayed stronger neural hierarchical coherence, highlighting a nuanced interplay between internal nesting of brain rhythms and external entrainment rooted in language-specific speech rhythms. Our cross-language findings reveal brain-speech synchronization is not purely a "bottom-up" but benefits from "top-down" processing from listeners' language-specific experience.

1. Introduction

A growing number of brain imaging studies suggest that speech is processed at multiple temporal windows operated by a set of neuronal oscillators whose frequencies are tuned to relevant features of the acoustic-linguistic signal (Ding et al., 2016; Ghitza, 2011; Gross et al., 2013; Hyafil et al., 2015; Kösem & Van Wassenhove, 2017; Poeppel, 2003; Rimmele et al., 2023; Teng et al., 2017). The oscillations associated with speech are spectrally distributed in the gamma (>30 Hz), theta (4-8 Hz), and delta (1-3 Hz) frequency bands of the EEG, roughly corresponding with the time spans of phonemic, syllabic, and supra-syllabic units. Presumably, the processing of speech might be realized through phase alignment of brain oscillations to the speech amplitude envelope, which segments/parses the continuum of speech sounds into linguistic representations (Doelling et al., 2014; Ghitza, 2012; Luo & Poeppel, 2007).

Such brain-to-speech synchronization is especially significant in terms of coding syllable rhythm. Theoretical and empirical work suggests brain activity imposes a constraint on processing such that auditory perception is optimized when theta band (4–8 Hz) oscillations coincide with the range of natural syllable rates (Ghitza, 2012; Houtgast & Steeneken, 1985; Luo & Poeppel, 2007; Poeppel & Assaneo, 2020). For instance, speech intelligibility is severely degraded with low-pass filtering below 2 Hz and is only marginally improved by adding modulations above 8 Hz (Drullman et al., 1994). Moreover, cortical-acoustic entrainment and intracranial auditory-motor coherence is enhanced at frequencies close to the dominant syllable rhythm which has been empirically found to be 4–5 Hz across languages (Assaneo & Poeppel, 2018; He et al., 2023). However, whether there are also optimal supra-syllabic frequencies within lower-frequency delta neural oscillations has not been explicitly tested, though several studies have begun to examine delta-neural entrainment.

Cycles of delta oscillations often align with repetitive complex sounds including frequency-modulated complex tones (Henry & Obleser, 2012), digit strings (Rimmele et al., 2021), noise-vocoded speech (Bröhl & Kayser, 2021), prosodic or lexical phrases (Cogan & Poeppel, 2011; Gross et al., 2013; Keitel et al., 2018; Lo et al., 2022), and sentences (Lu et al., 2022). However, the particular sound elements that

* Corresponding author at: Department of Speech, Language and Hearing Sciences, Indiana University, Bloomington, IN, USA.
E-mail address: gbidel@iu.edu (G.M. Bidelman).

https://doi.org/10.1016/j.bandl.2024.105463
Received 5 December 2023; Received in revised form 1 September 2024; Accepted 3 September 2024
Available online 7 September 2024

entrain delta oscillations remain elusive, being variably attributed to "intonation, prosody, and phrases" (Boucher et al., 2019; Ghitza, 2011; Giraud & Poeppel, 2012; Gross et al., 2013; Rimmele et al., 2021). Moreover, prominent theories in language processing, such as the asymmetric sampling in time (AST) or TEMPO hypotheses, have overlooked the potential hierarchical role of delta oscillations (Ghitza, 2011; Ghitza & Greenberg, 2009; Hickok & Poeppel, 2007; Poeppel, 2003). This has led to a growing debate, with some arguing ongoing delta oscillations modulate theta activity (Gross et al., 2013; Lakatos et al., 2005), and others asserting the master role of the theta oscillator (Ghitza, 2011, 2013). These discrepancies pinpoint the necessity for a more integrated exploration of delta oscillations, which have generally been overlooked in the literature.

However, there are several prominent suprasegmental features of speech that might be optimally coded by delta brain oscillations. One important feature that creates a natural hierarchy in speech is stress. Specifically, stress foot[1] is a supra-syllabic unit that organizes a group of syllables by assigning emphasis on the stressed syllable (Hogg et al., 1987; Leong, 2012; Selkirk, 1980). Approximately 85 % of English words begin with the first syllable stressed (Cutler & Carter, 1987), which is mostly signaled by higher amplitude and longer duration (Fry, 1955; Greenberg et al., 2003), and to a lesser extent, pitch (Arvaniti, 2009; Greenberg, 1999; Kochanski et al., 2005; Silipo & Greenberg, 1999, 2000). Importantly, acoustic research illustrates that English simultaneously carries syllable and stress foot rhythms in the speech signal, represented by frequency-specific amplitude envelopes that closely correspond to theta and delta brain oscillations (Greenberg et al., 2003; Leong, 2012; Leong et al., 2014; Tilsen & Arvaniti, 2013).

The hierarchical nature of stress assignment also creates the opportunity for nesting of different speech elements. For example, while the dominant syllable rhythm is at 4-5 Hz across languages, the rhythm of stress foot in English is centered at half this speed, nominally around 2 Hz (Ding et al., 2017; Greenberg, 1999; Greenberg et al., 1996; Tilsen & Arvaniti, 2013; Tilsen & Johnson, 2008). Indeed, faster syllable rhythms are embedded into slower stress foot constituents, creating hierarchical nesting (Goswami & Leong, 2013; Leong, 2012). Such hierarchy is quantitatively illustrated by cross-frequency phase coupling seen in different acoustic constituents of the speech amplitude envelope. For example, Goswami and Leong (2013) showed a phase hierarchy relationship, where the ridge in the stress foot envelope always aligns with the stressed syllable and away from the unstressed syllable. Such hierarchy constrains stressed syllables to occur only in certain phases of the stress foot envelope (Leong, 2012). This is of particular interest given that the relationship between delta and theta brain oscillations may provide one such mechanism that mirrors this hierarchical structure of speech. However, there remains an empirical gap on how multiscale brain oscillations lock to the hierarchical properties of stress rhythms. To our knowledge, whether such stress foot-syllable hierarchy seen in the speech signal is reflected neurobiologically in delta-theta brain rhythm coupling remains to be tested.

The hierarchical nature of stress also affords the opportunity to explore cross-language differences in delta-theta mechanisms of speech processing. Yet, the specifics of brain oscillatory dynamics that vary among speakers with distinct language backgrounds remains largely unexplored. Still, some studies show cortical oscillations reliably track the syllabic amplitude envelope even in a foreign or unintelligible language (Ding et al., 2016; Zou et al., 2019). However, this brain-speech tracking appears to falter at the suprasyllabic level, and foreign listeners struggle to understand the linguistic content conveyed by delta band activity (Blanco-Elorrieta et al., 2020; Ding et al., 2016). These findings suggest the possibility of language-specific tuning of brain oscillatory dynamics, particularly pertaining to processing at the suprasyllabic level.

Indeed, in the context of stress encoding, it is reasonable to assume that cross-linguistic differences in delta band oscillations might occur in native English vs. nonnative speakers owing to the relative importance of stress in English vs. other languages. In particular, a comparison between English and Mandarin Chinese listeners could elucidate experience-dependent changes in stress-related brain processing given the distinctive prosodic features in each language (Hogg et al., 1987; Jongman et al., 2006). Supporting this cross-language design, behavioral and EEG studies have in fact shown that intensity is a less reliable cue for Mandarin listener's perception of English stress given the lesser importance of this cue in their native language (Mandarin) (Chrabaszcz et al., 2014; Chung & Bidelman, 2016). Thus, one primary objective herein was to further characterize such cross-language differences in oscillatory stress processing.

The current study aimed to examine delta (stress foot level) and theta (syllable level) oscillations in terms of how these neural correlates of rhythmic stress processing vary with language experience and acoustic modulations. We analyzed multilevel EEG-acoustic phase synchrony and intra-brain cross-frequency coupling while English and Chinese listeners perceiving various rhythmic stress patterns. We hypothesized that brain oscillations in the delta and theta bands would concurrently synchronize to stress and syllable rhythms, given their putative role in coding these properties of speech. Furthermore, we hypothesized this brain-stress synchronization might be enhanced by the dominant natural stress patterns (e.g., amplitude-signaled high salient stress rhythm at 2 Hz) in English speech. We further posited the acoustic phase hierarchy between stress foot and syllable rhythms (Leong, 2012; Leong et al., 2014) would be paralleled in brain activity (EEG) as enhanced coupling of delta-theta oscillations. Lastly, we hypothesized Chinese speakers might have reduced neural responses coding stress rhythm (given the relative unimportance of stress in their native Mandarin), yet maintain neural entrainment to syllable rhythm—which is likely more discernable to even non-native speakers.

2. Materials & Methods

2.1. Participants

The study included N=34 young adults recruited from the University of Memphis student body and Greater Memphis area. N=17 were native speakers of American English (7 males and 10 females) and N=17 were native speakers of Mandarin Chinese (7 males and 10 females). The two groups were closely matched in age (English: 24.9 ± 4.6 years; Chinese: 27.3 ± 3.5 years), years of education (English: 18.5 ± 3.76 years; Chinese: 20.8 ± 2.56 years), and musical training (English: 6.9 ± 6.4 years; Chinese: 5.3 ± 8.0 years). The majority of participants were right-handed (English: 60 % ± 60 %; Chinese: 61 % ± 44 %), as evaluated using the Edinburgh Handedness Inventory (Oldfield, 1971). All participants had normal hearing sensitivity, defined as pure tone thresholds of ≤ 25 dB HL at octave frequencies from 500 Hz to 8000 Hz in both ears. There was no history of speech, language, or neuropsychiatric disorders reported among participants.

We used a language history questionnaire to assess language background (e.g., Li et al., 2006). Our inclusion criteria for native Mandarin speakers were consistent with prior cross-language EEG studies (Bidelman et al., 2011; Blanco-Elorrieta et al., 2020; Chung & Bidelman, 2016). Chinese listeners were born and raised in China, with first exposure to English beginning in school around the age of 7.8 ± 2.63 years. They resided in the United States (US) during the experiment, with a duration of stay of 4.9 ± 3.61 years. Their self-reported English proficiency was moderate to high (4.9 ± 1.09, with a score of 7

[1] Leong (2012) coined the term "stress foot", which is also known as metrical or prosodic foot, to emphasize its holistic role in speech, blending rhythmic segments with suprasegmental features. In the current study, stress foot rhythm, or stress rhythm, is used interchangeably to denote the continuous suprasyllabic rhythm that arises from the process of assigning stress to string together multiple syllables.

indicating native-like proficiency). Each reported using native Mandarin approximately 59 % ± 17 % of their daily communication. Two Mandarin speakers who also speak Cantonese were excluded from the study because of potential confounds related to Cantonese listeners' advantages in stress perception (Choi, 2021). All participants provided written informed consent in accordance with a protocol approved by the University of Memphis Institutional Review Board and received compensation for their involvement.

2.2. EEG stimuli

Audio tokens of the syllables 'ba' and 'ma' were recorded by a male talker (2nd author) spoken in isolation with natural and similar loudness and pitch. These syllables were similar to stimuli used in our previous study on neural speech entrainment (He et al., 2023). Each syllable underwent temporal compression to a fixed duration of 120 ms and then concatenated to form pairs (e.g., 'baba') in Praat (Boersma & Weenink, 2013). We then separately manipulated syllable amplitude and duration to create four different stress tokens (e.g., 'BAba') where the first syllable was stressed, conforming to the trochaic foot (Fig. 1).

Amplitude-signaled stress pattern (Fig. 1 A & B)**.** High salient tokens were characterized by a 50 % higher amplitude between stressed and unstressed syllables, while low stress tokens were reduced to a 25 % contrast. However, each syllable maintained a uniform 120 ms duration across both salience levels.

Duration-signaled stress pattern (Fig. 1 C & D)**.** Similar with the amplitude condition, the duration contrast between stressed and unstressed syllables was marked at 50 % and 25 % for high and low saliences, respectively. High salient tokens featured 180 ms stressed syllables paired with 120 ms unstressed syllables, while low salient tokens contained 150 ms stressed syllables alongside 120 ms unstressed counterparts, all maintaining uniform amplitude.

Finally, each stress disyllable was concatenated (by inserting silence) to generate a continuous speech train of 6 s at different rates of 1, 2, and 3 Hz. These rates were chosen because English stress rhythm typically unfold with a nominal rate around 2 Hz (Dauer, 1983; Leong, 2012; Tilsen & Arvaniti, 2013). Altogether, we generated 12 stimulus conditions, each featuring a distinct stress pattern due to manipulation of acoustic cue (amplitude or duration), salience (high or low), and rhythm (1, 2, and 3 Hz).

2.3. Data acquisition and preprocessing

During electrophysiological recordings, participants comfortably reclined in front of a PC monitor and performed speech perception tasks in an electro-acoustically shielded booth (Industrial Acoustics Company). Binaural auditory stimuli were presented at 84 dB SPL through ER-2 insert earphones (Etymotic Research). Stimulus intensity was calibrated using a Larson-Davis SPL meter measured in a 2-cc coupler (IEC 60126). The presentation of stimulus and task instructions was managed by MATLAB 2013 (The MathWorks, 2013) directed to a TDT RP2 signal processing interface (Tucker-Davis Technologies).

Participants were instructed to listen to the speech streams and identify whether they heard "STRONG weak" or "weak STRONG" syllable sequences using a keyboard (with keys labeled as 'AAbb' or 'aaBB'). There were no time limits for the behavioral response. A subsequent trial commenced after the listener's response. Each stress stimulus condition comprised 10 trials (each 6 s). The presentation order of the conditions was randomized both within and across participants. Our behavioral task was primarily designed to keep subjects attentive and awake rather than as a comprehensive assessment of stress perception, per se.

Continuous EEG signals were recorded using Ag/AgCl disc electrodes placed at the mid-hairline and referenced to linked mastoids (A1/A2), with the mid-forehead serving as the ground. This single-channel montage is highly effective in recording entrained, auditory neural responses to speech (He et al., 2023) generated from the supratemporal plane including (though not exclusively) auditory cortex (Bidelman et al., 2013; Momtaz & Bidelman, 2024; Picton et al., 1999). Inter-electrode impedance was maintained < 10 kΩ. Continuous EEGs were digitized at a sampling rate of 1000 Hz using SynAmps RT amplifiers (Compumedics Neuroscan) and an online passband filter of 0–400 Hz.

Subsequent preprocessing was conducted using a customized MATLAB script. To focus on the slow electrophysiological activities, neural signals were further passband filtered (0.9–30 Hz; 10th order Butterworth). First, EEGs were segmented into individual 6-s epochs[2]–conforming to the length of the audio stimulus—and concatenated, resulting in 60 s of EEG data per condition. To minimize eye blink artifacts, we applied a wavelet-based denoising algorithm to the continuous EEGs (Khatun et al., 2016). Fig. 2**A** shows examples of one trial of EEG data corresponding to delta (stress) and theta (syllable) neural responses from the English group for the 1, 2, and 3 Hz stress rates, respectively. Fig. 2 **B** presents the corresponding spectrum of 60-s of continuous EEG data.

2.4. Electrophysiological data analysis

Phase Locking Value (PLV) and n:m Phase Synchronization Index (nmPSI) are bivariate time-series measures that quantify the degree of phase synchronization between two oscillators or time series. PLV computes the phase synchrony of two time series (e.g., acoustic and EEG signals) at a singular frequency (Assaneo & Poeppel, 2018; He et al., 2023; Lachaux et al., 1999). In contrast, nmPSI evaluates the cross-frequency phase coupling between two oscillators with distinct frequencies described by n and m (e.g., delta and theta frequency bands of EEG signals), where n:m is an integer relation (Leong et al., 2017; Rosenblum et al., 1998; Schack & Weiss, 2005). Conceptually, both PLV and nmPSI capture the temporal consistency in phase difference (and, conversely, the coherence) between two signals. Their resulting values range from 0 (no synchronization) to 1 (complete synchronization). PLV and nmPSI were computed using the following formulas:

$$PLV = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i[\theta_1(t)-\theta_2(t)]}\right| \tag{1}$$

$$nmPSI = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i[n\theta_1(t)-m\theta_2(t)]}\right| \tag{2}$$

Here, t denotes the discretized time, T is the total number of time points, and $\theta_1(t)$ and $\theta_2(t)$ are the Hilbert phases of the first and second signals, respectively.

The current study assessed synchronization between neural and acoustic speech signals using PLV at frequencies corresponding to stress rhythm (i.e., 1, 2, and 3 Hz) and syllable rhythm (i.e., 2, 4, and 6 Hz), respectively. This results in PLV_{Stress} representing brain-acoustic synchronization at the stress level and $PLV_{Syllable}$ reflecting brain-acoustic synchronization at the syllable level. We measured nmPSI to quantify the cross-frequency coupling within the brain's theta and delta frequency bands, corresponding to the alignment of nested syllable and stress rhythms unfolding at a 2:1 ratio. Specifically, frequency-specific neural signals and acoustic inputs were computed by applying passband filters around the frequencies of interest (±0.5 Hz) (see Fig. 2). The phase was extracted as the imaginary part of the signal's Hilbert transform. PLV was then computed between the EEG signal and acoustic stimulus waveform within each narrow frequency band and averaged over time per individual according to Equation (1). In contrast, nmPSI

[2] Due to data logging error, one participant yielded 9 epochs for the condition of amplitude stress cue of low salience at 3 Hz, and another participant yielded in 11 epochs for the condition of amplitude stress cue of high salience at 1 Hz.

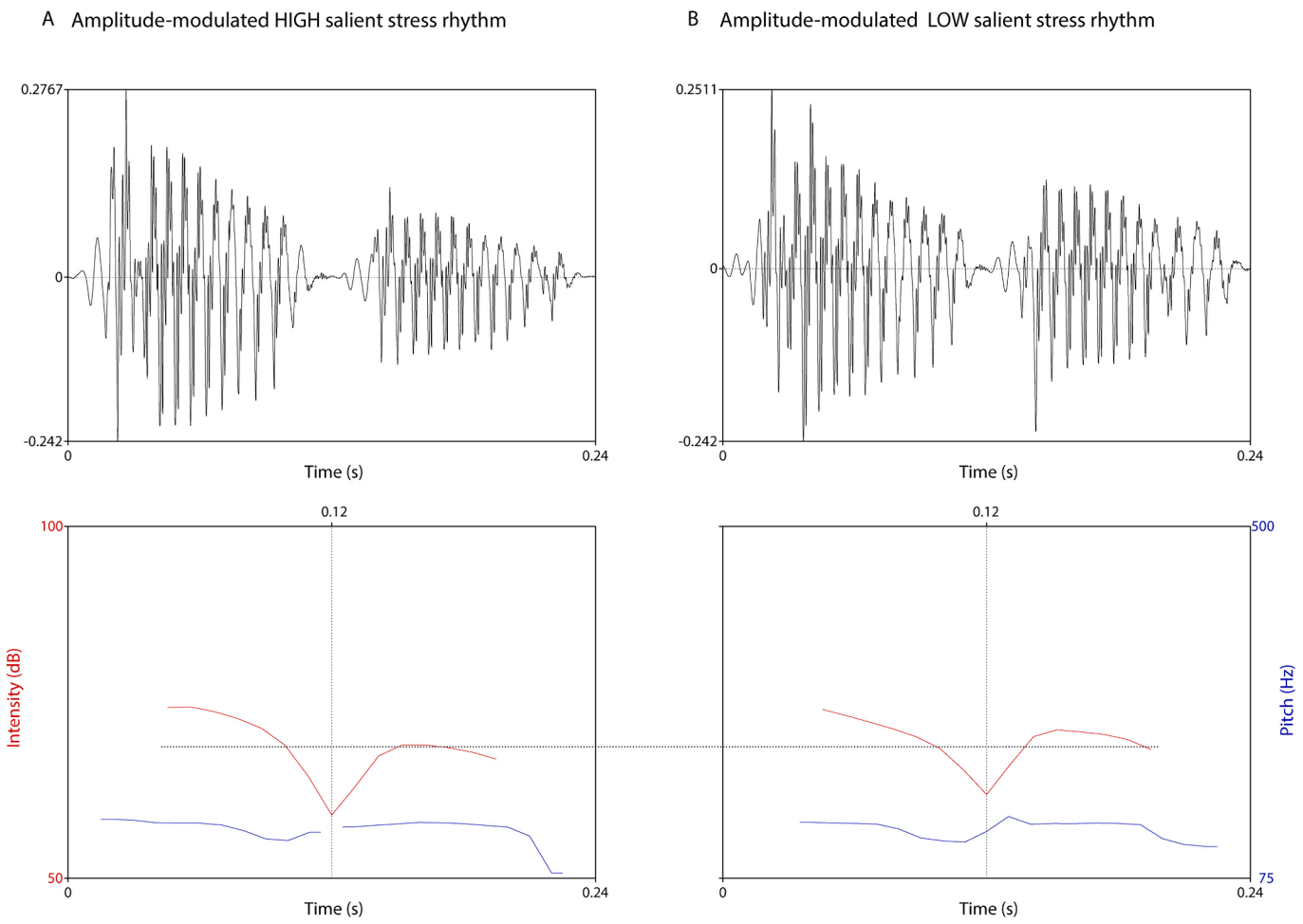


Fig. 1. Examples of stress stimuli used in the EEG study (i.e., a trochaic stress token). Disyllables were modulated by amplitude envelope (A & B) and syllable duration (C & D) with a high or low stress salience, respectively. black = speech sound stimulus; red = intensity contour; blue = pitch contour; dash lines mark the comparison between stressed and unstressed syllables in amplitude or duration. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

was computed by bandpass filtering the EEG data into two separate bands (i.e., m = 1, 2, 3 ± 0.5 Hz; n = 2, 4, 6 ± 0.5 Hz) to isolate phase-locked responses to the stress (m) and syllable (n) rhythm in the brain at a 2:1 ratio. To reduce noise in the metric, we quantified nmPSI in a moving window (6 sec; overlap ratio of 0.3) and averaged across windows for each condition according to Equation (2). Additionally, our study was concerned with relative comparisons between language groups and stimulus conditions rather than defining absolute values of PLV[3] and nmPSI, per se. Nevertheless, we established a noise floor of our PSI metric to account for potential baseline due to theta oscillations being twice the delta-stress rate (overlapping with its harmonics). We applied this identical nmPSI analysis to our previous EEG data evoked by similar syllable trains but devoid of any stress patterns (e.g., 'ba-ba-ba…') (He et al., 2023).

[3] We estimated the noise floor of the metric by computing PLV between two 60 sec samples of unique Gaussian noise. This revealed a PLV=0.0046, which can be taken as approximate noise floor of the metric. Actual PLV values measured in the observed EEG data were over an order of magnitude larger (see Figs. 3-4), suggesting brain-to-acoustic synchronization was both supra-threshold and non-random.

2.5. Statistical analysis

We conducted four-way mixed model analyses of variance (ANOVAs) in R (version 1.3.1073; 'lme4' package; Bates et al., 2014) to assess whether multi-scale brain-to-speech synchrony and cross-frequency coupling within the brain differed due to the acoustic stress patterns and language experience by measuring PLV_{Stress}, $PLV_{Syllable}$, nmPSI, and percent correct during behavioral stress perception. The model included within-subject factors of the stress cue (2 levels; amplitude vs. duration), stress salience (2 levels; high vs. low), and stress rate (3 levels; 1, 2, and 3 Hz) and a between-subject factor of group (2 levels; English vs. Chinese); subjects served as a random factor [e.g., PLV~cue*salience*rate*group+(1|sub)]. We used Tukey post hoc tests to correct for multiple comparisons (when applicable). Given our *a priori* hypothesis regarding potential enhancements of synchronization at 2 Hz (nominal English stress rhythm), following the initial omnibus ANOVA, we examined contrasts for nmPSI and PLV_{Stress} between 2 Hz versus the other stress rates. Similar contrasts were conducted for $PLV_{Syllable}$ between the nominal rate of 4 Hz versus others.

Furthermore, to assess associations between neural-neural and neural-acoustic synchrony measures, we used repeated measures correlations (rmCorr; Bakdash & Marusich, 2017). Unlike conventional correlations, rmCorr accounts for non-independence among each lis-

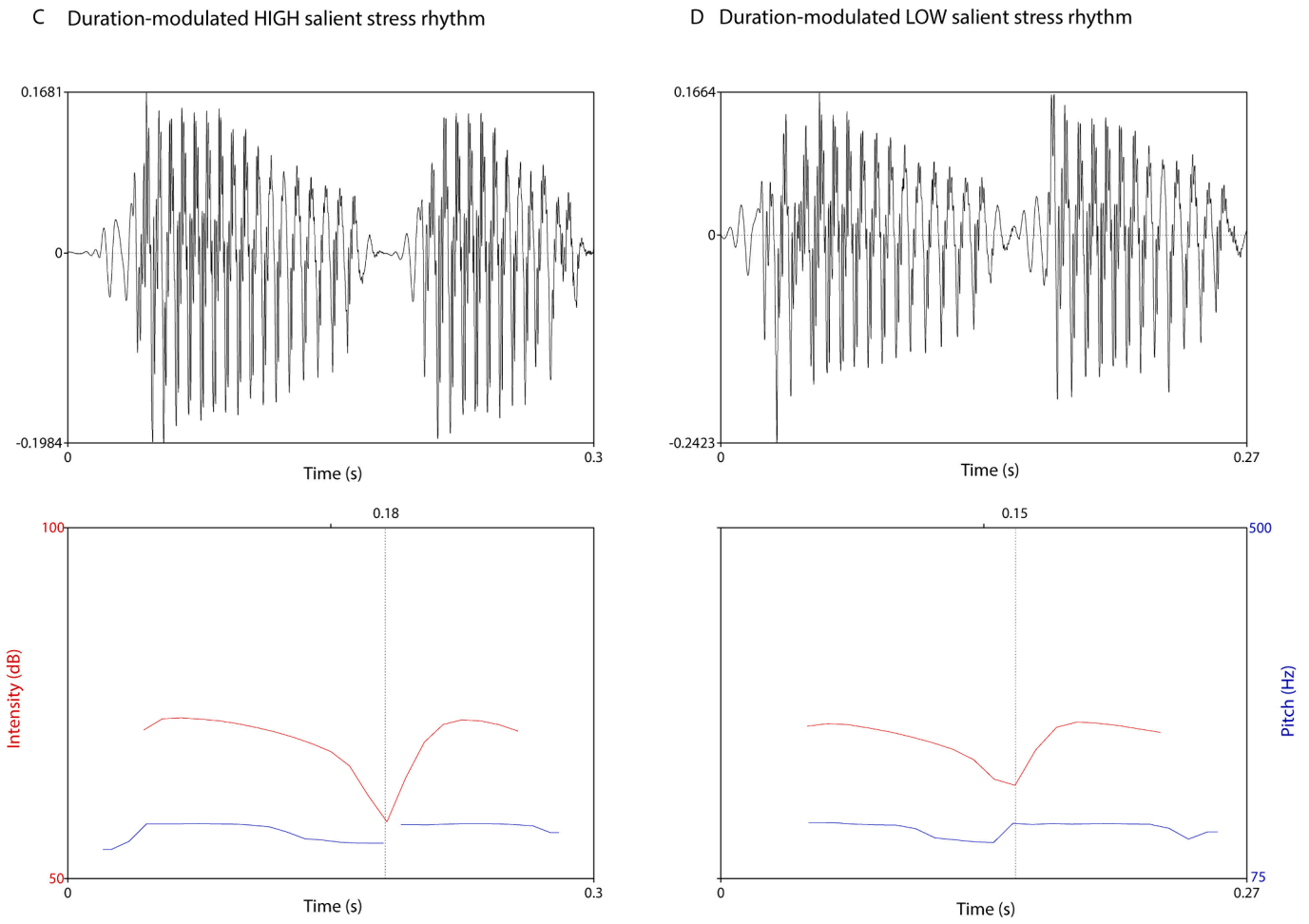


Fig. 1. (*continued*).

tener's observations and measures within-subject correlations by evaluating the common intra-individual association between two measures. Preliminary diagnostics (quantile–quantile plot and residual plots) were used to validate normality and homogeneity assumptions. Behavioral data from the EEG task (i.e., percentage of correctly perceived stress patterns) were rationalized arcsine transformed (Studebaker, 1985). A priori significance level was $\alpha = 0.05$. Effect sizes are presented as η_p^2.

3. Results

Correct percent performance on the behavioral task showed no significant group differences and results approached chance level (see Supplemental material; Fig. S1). However, this might be expected given that our task was primarily designed to keep subjects attentive and awake rather than as a comprehensive assessment of stress perception.

3.1. Brain to speech tracking at the stress level (PLV_{Stress})

We examined how neural oscillations phase lock to the external (i.e., acoustic) stress rhythms at rates of 1, 2, and 3 Hz (Fig. 3). An ANOVA conducted on PLV_{Stress} revealed significant main effects for group ($F_{1,32} = 4.69$, $p = 0.038$, $\eta_p^2 = 0.13$), stress rate ($F_{2,352} = 10.03$, $p = 0.0001$, $\eta_p^2 = 0.05$), cue ($F_{1,352} = 4.02$, $p = 0.046$, $\eta_p^2 = 0.01$), along with a two-way stress cue * salience interaction ($F_{1,352} = 8.65$, $p = 0.003$, $\eta_p^2 = 0.02$). Notably, English listeners demonstrated stronger brain-to-acoustic stress tracking than native Chinese speakers. The rate effect was attributed to stress rhythms at 1 and 2 Hz eliciting greater PLV_{Stress} ($p_{1\ vs.\ 3\ Hz} = 0.001$; $p_{2\ vs.\ 3\ Hz} = 0.0001$) compared to 3 Hz. The interaction of cue*salience arose from enhanced PLV_{Stress} for amplitude- ($p < 0.001$) compared to duration-signaled high salient stress, suggesting a neural preference of amplitude cues for both groups. Also, when stress was signaled by duration, we found higher PLV_{Stress} for low compared to high stress salience ($p < 0.001$). These results demonstrate differences in exogenous neural-acoustic synchronization across individuals' language experience and stress rhythm acoustic features.

Motivated by the predominant rate of 2 Hz in natural English stress rhythms (Dauer, 1983; Leong, 2012; Tilsen & Arvaniti, 2013) and the inverted-V rate pattern depicted in Fig. 3, we conducted an *a priori* contrast of 2 Hz against other rates by group and stress cue. Our assumption was confirmed in that PLV_{Stress} peaked at 2 Hz exclusively for English speakers ($p = 0.0001$) under duration-modulated stress. Interestingly, this enhancement was absent for Chinese whose native language does not include English-based stress patterns ($p_{amplitude} = 0.98$; $p_{duration} = 0.373$). These results demonstrate an enhancement of speech-to-brain phase-locking (PLV_{Stress}) at the frequency inherent to natural English stress rhythm (2 Hz) that differs by individuals' language

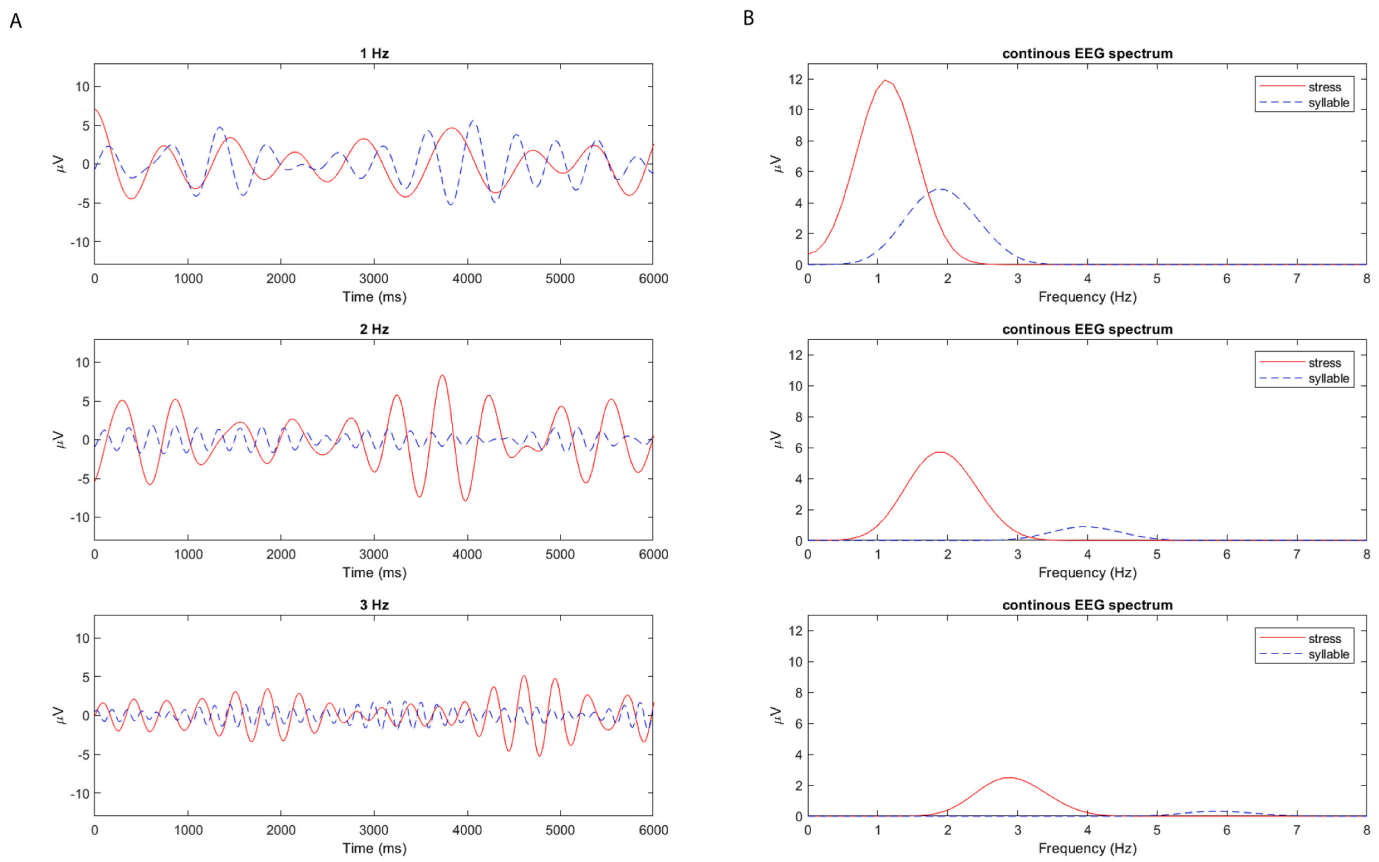


Fig. 2. Continuous EEGs show phase-locking to stress and syllable rhythms. Shown here is an example from EEGs of the initial trial (6 s) that were averaged across English individuals for the amplitude high stress condition. (A) neural phase coupling represents the phase hierarchy in speech rhythms. (B) Spectrum analysis quantifies the intensity of each frequency of interest. Red = frequencies of stress rhythm; blue = frequencies of syllable rhythm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

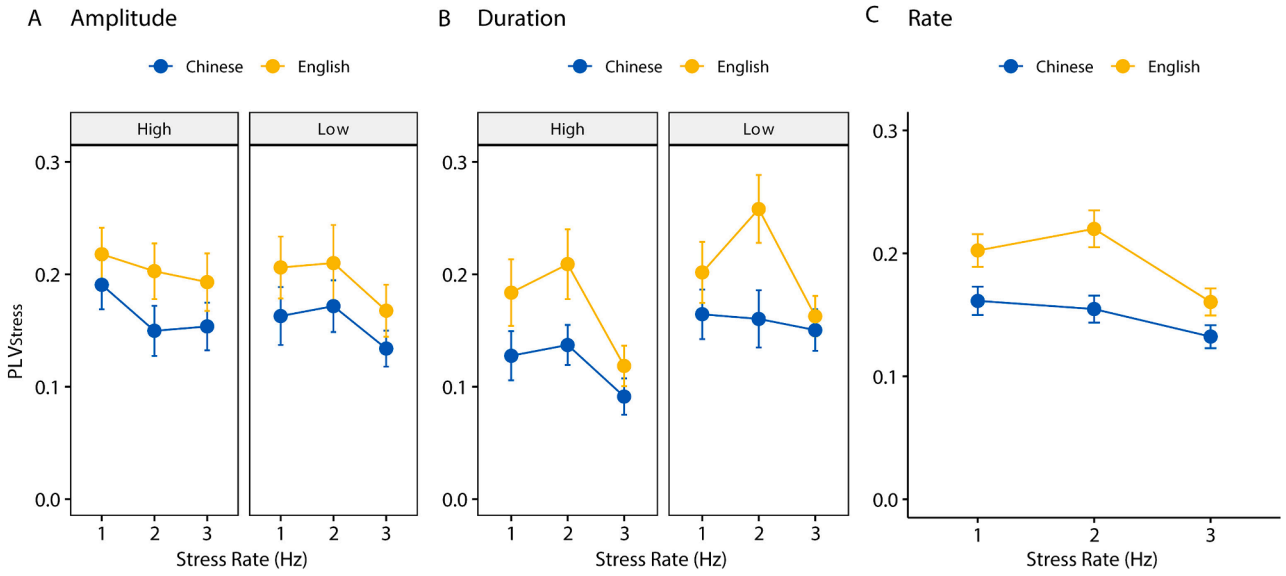


Fig. 3. Brain oscillations synchronize to the rate of *stress rhythms*. Cross-linguistic PLV_{Stress} comparisons by stress rate and salience signaled by (A) amplitude envelope and (B) syllable duration. (C) English differed from Chinese speakers across rates and exhibited PLV_{Stress} enhancement at 2 Hz—corresponding to the natural stress rate found in English. Panel C outlines the main effect of rate, aggregating data across cue and salience conditions. PLV_{Stress} refers to neural-to-stress rhythm phase locking; errorbars = ± 1 s.e.m.

exposure.

3.2. Brain to speech tracking at syllable level ($PLV_{Syllable}$)

As our stimuli simultaneously carried stress and syllable rhythms organized as hierarchical tiers, we next proceeded to test the extent to which neural oscillations phase lock to the acoustic syllable rates[4] of 2, 4, and 6 Hz, which are 2-times faster than stress rhythms. An ANOVA conducted on $PLV_{Syllable}$ revealed a main effect of syllable rate ($F_{2,352} = 5.93$, $p = 0.003$, $n_p^2 = 0.03$) and two-way interactions of stress cue * group ($F_{1,352} = 5.79$, $p = 0.017$,$n_p^2 = 0.02$) and stress cue * rate ($F_{2,352} = 5.53$, $p = 0.004$, $n_p^2 = 0.03$) (Fig. 4). Pairwise comparisons revealed English speakers had higher $PLV_{Syllable}$ than Chinese speakers under amplitude- ($p = 0.05$) but not duration-signaled stress. This group difference was further Tukey pairwise compared by rates and was only significant at a syllable rate of 4 Hz under amplitude-indicated stress patterns ($p = 0.005$) motivated by the cue* rate interaction. For English speakers, amplitude-signaled stress rhythm had a stronger $PLV_{Syllable}$ than duration-signaled stress ($p = 0.029$), consistent with PLV_{Stress}. The stress cue * rate interaction was driven by stronger $PLV_{Syllable}$ for duration cues at 3 Hz compared to other rates ($p_{3\text{ vs. }2\text{ Hz}} < 0.001$; $p_{3\text{ vs. }2\text{ Hz}} = 0.001$). Additionally, at 3 Hz, duration cues evoked higher $PLV_{Syllable}$ than amplitude cues.

Consistent with PLV_{Stress}, we performed an *a priori* contrast involving the 4 Hz-syllable rate (that is, 2 Hz-stress rate) versus other rates (i.e., 2 and 6 Hz syllable rates) by group and stress cue. We observed an enhanced $PLV_{Syllable}$ at 4 Hz ($p = 0.011$) only in English speakers for amplitude cues, and not for Chinese speakers. Notably, 4 Hz closely aligns with the mean syllable rate in English (Goswami & Leong, 2013; Greenberg et al., 2003; Tilsen & Johnson, 2008) and many other languages, including Chinese (Ding et al., 2017). These results demonstrate an enhancement of speech-to-brain phase-locking ($PLV_{Syllable}$) at the frequency inherent to natural English rhythm (2 Hz) that, like our stress findings, also differs by individuals' language exposure.

3.3. Cross-frequency coupling within the brain (nmPSI)

Fig. 5 illustrates delta-theta phase coupling *within the brain* as measured by nmPSI. Results yielded significant main effects, including group ($F_{1,32} = 90.42$, $p < 0.0001$,$n_p^2 = 0.74$), stress rate ($F_{2,352} = 157.82$, $p < 0.0001$,$n_p^2 = 0.47$), cue ($F_{1,352} = 91.03$, $p < 0.0001$,$n_p^2 = 0.21$), and salience ($F_{2,352} = 87.88$, $p < 0.001$,$n_p^2 = 0.20$). Post hoc analysis indicated peak nmPSI occurred at a stress rate of 1 Hz, declining gradually with faster rates (*all* $p < 0.0001$) for both groups. Moreover, English speakers had greater nmPSI than Chinese speakers across all stress rates (Fig. 5**C**). In addition, we found a significant three-way interaction of cue * salience * group ($F_{1,352} = 75.90$, $p < 0.0001$, $n_p^2 = 0.18$), which was attributed to English speakers having stronger nmPSI relative to Chinese speakers for low salient stress stimuli ($p_{amplitude} < 0.0001$; $p_{duration} = 0.0026$) (Fig. 5**A and B**, right panels). However, no group differences were observed for more salient (i.e., high) stress stimuli—true for both stress cues (Fig. 5**A & B**, left panels). Generally speaking, both Chinese and English speakers exhibited nmPSI above baseline, indicating internal delta-theta coupling represented the stress-syllable hierarchy of our stimuli above what would be expected by random variation alone. Consistent with neural-acoustic findings for both stress and syllables, the phase hierarchy in the brain was enhanced in English listeners only.

Fig. 6 illustrates the data broken down by stress cue (amplitude vs. duration), cue salience, and group to emphasize language-specific cue differences. Pairwise comparisons revealed that nmPSI differences in stress cue and salience were only observable for the Chinese group (Fig. 6**B**), where high salient stress resulted in higher nmPSI compared to low salience for both cues ($p_{amplitude} < 0.0001$; $p_{duration} = 0.043$). Also, under low stress salience, duration-related stress had higher nmPSI than amplitude stress ($p < 0.0001$) within Chinese. Critically, there were no significant nmPSI differences due to acoustic stress cue nor salience for native English speakers (Fig. 6**A**; 3-way ANOVA: $p_{cue} = 0.727$; $p_{salience} = 0.196$). These results support prior findings by observing that neural coherence, as measured by nmPSI, was equally robust in both acoustic parameters for English listeners, but that Chinese listeners' nmPSI was lower for amplitude modulated stress.

3.4. Correlations

To explore the association between internal brain cross-frequency synchronization and external brain-speech synchronization, we conducted within-subject correlations using rmCorr for all the feasible pairwise variables (i.e., nmPSI, PLV_{Stress}, $PLV_{Syllable}$). Fig. 7 depicts a positive correlation between nmPSI and PLV_{Stress} for English ($r = 0.23$, $p = 0.002$) but not Chinese ($r = 0.07$, $p = 0.352$) speakers. English individuals exhibiting stronger internal cross-frequency coupling also demonstrated better external brain tracking of stress rhythm. These results were corroborated by between-subject Pearson's correlation (English: $r = 0.170$, $p = 0.017$; Chinese: $r = 0.082$, $p = 0.245$).

4. Discussion

Here, we provide new evidence that neural oscillations across multiple time scales mirror the hierarchical nature of the acoustic stress rhythm in speech and do so in a language-dependent manner. Specifically, phase synchrony measures revealed five key findings: (i) brain oscillations at multiple temporal scales (delta and theta) concurrently phase locked to the rates of stress and syllable rhythms, (ii) amplitude was a more robust stress indicator than duration; (iii) only English speakers demonstrated enhanced multiscale brain-to-speech tracking at the dominant stress rate (2 Hz) and syllable rate (4 Hz) characteristic of natural English, while this phenomenon was absent in Chinese speakers; (iv) both English and Chinese individuals showed delta-theta phase coupling within the brain that mirrors the stress-syllable hierarchy in natural speech but such coupling was stronger in native English listeners; (v) individuals with superior nesting of neural oscillations (i.e., English listeners) showed enhanced cortical-acoustic tracking to stress. Collectively, our findings suggest brain entrainment mechanisms coding aspects of speech-language are not solely acoustic-induced responses but benefit from phonological knowledge gained from sustained experiences of speaking and listening to a stress-dominant language.

4.1. Cortical encoding of stress rhythm via delta phase-locking depends on language experience

Prominent oscillatory-based models (e.g., TEMPO, AST) of language temporal processing have generally overlooked the delta band of the EEG which corresponds to slower-than-syllable rhythms (Ghitza, 2011; Ghitza & Greenberg, 2009; Hickok & Poeppel, 2007; Poeppel, 2003). Our PLV_{Stress} findings show delta oscillations phase-lock to slower (<4 Hz) acoustic regularities, explicitly tagging the stress rhythms in English. Notably, such neural-audio synchronization is modulated by various acoustic attributes (i.e., stress rate and cue type) and diminishes in individuals with a foreign language background. Prior studies have assumed delta oscillation retain an analogous role as theta, parsing continuous speech into sequential delta-size chunks (Giraud & Poeppel, 2012; Rimmele et al., 2021). However, critical to our findings is the

[4] In the duration condition, while the stimuli preserve the overall syllable rhythm, the duration manipulation between stressed and unstressed syllables inevitably leads to non-isochronous syllables which might create jitter in the response and weaken $PLV_{Syllable}$. However, $PLV_{Syllable}$ magnitudes were, on average, similar between amplitude and duration-cuing stress (e.g., Fig. 4) suggesting any jitter introduced in our stimuli did not negatively impact $PLV_{Syllable}$.

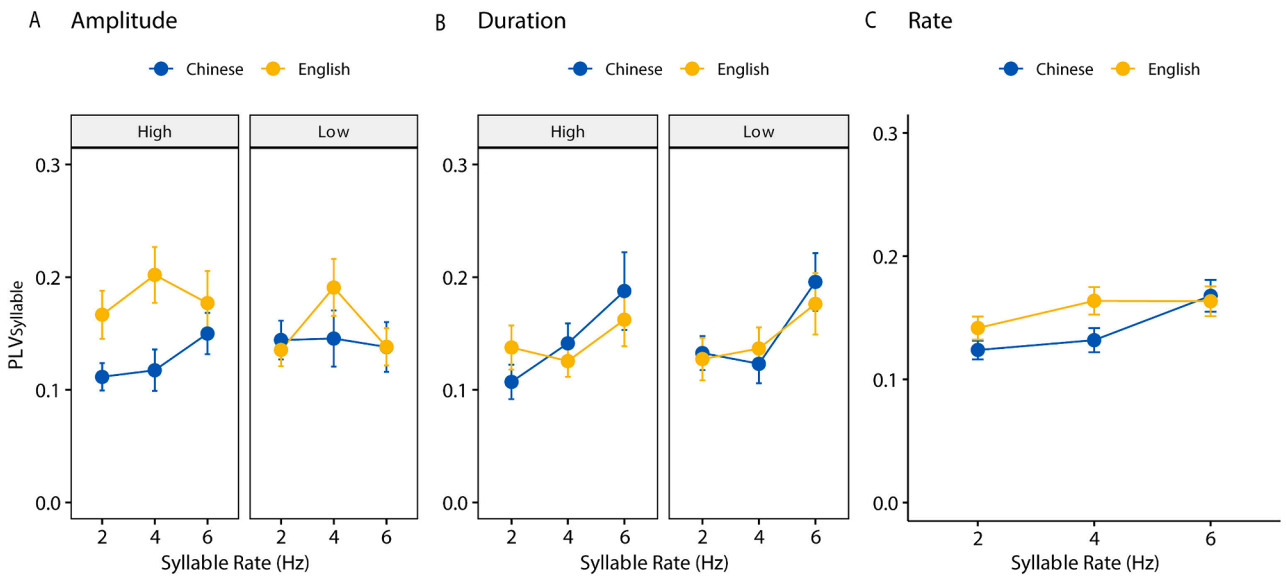


Fig. 4. Brain oscillations phase lock to the rate of *syllable rhythms*. Cross-linguistic $PLV_{Syllable}$ comparisons by stress rate and salience signaled by (A) amplitude envelope and (B) syllable duration. (C) $PLV_{Syllable}$ enhancement at syllable rhythm of 4 Hz, matching the center syllable rate across many languages, was exclusively observed in English speakers, not Chinese. Panel C outlines the main effect of rate, aggregating data across cue and salience conditions. $PLV_{Syllable}$ refers to neural-to-syllable rhythm phase locking; error bars = ± 1 s.e.m.

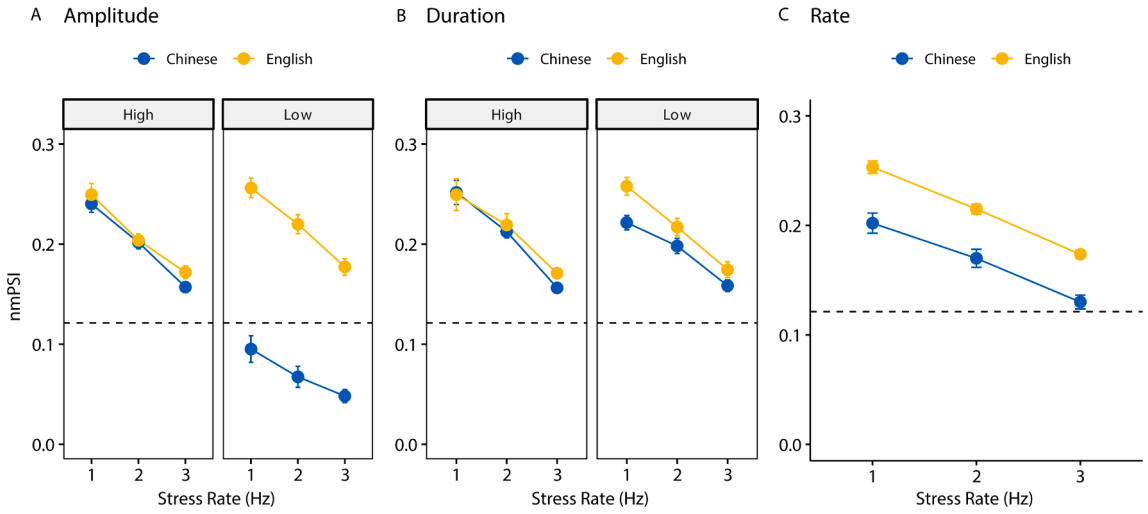


Fig. 5. Phase coupling of delta-theta neural oscillations represents the phase hierarchy of stress-syllable rhythms. Cross-language comparisons of nmPSI as a function of stress rate and salience modulated by (A) amplitude envelope and (B) syllable duration. (C) Significant group differences in nmPSI across stress rhythm rate with a dataset aggregated across cue and salience conditions. Dashed lines = nmPSI baseline computed for stress-free syllable trains from He et al. (2023). error bars = ± 1 s.e.m.

proposition that delta oscillations are associated with the hierarchical nesting role of stress rhythms.

Speculatively, we propose that delta oscillations might serve a higher-order mechanism, extending beyond simple stress segmentation to facilitate temporal integration and establish a cohesive phonological representation, possibly via delta modulation of theta activity (Gross et al., 2013; Lakatos et al., 2005; Morillon et al., 2019). Indeed, this nesting function of delta is confirmed by our cross-frequency phase coupling analysis (i.e., nmPSI), evident during the processing of stress patterns. Presumably, delta oscillations coordinate syllable nesting and stress segmentation to streamline ongoing speech processing. Our premise is particularly compelling given the large numbers of individual syllables in connected speech and the consequent cognitive demands on memory and attention, which are consistent with the increased delta activity in working memory where attention is focused on an internal representation (Bidelman et al., 2021; Harmony, 2013). Moreover, this converges with neuroimaging evidence pointing to lexical and semantic grouping via delta oscillatory activities, even in the absence of acoustic boundary cues (Ding et al., 2016; Lo et al., 2022; Meyer et al., 2017).

Furthermore, we found English listeners exhibited stronger phase encoding of stress patterns compared to Chinese speakers, independent of stress cue, rate, and salience. These findings are in line with previous cross-language, electrophysiological studies which show differential brain responses in native vs. nonnative listeners to stress information (Chung & Bidelman, 2016). English listeners' superior encoding and tracking of stress patterns could lie in their heightened perceptual sensitivity and detection accuracy of stress patterns (Chrabaszcz et al., 2014; Qin et al., 2017). Conversely, Mandarin speakers' poorer synchronization to ongoing acoustic cues that are essential for discerning English stress patterns is likely due to their more limited exposure and

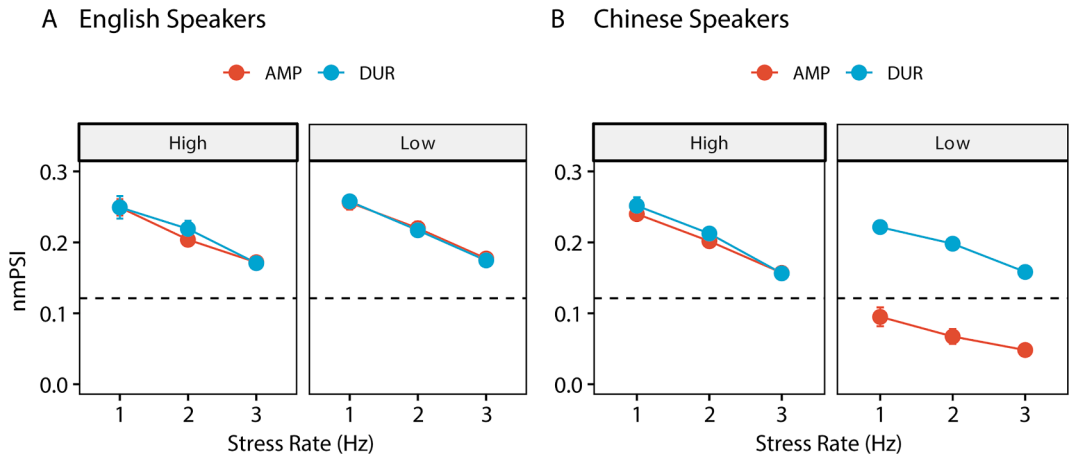


Fig. 6. Effects of acoustic stress cue and salience on cross-frequency coupling. (A) nmPSI in English speakers was invariant to acoustic stress manipulations (B) Contrastively, Chinese listeners' nmPSI was more prone towards high salience and duration cue, suggesting stronger coupling of their nested brain oscillations in these conditions. Dashed lines = nmPSI baseline computed for stress-free syllable train perception. error bars = ± 1 s.e.m.

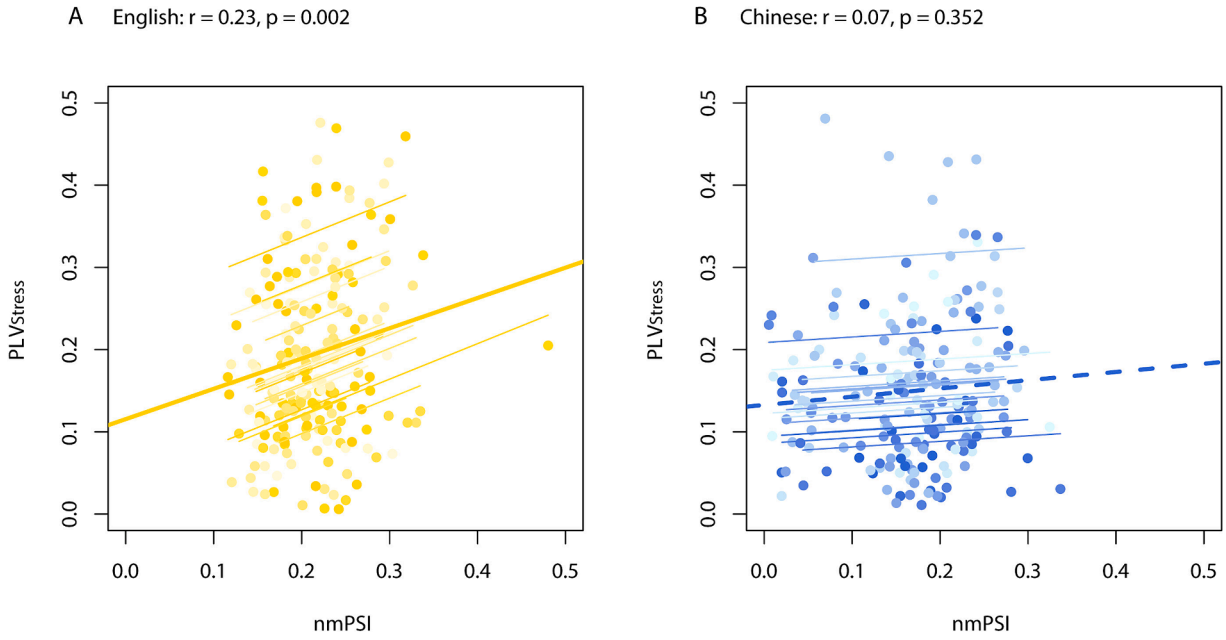


Fig. 7. Repeated measure correlations between internal cross-frequency coupling and external audio-neuro tracking in (A) English speakers, and (B) native Mandarin-Chinese speakers. PLV_{Stress} refers to neural-to-stress rhythm phase locking; nmPSI represents the phase coupling of delta-theta neural oscillations to hierarchical stress rhythms. Dots/thin lines = individual data; solid thick line = significant overall relation; dotted thick line = *n.s.* relation.

motor practice with a stress-dominant language (Chung & Jarmulowicz, 2017; Ding et al., 2020). Such experience-dependent effects emerge in both groups' EEG. Chinese responses were severely hindered by stress manipulations whereas English responses were largely impervious (Fig. 6). Consequently, Chinese speakers' struggle to capture acoustic stress regularities in ongoing speech and subsequent failure to segment delta-size chunks might be due to a "perceptual narrowing" of speech representations that are not behaviorally relevant cues in Mandarin (Jeng et al., 2011; Tierney & Nelson III, 2009). For example, compared to native English speakers, Chinese learners of English rely more heavily on fundamental frequency—which is more relevant to their native tonal language—than duration or intensity (Chung & Bidelman, 2021; Chung et al., 2021; Wang, 2008). Perceptual narrowing due to synaptic neural pruning could manifest at the macroscopic level in the less synchronized brain-to-speech oscillations we find in our EEG data. This lack of coherence, readily achieved by native English speakers, further suggests delta oscillatory synchronization is not merely a passive "bottom-up" mechanism. Rather, we suggest it is sculpted by "top-down" regulation fostered by a listener's lifetime of sensory experiences and accumulated phonological stress knowledge inherent to speaking a specific language.

4.2. Multilevel brain-to-speech synchronization is optimized for the natural rate of stress rhythms

Research has emphasized the importance of amplitude envelope in the brain's neural entrainment to speech at the syllable level (Assaneo & Poeppel, 2018; He et al., 2023). Our $PLV_{Syllable}$ findings further show that syllabic-theta synchronization is critical to suprasegmental processing of stress. Similar dual-frequency synchronization has been observed in intelligible story listening (Gross et al., 2013; Park et al., 2015) and other cognitive tasks (Palva & Palva, 2018). Such nesting of brain responses might be necessary for stress processing since it simultaneously occurs on two distinct timescales (theta-syllable; delta-stress) and each band could track different frequency-specific acoustic

information. However, such architecture does not necessary require there be an exhaustive linear division of the incoming speech signal into individual segments (Ghitza, 2013). Rather, a heterodyning of neural oscillation might help establish hierarchical time resolution windows that synchronize to different features of the input (e.g., syllable vs. stress). Corroborated by our PLV and nmPSI measures, our data converge with prior models of language processing that, at least theoretically, can be described as a series of coupled neural oscillators carrying different features of the linguistic signal (Ding et al., 2016; Ghitza, 2011; Hickok & Poeppel, 2007; Park et al., 2015). Our work extends such frameworks by implicating multi-time resolution and experience-dependent plasticity to these models.

Furthermore, we observed differences in cortical-acoustic synchronization across syllable and stress rates. English (but not Chinese) speakers demonstrated enhanced PLV_{Stress} at 2 Hz, closely aligning with the nominal speed of English stress (Dauer, 1983; Leong, 2012; Tilsen & Arvaniti, 2013). Interestingly, we found a similar phase-locking enhancement at 4 Hz, the dominant syllable rate typical for many languages (Ding et al., 2017; Greenberg et al., 2003; Greenberg et al., 1996; Tilsen & Johnson, 2008), that was evident in English speakers but absent in Chinese speakers. This contradicts previous assertions that cross-linguistic differences in neural-acoustic synchronization only appear at the supra-syllabic (but not syllabic) level (Blanco-Elorrieta et al., 2020; Ding et al., 2016; Rimmele et al., 2023), simply because the latter is similar across languages (Ding et al., 2017). However, our analysis further confirmed that the group differences at syllabic level were exclusively marked at the 4 Hz syllable rate. These findings suggest that neural enhancement at the universal syllable rate (4 Hz) might disappear when processing syllables within a foreign *stress context*. As evidenced by our PLV_{Stress} results, the absence of 4 Hz syllabic enhancements in Chinese speakers presumably results from their limited neural coding of stress patterns at the supra-syllabic level (here 2 Hz). Alignment of brain activity to dominant natural rhythms is the key to observing enhancements in neural-speech entrainment (He et al., 2023). The lack of such effect in nonnative listeners implies that a failure to synchronize with higher-order properties of the speech signal (i.e., stress rhythm) might actually impede essential neural processing at lower levels of the hierarchy (i.e., syllable tracking). Future studies are needed to fully test this possibility.

4.3. Neural coupling of delta-theta oscillations mirrors phase hierarchy between speech rhythms

To empirically test for hierarchical relations between frequency-specific neural oscillations, we measured *n:m* phase synchrony within the EEG, which can be intuitively described as the ongoing phase of *n*-cycles of an oscillation synchronizing with *m*-cycles of another oscillation (Leong, 2012; Schack & Weiss, 2005). Unlike our PLV analysis, which reflects the brain's tracking of sound features of the external acoustic signal, nmPSI reflects oscillatory coupling internal to the brain (*brain-to-brain* synchronization). Most of our nmPSI results uncovered significant phase-phase coupling between delta and theta neural oscillations that were above the noise floor computed from non-stressed stimuli for both English and Chinese speakers,[5] closely mirroring the phase hierarchy carried by acoustic stress and syllable envelopes. Moreover, Chinese listeners demonstrated similar delta-theta coherence as English speakers under high stress salience, which was not observed in external neuro-stress tracking (i.e., PLV). These findings indicate a robust neural hierarchy of delta and theta oscillations, even when listeners are less experienced with stress rhythm. Notably, native English speakers showed enhanced nmPSI compared to Chinese speakers, suggesting that delta-theta coherence cannot be solely attributed to passive harmonic nesting. In contrast, our results highlight the language-specific effects on the hierarchical neural processing of speech temporal information. Additional examples of hierarchical coupling stems from studies showing increased delta-theta phase-amplitude coupling during intelligible story perception (Gross et al., 2013). Thus, the existence of such nesting in multiple domains of speech processing suggests delta oscillations might play a higher-order role, reorganizing both the phase and amplitude behaviors of theta oscillators that code different properties of the linguistic signal, stress or otherwise.

Converging with our multilevel PLV results, nmPSI measures also demonstrated hierarchical nesting between neural oscillations. These findings demonstrate that ongoing auditory delta oscillations become synced with the *external* acoustic stress regularities which might then formulate an oscillatory hierarchy *internal to the brain* during speech processing, or vice versa. Supporting this notion, we found significant correlations between nmPSI and PLV_{Stress}, indicating that a higher degree of internal hierarchical coherence predicts the external alignment of auditory oscillations with stress patterns, or vice versa. Our findings establish a new, heretofore unrecognized relationship between internal neural coherence and external neural tracking across multiple scales, that also varies in a language-dependent manner.

4.4. Amplitude cues dominate the neural encoding of stress

Another aim of our study was to evaluate how different acoustic attributes of stress entrain brain oscillations in native vs. non-native speakers. Though English listeners outperformed Chinese listeners in PLV_{Stress}, brain-to-acoustic tracking was generally enhanced for stress patterns carried by amplitude compared to duration cues regardless of group. Additionally, English listeners showed more robust syllable tracking ($PLV_{Syllable}$) than Chinese individuals for amplitude cues. These findings imply that amplitude-signaled stress more effectively fosters delta-stress synchronization and, at least in English speakers, improves syllabic neural tracking. In general, our data suggest that amplitude cues are more perceptually salient to distinguish stress patterns for both English and Chinese speakers, consistent with prior studies (Chrabaszcz et al., 2014; Zeng et al., 2022). Furthermore, our findings reinforce the "iambic-trochaic law" in linguistics, which posits an innate tendency for intensity-contrasting elements to be perceived as trochaic stress (Strong-weak patterns)—the characteristic of our stimuli—whereas duration-varying components lean towards iambic perception (Crowhurst, 2020; Hay & Diehl, 2007; Hayes, 1995).

However, cross-frequency coupling within the brain also revealed distinct acoustic preferences between language groups. For English speakers, nmPSI values were invariant to acoustic stress cue type (amplitude ≈ duration) and salience (high ≈ low), indicating remarkable stability in delta-theta brain coherence among native speakers even in scenarios of weak stress cues. Contrastively, Chinese speakers showed significant acoustic-driven effects in nmPSI, with stronger coherence for more salient stimuli and duration vs. amplitude cues. This indicates that their neural coherence induced by (English) stress patterns is perhaps more vulnerable to acoustic variations.

Alternatively, this acoustic modulation on nmPSI observed in native Chinese speakers who are also second language learners of English may reflect a less ingrained but more flexible mechanism that allows a listener to adapt to unfamiliar rhythms. This could be highly beneficial to non-native language learning, as is the case for our Chinese listeners. Indeed, it is worth noting that for duration-based stimuli, nmPSI in Chinese listeners exceeded the baseline nmPSI to stress-free syllable trains. Speculatively, this implies that even non-native listeners may have attempted to construct an internalized hierarchy for duration-based stress stimuli. A possible explanation may lie in the inherent structure of the Chinese language. Chinese might possess a phonological hierarchy similar to English, but one that is organized by tonal instead of stress rules (Duanmu, 2007; McCawley, 1978). For instance, Chinese syllables that carry lexical tone are longer than their weak neutral (non-

[5] The nmPSI responses in Chinese speakers were below baseline for low salience, amplitude-signaled stress.

tone) counterparts. And a supra-syllabic unit emerges by following the rule that neutral tones only present after those with tones (Duanmu, 2004; Li et al., 2014). Such duration-related phonology in Chinese may transfer as a cue-weighting strategy to process an unfamiliar stress hierarchy (Holt & Lotto, 2006; Zhang & Francis, 2010), a leading to effective but diminished delta-theta brain coherence.

Unfortunately, our single-channel EEG montage did not allow the disentanglement and localization of the sources of brain oscillations. Future research obtaining neural recordings for multi-rhythmic stimuli with enhanced spatial and temporal resolutions may help to isolate varying responses to these rhythms. Additionally, we acknowledge that our interpretations are somewhat speculative given the use of a priori contrasts between 2 Hz vs. other stress rates. Future confirmatory studies are needed to verify the robustness of our findings regarding neural tracking at the dominant stress rhythm. Similarly, our phase analysis prioritized cross-language group comparison over absolute degree of synchrony. Though baselining PLV responses is common when assessing absolute phase-locking (e.g., Assaneo & Poeppel, 2018; He et al., 2023), our analyses were concerned with *relative* comparisons between groups that render absolute measures requiring baseline correction moot. The lack of baseline adjustments in the current PLV analyses may weaken the robustness and increase the variability of our brain-acoustic synchronization magnitudes. Still, the large group differences we find coupled with low intra-group variability suggests this was not problematic. Moreover, we note that all PLV values were nearly an order of magnitude above noise floor estimates for the metric (see footnote #3). While unlikely, we cannot rule out the possibility that the observed cross-frequency coupling was instead driven by the processing of stress, especially in English speakers, rather than by phase relations between syllables and stress.

4.5. Conclusions

Collectively, our data demonstrate an intricate interplay between neural oscillations, speech rhythms, and stress hierarchical phonology, providing a new dimension to our understanding of perceptual speech processing. Our findings bridge several gaps, showing multiple timescales of neural oscillations internally cohere and externally synchronize with syllable and stress rhythm. Crucially, individual variations in hierarchical coherence internal to the brain predict their external entrainment ability, and vice versa, essentially reshaping the brain's engagement with the rhythmic essence of speech. English speakers displayed native advantages in oscillatory synchrony during stress encoding, emphasizing benefits from "top-down" processing rooted in their lifetime exposure to a stress-dominant language. Our results highlight the critical role of brain oscillations in tracking and encoding stress and syllable rhythms in a language-dependent manner.

CRediT authorship contribution statement

Deling He: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Conceptualization. **Eugene H. Buder:** Writing – review & editing, Supervision, Conceptualization. **Gavin M. Bidelman:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Funding acquisition, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Requests for data and materials should be directed to G.M.B [gbidel@iu.edu]. This work was supported by the Institute for Intelligent Systems Student Organization Thesis/Dissertation Award funding awarded to D.H. and the National Institute on Deafness and Other Communication Disorders (R01DC016267) awarded to G.M.B.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bandl.2024.105463.

References

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica, 66*(1–2), 46–63.
Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science advances*, *4* (2), eaao3842.
Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.
Bidelman, G. M., Brown, J. A., & Bashivan, P. (2021). Auditory cortex supports verbal working memory capacity. *NeuroReport, 32*(2), 163.
Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of cognitive neuroscience, 23*(2), 425–434.
Bidelman, G. M., Moreno, S., & Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage, 79*, 201–212.
Blanco-Elorrieta, E., Ding, N., Pylkkänen, L., & Poeppel, D. (2020). Understanding requires tracking: Noise and knowledge interact in bilingual comprehension. *Journal of cognitive neuroscience, 32*(10), 1975–1983.
Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.51) [Computer software]. https://www.fon.hum.uva.nl/praat.
Boucher, V. J., Gilbert, A. C., & Jemel, B. (2019). The role of low-frequency neural oscillations in speech processing: Revisiting delta entrainment. *Journal of cognitive neuroscience, 31*(8), 1205–1215.
Bröhl, F., & Kayser, C. (2021). Delta/theta band EEG differentially tracks low and high frequency speech-derived envelopes. *Neuroimage, 233*, Article 117958.
Choi, W. (2021). Cantonese advantage on English stress perception: Constraints and neural underpinnings. *Neuropsychologia, 158*, Article 107888.
Chrabaszcz, A., Winn, M., Lin, C. Y., & Idsardi, W. J. (2014). Acoustic cues to perception of word stress by English, Mandarin, and Russian speakers. *Journal of Speech, Language, and Hearing Research, 57*(4), 1468–1479.
Chung, W.-L., & Bidelman, G. M. (2016). Cortical encoding and neurophysiological tracking of intensity and pitch cues signaling English stress patterns in native and nonnative speakers. *Brain and language, 155*, 49–57.
Chung, W.-L., & Bidelman, G. M. (2021). Mandarin-speaking preschoolers' pitch discrimination, prosodic and phonological awareness, and their relation to receptive vocabulary and reading abilities. *Reading and Writing, 34*(2), 337–353. https://doi.org/10.1007/s11145-020-10075-9
Chung, W.-L., & Jarmulowicz, L. (2017). Stress judgment and production in English derivation, and word reading in adult Mandarin-speaking English learners. *Journal of Psycholinguistic Research, 46*, 997–1017.
Chung, W.-L., Jarmulowicz, L., & Bidelman, G. M. (2021). Cross-linguistic contributions of acoustic cues and prosodic awareness to first and second language vocabulary knowledge. *Journal of Research in Reading, 44*(2), 434–452. https://doi.org/10.1111/1467-9817.12349
Cogan, G. B., & Poeppel, D. (2011). A mutual information analysis of neural coding of speech by low-frequency MEG phase information. *Journal of neurophysiology, 106*(2), 554–563.
Crowhurst, M. J. (2020). The iambic/trochaic law: Nature or nurture? *Language and Linguistics Compass, 14*(1), e12360.
Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *J Computer Speech & Language, 2*(3–4), 133–142.
Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*(1), 51–62.
Ding, H., Lin, B., Wang, L., Wang, H., & Fang, R. (2020). A Comparison of English Rhythm Produced by Native American Speakers and Mandarin ESL Primary School Learners. INTERSPEECH.
Ding, N., Melloni, L., Zhang, H., Tian, X., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature neuroscience, 19*(1), 158.
Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews, 81*, 181–187.
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*, 761–768.
Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *The Journal of the Acoustical Society of America, 95*(2), 1053–1064.
Duanmu, S. (2004). Left-headed feet and phrasal stress in Chinese. *Cahiers de Linguistique Asie Orientale, 33*(1), 65–103.
Duanmu, S. (2007). *The phonology of standard Chinese*. OUP Oxford.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *The Journal of the Acoustical Society of America, 27*(4), 765–768.
Ghitza, O. (2011). Linking speech perception and neurophysiology: Speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in psychology, 2*, 130.
Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in psychology, 3*, 238.
Ghitza, O. (2013). The theta-syllable: A unit of speech information defined by cortical function. *Frontiers in psychology, 4*, 138.
Ghitza, O., & Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica, 66*(1–2), 113–126.
Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature neuroscience, 15*(4), 511–517.
Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *J Laboratory Phonology, 4*(1), 67–92.
Greenberg, S. (1999). Speaking in shorthand–A syllable-centric perspective for understanding pronunciation variation. *Speech Communication, 29*(2–4), 159–176.
Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics, 31*(3), 465–485. https://doi.org/10.1016/j.wocn.2003.09.005
Greenberg, S., Hollenback, J., & Ellis, D. (1996). Insights into spoken language gleaned from phonetic transcription of the Switchboard corpus. Proc. ICSLP.
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.
Harmony, T. (2013). The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience, 7*, 83.
Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics, 69*(1), 113–122.
Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
He, D., Buder, E. H., & Bidelman, G. M. (2023). Effects of Syllable Rate on Neuro-Behavioral Synchronization Across Modalities: Brain Oscillations and Speech Productions. *Neurobiology of Language, 4*(2), 344–360.
Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *J Proceedings of the National Academy of Sciences, 109*(49), 20095–20100.
Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393–402. https://doi.org/10.1038/nrn2113
Hogg, R., Hogg, R. M., Hogg, R. M., & McCully, C. (1987). *Metrical phonology: A course book*. Cambridge University Press.
Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America, 119*(5), 3059–3071.
Houtgast, T., & Steeneken, H. J. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America, 77*(3), 1069–1077.
Hyafil, A., Fontolan, L., Kabdebon, C., Gutkin, B., & Giraud, A.-L. (2015). Speech encoding by coupled cortical theta and gamma oscillations. *J Elife, 4*, e06213.
Jeng, F.-C., Hu, J., Dickman, B., Montgomery-Reagan, K., Tong, M., Wu, G., & Lin, C.-D. (2011). Cross-linguistic comparison of frequency-following responses to voice pitch in American and Chinese neonates and adults. *Ear and hearing, 32*(6), 699–707.
Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). *Perception and production of Mandarin Chinese tones*. na.
Keitel, A., Gross, J., & Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS biology, 16*(3), e2004473.
Khatun, S., Mahajan, R., & Morshed, B. I. (2016). Comparative study of wavelet-based unsupervised ocular artifact removal techniques for single-channel EEG data. *IEEE journal of translational engineering in health and medicine, 4*, 1–8.
Kochanski, G., Grabe, E., Coleman, J., & Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *The Journal of the Acoustical Society of America, 118*(2), 1038–1054.
Kösem, A., & Van Wassenhove, V. (2017). Distinct contributions of low-and high-frequency neural oscillations to speech comprehension. *Language, cognition and neuroscience, 32*(5), 536–544.
Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping, 8*(4), 194–208. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C
Lakatos, P., Shah, A. S., Knuth, K. H., Ulbert, I., Karmos, G., & Schroeder, C. E. (2005). An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex. *Journal of neurophysiology, 94*(3), 1904–1911.
Leong, V. (2012). *Prosodic rhythm in the speech amplitude envelope: Amplitude modulation phase hierarchies (AMPHs) and AMPH models*.
Leong, V., Kalashnikova, M., Burnham, D., & Goswami, U. (2017). The temporal modulation structure of infant-directed speech. *Open Mind, 1*(2), 78–90.
Leong, V., Stone, M. A., Turner, R. E., & Goswami, U. (2014). A role for amplitude modulation phase relationships in speech rhythm perception. *The Journal of the Acoustical Society of America, 136*(1), 366–381.
Li, A., Gao, J., Jia, Y., & Wang, Y. (2014). Pitch and duration as cues in perception of neutral tone under different contexts in Standard Chinese. Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific.
Li, P., Sepanski, S., & Zhao, X. (2006). Language history questionnaire: A web-based interface for bilingual research. *Behavior research methods, 38*(2), 202–210.
Lo, C.-W., Tung, T.-Y., Ke, A. H., & Brennan, J. R. (2022). Hierarchy, not lexical regularity, modulates low-frequency neural synchrony during language comprehension. *Neurobiology of Language, 3*(4), 538–555.
Lu, Y., Jin, P., Pan, X., & Ding, N. (2022). Delta-band neural activity primarily tracks sentences instead of semantic properties of words. *Neuroimage, 251*, Article 118979.
Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *J Neuron, 54*(6), 1001–1010.
McCawley, J. D. (1978). IV - What Is a Tone Language? In V. A. Fromkin (Ed.), *Tone* (pp. 113-131). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-267350-4.50009-1.
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.
Momtaz, S., & Bidelman, G. M. (2024). Effects of stimulus rate and periodicity on auditory cortical entrainment to continuous sounds. *Eneuro, 11*(3).
Morillon, B., Arnal, L. H., Schroeder, C. E., & Keitel, A. (2019). Prominence of delta oscillatory rhythms in the motor cortex and their relevance for auditory and speech perception. *Neuroscience & Biobehavioral Reviews, 107*, 136–142. https://doi.org/10.1016/j.neubiorev.2019.09.012
Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4
Palva, J. M., & Palva, S. (2018). Functional integration across oscillation frequencies by cross-frequency phase synchronization. *European Journal of Neuroscience, 48*(7), 2399–2406.
Park, H., Ince, R. A., Schyns, P. G., Thut, G., & Gross, J. J. C. B. (2015). *Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners., 25*(12), 1649–1653.
Picton, T., Alain, C., Woods, D., John, M., Scherg, M., Valdes-Sosa, P., Bosch-Bayard, J., & Trujillo, N. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiology and Neurotology, 4*(2), 64–79.
Poeppel, D. (2003). The analysis of speech in different temporal integration windows: Cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication, 41*(1), 245–255.
Poeppel, D., & Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *J Nature reviews neuroscience, 21*(6), 322–334.
Qin, Z., Chien, Y.-F., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics, 38*(3), 541–570.
Rimmele, J. M., Poeppel, D., & Ghitza, O. (2021). Acoustically driven cortical δ oscillations underpin prosodic chunking. *J Eneuro, 8*(4).
Rimmele, J. M., Sun, Y., Michalareas, G., Ghitza, O., & Poeppel, D. (2023). Dynamics of functional networks for syllable and word-level processing. *Neurobiology of Language, 4*(1), 120–144.
Rosenblum, M. G., Kurths, J., Pikovsky, A., Schafer, C., Tass, P., & Abel, H.-H. (1998). Synchronization in noisy systems and cardiorespiratory interaction. *IEEE Engineering in Medicine and Biology Magazine, 17*(6), 46–53.
Schack, B., & Weiss, S. (2005). Quantification of phase synchronization phenomena and their importance for verbal memory processes. *Biological cybernetics, 92*(4), 275–287.
Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic inquiry, 11*(3), 563–605.
Silipo, R., & Greenberg, S. (1999). Automatic transcription of prosodic stress for spontaneous English discourse. Proc. of the XIVth International Congress of Phonetic Sciences (ICPhS).
Silipo, R., & Greenberg, S. (2000). Prosodic stress revisited: Reassessing the role of fundamental frequency. Proc. NIST Speech Transcription Workshop.
Teng, X., Tian, X., Rowland, J., & Poeppel, D. (2017). Concurrent temporal channels for auditory processing: Oscillatory neural entrainment reveals segregation of function at different scales. *PLoS biology, 15*(11), e2000812.
Tierney, A. L., & Nelson, C. A., III (2009). Brain development and the role of experience in the early years. *Zero to three, 30*(2), 9.
Tilsen, S., & Arvaniti, A. (2013). Speech rhythm analysis with decomposition of the amplitude envelope: Characterizing rhythmic patterns within and across languages. *The Journal of the Acoustical Society of America, 134*(1), 628–639.
Tilsen, S., & Johnson, K. (2008). Low-frequency Fourier analysis of speech rhythm. *The Journal of the Acoustical Society of America, 124*(2), EL34-EL39.
Wang, Q. (2008). *Perception of English stress by Mandarin Chinese learners of English: An acoustic study*.
Zeng, Z., Liu, L., Tuninetti, A., Peter, V., Tsao, F.-M., & Mattock, K. (2022). English and Mandarin native speakers' cue-weighting of lexical stress: Results from MMN and LDN. *Brain and language, 232*, Article 105151.
Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics, 38*(2), 260–271.
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.