Acoustic-driven and cross-language effects on neuro-behavioral synchronization to speech rhythms

Deling He

Follow this and additional works at: https://digitalcommons.memphis.edu/etd

Recommended Citation

This Dissertation is brought to you for free and open access by University of Memphis Digital Commons. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of University of Memphis Digital Commons. For more information, please contact khggerty@memphis.edu.
ACOUSTIC-DRIVEN AND CROSS-LANGUAGE EFFECTS ON NEURO-BEHAVIORAL SYNCHRONIZATION TO SPEECH RHYTHMS

by

Deling He

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Major: Communication Sciences and Disorders

The University of Memphis

December 2023
Dedication

Dedicated to my dear mom, Yongqin Yuan, and my dear sister, Tongtong He,
the kindest, strongest, and most beautiful souls in the world.

Without you, I would have never been more truly and happily myself.
You achieved me.
Acknowledgements

I would like to express my deepest appreciation to Dr. Eugene Buder for seeing my potential and bringing me into the world of academia. His kindness, enthusiasm, and witty humor are the catalyst, fostering my vision and freedom to explore my research interests.

My profound gratitude particularly goes to Dr. Gavin Bidelman, who guided and armed me with vital research minds and skills. I believe every one of his students would concur with me that he is the model of a professional and genuine scientist, supportive mentor, and dedicated educator, consistently going above and beyond in his commitment to science and nurturing the next generation of scholars.

I also sincerely thank Dr. Lynda Feenaughty and Dr. David Kimbrough Oller. They are the mentors who consistently advocate for and encourage you to pursue your ambitions with courage and tenacity, fostering an environment of innovation and critical thinking.

My gratitude extends to my friends, seniors, and peers, with whom I have forged enduring friendships, experienced personal growth, and celebrated mutual progress.

The successful completion of the dissertation and my own personal development would not have been possible without them.
Preface

Chapter 2 was published as a manuscript in *Neurology of Language*. Chapter 3 will be submitted for publication at *Brain and Language*.


Abstract

Speech rhythms play important roles in facilitating effective communication. This dissertation aimed to examine the neuro-behavioral synchrony related to hierarchical speech rhythms. Study #1 investigated relations between entrainment in the perceptual and production domains by measuring individuals’ brain oscillatory tracking to speech (EEG) and their simultaneous and non-simultaneous vocal productions synchronized to syllable rates between 2.5 and 8.5 Hz. We show that neural oscillations flexibly phase lock to speech in a rate-dependent manner, but is boosted near ~4.5 Hz, the dominant syllable rate. Cued speech productions (recruiting sensorimotor interaction) were optimal between 2.5 and 4.5 Hz, suggesting a low-frequency constraint on sensorimotor integration. Contrarily, “pure” motor productions (without concurrent sound cues) were most precisely generated at rates of 4.5–5.5 Hz, paralleling the neuroacoustic tracking. Notably, individuals demonstrating stronger auditory-perceptual entrainment exhibited better synchronization with speech rhythms motorically, supporting an intimate link between exogenous and endogenous rhythmic processing that is optimized at 4–5 Hz in both auditory and motor systems.

Study #2 extended Study #1 to examine how neural oscillations code the hierarchical nature of stress rhythms depending on language experience. By measuring multilevel EEG-acoustic synchrony and intra-brain cross-frequency phase coupling, we show encoding of stress involves different EEG signatures (delta rhythms = stress rate; theta rhythms = syllable rate), is stronger for amplitude vs. duration stress cues, and induces nested delta-theta coherence that reflects the stress-syllable hierarchy in speech. English speakers exhibited enhanced synchronization at the central stress (2 Hz) and syllable (4 Hz) rates intrinsic to
natural English suggesting plasticity in stress processing. Significant group differences across
English and Mandarin speakers emphasize that the brain-speech enhancement is not purely a
“bottom-up” response but is improved by “top-down” processing rooted in listeners’
language-specific knowledge. Critically, English individuals with superior cortical-acoustic
stress tracking capabilities also displayed pronounced neural hierarchical coherence,
highlighting a nuanced interplay between internal neural coherence and external entrainment.

Collectively, our studies reveal intricate relationships between synchronization in the
auditory and motor systems (i.e., neural oscillations) and multilevel speech rhythms and
highlight the critical role of brain oscillations in tracking and encoding speech rhythms
hierarchically and language-dependently.
# Table of Contents

List of Figures .................................................................................................................. ix
Chapter 1: General Introduction ......................................................................................... 1
  Brain oscillations and synchronization ......................................................................... 2
    Biophysical properties ............................................................................................... 2
    Functional significance ............................................................................................. 4
    Multiscale temporal organization .............................................................................. 7
Speech rhythms .................................................................................................................. 10
  Syllable rhythm ........................................................................................................... 12
  Stress foot rhythm ....................................................................................................... 13
Statement of research objectives ...................................................................................... 17
Chapter 2: Effects of Syllable Rate on Neuro-Behavioral Synchronization Across Modalities:
  Brain Oscillations and Speech Productions ............................................................... 19
  Introduction ................................................................................................................... 19
Materials & Methods ........................................................................................................ 23
  Participants .................................................................................................................... 23
  Stimuli ........................................................................................................................... 23
  Data acquisition and preprocessing ............................................................................. 24
  Data analysis: Quantifying synchronization and rate accuracy .................................... 28
  Statistical analysis ....................................................................................................... 30
Results ............................................................................................................................. 31
  Cortical oscillation synchrony is enhanced at ~4.5 Hz syllable rate .......................... 31
  Spontaneous speech synchronization is restricted to slower rates ............................ 33
  Correspondence between syllable perception and production ................................... 33
  Brain-behavior correlations between production and neural speech entrainment
    accuracy ..................................................................................................................... 35
Discussion ......................................................................................................................... 36
  Cortical oscillation synchrony is modulated by the heard syllable rates .................... 37
  Simultaneous speech-audio synchronization is rate restricted .................................... 39
  Non-simultaneous productions highlight an intrinsic rhythm at 4-5 Hz ..................... 40
  Differences and limitations to related studies .......................................................... 42
Chapter 3: Acoustic-driven and cross-language effects on the neuro-behavioral synchronization to stress rhythms

Introduction ............................................................................................................................................. 45

Materials & Methods ............................................................................................................................ 50
  Participants ........................................................................................................................................ 50
  EEG Stimuli ........................................................................................................................................ 50
  Data acquisition and preprocessing ................................................................................................. 51
  Statistical analysis ............................................................................................................................ 53

Results .................................................................................................................................................. 58
  Brain to speech tracking at the stress level (PLVstress) ............................................................. 58
  Brain to speech tracking at syllable level (PLVsyllable) ........................................................... 60
  Cross-frequency coupling within the brain (nmPSI) ....................................................................... 62
  Correlations .................................................................................................................................... 64

Discussion .......................................................................................................................................... 65
  Cortical encoding of stress rhythm via delta phase-locking depends on language experience .... 66
  Multilevel brain-to-speech synchronization is optimized for the natural rate of stress rhythms .... 68
  Neural coupling of delta-theta oscillations mirrors phase hierarchy between speech rhythms .... 70
  Amplitude cues dominate the neural encoding of stress ............................................................. 72
  Conclusions ..................................................................................................................................... 73

Chapter 4: General Discussion ........................................................................................................... 75
  Limitations ....................................................................................................................................... 76
  Future directions ............................................................................................................................... 77

Reference ............................................................................................................................................. 78

Appendix .............................................................................................................................................. 93
List of Figures

Figure 1: Examples of EEG and speech tasks. .................................................................27
Figure 2: Phase-locked neural oscillations synchronize to the rate of the syllable envelope. 32
Figure 3: Simultaneous speech synchronization to syllable trains is modulated by rate. ........33
Figure 4: Participants’ produced speech rate compared to the target rate of auditory stimuli. .................................................................35
Figure 5: Correlations between brain and production synchronization to speech. ..............36
Figure 6: Examples of stress stimuli used in the EEG study (i.e., a trochaic stress token). ...53
Figure 7: Continuous EEGs show phase-locking to stress and syllable rhythms. ..............55
Figure 8: Brain oscillations synchronize to the rate of stress rhythms. .............................59
Figure 9: Brain oscillations phase lock to the rate of syllable rhythms. .............................61
Figure 10: Phase coupling of delta-theta neural oscillations represents the phase hierarchy of stress-syllable rhythms. .................................................................63
Figure 11: Effects of acoustic stress cue and salience on cross-frequency coupling ..........64
Figure 12: Repeated measure correlations between internal cross-frequency coupling and external audio-neuro tracking .................................................................65
Chapter 1: General Introduction

Neural correlates of spoken communication are ubiquitous in brain activity. Speech sounds are the outcome of motor production, the acoustic input to auditory perception, and related cognitive processes. It is conceivable that some aspects of speech processing arise from common brain mechanisms that should be accessible for both perception and production. Theoretically, if the evolution of spoken language was constrained by endogenous neural mechanisms, the comprehension and production of speech must be compatible with the limits of auditory and motor neural processing (Gentilucci & Corballis, 2006; Giraud et al., 2007; Liberman & Whalen, 2000). Moreover, some have posited that speaking and listening performance might be optimized if the acoustic structure of speech aligns with intrinsic rhythms of the brain (i.e., “brain oscillations”) that are preferred by the auditory and motor nervous systems (Assaneo et al., 2021; Keitel et al., 2018; Keitel et al., 2017). Indeed, studies have shown that ongoing neural oscillations in both auditory and motor cortex faithfully track the continuous speech amplitude envelope (i.e., biomechanical oscillations of the speech apparatus) (Assaneo & Poeppel, 2018; Keitel et al., 2018). Moreover, the frequency of these rhythms becomes entrained (i.e., selectively tuned) to the natural rhythms of speech and its various featural properties (Assaneo & Poeppel, 2018; Doelling et al., 2014). This alignment indicates the dynamic oscillations at speech-relevant frequencies might be the key to investigating a common organizing principle or a mechanism of sensorimotor integration between auditory processing and motor speech production.
Brain oscillations and synchronization

Biophysical properties

Brain oscillations are commonly studied in the framework of neurodynamics, in which the brain is considered a dynamic system consisting of a set of oscillators. A “brain oscillator” is defined as a non-linear biophysical model that is capable of generating sustained rhythmic behavior by itself but also responding to external stimuli (Doelling & Assaneo, 2021). Biophysically, this non-linear model can be described by a set of differential equations whose parameters can be adjusted to yield different input/output patterns. In a broad sense, the (quasi)periodic behavior of interacting neural ensembles creates repetitive electrical fluctuations in brain activity that have been called “brain (or neural) oscillations.” Periodic components dominate the output of most cortical networks and are assumed to reflect neural excitability generated by the simultaneous firing of local neuronal populations and/or interactions between neurons across multiple spatiotemporal scales (Buzsaki & Draguhn, 2004; Canolty & Knight, 2010; Jensen & Colgin, 2007). In contrast, the aperiodic components appearing as 1/f power-law relationships are usually considered background noise and/or relaxed resting states that reflect unsynchronized, random neuronal activity (Freeman et al., 2000). Therefore, we define brain oscillations as rhythmic and repetitive fluctuations in the brain’s ongoing activity recorded by electrophysiological signals.

Brain oscillations are commonly distributed at distinct frequency scales, including delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), low gamma (30-80 Hz), and high gamma (80-200 Hz) rhythms (Penttonen & Buzsáki, 2003). On a spatial scale, high-
frequency oscillations are confined to a small and local neuron group, whereas slow
oscillations are confined to a large and long-range network (Von Stein & Sarnthein, 2000).
On the time scale, longer-period (slow frequency) oscillations represent a larger temporal
window for the intervals between neural activation, therefore allowing integrations with
longer delays and greater variability from large involved areas (Penttonen & Buzsáki, 2003).
Importantly, the non-linearity of the brain enables small perturbations in the widespread slow
oscillations to affect/modulate faster local events. Such hierarchical organization of brain
rhythms has been proposed as a mechanism of “top-down” processing (Gross et al., 2013;
Lakatos et al., 2005; Von Stein & Sarnthein, 2000). Together, brain oscillations might be the
key mechanism to enable multiscale spatiotemporal integrations and facilitate functional
processes by nesting faster events of neural excitability into slower ones.

Brain oscillations have been extensively observed in neurocognitive speech-language
studies, where electroencephalogram (EEG) and magnetoencephalography (MEG) responses
follow the rhythmic structure of the stimulus, thereby presenting a rhythmic oscillatory
pattern (Casas et al., 2021; Henry & Obleser, 2012; Luo & Poeppel, 2007; Will & Berg,
2007). This phenomenon is commonly assumed to reflect a dynamic synchronization between
the internal neural oscillations and external physical oscillations of stimuli, known as “neural
entrainment,” “cortical tracking,” or “brain synchronization.” The underlying mechanism is
that the synchronization arises from a brain oscillator, which is a truly rhythmic generator and
can be entrained by external stimulation (Assaneo & Poeppel, 2018; Doelling & Assaneo,
2021; Ghitza, 2011). Such systems could be modeled by Hopf-Andronov (HA) bifurcation to
explain the oscillatory behaviors of brain in both bottom-up and top-down scenarios, where the system constantly balances the interplay between these processing scenarios (Doelling & Assaneo, 2021). HA bifurcation occurs when the parameters are continuously increasing or decreasing until surpassing a critical point, resulting in a switching of oscillatory regimes, such as occurs in transitions between resting and stimulation states. External coupling through bottom-up stimulation forces neural oscillators into a rhythmic regime, possibly modulated by their coupling strength. Internal top-down processes (e.g., attention, prediction) push the system into an oscillatory regime, possibly by shifting an existing parameter in the equation. Importantly, oscillators may not respond identically to the same input that arrives at different points of time (i.e., phase). Stimuli during downward phases of neural oscillations may inhibit another natural progression or tendency, and stimuli during upward phases may advance it. Therefore, the behavior of neural oscillators is characterized by phase resetting and phase synchrony and depends on mechanisms by which they interact with the external world—bottom-up and top-down cues—where internal processes actively modify the oscillation.

**Functional significance**

The phenomenon of brain oscillations synchronizing to speech rhythms has sparked cutting-edge, neurocognitive theories of language and the brain (Doelling & Assaneo, 2021; Obleser & Kayser, 2019; Poeppel & Assaneo, 2020). The processing of continuous speech is realized through the phase alignment of brain oscillations to the speech amplitude envelopes, which presumably segment/parse the streams of speech into discrete linguistic representations with
a temporal window of ~200-250 ms (Assaneo & Poeppel, 2018; Ghitza, 2012; Luo & Poeppel, 2007). Specifically, the sharpness of the speech envelope acts as a temporal cue for syllable rate, driving syllabic brain synchronization and facilitating intelligibility (Doelling et al., 2014). Conversely, a flat envelope prevents brain oscillations from tracking syllables, resulting in reduced intelligibility possibly because of the failure of syllable parsing (Ghitza, 2012). Critically, such brain synchronization is not equally efficient at all syllable rates. The physiology imposes a constraint such that speech perception is optimal at syllable rates that fall within the range of intrinsic brain rhythm in the theta band (Ghitza, 2012; Houtgast & Steeneken, 1985; Poeppel & Assaneo, 2020). This also leads to an enhancement of auditory neural tracking at frequencies close to the natural syllable rhythm at 4-5 Hz across languages (He et al., 2023) and stronger coupling between auditory and motor cortices at 4.5 Hz (Assaneo & Poeppel, 2018). These findings imply that the temporal rhythm of speech may emerge as a consequence of the intrinsic rhythm in the brain, possibly via a sensorimotor interface. However, it remains unclear how necessary it is for the brain to parse speech sounds into isolated segments before accessing lexical representations. Possibly, the phase synchronization behavior itself is a sufficient procedure to code real-life complex speech and is not amenable to a linear segmentation analysis.

In addition, brain oscillations have been associated with higher-order, top-down mechanisms, resulting in boosted processing of upcoming events. The instantaneous phase of neural oscillation is often assumed to reflect the excitability level of the neural network at a given moment (Bishop, 1932; Schroeder & Lakatos, 2009). In the scenario of sensory
processing, when the timing of sensory input coincides with high-excitability phase of neural oscillation, the possible gain (e.g., excitability) of local neuron ensembles would be higher than that at low-excitability phase (see Obleser & Kayser, 2019, Fig.1 panels C & D). Empirical evidence from visual perception has shown that the pre-stimulus ongoing EEG of near-threshold stimuli exhibited different phase angles for detected and missed stimuli (Busch et al., 2009), which account for 16% trial-to-trial variability and allow prediction of detection performance at the single-trial level (Busch & VanRullen, 2010; Drewes & VanRullen, 2011; VanRullen et al., 2011).

Additionally, the phase of synchronized brain oscillations may provide valuable information for a probabilistic prediction of upcoming events. Studies found significant phase consistency across trials with subthreshold rhythmic sounds (Ten Oever et al., 2017) and periodic oscillatory patterns tracking to the subsequently missed stimuli (Zoefel & Heil, 2013; Zoefel et al., 2018), indicating a predictive process for the upcoming events. These findings suggest that the phase of ongoing neural oscillations may shape sensory perception, possibly by providing a temporal reference for the excitability level of neural networks. And the strength of phase synchronization may be modulated by top-down mechanisms, so the brain is able to actively select important inputs by aligning high-excitability phases with their predicted timing.

Overall, brain oscillations reflect complex, dynamic processes that are characterized by instantaneous frequencies and relative phases. Phase resetting and synchronization may be a fundamental mechanism allowing the brain’s intrinsic rhythms to adapt to the varieties of
inputs with efficiency and flexibility.

**Multiscale temporal organization**

A growing number of studies suggest that speech is processed at multiple temporal windows operated by a set of neuronal oscillators at frequencies particularly relevant to speech (Ding et al., 2016; Ghitza, 2011; Gross et al., 2013; Hyafil et al., 2015; Kösem & Van Wassenhove, 2017; Poeppel, 2003; Rimele et al., 2023). The oscillations involved in speech are spectrally distributed in the gamma (>30 Hz), theta (4-8 Hz), and delta (1-3 Hz) frequency bands, corresponding roughly with the time spans of phonemic, syllabic, and supra-syllabic units. However, a critical question is how brain oscillations at multiple timescales are organized as a response to the temporal hierarchy of speech information.

Poeppel (2003) proposed asymmetric sampling in time (AST), assuming that speech is processed by both left and right auditory cortices but with differential sensitivity to specific temporal structures in the acoustic signal. Specifically, speech is processed concurrently on at least two different time scales: the right auditory areas preferentially extract information from a theta-syllabic temporal window, while the left auditory cortex decodes from a gamma-phonemic window. This temporal hemispheric asymmetry is supported by neuroimaging evidence (Boemio et al., 2005; Giroud et al., 2020; Luo & Poeppel, 2012) and resting state brain activity without any auditory inputs (Giraud et al., 2007). Furthermore, an innovative aspect of AST—the multi-time resolution process, is integrated into a comprehensive dual-stream model of language processing proposed by Hickok and Poeppel (2007). It advances that speech is processed in two temporal windows by two separate dorsal (articulatory motor
representations) and ventral (lexical conceptual representations) streams. While each stream can independently initiate lexical access, optimal performance occurs when syllabic- and phonemic-rate information is concurrently involved and combined.

Following AST, Ghitza and Greenberg (2009) proposed the TEMPO model after observing that the insertion of silent gaps between speech elements restores intelligibility. They interpreted the inserted silences as providing “needed” time for an artificial theta-syllabic rhythm. Ghitza (2011, 2012, & 2013) refined this model by incorporating two paths of decoding and parsing for processing auditory streams. The parsing path segments speech streams through an array of cascaded oscillators of theta, beta, and gamma that synchronize with speech rhythm. Guided by the parsing path, the decoding path links the multilevel “packets” or “chunks” of speech with stored linguistic representations. Importantly, TEMPO established a hierarchical structure of temporal windows, where gamma frequency is a constant multiple of beta and so is beta embedded inside a theta oscillator. Theta is believed to provide syllabic parsing and act as the master oscillator due to its robust presence of energy fluctuations in the brain and close association with intelligibility (Ghitza, 2011; Luo & Poeppel, 2007). The hierarchical tier of theta-gamma is confirmed with neural microcircuit modeling that show theta oscillations organize gamma neurons spiking activities (i.e., excitatory and inhibitory) (Hyafil et al., 2015).

Both TEMPO and AST models hold the assumption that speech has a neural representation related to its rhythmic structures and that temporal processing of speech involves multi-scale time sampling or parsing. However, AST holds the view that neither of
the oscillators takes place with privileged analysis at any level, while TEMPO assumes a hierarchical structure with theta oscillator mastering the process followed by faster beta and gamma analysis. Empirical data may support a hierarchy of brain oscillations in both animals and humans (Gross et al., 2013; Lakatos et al., 2005). Gross et al. (2013) used mutual information to calculate cross-frequency coupling and found that the hierarchy is realized by delta-phase modulating theta-amplitude and theta-phase modulating gamma-amplitude. Importantly, this hierarchical structure is enhanced with speech edges in the envelope and attenuated by backward display, suggesting both bottom-up and top-down modulation.

Critically, neither of the theories addresses slower-than-syllable rhythms, such as delta oscillations at the suprasyllabic level. Other studies often assumed that a delta oscillator retains an analogous role to that of the theta oscillator in segmenting continuous speech, with its segmentation occurring at phrasal or prosodic level (Keitel et al., 2018; Rimmele et al., 2021). However, studies showed that delta oscillations are more susceptible to acoustic manipulations (Bröhl & Kayser, 2021), reflect high-order syntactic information (Ding et al., 2016; Lu et al., 2022a; Lu et al., 2022b), suggest interactions between auditory and motor systems which also predict the timing of speech (Keitel et al., 2018). These findings suggest that delta may differentially contribute to speech comprehension, implying a higher-order computational mechanism. For instance, Gross et al. (2013) found increased delta-theta phase-amplitude coupling during intelligible story perception, suggesting the regulating role of delta oscillation on the amplitude of the theta band. Crucially, it also raises the question of whether theta still holds the role of master oscillator if delta is included into the hierarchical
structure of brain oscillations.

**Speech rhythms**

Speech sounds unfold over time as a continuous stream. However, constituent elements punctuate the stream and exhibit rhythmic regularities in occurrence. These temporal regularities may be referred to as *speech rhythm*. In 1945, Pike suggested a distinction in linguistic rhythms, identifying them as syllable-timed vs. stress-timed language. However, it’s important to note that this classification is a simplification and doesn’t encompass the full diversity of languages. It was drawn from intuitive perceptions of rhythmical tendencies, potentially linked to the assumed recurrence of isochronous elements (Abercrombie, 1967).

In syllable-timed languages (e.g., Spanish and Italian), syllables always recur at the same intervals, resulting in a “machine-gun” like rhythm. In contrast, stress-timed languages (e.g., English and Dutch) exhibit a “Morse-code” like rhythm by alternating long and short pulses. The recurring constituent in stress-timed language is stress foot, which is composed of stressed and unstressed syllables since syllables do not occupy the same amount of time and the stressed syllables take up a longer duration. Abercrombie (1967) further claimed that all languages throughout the world could be dichotomously categorized based on the durational isochrony of syllables or stress-feet, though further research hypothesized a third category of mora-timed languages such as Japanese and Tamil (Port et al., 1987).

Despite the fact that this classification has been widely used, it remains a challenge, and in fact, it has failed to provide empirical evidence of isochronous elements in speech (Dauer, 1983; Ramus et al., 1999; Roach, 1982). Acoustical data fails to provide objective
rules for assigning languages to one of the categories based on standard deviation of syllable durations, percentage deviations of inter-stress intervals, etc. (Roach, 1982). In English, the inter-stress intervals are not equidistant but increase with the numbers of syllables they contain (Dauer, 1983; Nakatani et al., 1981). Dauer (1983) suggests that the rhythm of speech measured by durational variability depends on a number of factors, such as syllable structure, stress influence on phoneme, stress-prominent realization, etc. Rather than focusing on isochrony of phonological elements, other studies applied phonetic durational account for the rhythmic categorization by measuring the statistical properties of duration of vocalic (i.e., onset and offset of a vowel or a cluster of vowels) and consonantal intervals, such as relative proportions of vocalic intervals (%V), standard deviation of vocalic (∆V) and consonantal durations (∆V), pairwise variability indices (PVI) (Dellwo et al., 2003; Grabe & Low, 2002; Ramus et al., 1999). Though these measurements were successful in differentiating languages to some extent, the underlying mechanism linking statistical properties of phonetic duration to the perceived rhythmic isochrony remains undetermined.

It is perhaps impossible to devise an objective test to identify the rhythmic categories of language based on isochronous segment duration, and the idea of isochrony is auditive and subjective (Dauer, 1983; Roach, 1982). Speech displays both (quasi)periodic and non-periodic tendencies across different timescales. Consequently, the term isochrony is now avoided to describe language rhythm in speech production. Instead, the descriptive categories of stress-timed and syllable-timed language have been adopted. However, there is no doubt that both syllables and stress-feet play crucial roles in governing the rhythmical tendencies of
spoken language.

**Syllable rhythm**

Despite the expectation that different languages would have distinct rhythms, empirical acoustic data with large samples point toward a universal rhythm at the syllabic level across languages. Greenberg et al. (1996) conducted durational analysis of English speech which contains 72-minute-long spontaneous conversations from 750 speakers. The results showed that the mean syllable duration of English is 190 ms, with the 20th to 80th percentiles of the histogram distribution including syllable duration ranging from 107.5 ms to 260 ms (Greenberg et al., 1996). A similar study was reported for conversational Japanese, where the mean syllable duration is 166 ms with a standard deviation of 73 ms (Arai & Greenberg, 1997). Importantly, both languages demonstrate a similar shape in the histogram distribution of the occurrence of syllable duration (Arai & Greenberg, 1997). Further, Greenberg et al. (1996) proposed that syllable duration can be conceptualized in terms of modulation frequency because of their reciprocal relationship if a behavior (e.g., syllables) is assumed to reoccur, at least, regularly. Both English and Japanese exhibit a similar spectral pattern with a spectrum peak around 4 Hz (i.e., syllable duration around 250 ms) and a bandwidth of 2-10 Hz (i.e., syllable duration ranging from 100–500 ms) (Arai & Greenberg, 1997; Greenberg et al., 2003; Greenberg et al., 1996). Critically, a recent study further showed a remarkable consistent pattern of speech modulation spectra across 9 languages, with a spectral peak between 4-5 Hz and greatest power within a bandwidth of 2-10 Hz (Ding et al., 2017). The similarities in patterns of syllable duration and modulation spectra observed in languages
with typologically different rhythmic characteristics suggest a universal syllabic rhythm of spoken language, in which syllables reoccur regularly every ~200-250 ms with variations within 100 ms -500 ms. In other words, speech is characterized by pseudo-periodic syllable rhythm, which generally unfolds at a similar speed independent of a speaker’s language.

Importantly, syllable rhythm plays a crucial role in language comprehension. A line of research indicates that speech intelligibility closely depends on the integrity of the modulation spectrum within the frequency range of 2 to10 Hz (Ghitza, 2012; Houtgast & Steeneken, 1985; Poeppel & Assaneo, 2020). The peak of the modulation spectrum is attenuated and shifted down to 1–2 Hz for speech recorded in highly reverberant environments where speech is difficult to understand (Houtgast & Steeneken, 1985). Also, the intelligibility is severely degraded if the speech is artificially low-pass filtered below 2 Hz and is only marginally improved by adding components of modulation frequencies above 8 Hz (Drullman et al., 1994). Critically, it has been highlighted that slow brain oscillations (<10 Hz) might track syllable rhythm to facilitate the brain’s processing of speech (Assaneo & Poeppel, 2018; Ghitza, 2012; Luo & Poeppel, 2007). Therefore, syllable has been proposed as a primitive unit of language and plays a critical role in its comprehension.

**Stress foot rhythm**

Spoken language is not merely a linear sequence of individual syllables, but rather the syllables are grouped together into a higher-order constituent—the stress foot (in the case of English). Metrical phonology is a subfield of phonology that specifically studies the representation of stress patterns in English and the rules that govern stress assignment (Hogg
et al., 1987). The foot is a primary metrical unit of rhythm, consisting of a group of syllables with only one stress. For instance, a binary foot may have either a ‘Strong-weak’ or ‘weak-Strong’ pattern, with the ‘S-w’ constituent known as a trochaic foot and the ‘w-S’ as an iambic foot. The ‘S-w’ motifs can be clearly perceived and illustrated at the syllable level, even by naïve listeners. Theoretically, this ‘S-w’ pattern can occur at any level, including syllable, stress foot, word (i.e., primary lexical stress), prosodic phrase (i.e., phrasal accent), etc. For example, PENN-syl-VA-nia contains four syllables that alternate as ‘S-w-S-w’ pattern, which can be further grouped into ‘Strong feet-weak feet’. Such grouping forms a hierarchical structure like a branching tree and is usually illustrated by the metrical tree or metrical grid (see examples from Hogg et al., 1987). Therefore, metrical phonology posits a hierarchical organization of language in which bounding or nesting relations are built between hierarchical tiers (e.g., the tier of syllable and foot) through cyclic stress assignment.

Conventionally, the prominence of stressed syllables is thought to be attributed to variations in fundamental frequency ($f_o$), intensity, and duration. Fry (1958) studied contrastive stress by stepping up or down $f_o$ from the first to the second syllable with a varied size of $f_o$ step. The results showed that while $f_o$ did contribute to the perception of stress, the effect was modest, suggesting that factors other than $f_o$ are likely to play a more decisive role. And the size of the $f_o$ step did not consistently influence stress pattern, indicating the relation of $f_o$ to stress is not as direct as: the higher the $f_o$, the higher the stress. Yet, more and more studies suggest that acoustic cues of amplitude and duration better discriminate stressed from unstressed syllables than $f_o$ (Kochanski et al., 2005; Silipo & Greenberg, 1999, 2000). Fry
found that higher peak intensity and longer duration of vowels serve as cues for the perception of stressed syllables. Specifically, the duration and intensity ratios between stressed and unstressed vowels are reliable cues for the judgment of stress, with duration ratio proving to be a more effective cue than intensity ratio.

An alternative approach to identify speech rhythm is envelope-modulation analysis which analyzes the low-pass filtered temporal fluctuation of speech amplitude envelope with a power spectrum. Greenberg et al. (2003) found an overlapping distribution of stressed and unstressed syllables in the modulation spectrum, where stressed syllables belong to the left side of the spectrum with slower frequency and unstressed syllables display on the right with faster frequency. Moreover, to identify the stress foot as an entity, Tilsen and Arvaniti (2013) applied a non-linear empirical mode decomposition of the amplitude envelope for conversational English speech. Stress rhythm is captured at a frequency of 2-2.5 Hz in the amplitude envelope. Indeed, this result is consistent with acoustic measurements in which the average interstress interval in English is around 450 ms with a standard deviation of 150 ms (Dauer, 1983).

Leong (2012) introduced the term “stress foot”, an alternative to as metrical or prosodic foot, to emphasize its integral function in both the segmental and suprasegmental domains. In their study, separate syllable and stress foot envelopes were established by using amplitude modulation and discovered a phase hierarchy relationship between them. The syllable rhythm is carried by an envelope with high-low amplitude alternation to represent stressed and unstressed syllables, while the stress foot rhythm is carried by a pseudo-
sinusoidal envelope. The ridge in the stress envelope is shifted toward the stressed syllable and away from the unstressed syllable, indicating that stressed syllables are constrained to occur at a certain phases of the stress envelope (Leong, 2012). Furthermore, Leong et al. (2014) reported that the perception of a stress pattern (e.g., ‘S-w’) can be reversed with a \( \pi \) radian phase shift between syllable and stress envelope, while the original stress pattern is retained with a \( 2\pi \) phase shift. Indeed, the phase relationship between syllable and stress rhythm is a critical element in the hierarchical structure of spoken language.

It is noteworthy that speech sounds could also be modeled as biomechanical oscillations that resemble some of the dynamic-physical properties of brain oscillations, such as the \( 1/f \) power relationship (Buzsaki, 2006; Voss, 1975) and bifurcation dynamics (Skljarov, 1996). O’Dell and Nieminen (1999) developed a coupled-oscillator model of speech rhythm, postulating syllable and stress foot as two coupled-oscillators. The model assumes that, if in isolation, each of the oscillators would exhibit simple oscillatory (e.g., linear) behavior. However, when the faster syllable oscillation is nested into the slower stress oscillation, a larger dynamic system with much more complex behavior is built, including hierarchical and intra- and inter-level coordination (O’Dell & Nieminen, 1999; O’Dell & Nieminen, 2009). In the case of English, the frequency of syllable-related oscillations may be modulated by the phase of stress-related oscillations. The latter might slow down the stressed syllable in the vicinity of some particularly phase representing stress and speed up the unstressed syllable. Relatedly, as syllables are bound into a stress foot, the syllable internal coarticulation would also be expected to increase.
Stress foot is also known as metrical foot or prosodic foot with retrospective explanations. Linguists originally borrowed the term “foot” from poetic meter to specifically refer to the rhythmic arrangement of segments, whereas phonologists preferred a quotation of “prosodie” to emphasize that stress is a supra-segmental feature of spoken language (Selkirk, 1980). However, the term “stress foot” is adopted from Leong (2012) as the assumption is that stress foot unit, containing one stressed syllable, is an integral speech constituent that encompasses both rhythmic segmental properties and suprasegmental prominences and extends beyond simple temporal structures. This rhythmic continuity underscores the profound interplay between stress and syllabic elements, illuminating the nuanced intricacies of linguistic prosody. Within the present dissertation, the terminologies "stress foot rhythm" and "stress rhythm" are employed interchangeably.

In conclusion, it is clear that speech rhythm is not a simple isochronous sequence of temporal units, but rather a complex dynamic system comprised of hierarchically organized multi-level units, possibly through the mechanism of stress-phase coding of syllables. Importantly, it is conceivable that such a hierarchical structure of speech would exert significant influence on speech rhythm perception as detected neurally.

**Statement of research objectives**

Considerable studies suggest that speech temporal processing is conducted at multi-time scales through brain synchronization with speech rhythms (Ding et al., 2016; Ghitza, 2011; Gross et al., 2013; Hyafil et al., 2015b; Kösem & Van Wassenhove, 2017; Poeppel, 2003). However, it remains uncertain how the neuro-behavioral synchronization during speech
perception and production adapts to the acoustic and phonological properties in speech signal and how it varies for speakers with different native languages that have distinct rhythmic structures. The current dissertation used behavioral (perception, production) and neural (EEG) measures to characterize synchronization skills in the acoustic, motor, and neural domains. In two studies, we tested for flexibility in these skills by manipulating acoustic properties of the speech signal itself (Study 1: syllable rates; Study 2: stress patterns within the stress foot-syllable hierarchy). Moreover, we examined how brain and speech synchrony vary due to the different language structure and experiences of the listener (Study 2: English and Chinese). Collectively, this work aims to shed light on how oscillatory mechanisms in the brain mirror the rhythmic fluctuations in speech amplitude and their phonological structure as organized by syllable-stress relationships, and how these properties are changed by language experience.
Chapter 2: Effects of Syllable Rate on Neuro-Behavioral Synchronization Across Modalities: Brain Oscillations and Speech Productions

Introduction

The auditory cortex faithfully tracks amplitude modulations in continuous sounds, regardless of whether those acoustic events are speech (Ahissar et al., 2001; Casas et al., 2021; Luo & Poeppel, 2007), modulated white noise (Henry & Obleser, 2012), or clicks (Will & Berg, 2007). This phenomenon, whereby a listener’s rhythmic brain activity (i.e., “oscillations”) entrains to the physical signal, is described as “neural synchronization” or “cortical tracking.” Neurocognitive models suggest that the phase of ongoing brain oscillations, especially within the low theta band (4 to 8 Hz), lock to the slowly varying amplitude envelope to parse continuous sounds into discrete segments necessary for speech comprehension (Doelling et al., 2014; Ghitza, 2011, 2012; Giraud & Poeppel, 2012; Luo & Poeppel, 2007). In particular, speech syllable rhythms, which exhibit a quasi-regularity in their envelope modulation (Ding et al., 2017; Tilsen & Johnson, 2008), have been used to study how the brain parses the continuous speech stream (Ghitza, 2012; Hyafil et al., 2015). However, such brain entrainment is not solely low-level neural activity that simply mirrors the acoustic attributes of speech. Rather, entrained responses also serve a casual role in facilitating speech comprehension (Doelling et al., 2014; Luo & Poeppel, 2007; Peelle et al., 2013). These studies demonstrate that the degree to which auditory cortical activity tracks acoustic speech (and non-speech) signals provides an important mechanism for perception.
Syllable rhythms in speech range in speed from 2-8 Hz (Ding et al., 2017). With this variability in mind, it is natural to ask whether the brain’s speech systems are equally efficient across syllable rates, or instead are tuned to a specific natural speech rhythm. Indeed, the majority of the world’s languages unfold at rates centered near 4-5 Hz and neuroacoustic entrainment is enhanced at these ecological syllable speeds (Ding et al., 2017; Poeppel & Assaneo, 2020). In their neuroimaging study, Assaneo and Poeppel (2018) demonstrated that auditory entrainment (i.e., sound-to-brain synchronization) is modulated by speech rates from 2.5 to 6.5 Hz but declines at faster rates. In contrast, a more restricted 2.5-4.5 Hz frequency coupling was found in phase-locked responses to speech between auditory and motor cortices (i.e., brain-to-brain synchronization) (Assaneo & Poeppel, 2018). This suggests that when neural oscillations can entrain to a wider band of external rhythms (e.g., 2.5-6.5 Hz), motor cortex resonates at select frequencies to emphasize syllable coding at 4.5 Hz. A neural model was proposed accordingly: speech-motor cortical function is modeled as a neural oscillator, an element capable of generating rhythmic activity, with maximal coupling to auditory system at 4.5 Hz. Such studies suggest, at least theoretically, a convergence of the frequency of endogenous brain rhythms during speech production and the cortical encoding of speech at its input.

In parallel with auditory-motor cortex coupling, behavioral sensorimotor synchronization has been extensively characterized by having individuals produce certain movements in time along with external physical events. Sensorimotor skills have most often been studied in the form of tapping to a periodic stimulus (Repp, 2005). The rate limits of
synchronization in beat tapping approximately correspond with inter-onset intervals between 100 ms (Pressing & Jolley-Rogers, 1997) and 1800 ms (Miyake et al., 2004; Repp, 2005). However, these examples of non-speech motor synchronization may not generalize to speech considering its unique nature in human cognition. The therapeutic benefits of synchronizing to audio or visual speech productions, referred to speech entrainment, has been demonstrated in patients with Broca’s aphasia (Fridriksson et al., 2012; Thors, 2019). However, experience-based rates (i.e., patient’s most comfortable rate) have been implicitly used in speech entrainment tasks rather than systematically verified. Additionally, Assaneo et al., (2019) found some listeners involuntarily match their speech with external rhythm while others remain impervious by using a spontaneous speech synchronization (SSS) task. Listeners were instructed to freely produce syllable trains while hearing syllables at rates of 4.5 syll/sec with the goal of monitoring the occurrence of syllables. Their data established the link between word learning capabilities and sensorimotor speech synchrony. Critically, the optimal rate of the speech sounds in those studies was assumed to be close to the natural/normal speaking rate (i.e., ~4-5 Hz). Uncertainty also persists regarding how wider ranges of syllable rates might affect speech synchronization. Further, studies showed that better rhythm perception abilities are indicative of increased conversational quality mediated by better speech entrainment (Wynn et al., 2022). Thus, it is highly plausible that an individual’s preference for certain stimulus rates perceptually might facilitate their successfully entrainment at similar preferred rates during production. To address this knowledge gap and explicitly test for frequency-specific coupling in speech perception and
production, sensorimotor and auditory synchronization must be measured in a common paradigm.

In the present study, we aimed to empirically compare syllable rate sensitivity of the auditory-perceptual and (sensori)motor systems. In doing so, we ask whether brain and speech entrainment is or is not selectively tuned to the fundamental periodicity inherent to speech (~4.5 Hz) and thus represents a common organizing principle of processing across modalities. This notion has been suggested, but to our knowledge has remain largely untested, in prominent neurocognitive models of speech processing (Assaneo & Poeppel, 2018; Assaneo et al., 2021; Poeppel & Assaneo, 2020). To this end, we measured neuroacoustic tracking of listeners’ EEG to speech syllable trains to quantify their perceptual entrainment to speech. To quantify motor entrainment, we measured speech productions where participants synchronized to a wide range of syllable rates between 2.5-8.5 Hz along with (simultaneous production) or without (non-simultaneous production) a concurrent auditory speech stimulus. Employing both production tasks allowed us to isolate more or less “pure” measures of motor system by including/excluding external auditory stimuli. Brain-behavior correlations and comparison of rate profiles across EEG and production data allowed us to explicitly characterize possible links between auditory neural and motor production entrainment mechanisms of speech processing.
Materials & Methods

Participants

Fifteen young adults participated in the study (mean age 26.7±3.4 years; 10/5 females/males). (One additional participant completed the experiment but their data were lost due to a logging error). They were of mixed race and ethnicity. Ten were native English speakers and five were bilingual with English as a second language. Several participants had musical training (mean 9.9±3.8 years). All participants were right-handed (Oldfield, 1971a) and reported no history of neuropsychiatric disorders. All had normal hearing sensitivity, defined as air-conduction pure tone thresholds ≤ 25 dB HL at octave frequencies from 500 Hz to 4000 Hz. Listeners were provided written informed consent in compliance with a protocol approved by the IRB and were monetary compensated for their time.

Stimuli

EEG stimuli

We used stimuli inspired by Assaneo and Poeppel (2018) to characterize brain synchrony to rhythmic speech. Each consisted of trains of a single repeating syllable from the set /ba/, /ma/, /wa/, /va/ (random draw). Individual tokens were synthesized from online text-to-speech software (www.fromtexttospeech.com) using a male voice, and time compressed in Praat to 120 ms durations (https://www.fon.hum.uva.nl/praat/; P. Boersma & D. Weenink, 2013). Tokens were concatenated to create syllable trains of 6 s duration. To vary syllables’ rate, we parametrically varied the silent gap between tokens from 0 to 280 ms to create 7 continuous streams of speech syllables with rates of 2.5, 3.5, 4.5, 5.5, 6.5, 7.5, and 8.5
syll/sec. In practice, the 8.5 Hz condition was presented at a nominal rate of 8.33 Hz to achieve the fastest presentation speed possible given the 120 ms duration of our individual speech tokens.

**Speech production stimuli**

To assess simultaneous (cued) and non-simultaneous (un-cued) speech production synchronization, we generated another two sets of stimuli adapted from the spontaneous synchronization of speech task (SSS; Assaneo et al., 2019). To study the non-simultaneous rhythm production, we used syllable trains of continuous repetition of /ta/ lasting for 10 s.

For simultaneous rhythm production, we used 60 s long syllable streams with 16 distinct syllables (unique consonant-vowel combinations) that were randomly concatenated. We generated 7 rate conditions (2.5 ~ 8.5 syll/sec). This was achieved by temporally compressing/expanding the 4.5 Hz syllable stream from Assaneo et al. (2019) by the appropriate scale factor using the “Lengthen” algorithm in Praat (P. Boersma & D. Weenink, 2013).

**Data acquisition and preprocessing**

Participants were seated comfortably in front of a PC monitor and completed the three experimental tasks in a double-walled, sound-attenuating booth (Industrial Acoustics Company, Inc., Bronx, New York). Auditory stimuli were presented binaurally at 82 dB SPL via electromagnetically shielded ER-2 insert earphones. Stimuli and task instructions were controlled by MATLAB ® 2013 (The MathWorks, Inc.) routed to a TDT RP2 signal processing interface (Tucker-Davis Technologies). Speech production samples were recorded
digitally with a professional microphone (Blue Yeti USB, Logitech; 44100 Hz; 16 bits; cardioid pattern).

**EEG data**

During neural recordings, participants listened to rhythmic syllable trains (Fig.1A). To maintain attention, they were instructed to identify which syllable (i.e., /ba/, /ma/, /wa/, /va/) was presented at the end of the trial via button press. There was no time constraint to respond, and the next trial started after the button press. Listeners heard 10 trials of each 6 sec syllable train per syllable rate condition. Rate and syllable token were randomized within and between participants.

Continuous EEGs were recorded differentially between Ag/AgCl disc electrodes placed on the scalp at the mid-hairline referenced to linked mastoids (A1/A2) (mid-forehead = ground). This single channel, sparse montage is highly effective for recording auditory cortical EEG given their fronto-central scalp topography (Bidelman et al., 2013; Picton et al., 1999). Inter-electrode impedance was kept ≤ 10 kΩ. EEGs were digitized at 1000 Hz (Neuroscan SynAmps RT amplifiers) using an online passband of DC - 400 Hz. Neural signals were bandpass filtered (0.9- 30 Hz; 10th order Butterworth), epoched into individual 6 sec trial segments synchronized to the audio stimuli, and concatenated. This resulted in 60 sec of EEG data per rate condition. Eyeblinks were then nullified in the continuous EEG via a wavelet-based denoising algorithm (S. Khatun et al., 2016). Trials were averaged in the time domain to derive cortical neural oscillation for each condition. We measured synchronization between brain and acoustic speech signals via phase-locking values (PLV; see Section 2.4.1).
**Speech production data**

*Non-simultaneous syllable rhythm synchronization* (Fig.1B). Participants first listened to rhythmic syllable trains (/ta/ repeated for 10 sec). They were instructed to then whisper /ta/ with the same pace as the previous stimulus for 10 s (i.e., without a concurrent audio stimulus). With this explicit instruction and whispering articulation requirement, we aimed to investigate intentional speech rhythm production with internal rhythmic cues instead of guidance from self-auditory feedback. The procedure was repeated twice for each rate condition. Two runs were conducted in anticipation of avoiding possible practice effects. However, data from the two runs were highly correlated ($r_{2.5} = 0.75$, $r_{3.5} = 0.88$, $r_{4.5} = 0.80$, $r_{5.5} = 0.91$, $r_{6.5} = 0.86$, $r_{7.5} = 0.77$, $r_{8.5} = 0.82$, $p < 0.001$), indicating good test-retest repeatability. Moreover, paired *t*-tests further confirmed the two runs did not differ at any of the rates ($p_{2.5} = 0.85$, $p_{3.5} = 0.66$, $p_{4.5} = 0.22$, $p_{5.5} = 0.17$, $p_{6.5} = 0.23$, $p_{7.5} = 0.94$, $p_{8.5} = 0.17$).

*Simultaneous syllable rhythm synchronization* (Fig.1C). We adapted the SSS test (Assaneo et al., 2019) to measure cued motor speech to auditory synchronization. Participants were instructed to continuously whisper /ta/ while *concurrently* listening to a rhythmic syllable stream for 60 s. By employing whisper and insert earphones, we aimed to avoid participants’ using their own production sounds as auditory feedback to their speech output. After each trial, listeners indicated whether a target syllable were presented in the previous stream. Four target syllables were randomly chosen from a pool of eight (50% were from the syllable stream). Importantly, we did not explicitly instruct participants to synchronize to the external audio rhythm and we also removed their training session. In that study, listeners first
heard a fixed syllable rate at 4.5 Hz presented auditorily. This may have primed them to produce syllables with the same pace leading to an artificial increase in performance at 4.5 Hz. Participants were informed the goal was to correctly identify the target syllable and that the speech they heard was only to increase task difficulty. The purpose of this behavioral task was to prevent participants from intentionally matching their speech to the aural inputs by directing their attention to the syllable identification task.

Figure 1: Examples of EEG and speech tasks. (A) brain entrainment to speech envelope for a slower (2.5 syll/sec) and higher (8.5 syll/sec) syllable rate. Black, cortical EEG responses, green, schematized EEG envelope; red, stimulus waveform, pink, speech fundamental envelope. (B) Schematic of the non-simultaneous (un-cued) speech production task (2.5 Hz rate). (C) Schematic of the cued (simultaneous) production synchronization task (2.5 Hz rate). Pink=auditory stimuli, light blue=speech production samples.
Data analysis: Quantifying synchronization and rate accuracy

We performed analyses using custom scripts written in MATLAB and used TF32 software to examine the rate of acoustic signals (Milenkovic, 2002).

Phase locking value

We measured brain-to-stimulus synchronization (and similarly speech-to-stimulus synchronization) as a function of frequency via phase-locking value (PLV) (Lachaux et al., 1999). Neural and auditory signals were bandpass filtered (±0.5 Hz) around each frequency bin from 1-12 Hz (0.5 Hz steps). The envelope was calculated as the absolute value of the signal’s Hilbert transform. PLV was then computed in each narrow frequency band according to Eq. 1.

\[
PLV = \frac{1}{T} \left| \sum_{t=1}^{T} e^{i[\theta_1(t) - \theta_2(t)]} \right| \quad (\text{Eq. 1})
\]

where \( \theta_1(t) \) and \( \theta_2(t) \) are the Hilbert phases of the EEG and stimulus signals, respectively.

Intuitively, PLV describes the consistency in phase difference (and by reciprocal, the correspondence) between the two signals over time. PLV ranges from 0-1, where 0 represents no (random) phase synchrony and 1 reflects perfect phase synchrony between signals. The PLV was computed for windows of 6 seconds length and averaged within each rate condition. Repeating this procedure across frequencies (1-12 Hz; 0.5 Hz steps) resulted in a continuous function of PLV describing the degree of brain-to-speech synchronization across the bandwidth of interest (e.g., Assaneo et al., 2019). PLVs were then baselined in the frequency domain by centering each function on 0 by subtracting the value of the first (i.e., 1 Hz) frequency bin which is not a stimulation rate. This allowed us to evaluate the relative
change in stimulus-evoked PLV above the noise floor of the metric. We then measured the peak magnitude from each PLV function to trace changes brain-to-speech synchronization with increasing syllable rate.

For speech production-to-stimulus synchronization (which are both acoustic signals), we processed the recordings using the speech modulation procedure described by Tilsen and Johnson (2008). We first discarded the first/last 5 s of each recording to avoid onset/offset artifacts and then normalized the amplitude. We then bandpass filtered the signal (3000-4000 Hz; 4th order Butterworth) to highlight the voiceless whispered energy followed by half-wave rectification to extract the speech envelope. We then lowpass filtered (fc=30 Hz), downsampled (Fs=80 Hz), windowed (Tukey window), and de-meaned the envelope modulated signal to isolate slower speech rhythms. As in the brain-to-stimulus synchronization analysis, we then measured PLV between the acoustic productions and speech stimulus for each rate.

**Speech rate**

As an alternate approach to corroborate the automatic rate measures, we manually counted syllables for each 10 s recording of participants’ non-simultaneous productions from wideband spectrograms computed in TF32. Speech rate was calculated as the number of syllables per second; onset and offset silences were not included in these calculations. Since the audio recordings of implicit speech rate production were 60 s each, we further validated the reliability of syllable counting by applying an automatic peak finding algorithm. Again, the first/last 5 s were discarded to avoid transient onset/offset effects. We then extracted the
Hilbert envelope and smoothed the signal using a 30 ms moving average. The amplitude was normalized before and after envelope extraction. At last, we employed MATLAB’s ‘findpeaks’ function (minpeakheight=0.08, minpeakprominence=0.01, minpeakdistance=117ms) to automatically detect and measure syllable peaks. Visual inspection and auditory playback were used to determine these optimal parameters. The speech rate calculated from the spectrogram and peak finding algorithm were highly correlated ($r=0.95$; $p<0.0001$) confirming the reliability of the automatic analysis approach.

**Statistical analysis**

Unless noted otherwise, we analyzed the data using one-way, mixed-model ANOVAs in R (v1.3.1073) (lme4 package; Bates et al., 2015) with rate (7 levels; 2.5-8.5 Hz) as a categorical fixed effect and subjects as random factor [e.g., PLV~ rate +(1| subject)] to assess whether the brain-to-stimulus and speech-to-stimulus synchrony differed across syllable rate. The Tukey post hoc test for multiple comparisons was used. Moreover, to test whether the PLV at syllable rhythm of 4.5 Hz is enhanced for neural synchrony, following the omnibus ANOVA, we used an *a priori* contrast to compare neural PLV at 4.5 Hz vs. other syllable rates. For production data, we tested whether participants’ produced rate achieved the target syllable rate using one-sample Shapiro *t*-test and Wilcoxon singed rank test for the simultaneous (implicit) and non-simultaneous (explicit) rate production tasks, respectively. Significance in these tests indicates participant’s production speed deviated (e.g., was slower/faster than) the nominal stimulus rate. To assess brain-behavior associations, we first used Pearson’s correlations to test the across individual association after aggregating across rates between
neural and production PLV. We then used repeated measures correlations (rmCorr) (Bakdash & Marusich, 2017) to assess within-subject relations between neural and acoustic synchrony measures. Unlike conventional correlations, rmCorr accounts for non-independence among observations, adjusts for between subject variability, and measures within-subject correlations by evaluating the common intra-individual association between two measures. Initial diagnostics (QQ- and residual plots) were used to verify normality and homogeneity assumptions. Consequently, PLV measures were SQRT transformed to allow for parametric ANOVAs. Behavioral data from the EEG task (i.e., percentage of correctly perceived syllables) were rationalized arcsine transformed (Studebaker, 1985). A priori significance level was set at $\alpha = 0.05$. Effect sizes are reported as $n_p^2$.

Results

**Cortical oscillation synchrony is enhanced at ~4.5 Hz syllable rate**

The percentage of correctly perceived syllables during EEG recordings showed no significant difference [$F_{6,90} = 1.76, p = 0.1162, n_p^2 = 0.11$] across conditions, suggesting participants were equally engaged in the listening task across rates. We evaluated neural-speech PLV (Fig. 2) to assess how ongoing brain activity synchronized to speech (Assaneo & Poeppel, 2018) over an expanded range of ecologically valid syllable rates (2.5-8.5 Hz) characteristic of the most languages (Ding et al., 2017; Poeppel & Assaneo, 2020). Each PLV plot shows a strong peak at the fundamental frequency surrounding the rate of the stimulus as well as additional peaks at harmonic frequencies. Harmonic energy was also present in the acoustic stimuli. An
ANOVA conducted on neural PLV revealed a main effect of syllable rate [$F_{6,90} = 3.76$, $p=0.0022$, $n_p^2 = 0.2$]. An a priori contrast showed that PLV was stronger for 4.5 Hz compared to all other rates ($p=0.026$). Interestingly, 4.5 Hz corresponds with the mean syllable rate in English (Goswami & Leong, 2013; Greenberg et al., 2003) as well as most other languages (Ding et al., 2017; Varnet et al., 2017). Our results reinforce the notion that neural oscillations synchronize to the speech envelope and are modulated by syllable rate. More critically, we observed an enhancement of PLV at the frequency close to the predominant syllable rhythm (4.5 syll/sec) inherent to most languages, suggesting a preferred rate of neural oscillation coherent with listeners’ long-term listening experience.

**Figure 2: Phase-locked neural oscillations synchronize to the rate of the syllable envelope.** The PLV increment from baseline between neuroelectric activities and the stimuli envelope across frequency are enhanced at 4.5 Hz. Note the peak in the PLV close to the nominal syllable rate as well as higher harmonics. Similar harmonics were observed in the spectra of the acoustic stimulus envelopes, owing to the non-sinusoidal nature of speech waveforms. The bottom right panel represents the distribution of peak PLV ($\pm 0.5$ Hz) as a function of stimulus syllable rate. Shading = $\pm 1$ s.e.m.
Spontaneous speech synchronization is restricted to slower rates

Our experiment next examined whether listeners’ cued speech productions were synchronized to the simultaneous audio track at various syllable rates (Fig. 3). Speech-to-stimulus PLVs showed selective peaks at the audio speech rhythm that declined with increasing rate above ~6.5 Hz [main effect of syllable rate: $F_{6,90} = 14.355$, $p < 0.0001$, $n_p^2 = 0.49$]. Post hoc analysis revealed stronger PLV for slower (2.5-4.5 Hz) vs. faster (5.5-8.5 Hz) rates [all $p$-values <0.05]. These results suggest that participants can only synchronize their speech productions to relatively slow syllable rate (i.e., motor performance is akin to a lowpass filter).

Figure 3: Simultaneous speech synchronization to syllable trains is modulated by rate. As in Figure 2 but for PLV computed between the acoustic stimuli and participants’ speech productions.

Correspondence between syllable perception and production

To explore the link between syllable rhythm entrainment in perception and production, we measured participants’ accuracy for producing target syllables under the two experimental
settings: one following an explicit instruction to replicate a previously heard rhythm (“non-simultaneous/un-cued” productions) and the other with implicit instruction to mirror a concurrently presented syllable train (“simultaneous/cued” production). One sample $t$-tests showed that for non-simultaneously produced syllable rate (NSR; Fig. 4A), participants only hit target rates at 4.5 and 5.5 syll/sec [4.5 Hz: $t(14)= -1.49, p=0.16$; 5.5 Hz: $t(14)= -1.74, p=0.10$]. However, the variability in productions also appeared to differ across rates. Indeed, measuring the mean absolute deviation of responses, we found smaller variability in productions at rates of 2.5 and 3.5 Hz vs. 4.5 and 5.5 Hz ($p=0.003$, one-way ANOVA). This suggests at least part of the effect at 4.5-5.5 Hz in Fig. 4A might be attributed to more/less precise productions across rates. Notably, productions deviated from (were slower than) the target speeds above 6.5 Hz indicating they failed to keep pace with the audio stimulus.

Simultaneously produced rate (SSR; Fig. 4B) measures showed highly accurate reproductions for 2.5~4.5 Hz ($p_{2.5}=0.46$, $p_{3.5}=0.13$, $p_{4.5}=0.26$), with slowing of production at higher rates. The results of SSR were consistent with the enhanced speech-to-stimulus PLV at 2.5-4.5 Hz (see Fig. 3).
Figure 4: Participants’ produced speech rate compared to the target rate of auditory stimuli. (A) Speech rate was produced after rhythmic syllable trains were presented (non-simultaneous) with explicit instructions of pace duplication. (B) Participants produced syllables while simultaneously listening to rhythmic streams with implicit rate synchronization. *p < 0.05, significant deviations from the expected rate (red +) based on one sample tests against the nominal (target) rate value. Shaded region = ±1 SD.

Brain-behavior correlations between production and neural speech entrainment accuracy

To explore the relationship between auditory and motor (production) responses, we conducted between and within-subject correlations. Fig.5A suggests a non-significant relation between neural and production PLV when the data are considered on the whole, with respect to each individual. Indeed, rmCorr correlations assessing within-subject correspondence revealed a positive correlation between neural and speech PLV ($r = 0.25$, $p = 0.019$, Fig.5B), indicating an auditory-motor relation in rhythmic synchronization abilities at the individual level.
Figure 5: Correlations between brain and production synchronization to speech. (A) Pearson correlation (between-subjects) aggregating across rate conditions between neural and production PLV. PLV_EEG: neural-to-stimulus PLV; PLV_pro: speech-to-stimulus PLV; (B) Repeated measures correlations (within-subjects) between neural and production PLV. Dots: individual participants’ responses; Solid lines: within-subject fits to each individual’s data across the 7 rates; Dashed line: linear fit across the aggregate sample. * p<0.05, ** p<0.01, *** p<0.001

Discussion

By measuring EEG oscillations and acoustical speech productions in response to syllable trains presented at various rates, the current study aimed to evaluate syllable rate-dependencies in auditory neural entrainment and simultaneous speech synchronization, and possible dynamic relations between these domains. We first confirmed that auditory brain activity robustly synchronizes to the ongoing speech envelope and flexibly adapts to the speed of syllable trains in a rate-dependent manner (Assaneo & Poeppel, 2018; Ding et al., 2016; Rimmele et al., 2021; Will & Berg, 2007). More interestingly, we found that neuroacoustic phase locking was boosted at rates of ~4.5 Hz, corresponding to the putative
dominant syllable rate observed across languages (Ding et al., 2017). Production data showed that simultaneous speech synchronization to audio rhythms was largely restricted to slower syllable rates (2.5-4.5 Hz). In contrast, and converging with neural data, we found “pure” motor rate productions were produced more accurately; participants more precisely matched syllable rates between 4-5 syll/sec even without concurrent auditory cuing. Lastly, correlations between brain and production PLV data extend prior work (Assaneo & Poeppel, 2018; Assaneo et al., 2019) by explicitly linking auditory and motor entrainment skills. We find that individuals with superior auditory entrainment to speech also show enhanced motor speech capabilities in speech-audio synchronization.

Cortical oscillation synchrony is modulated by the heard syllable rates

Corroborating previous M/EEG studies (Assaneo & Poeppel, 2018; Ding et al., 2016; Keitel et al., 2018; Teng et al., 2017), our data reveal that low frequency neural oscillatory signals (2.5-8.5 Hz) robustly phase-lock and closely mirror the rate of auditorily presented speech. Neuroacoustic phase-locking did diminish with increasing rate, consistent with previous findings showing cortical activity fails to synchronize with the envelope of accelerated speech (Ahissar et al., 2001; Nourski et al., 2009). However, entrainment remained above the noise floor even for the fastest syllable rate (8.5 Hz). Accurate neural entrainment to a larger range of frequencies, even some of which are well beyond the regular speeds of intelligible speech (Adams & Moore, 2009; Momtaz et al., 2021; Viemeister, 1979), is perhaps not surprising given the ease at which the auditory system tags temporal acoustic landmarks of
speech and non-speech signals (Doelling et al., 2014; Luo & Ding, 2020; Momtaz et al., 2021; Viemeister, 1979). In order to cope with the varying timescales of temporal patterns in speech, neuronal processing must demonstrate rate flexibility (Saltzman & Munhall, 1989; Van Lieshout et al., 2004). Indeed, neural entrainment to external rhythmicity helps ensure proper signal detection (Besle et al., 2011; Stefanics et al., 2010) and facilitates speech comprehension (Doelling et al., 2014; Giraud & Poeppel, 2012; Luo & Poeppel, 2007). One hypothesis of these phenomena is that continuous speech is discretized and segmented on multi-scale temporal analysis windows formed by cortical oscillation locking to the input speech rhythm (Ghitza, 2011, 2012, 2014; Giraud & Poeppel, 2012). Our data support these general notions that low-frequency activity of auditory cortex flexibly tracks the speed of the speech envelope via phase synchronization of cortical activity.

Interestingly, cortical responses also showed enhanced phase-locking for speech rates proximal to 4.5 Hz. Notably, we observed a bell-shaped rate-dependence with the maximum gain in neural phase locking near 4.5 Hz, which aligns with the dominant spectral profile of syllable rates across languages (Ding et al., 2017). This finding suggests that neural excitability is adjusted to align the acoustic temporal structure of speech such that neural oscillations are tuned to track the acoustics proclivities of natural languages. This is probably coherent with listeners’ long-term listening and speaking experience with the dominant speech rhythms in their language. This supports the notion that neural oscillations coding speech reflect an interplay of input processing and output generation in which the associated neural activities are shaped over time by the statistical structure of speech (Poeppel, 2003).
**Simultaneous speech-audio synchronization is rate restricted**

Paralleling our brain-audio synchronization data, we further asked whether simultaneous speech-audio synchronization is affected by syllable rates from 2.5-8.5 syll/sec. Importantly, we did not explicitly instruct participants to match the audio rate nor did we provide practice on the task, which we speculate can lead to priming effects and apparent enhancements in synchronization at certain rates (cf. Assaneo et al., 2019). The resulting production data demonstrate that participants’ rhythmic speech output does not uniformly synchronize across rates but is instead severely restricted to slower frequencies from 2.5-4.5 Hz. Because the simultaneous production task implicitly instructed listeners to align their self-speech production to heard audio, it necessarily evoked sensorimotor integration. The fact such productions are limited to low rates is consistent with neuroimaging results indicating selective coupling between auditory and motor cortices between 2.5-4.5 Hz (Assaneo & Poeppel, 2018). Moreover, the lack of entrainment at higher frequencies as observed in our EEG data perhaps suggests the sensorimotor effects of producing while also listening to speech might create a mixture of entrained brain processes which interfere or are at least distinct from one another. The shift to slow rates preference in motor speech synchronization also seems reasonable given the risk of articulatory undershooting when speaking fast (Gay et al., 1974), and the speed of articulation is constrained by the biomechanical limits of articulators. Alternatively, this rate-constriction could result from the oscillator tuning of the motor system in which it involuntarily entrains to (i.e., “resonates” with) the auditory stimuli when rates are close to its intrinsic rhythm. It is conceivable that auditory-motor interaction
has adapted its sensitivity to both forms of natural constraints imposed by the articulatory and motor systems.

Neurophysiologically, this low-pass filter shape could also result if motor responses are dominated by lower-frequency rhythms of the brain. Indeed, delta (0.5–4 Hz) oscillations are thought to reflect endogenous rhythms from primary motor cortex (Keitel & Gross, 2016; Morillon et al., 2019) which can emerge in the absence of acoustic stimulation (Ding et al., 2016; Rimmele et al., 2021). Other possible explanations could be due to the cognitive demands of this task, which consumes heavier cognitive load (Zhou et al., 2018) and requires extra neurocomputational time to match the motor program with the auditory input. Higher task demands would tend to result in successful synchronization only at the easiest (slowest) rate conditions. Low-frequency components of the EEG have been linked to cognitive operations such as sustained attention and working memory (Bidelman et al., 2021; Kirmizi-Alsan et al., 2006). However, this explanation seems speculative since we could not explicitly measure brain oscillations during the production tasks. Instead, the lowpass nature of the simultaneous production data seems parsimoniously described in terms of limits to sensorimotor processing, with more severe constraints imposed by the motor component.

Non-simultaneous productions highlight an intrinsic rhythm at 4-5 Hz

Under an oscillatory framework, different aspects of spoken communication arise from neural oscillations that are accessible for both perception and production. Such oscillations could emerge in the context of input processing and output generation and result in the associated
auditory and motor activities that would reflect the structure of speech (Giraud et al., 2007; Giraud & Poeppel, 2012; Liberman & Whalen, 2000).

A second aspect of our study design examined natural speech rate productions via non-simultaneous productions. Some conditions were quite challenging given the rapid production speeds required of the task. This paradigm provided listeners with minimal auditory feedback and thus, better isolated more “pure” motor system responses during speech output. Without concurrent auditory feedback either from their own speech or external stimuli, possible interference confounds from sound-evoked auditory oscillations mentioned earlier are minimized. Surprisingly, we found participants’ productions under these conditions hit target speeds (statistically speaking) only for rates of 4.5 and 5.5 syll/sec. Productions failed to meet targets (i.e., were slower than the nominal rates) at all lower and higher syllable speeds. However, we also note production variability differed as speeds increased (Fig. 4A). While we interpret the non-simultaneous data to reflect motor speech function during limited auditory involvement, an alternate interpretation might be the more explicit instruction of rate imitation. Nevertheless, those findings align with our EEG results on auditory entrainment which similarly showed maximum synchronization at 4.5 Hz and flexibility with wide range of speech rates. This frequency specialization in both the speech perception and production data is suggestive of a “resonance” of intrinsic neural oscillations representing syllable rhythm (Assaneo & Poeppel, 2018; Luo & Poeppel, 2007; Poeppel & Assaneo, 2020).
The notion of an intrinsic 4-5 Hz rhythm receives further support from several other observations: the predominant peak in speech envelope spectra for many languages and speaking conditions (Ding et al., 2017; Goswami & Leong, 2013); the mean syllable duration in English (~200 ms) (Greenberg et al., 2003; Pellegrino et al., 2011); the coordinated articulation or motor gesture trajectory in sound production (Poeppel & Assaneo, 2020); the lips, tongue, and hyoid move with a 5 Hz rhythm during lip-smacking in monkey (Ghazanfar et al., 2012). Neurologically, continuous speech is processed through a temporal integration window of ~200 ms (period of 4-5 Hz) (Luo & Poeppel, 2007). Studies using transcranial alternating current stimulation (tACS) further show that 5 Hz stimulation enhances cortical entrainment and results in better sentence comprehension (Wilsch et al., 2018). The striking coherence between these divergent methodologies, along with the present data, supports the notion of an intrinsic rhythm at ~4-5 Hz, a computational primitive in cortical speech processing that also seems to link input and output processing.

**Differences and limitations to related studies**

Our stimulus paradigm was adapted from previous neuroimaging studies on neural entrainment to speech rhythm (e.g., Assaneo and Poeppel, 2018; Assaneo et al., 2019). However, there are several distinct aspects of the findings presented here. First, our cortical tracking data observed a stronger brain-to-speech phase synchronization at 4.5 syllables/sec which contrasts with previous reports suggesting auditory cortex is invariant in syllable tracking across rates (Assaneo and Poeppel, 2018). Although listening to rhythmic sounds induces motor cortex (Bengtsson et al., 2009; Wilson et al., 2004), our single channel EEG
recordings do not allow us to localize our effects to auditory vs. motor cortex responses, *per se*. In this regard, high-density neural recordings (Assaneo and Poeppel, 2018) revealed enhanced *brain-to-brain* coupling of speech-evoked oscillations between auditory and motor cortices specifically at 4.5 Hz. It is possible then that the gain in cortical phase-locking at 4.5 Hz observed in our data reflects neural entrainment in motor-related regions (Assaneo and Poeppel, 2018). Accordingly, other neuroimaging studies have shown that oscillation power in motor areas modulates auditory cortex tracking of acoustic dynamics to facilitate comprehension (Keitel et al., 2018; Keitel et al., 2017). Given the scalp EEG reflects a mixture of intracranial sources, the effects we observed probably reflect a mixture of entrained oscillations in auditory and motor cortex as suggested by previous MEG studies (Bengtsson et al., 2009; Wilson et al., 2004). Multichannel EEG recordings with source reconstruction analysis could test this hypothesis in future studies. Privileged recruitment of motor brain regions induced by concurrent auditory entrainment may account for the local enhancements in PLV we observe near 4.5 Hz in both our EEG and production data.

Second, we observed a more complex syllable rate-constrained pattern in speech-audio responses (simultaneous productions) but a preferred syllable rhythm for isolated motor synchronization (non-simultaneous productions). To our knowledge, these novel findings have not been observed previously and are only revealed by comparing speech productions with varying degrees of sensory and motor involvement. By explicitly examining multiple modes of production and tasks which tease apart sensory from motor processes, our data establish a link between exogenous and endogenous speech entrainment mechanisms and
further reveal unique specialization at 4-5 Hz in both the auditory and motor modalities. These parallel effects likely trace back to the long-term experience of the listener and dominant syllable rates for input processing and output production. In contrast, with concurrent auditory inputs, the rate-restricted pattern could emerge from the tuning of motor oscillator and its interaction with the sensory system. Future studies are also needed to test whether this oscillator tuning is mediated by the better vs. worse synchronization performance. It is possible the bimodal distribution in speech-rate synchronization observed in prior work (Assaneo et al., 2019) is only apparent with a very large number of participants or those with more heterogeneous backgrounds.

In conclusion, our data establish a positive speech perception-production link for rate synchronization. Both perceptual and motor entrainment for speech processing seem optimized for rates between 4-5 Hz, the putative nominal speech rate across languages. Still, these links are only identifiable when carefully considering the nature of speech production and tasks that isolate motor from sensorimotor processes. Moreover, we find synchronization skills are subject to individual differences, with performance in the perceptual domain predicting skills in motor domain and vice versa. As such, our findings provide support for theoretical notions of an oscillation-based account of speech processing which organize both input and output domains of speech processing.
Chapter 3: Acoustic-driven and cross-language effects on neuro-behavioral synchronization to stress rhythms

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; Rimele 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 the phase alignment of brain oscillations to the speech amplitude envelopes, which segment/parse the continuum speech signals 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 in theta band oscillations which fall within 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 modulation components 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 entrain delta oscillation 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, current dominant theories in language processing, such as Poeppel’s asymmetric sampling in time or Ghitza’s TEMPO 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 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 in English (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 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).

\(^1\) 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.
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 (syllable level) and theta (stress foot 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 concurrently synchronize to stress and syllable rhythms. And the 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. Furthermore, we posited the acoustic phase hierarchy between stress foot and syllable rhythms (Leong, 2012; Leong et al., 2014) would be paralleled in the brain (EEG) as enhanced coupling of delta-
theta oscillations. Lastly, we hypothesized Chinese speakers might have reduced neural responses coding stress foot 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. Such findings would open a window to closely link neural mechanisms with linguistic structure for a more cohesive and comprehensive understanding of speech temporal processing.

Materials & Methods

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, 1971b). 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 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.

**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 (P. Boersma & D 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 (Figure 6).

*Amplitude-signaled stress pattern* (Fig. 6 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. 6 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 seconds 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).
Figure 6: 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.

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 the keyboard (labeled as ‘AAbb’ or ‘aaBB’). There were no time limits for behavioral responses. The 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
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 auditory cortex (Bidelman et al., 2013; Picton et al., 1999). Inter-electrode impedance was maintained below 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—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 (Saleha Khatun et al., 2016). Figure 7A 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. Figure 7B presents the corresponding spectrum of 60-s continuous EEG data.

---

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 11 epochs for the condition of amplitude stress cue of high salience at 1 Hz.
Figure 7: 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 dynamic 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;

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:

\[
PLV = \frac{1}{T} \left| \sum_{t=1}^{T} e^{i(\theta_1(t) - \theta_2(t))} \right| \quad \text{(Eq. 1)}
\]

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

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\textsubscript{Stress} representing brain-acoustic synchronization at the stress level and PLV\textsubscript{Syll} 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 was computed by bandpass filtering the EEG data into two separate bands (i.e., \(m = 1, 2, 3 \pm 0.5\) Hz; \(n = 2, 4, 6 \pm 0.5\) 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.
establish the noise floor of our PSI metric, we applied this identical 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).

**Statistical analysis**

We conducted four-way mixed model analyses of variance (ANOVAs) in R (version 1.3.1073; ‘lme4’ package; Bates et al., 2015) 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\textsubscript{Stress}, PLV\textsubscript{Syll}, and nmPSI. 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. 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\textsubscript{Stress} between 2 Hz versus the other stress rates. Similar contrast was conducted for PLV\textsubscript{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
listener’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 syllables) were rationalized arcsine transformed (Studebaker, 1985). A priori significance level was $\alpha = 0.05$. Effect sizes were presented as $n_p^2$.

Results

Our behavioral task was primarily designed to keep subjects attentive and awake rather than assess stress perception, per se. Indeed, correct percent performance showed no group differences and results approached chance level (see Appendix; Fig. S1).

Brain to speech tracking at the stress level (PLV$_{\text{Stress}}$)

We examined how neural oscillations phase lock to the external (i.e., acoustic) stress rhythms at rates of 1, 2, and 3 Hz (Fig. 8). An ANOVA conducted on PLV$_{\text{Stress}}$ revealed significant main effects for group ($F_{1,32} = 4.69, p = 0.038, n_p^2 = 0.13$), stress rate ($F_{2,352} = 10.03, p = 0.0001, n_p^2 = 0.05$), cue ($F_{1,352} = 4.02, p = 0.046, n_p^2 = 0.01$), along with a two-way stress cue * salience interaction ($F_{1,352} = 8.65, p = 0.003, n_p^2 = 0.02$). 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$_{\text{Stress}}$ ($p_1$ 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$_{\text{Stress}}$ 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, higher PLV\textsubscript{Stress} was found for low compared to high stress salience ($p < 0.001$). Our results highlight the differences in exogenous neural-acoustic synchronization across individuals’ language experience and stress rhythm rates.

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 Figure 3, we conducted an \textit{a priori} contrast of 2 Hz against other rates by group and stress cue. Our assumption was confirmed in that PLV\textsubscript{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 ($p\text{amplitude} = 0.98$; $p\text{duration} = 0.373$). Our results demonstrate an enhancement of speech-to-brain phase-locking (PLV\textsubscript{stress}) at the frequency inherent to natural English stress rhythm (2 Hz), that is also shaped by individuals’ language exposure.

Figure 8: \textbf{Brain oscillations synchronize to the rate of stress rhythms.} Cross-linguistic PLV\textsubscript{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\textsubscript{Stress} enhancement at 2 Hz—corresponding to the natural stress rate found in English. Panel C outlines the main effects of rate, arrogating data across cue and salience conditions. PLV\textsubscript{Stress} refers to neural-to-stress rhythm phase locking; errorbars = ± 1 s.e.m.
Brain to speech tracking at syllable level (PLV<sub>syllable</sub>)

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<sup>3</sup> of 2, 4, and 6 Hz, which are 2-times faster than stress rhythms. An ANOVA conducted on PLV<sub>Syllable</sub> revealed a main effect of stress rate ($F_{2,352} = 5.93, p = 0.003, \eta^2_p = 0.03$) and two-way interactions of stress cue * group ($F_{1,352} = 5.79, p = 0.017, \eta^2_p = 0.02$) and stress cue * rate ($F_{2,352} = 5.53, p = 0.004, \eta^2_p = 0.03$) (Fig. 9).

Pairwise comparisons revealed English speakers had higher PLV<sub>Syllable</sub> 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<sub>Syllable</sub> than duration-signaled stress ($p = 0.029$), consistent with PLV<sub>Stress</sub>. The stress cue * rate interaction was driven by stronger PLV<sub>Syllable</sub> 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<sub>Syllable</sub> than amplitude cues.

Consistent with PLV<sub>Stress</sub>, we performed an <i>a priori</i> 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

---

<sup>3</sup> 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. However, PLV magnitudes were, on average, similar between amplitude and duration-cuing stress (e.g., Fig. 8) suggesting any jitter introduced in our stimuli did not negatively impact PLV.
group and stress cue. We observed an enhanced $PLV_{\text{Syllable}}$ at 4 Hz ($p = 0.011$) only in English speakers for amplitude cues. 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). Surprisingly, this preferred syllable rate of 4 Hz in natural speech failed to result in neural coding enhancement for Chinese speakers, possibly due to their limited brain tracking at the stress level (cf. Fig. 8).

Together, these findings validate that neural-acoustic synchronization is modulated by syllable rates and enhanced because of alignment to the natural syllable rate and, critically, the presence of stress-level brain tracking.

**Figure 9:** Brain oscillations phase lock to the rate of syllable rhythms. Cross-linguistic $PLV_{\text{Syllable}}$ comparisons by stress rate and salience signaled by (A) amplitude envelope and (B) syllable duration. (C) $PLV_{\text{Syllable}}$ 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 effects of rate, arrogating data across cue and salience conditions. $PLV_{\text{Syllable}}$ refers to neural-to-syllable rhythm phase locking; error bars $= \pm 1$ s.e.m.
Cross-frequency coupling within the brain (nmPSI)

Figure 10 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. 10C). 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 ($p_{amplitude} < 0.0001$; $p_{duration} = 0.0026$) (Figs. 10A and B, right panels). However, no group differences were observed for more salient (i.e., high) stress stimuli—true for both stress cues (Figs. 10A & B). Generally speaking, both Chinese and English speakers also 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. Crucially, this phase hierarchy in the brain was enhanced in English listeners who possess extensive experience in a language specifically structured by stress patterns.

Figure 11 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. 11B), 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. 11A; 3-way ANOVA: $p_{cue} = 0.727; p_{magnitude} =
0.196$). These findings suggest that English listeners coded and constructed consistent neural
coherence that was equally robust across varying acoustics. That is, English speakers’ neural
responses effectively entrained to the hierarchical stress rhythms even in scenarios where the
aural information cuing stress was relatively weak in perceptual salience. In stark contrast,
Chinese listeners’ hierarchical entrainment nmPSI was more susceptible to acoustic
variations signaling stress patterns.

**Figure 10:** 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 arrogated
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.
Figure 11: Effects of acoustic stress cue and salience on cross-frequency coupling. (A) nmPSI in English 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.

**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\text{Stress}, PLV\text{Syllable}). Figure 7 depicts a positive correlation between nmPSI and PLV\text{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: $r = 0.082, p = 0.245$). Collectively, these findings highlight the close link between neuro-audio synchronization and neural coupling of delta-theta oscillations coding the stress patterns in speech that are evident exclusively in English.
speakers who have extensive stress-language exposure.

Figure 12: Repeated measure correlations between internal cross-frequency coupling and external audio-neuro tracking in (A) English speakers, and (B) native Mandarin-Chinese speakers. PLVStress 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.

Discussion

Here, we provided 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, analyses of 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 (as in English listeners) also showed enhanced cortical-acoustic tracking to stress. Collectively, our findings suggest 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. The following sections interpret the results in further detail, addressing the study’s main objectives.

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

Prominent oscillatory-based models (e.g., TEMPO or asymmetric sampling in time) 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; Rimmel et al., 2021). However, critical to our findings is the proposition that delta oscillations are associated with the hierarchical nesting role of stress rhythms. Hence, we
propose that delta oscillations serve a higher-order mechanism, extending beyond simple stress segmentation to facilitate temporal integration and establish a cohesive phonological representation. 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 neuroimaging 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 essential for discerning English stress patterns is likely due to their more limited exposure and motor practice of a stress-dominant language.
Such experience-dependent effects emerge in both groups’ EEG. Chinese responses were severely hindered by stress manipulations whereas English responses were largely impervious (Fig. 5). Consequently, Chinese speakers struggle to capture acoustic stress regularities in ongoing speech and failure to segment delta-size chunks might be due to “perceptual narrowing” of speech representations that are not behaviorally relevant cues in Mandarin (Jeng et al., 2011; Tierney & Nelson III, 2009). Such 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 coherence, readily achieved by native English speakers, further suggests delta oscillatory synchronization is not merely a passive “bottom-up” mechanism. Rather, it is sculpted by a “top-down” regulation fostered by a listener's lifetime of sensory experiences and accumulated phonological stress knowledge inherent to speaking a specific language.

**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 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 and delta-stress—and each
band could track different frequency-specific acoustic information. However, such architecture does not necessarily 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\textsubscript{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; Rimele 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
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 PLVStress results, the absence of 4 Hz syllabic enhancements in Chinese speakers presumably results from their limited neural coding of stress patterns at the suprasyllabic 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 test this possibility.

**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). Our nmPSI results uncovered significant phase-phase coupling between delta and theta neural oscillations for
both English and Chinese speakers\(^4\), 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. 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\(_{\text{Stress}}\), 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

\(^4\) The nmPSI responses in Chinese speakers were below baseline under for low salience, amplitude-signaled stress.
external neural tracking across multiple scales, that also varies in a language-dependent manner.

**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\textsubscript{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\textsubscript{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, which posits an innate tendency for intensity-contrast 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 English and Chinese speakers. 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. However, 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 attempted to construct an internalized hierarchy for duration-based stress stimuli. A possible explanation may lie in the inherent structure of the Chinese language. It can be argued that 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-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), leading to effective but diminished delta-theta brain coherence.

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.
Chapter 4: General Discussion

Across two studies, the current dissertation demonstrates a nuanced interrelation between speech rhythm hierarchies, neural oscillations, individual linguistic experiences, and acoustic-modulation effects.

The first study elucidated a notable parallel between auditory neural entrainment and motor-speech production, with a distinctive optimization at the syllable rates of 4-5 Hz, the purported dominant syllable rate across various languages (Ding et al., 2017; Greenberg, 1999; Tilsen & Johnson, 2008). Furthermore, it underlined individual variations in synchronization skills, highlighting an intimate relationship between perceptual (EEG) and motor (vocal) domains in speech processing. The parallels across modalities could result from the dynamics of speech-motor system coupled with experience-dependent tuning of the perceptual system via the sensorimotor interface.

The second study examined the role of multi-scale neural oscillations in delineating the hierarchical structure of stress rhythm and their dependence on language experience. Remarkably, native English speakers exhibited optimal synchronization at dominant stress and syllable rates pertinent to their linguistic background (Dauer, 1983; Ding et al., 2017; Leong, 2012; Tilsen & Arvaniti, 2013; Tilsen & Johnson, 2008), underscoring a facilitation in their brain responses from their lifelong exposure to a stress-dominant language. This finding not only highlights the role of language experience in fine-tuning neural synchronization but also implicates more intricate neural mechanisms that foster phonological representations. In essence, the data suggest mechanisms for encoding stress
operate through a structured hierarchy of neural oscillators functioning at various temporal scales. Delta oscillations assume a nesting role to orchestrate the simultaneous phase tracking of theta oscillations to speech. In turn, such coupling facilitates speech processing at both supra-syllabic and syllabic levels. Furthermore, the linguistic experience-dependent plasticity on neural entrainment emphasizes the critical role of “top-down” mechanisms that shape ongoing sensory representations for speech.

Collectively, this dissertation offers converging support for the oscillation-based theories of speech processing (Ghitza, 2011; Ghitza & Greenberg, 2009; Hickok & Poeppel, 2007; Poeppel, 2003; Poeppel & Assaneo, 2020), and provides a window into the dynamics between external auditory entrainment, internal hierarchical coherence, and motor synchronization in the brain that is shaped by one's linguistic background. Our findings also pave the way for a more cohesive and comprehensive understanding of speech perception-production links in the broader framework of synchronization phenomena.

**Limitations**

While our studies provide new characterizations of speech rhythm synchronization and neural mechanisms, several limitations merit discussion. Firstly, the generalizability of the findings is somewhat restricted due to the synthesized and high-rhythmic nature of our stimuli, which also carried no linguistic meaning. It remains to be seen whether the observed effects hold when studying more naturalistic stimuli. Secondly, the studies predominantly emphasized syllable rate and stress rhythm, potentially overlooking other linguistic elements (e.g., phonemic gamma processing) that might influence neural entrainment (Hyafil et al., 2015;
Lizarazu et al., 2023). Moreover, study #1 (He et al., 2023) revealed significant individual variation in neural and production synchronization skills which was not fully explained by differences in auditory entrainment alone. This points to yet unidentified factors that may play a significant role in modulating real time speech processing beyond mere entrainment processes.

**Future directions**

Future research should broaden the linguistic spectrum to encompass a more diverse array of languages, potentially uncovering new, universal or distinctive synchronization patterns, providing insights on language processing and development. Moreover, studies using real life communicative speech could delve into a realistic intricate neural oscillatory dynamic, providing insights into the adaptability and plasticity of the human brain in response to environment demands and linguistic exposure. Integrating advanced neuroimaging techniques (e.g., fMRI) may offer finer granularity in understanding the neural source locations of the observed entrainment processes. Furthermore, exploring other linguistic elements and their interactions with neural entrainment mechanisms could pave the way for a more comprehensive theory of speech perception and production.
Reference


Arai, T., & Greenberg, S. (1997). The temporal properties of spoken Japanese are similar to those of English. EUROSPEECH,


attended events. *Journal of Neuroscience, 31*(9), 3176-3185.


Drewes, J., & VanRullen, R. (2011). This is the rhythm of your eyes: the phase of ongoing electroencephalogram oscillations modulates saccadic reaction time. Journal of Neuroscience, 31(12), 4698-4708.


Fridriksson, J., Hubbard, H. I., Hudspeth, S. G., Holland, A. L., Bonilha, L., Fromm, D., &


Leong, V. (2012). *Prosodic rhythm in the speech amplitude envelope: Amplitude modulation phase hierarchies (AMPHs) and AMPH models*


lateralization as ‘asymmetric sampling in time’. *Speech Communication, 41*(1), 245-255.


Skljarov, O. (1996). The bifurcation model of the speech rhythm and stuttering. 1st ETRW on Speech Production Modeling: From Control Strategies to Acoustics; 4th Speech Production Seminar: Models and Data,


Voss, R. P. (1975). I/F NOISE"" IN MUSIC and SPEECH.


Appendix

Behavioral responses

Figure S1 illustrates the correct percentage of stress pattern perception during EEG recording. A mixed model ANOVA showed significant three-way interaction effects of stress cue * magnitude * group ($F_{1,352} = 4.55, p = 0.034, \eta^2_p = 0.01$) and stress cue * rate * group ($F_{2,352} = 3.64, p = 0.027, \eta^2_p = 0.02$). Post hoc tests revealed the stress cue * magnitude * group interaction was attributed to duration-modulated stress rhythm under low stress magnitude, resulting in higher correct percentages compared with amplitude-sigaled stress ($p = 0.363$) within the Chinese group. Meanwhile, the cue * rate * group interaction resulted from duration-modulated stress rhythm at a rate of 3 Hz, leading to higher accuracy relative to amplitude-sigaled stress ($p = 0.000$) within the English group. However, it is important to recognize that neither interaction was driven by group differences, highlighting that both English and Chinese speakers performed comparably (and near chance level) in behaviorally identifying the stressed syllables (English: 55.93% ± 22.67%; Chinese: 59.30% ± 26.77%). While we hesitate to interpret null findings, a plausible explanation for these floors effects might be that both native and non-native speakers found the task too difficult to perceive the location of stressed syllables (or stimuli were too perceptually ambiguous) or participants did not fully understand the task instructions. Future studies using two-alternative forced choice approaches might also produce more definitive perceptual results.

Figure S1: Group comparisons in correctly perceiving stress patterns by stress rate and magnitude modulated by (A) amplitude envelope and (B) syllable duration. (C) no group differences in correct percentage by stress rates with a dataset arrogated across across cue and salience conditions. error bars = ± 1 s.e.m.
Institutional Review Board
Division of Research and Innovation
Office of Research Compliance
University of Memphis
315 Admin Bldg
Memphis, TN 38152-3370

PI: Gavin Bidelman
Co-Investigator:
Advisor and/or Co-PI:
Department: IIS-Institute For Intelligent Systems
Study Title: Neural correlates of complex auditory perception
IRB ID: 2370
Submission Type: Renewal
Level of Review: Expedited

IRB Meeting Date:
Decision: Approved
Approval Date: April 15, 2020

Research Notes:
Findings:

The IRB has reviewed the renewal request. The University of Memphis Institutional Review Board, FWA00006815, has reviewed your submission in accordance with all applicable statuses and regulations as well as ethical principles.

Approval of this project is given with the following obligations:

1. If this IRB approval has an expiration date, an approved renewal must be in effect to continue the project prior to that date. If approval is not obtained, the human subjects consent form(s) and recruiting material(s) are no longer valid and any research activities involving human subjects must stop.
2. When the project is finished a completion form must be completed and sent to the board.
3. No change may be made in the approved protocol without prior board approval, whether the approved protocol was reviewed at the Exempt, Expedited or Full Board level.
4. Exempt approval are considered to have no expiration date and no further review is necessary unless the protocol needs modification.
5. Human subjects training is required every 2 years and is to be kept current at citiprogram.org.

Thank you,
James P. Whelan, Ph.D.
Institutional Review Board Chair