

# Testing Low-Frequency Neural Activity in Sentence Understanding

by

Chia-Wen Lo

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
(Linguistics)  
in The University of Michigan  
2021

Doctoral Committee:

Associate Professor Jonathan R. Brennan, Chair  
Assistant Professor Natasha Abner  
Assistant Professor Nai Ding, Zhejiang University  
Professor Richard Lewis

Chia-Wen Lo

chiawenl@umich.edu

ORCID iD: 0000-0002-5326-7521

© Chia-Wen Lo 2021

## DEDICATION

This dissertation is dedicated to my grandmother and Samuel Epstein.

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere appreciation to Dr. Jonathan Brennan, my advisor, for his valuable insights, enthusiasm, and encouragement of this dissertation. His instructive guidance and knowledgeability have served as an indispensable inspiration for my research journey in neurolinguistics. I would also like to thank Dr. Natasha Abner, Dr. Nai Ding, and Dr. Richard Lewis, the members of my committee, for their constructive suggestions and thoughtful critiques. My special thanks are also extended to Meredith Belloni, Haoyu Du, Samia Elahi, Ibrahim Lubis, and Maya Roberts, the research assistants in the Computational Neurolinguistics Lab, for their help in data collection. I am also grateful for the support from the faculty and colleagues in the Linguistics department and the psycholinguistics discussion group. Special thanks also goes to Yu-han Kao and Lisa Lau for their advice in signal processing and dissertation writing. Finally, I wish to thank my family and my friends for their unwavering support and encouragement.



## TABLE OF CONTENTS

DEDICATION . . . . .	ii
ACKNOWLEDGEMENTS . . . . .	iii
LIST OF FIGURES . . . . .	vi
LIST OF TABLES . . . . .	vii
LIST OF APPENDICES . . . . .	viii
ABSTRACT . . . . .	ix
CHAPTER	
<b>I. Introduction . . . . .</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Research questions . . . . .	9
1.2.1 Hierarchical representation vs. Lexical representation . . . . .	10
1.2.2 Temporal dynamics of cortical tracking in delta oscillations . . . . .	11
1.2.3 Semantic compositionality . . . . .	13
1.3 Methodology . . . . .	14
1.4 Dissertation outline . . . . .	14
<b>II. Lexical Representation Can't Fully Capture the Functional Role of Low-Frequency Neural Activity During Language Comprehension . . . . .</b>	<b>16</b>
2.1 Introduction . . . . .	16
2.1.1 Cortical tracking of abstract linguistic structure . . . . .	18
2.1.2 Supporting evidence for the delta oscillations in language comprehension . . . . .	24
2.2 Experiment . . . . .	26
2.2.1 Participants . . . . .	27
2.2.2 Materials . . . . .	27
2.2.3 Procedure . . . . .	30
2.2.4 EEG recording and data analysis . . . . .	31
2.2.5 Predictions . . . . .	33
2.2.6 Simulations . . . . .	33
2.3 Results . . . . .	35
2.3.1 Results of the simulations . . . . .	35
2.3.2 Results of the experiment . . . . .	36
2.4 Discussion . . . . .	42
2.5 Summary of the study . . . . .	45

<b>III. Testing Temporal Boundaries of Composition in Low-Frequency Neural Activity</b> . . . . .	<b>46</b>
3.1 Introduction . . . . .	46
3.1.1 Comprehension in low-frequency oscillations . . . . .	48
3.1.2 Oscillation-based model vs. Evoked-response model . . . . .	55
3.1.3 Speech rhythmicity in low-frequency oscillations . . . . .	59
3.1.4 Summary . . . . .	60
3.2 Experiment . . . . .	62
3.2.1 Participants . . . . .	62
3.2.2 Materials . . . . .	63
3.2.3 Procedure . . . . .	66
3.2.4 EEG recording and data analysis . . . . .	67
3.2.5 Predictions . . . . .	68
3.3 Results . . . . .	69
3.3.1 Comprehension score . . . . .	69
3.3.2 EEG results . . . . .	70
3.4 Discussion . . . . .	79
3.5 Summary of the study . . . . .	83
<b>IV. Testing Semantic Compositionality in Low-Frequency Neural Activity</b> . . . . .	<b>85</b>
4.1 Introduction . . . . .	85
4.1.1 Delta oscillations correlate with syntactic and semantic processing . . . . .	86
4.1.2 Semantic variables in semantic processing . . . . .	91
4.2 Experiment . . . . .	97
4.2.1 Participants . . . . .	97
4.2.2 Materials . . . . .	98
4.2.3 Procedure . . . . .	104
4.2.4 EEG recording and data analysis . . . . .	105
4.2.5 Predictions . . . . .	106
4.3 Results . . . . .	106
4.4 Discussion . . . . .	109
4.5 Summary of the study . . . . .	112
<b>V. Conclusion</b> . . . . .	<b>114</b>
5.1 Summary of contributions . . . . .	114
5.2 Directions for future work . . . . .	115
<b>APPENDICES</b> . . . . .	<b>118</b>
<b>BIBLIOGRAPHY</b> . . . . .	<b>123</b>

## LIST OF FIGURES

### Figure

1.1	An illustration of neural entrainment to speech . . . . .	9
2.1	Power spectrum of speech stimuli . . . . .	29
2.2	Simulated power spectra for three models . . . . .	36
2.3	Evoked power for each condition . . . . .	37
2.4	Inter-trial phase coherence for each condition . . . . .	38
2.5	Induced power for each condition . . . . .	39
2.6	EPn at each frequency of interest . . . . .	40
2.7	ITPC at each frequency of interest . . . . .	41
2.8	IP at each frequency of interest . . . . .	42
3.1	Power spectrum of speech stimuli for different speech rates . . . . .	65
3.2	Power spectrum of speech stimuli for different rhythmicity . . . . .	65
3.3	Accuracy percentage for each condition (Red dot: mean; Red line: +/-1 SD) . . . . .	70
3.4	Inter-trial phase coherence for conditions with different speech rates. Regular:75ms shows the results from the subset of N = 17. . . . .	71
3.5	Normalized evoked power for conditions with different speech rates. Regular:75ms shows the results from the subset of N = 17. . . . .	72
3.6	ITPC and EPn across conditions with different speech rates. Peaks at 7-Hz in the 250/syl condition serves as a baseline activation. . . . .	74
3.7	Inter-trial phase coherence for conditions with different rhythms. . . . .	75
3.8	Evoked power for conditions with different rhythms. . . . .	76
3.9	ITPC and EPn across conditions with different rhythmicities. Peaks at 7-Hz in the 250/syl condition serves as a baseline activation. . . . .	77
3.10	The comparison of inter-trial phase coherence and evoked power for conditions with different speech rates (17 subjects included). Peaks at 7-Hz in the 250/syl condition serves as a baseline activation. . . . .	79
4.1	An example for word relations in E-HowNet. . . . .	99
4.2	Results of plausibility and specificity . . . . .	100
4.3	Results of predictability . . . . .	101
4.4	Results of word surprisal . . . . .	102
4.5	Results of semantic similarity . . . . .	103
4.6	ITPC for each condition . . . . .	107
4.7	ITPC at 1 Hz . . . . .	108
4.8	EPn for each condition . . . . .	108
4.9	EPn at 1 Hz . . . . .	109

## LIST OF TABLES

### Table

2.1	Summary of two accounts and predictions for reversed phrases . . . . .	24
2.2	Stimuli design. . . . .	28
2.3	Predictions of each condition . . . . .	33
3.1	Experimental conditions and frequency of interest . . . . .	64
3.2	Example stimuli . . . . .	64
4.1	Example stimuli for the norming process. . . . .	98
4.2	An example for extracting the value from Google BERT . . . . .	101
4.3	An example for semantic similarity . . . . .	102
A.1	Stimuli for Chapter II and III . . . . .	120
B.1	Stimuli for Chapter IV . . . . .	122

## LIST OF APPENDICES

### Appendix

- A. Supplementary material for Chapter II and III . . . . . 119
- B. Supplementary material for Chapter IV . . . . . 121

## ABSTRACT

Human language has the unique characteristic where we can create infinite and novel phrases or sentences; this stems from the ability of composition, which allows us to combine smaller units into bigger meaningful units. Composition involves us following syntactic rules stored in memory and building well-formed structures incrementally. Research has shown that neural circuits can be associated with cognitive faculties such as memory and language and there is evidence indicating *where* and *when* the neural indices of the processing of composition are. However, it is not yet clear “how” neural circuits actually implement compositional processes. This dissertation aims to probe “how” composition of meaning is represented by neural circuits by investigating the role of low-frequency neural activity in carrying out composition. Neuroelectric signals were recorded with Electroencephalography (EEG) to examine the functional interpretation of low-frequency neural activity in the so-called delta band of 0.5 to 3 Hz. Activities in this band have been associated with the processing of syntactic structures (Ding et al. 2016). First, whether these activities are indeed associated with hierarchy remains under debate. This dissertation uses a novel condition in which the same words are presented, but their order is changed to remove the syntactic structure. Only entrainment with syllables was found in this “reversed” condition, supporting the hypothesis that neural activities in the delta band entrain to abstract syntactic structures. Second, we test the timing for language users to

combine words and comprehend sentences. How comprehension correlates with this low-frequency neural activity and whether it represents endogenous neural response or evoked response remains unclear. This dissertation manipulates the length of syllables and regularity between syllables to test the hypotheses. The results support the view that this neural activity reflects endogenous response and suggest that it reflects top-down processing. Third, what semantic information modulates this low-frequency neural activity is unknown. This dissertation examines several semantic variables typically associated with different aspects of semantic processing. The stimuli are created by varying the statistical association between words, world knowledge, and the conceptual results of semantic composition. The current results suggest that low-frequency neural activity is not driven by semantic processing. Based on the above findings, we propose that neural activities in the delta band reflect top-down predictive processing that involves syntactic information directly but not semantic information.

## CHAPTER I

### Introduction

When listening to speech input, language users need to decompose connected speech into sounds and then combine sounds into syllables, syllables into words, words into phrases, and phrases into sentences. This dissertation aims to probe how human brains manage these concurrent processes by investigating low-frequency neural oscillations. This chapter begins by introducing the neural mechanisms regarding sentence processing to be examined and provides an overview of the major research questions in the dissertation.

#### 1.1 Motivation

Language is a complex system that involves multiple processes interacted (Hagoort, 2019). When listening to an utterance, a parser needs to make pragmatic inference and integrate linguistic and non-linguistic information to achieve intended interpretation of the utterance. As Hagoort (2019) points out, language is not merely a single-word processing but involves multiple networks interacted. This multiple network model is contrary to the classical models of language processing, which strictly specify that the Broca's area in the left inferior frontal cortex relates to language production while the Wernicke's area in the temporal cortex correlates with language comprehension. In the multiple network model, specifically, language is de-



composed and encoded as linguistic units such as the acquired linguistic knowledge (e.g. the meaning of the lexicon and the syntactic rules) and linguistic operations (e.g. the ability of generating larger or new sentences and the ability of memory retrieval). These processes are supported by the left inferior frontal gyrus (LIFG) and other sub-regions in the temporal cortex. Next, in order to know whether the utterance includes new information and integrate the current utterance into connected utterances, the involvement of attentional control systems is also required. The evidence of the above processes is observed in the right inferior frontal gyrus and the right angular gyrus (Menenti et al., 2009). Finally, the parser needs to integrate linguistic and non-linguistic information in order to get the core meaning of the utterance.

From the above, language processing involves complex and hybrid processes. Brain areas involved in language processing are indeed more extended to both hemispheres but biased with left. However, how computation for each process implemented in the brain still needs further investigation. This dissertation aims to focus on the compositional processes during sentence comprehension. The operations of syntactic rules and the retrieval of semantic knowledge are required when language users combine words into phrases/sentences. How human brains manage these compositional linguistic processes concurrently and achieve successful sentence comprehension remain unclear. Previous studies have provided fruitful evidence for “when” and “where” these processes occur (see review in Pykkänen and Brennan, 2019; Lau, 2018). Many studies examine this compositional process by comparing well-formed sentences and random word lists. For example, Bemis and Pykkänen (2011) examined the processing of two-word simple phrases by comparing the neural activities between the phrase (e.g. red boat), the non-phrase (e.g. xkq boat) and

the word list (e.g. cup boat) during an MEG recording. Their results have shown that an increased activity was found in the second word at roughly 250 ms in the anterior temporal lobe (ATL) and at about 400 ms in the ventromedial prefrontal cortex (vmPFC) for the phrase condition. In line with other relevant studies (e.g. Bemis and Pylkkänen, 2013a; Bemis and Pylkkänen, 2013b), these results suggest that the ATL may be a key region involved in linguistic combinatorics. In addition, several recent fMRI studies have shown greater activation for sentences vs. word lists in brain regions such as inferior frontal gyrus (Pallier et al., 2011; Schell et al., 2017; Zaccarella et al., 2017), posterior superior temporal sulcus (Zaccarella et al., 2017), anterior temporal lobe (Humphries et al., 2006; Matchin et al., 2017), angular gyrus (Humphries et al., 2006; Matchin et al., 2017), and temporal parietal junction (Matchin et al., 2017).

The above studies have shown the possible timing and location for compositional processes. As mentioned before, compositional processes involve operations of syntactic rules and semantic knowledge. Previous studies have also investigated to what extent syntactic and semantic knowledge involve in the modulation of neural responses. Ample studies address this issue by adopting jabborwocky sentences, which preserve grammatical elements but replace content words with pseudowords. For example, Humphries et al. (2006) compared normal sentences with jabborwocky sentences (e.g. the solims on a sonting grilloted a yome and a sovir) using fMRI scanning. Greater blood-oxygen-level-dependent (BOLD) signals were observed in the left inferior frontal gyrus, left angular gyrus, and left superior temporal gyrus for the normal sentence, compared to the jabborwocky sentences, suggesting that these regions involve word-level semantic access. They also found a greater activation in the left temporal lobe and the left angular gyrus for the normal sentences, compared

to the semantic incongruent sentences, suggesting that a combinatorial semantic effect involves in these regions. As Pylkkänen and Brennan (2019) point out, jaborwocky sentences are marked as “minus semantics” but still have rich argument structure and grammatical relationships. However, it is still unclear how participants comprehend jaborwocky sentences. In addition, several studies demonstrated various results: some studies found a greater activation in the LATL for jaborwocky sentences (Friederici et al., 2000; Humphries et al., 2006) while other studies found a greater activation in the LATL only for the normal sentences (Pallier et al., 2011; Matchin et al., 2017). Another approach is to use artificial grammar to examine pure syntactic processing, which requires participants to learn new syntactic rules and semantics for a new language. Several studies have shown that the left inferior frontal region (BA 44 and 45) involve in the processing of artificial grammar (Friederici et al., 2006; Petersson et al., 2012). For example, Petersson et al. (2012) have shown that a greater BOLD activation in BA 44 and 45 when participants read the sentences with syntactic violation in an artificial grammar paradigm, relative to the well-formed sentences. These results suggest that the hierarchical processing may involve in BA 44 and 45. However, as Pylkkänen and Brennan (2019) point out, activation for structure building is not robustly observed in these regions, contrary to the hypothesis that these regions involve structure building.

The above studies have shown promising findings for examining the timing and location of compositional processing from a formal experimental design. One critical issue is that how to generalize the above findings in the lab to a more naturalistic setting. Recent studies have built computational models to explicitly account for how these compositional processes occur in a naturalistic stimuli. For example, different from the previous formal approach, Brennan et al. (2012) had participants listen to

a chapter from *Alice's Adventures in Wonderland*. They used linear regression to estimate the relationship between the BOLD signals and the number of phrase completion. Their results reveal that a positive correlation between the BOLD signals and the number of word-by-word phrase completion in the LATL, but not the LIFG. Moreover, Brennan and Pylkkänen (2016) also show that a correlation between the number of parsing steps from a left-corner parsing strategy and the MEG activity was found at around 350-500 ms post word onset in the LATL when participants listened to the story. Align with the results from the formal experimental design, these computational models provide evidence from a naturalistic setting that the LATL involves the processing of composition. However, as Pylkkänen and Brennan (2019) mention, although computational models can explicitly quantify the fit between neural signals and predictors from the speech stimuli, these computational models indeed require several assumptions such as what measures of predictors adopted may affect model results. How to set parameters in computational models still requires more future research.

Neural basis of composition may be affected by how parsers retrieve lexicon and build structure incrementally. From the above studies, we see consistent results that the LATL involves in compositional processing. However, the LIFG is not robustly sensitive to this processing. Divergent results were also observed in control experiments with jaborwocky sentences included. It remains unclear how to disentangle syntactic and semantic processing to account for the neural activation in compositional processing. Moreover, results from computational models may be affected by the assumptions of parameters. It still requires more fine-grained research for how to use computational models to quantify the relationship between neural signals and predictors we are interested in composition. The dissertation aims to provide neural

evidence from the frequency-domain. Similar to the basic two-word composition, the dissertation uses four-syllable sentences and manipulates syntactic structure and semantics in the stimuli to tease apart whether syntactic or semantic information plays a role in modulating low-frequency neural responses, as detailed later.

From the above, we see some evidence for when and where the compositional processes occur. However, it is still unclear “how” these compositional processes are reflected on neural responses when we parse a continuous speech and segment it into meaningful units. Some studies have proposed formal models to account for how the computation of hierarchical and compositional processes integrate and modulate neural activities. Specifically, Martin (2020) takes linguistic representations as N-dimensional manifolds of neural trajectories. That is, multiple cell assemblies activate over time and reflect different spatiotemporal patterns of neural signals from acoustic realization to abstract high-level information during language comprehension. The system coordinates the hierarchical information by neural gain modulation and then neural representations are able to transform from one to another dimension. When a parser listens to a speech input, the speech input would be transformed into acoustic realization and then reduced to neural coding space by using high-level linguistic knowledge. Gain modulation is used to combine the processes in between. For example, when a speech envelope is recognized as a syllable or phoneme successfully, the gain modulation would inhibit the process of edge detection of speech envelope and then move forward to next process - lexical and morphological operations. These processes are realized by neural oscillations. As Martin (2020) points out, neural oscillations can offer insights from either latency in the time-domain and power and phase in the frequency-domain. As the compositional processes involve multiple concurrent processes, examining from neural oscillations serves as a critical step to

see how these processes modulate neural responses across different timescales. It is clear that our brains indeed use information from different timescales. However, as Martin (2020) mentions, it remains unclear whether neural oscillations truly reflect these computations and whether these oscillatory activities are endogenous neural responses. Additionally, the functional interpretation of each frequency band remains to be further investigated. The dissertation will offer neural evidence as the first steps towards connecting these issues. The first steps involve connecting rhythmic patterns of neural activity with specific aspects of sentence comprehension (syntax and semantics) and also with whether the rhythmic activity is endogenous or evoked response. By narrowing down the properties of that rhythmic activity, the way is paved for explicit accounts of how neural oscillations might carry out such functions.

Moreover, there are some critiques for the traditional ERP studies and examining from different approaches may be necessary for mapping the processing of linguistic representations and neural activity. As Poeppel and Embick (2005) articulate, studies examining ERP components associates these components with broad linguistic domains. For example, the N400 is associated with “semantic processing” and the P600 is usually associated with “syntactic processing”. However, the N400 cannot represent semantics, given that semantic processing indeed involves several sub-processes with other linguistic levels such as phonological and hierarchical processing and these sub-processes interact with each other. On the other hand, the core components of linguistic theories, such as the syntactic operation of Merge, can not be directly mapped to neurons. As Benítez-Burraco and Murphy (2019) point out, it may be feasible to decompose linguistic operations and map them to cross-frequency patterns, which denotes the association across multiple frequency bands of neural oscillations. This idea builds on a growing trend that holds that oscillat-

ing patterns of neuronal circuits to be a computational primitive (e.g. Buzsáki and Draguhn, 2004). Consequently, examining from neural oscillations offer a good way to see how linguistic processes reflect on brain signals concurrently when continuous speech unfolds.

Previous studies have shown that neural responses can be entrained to a rhythmic input (e.g. pure tones) or quasi-rhythmic input (e.g. speech). That is, neural activities can become synchronized to the rhythmicity of external input. Regarding speech, as illustrate in Figure 1.1, previous studies have reported a robust neural entrainment (i.e., the blue line in Figure 1.1) to speech envelope (i.e., the red line) in syllable processing (e.g. Luo and Poeppel, 2007; Giraud and Poeppel, 2012). However, how neural signals entrain to the structures above syllables (i.e., the green lines) is still vague since the connected speech does not have physical boundaries to indicate hierarchical information. In addition, how high-level information such as semantic and syntactic information reflects on neural entrainment and guide language comprehension is still unknown. The goal of the dissertation aims to examine how high-level information modulate low-frequency oscillations and determine the functional significance of low-frequency oscillations.

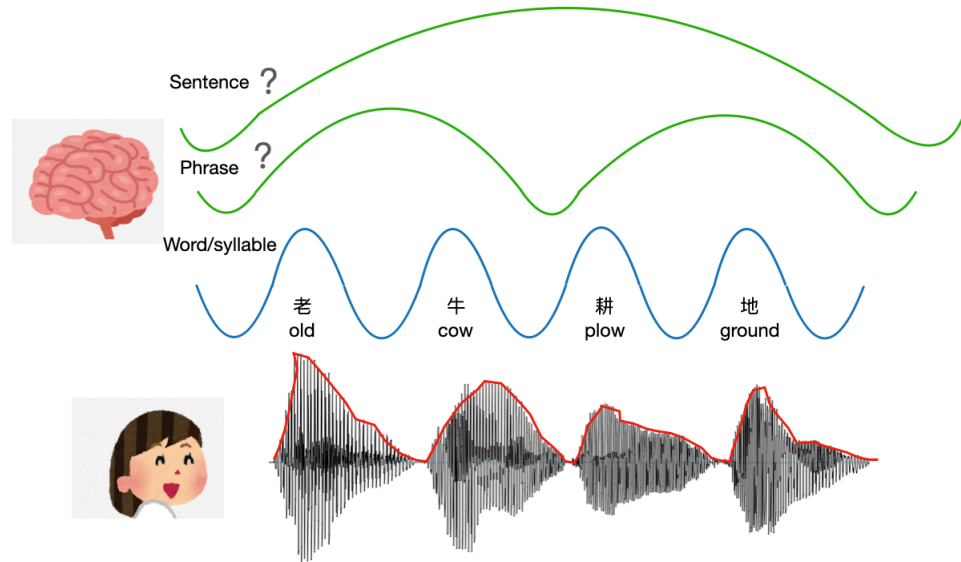


Figure 1.1: An illustration of neural entrainment to speech

## 1.2 Research questions

This dissertation aims to examine the functional significance of delta-band (0.5 - 4 Hz), which fits the usual timing of processes above the syllable-level across languages. Delta-band activities are low-frequency brainwaves with high amplitudes, which are usually observed during the stage of deep sleep. In terms of language processing, neural activities in the delta band have been associated with auditory processing or the process of chunking continuous speech into meaningful units. Previous studies also show that neural activities in the delta band can track abstract syntactic structure, yielding hierarchical oscillations in delta frequencies (Ding et al., 2016; Ding et al., 2017). However, whether those neural responses are modulated by syntactic information, and to what extent those neural activities are affected by semantic information, are both unclear. Indeed, what information guides these neural indices of the phrasal/sentence processing? Moreover, it is not well understood what cognitive operations are indexed by those low-frequency neural activities. The



following sections describe the three research questions of this dissertation.

### **1.2.1 Hierarchical representation vs. Lexical representation**

Speech signals are complex; their envelope and fine structure change quickly over time (Arnal et al., 2016). Previous studies have shown that the size of units correspond to certain neural frequencies. That is, bigger units are signaled at slower frequencies. To be specific, word and syllable boundaries are signaled at slow-wave frequencies (delta-theta) and can be perceived as discrete events while phonemes are signaled and perceived as discrete events at faster frequencies (alpha-beta). Furthermore, recent studies have associated neural activities in the delta band with the processing of different levels of hierarchical syntactic structures. In particular, Ding et al. (2016) shows that neural responses entrained to the speech at 4 Hz reflect the processing of syllables, 2 Hz reflects the processing of phrases, and 1 Hz reflects the processing of sentences when native speakers of Mandarin listened to continuous speech presented at 4 Hz for each syllable without prosodic cues and indication for phrasal/sentence boundaries. Crucially, English speakers without Mandarin linguistic knowledge only show entrainments at the syllable rate (4 Hz), but not at the phrase (2 Hz) and sentence rate (1 Hz). In addition, the result that the low-frequency neural activities entrained with hierarchical structures was replicated successfully in MEG experiments using English structures and was able to be reflected cross-linguistically.

In line with Ding et al. (2016), growing studies have shown that neural activities in the delta band reflect the processing of chunking continuous speech into syntactic phrases (e.g. Bonhage et al., 2017; Meyer et al., 2016). For example, Meyer et al. (2016) shows that the internal grouping of words in ambiguous sentences modulates phase in delta-band frequency, regardless of prosodic cues. Consistent with the

results in Ding et al. (2016), these studies have demonstrated that neural activities in the delta band reflect the process of chunking continuous speech into meaningful units. The use of isochronous speech is able to dissociate the effects of prosody and syntactic processing in the neural activities of the delta band.

However, whether these delta frequencies can be reliably interpreted as the entrainment of abstract hierarchical information is still under debate. The lexical representation approach, a model built by Frank and Yang (2018), provides different ways of interpreting the results from Ding et al. (2016). The model they built encodes word similarities and part of speech for each word and represents stimuli from Ding et al. (2016) as sequences of high-dimensional numerical vectors. No further syntactic information was encoded in their model. The results from their model prediction show similar power spectra to the results from Ding et al. (2016). Their simulation results suggest that the findings in Ding et al. (2016) may follow from the tracking of lexical or part-of-speech sequence information and are not necessarily interpreted as the cortical entrainment for hierarchical structures. For example, verbs occur at the frequency of 1 Hz while nouns occur at a rate of 2 Hz. The simulation from their model posits a question about what lexical properties are involved in modulating low-frequency neural oscillations. The first issue addressed in the dissertation is whether delta entrainment reflects sequence-based regularities or hierarchical structure.

### **1.2.2 Temporal dynamics of cortical tracking in delta oscillations**

The second goal of this dissertation aims to investigate the dynamics of cortical tracking in neural activities in the delta band. Different timescales carry different amount of information. Previous studies have shown that the rate for understandable speech is between 2-10 Hz (Köseme and van Wassenhove, 2017) and speech with too

fast or slow rate might cause the impairment of comprehension (Ghitza, 2014; Ghitza and Greenberg, 2009). Previous studies have developed several models to explain how syllables are perceived in speech (Ghitza, 2011; Giraud and Poeppel, 2012). For example, the template matching model called TEMPO (Ghitza, 2011; Ghitza and Greenberg, 2009), proposes that the cortical oscillations reflect the segmentation of speech occurs roughly at 150-300 ms, which falls in theta cycles. In TEMPO, phonetic features are matched within a beta cycle (about 50 ms) while syllables are matched in a theta cycle (about 200 ms). This is basically in line with the other “asymmetric sampling in time” model (Poeppel, 2003; Giraud and Poeppel, 2012), suggesting that the auditory system integrates multiple temporal windows with fast and slow rates concurrently. The fine-grained acoustic features are processed in a shorter time window (~30 ms) while the acoustic structures are processed in a longer time window (~200 ms).

However, few studies have addressed the issue about when a parser integrates high-level information such as syntactic and semantic information and thus delta entrainments are elicited. Ghitza (2017) has shown that TEMPO model can be generalized to explain the association of the delta and the word/phrasal processing by demonstrating behavioral results from two context-free digit recall tasks, as detailed in Chapter III. There is still a lack of neural evidence for how temporal dynamics interacting with high-level linguistic information is reflected in neural activities of the delta band. Furthermore, how top-down high-level information interacts with low-level acoustic information in the modulation of neural activities in the delta band remains elusive. In addition, whether the entrainments found in the previous literature are merely the results of evoke responses to external stimulus or the reflection of endogenous entrainments is still under debate (Zoefel et al., 2018). This dissertation

manipulates speech rate and regularity between syllables to determine the temporal boundary for integration of syntactic and semantic information.

### 1.2.3 Semantic compositionality

Whether neural activities in the delta band are sensitive to semantic information is still unknown. Previous studies have demonstrated that semantic information can modulate high-frequency oscillations (e.g. Bastiaansen et al., 2010; Lewis and Bastiaansen, 2015; Lewis et al., 2016; Lewis et al., 2017). For example, Hald et al. (2006) showed that gamma-band power ( $\approx 35 - 45$  Hz) increased in the semantically-coherent condition (e.g. The Dutch train is YELLOW and blue.), as compared to the semantically-incoherent condition (e.g. The Dutch train is SOUR and blue.), suggesting the role of gamma-band oscillations reflects the semantic unification of lexical items into sentence. Few studies have linked delta oscillations to the processing of semantics and the semantic information examined in those studies vary a lot. Additionally, low-frequency neural activities can be contributed by slow event-related responses such as the P300, N400, and P600 (Zhou et al., 2016). If low-frequency neural activities are contributed by slow event-related responses such as the N400, which is sensitive to semantic processing, we should find an association of what triggers both low-frequency neural activities and the N400, in terms of semantic processing. In addition, previous studies did not disassociate the potential factors that modulates the N400 such as predictability or plausibility to see which aspects of semantic processing can modulate neural activities in the delta band. Furthermore, based on Frank and Yang' s model, the pattern of delta observations reported by Ding et al. (2016) can reflect lexical information only. To some degree, semantic information plays a role in modulating neural activities in the delta band. This dissertation manipulates semantic and hierarchical content by varying predictability, plausibility, semantic

relatedness, and specificity to examine whether neural activities in the delta band can be driven by these semantic variables.

### **1.3 Methodology**

Three electroencephalogram (EEG) experiments were designed following the frequency-tagging paradigm in Ding et al. (2016). That is, all syllables in the stimuli were presented in a fixed speech rate and then power and phase coherence can be examined at the relevant frequencies of interest. All experiments are auditory experiments and stimuli are isochronous speech recorded from a computer program to remove prosodic cues. In addition, simulations from different computational models were conducted to demonstrate the predictions from Frank and Yang (2018).

### **1.4 Dissertation outline**

This dissertation aims to address the following issues: (1) whether the findings from Ding et al. (2016) reflects the lexical sequence information or hierarchical information? (2) how temporal properties such as speech rate and rhythmicity are reflected in low-frequency oscillations regarding sentence comprehension? (3) what high-level information modulates low-frequency neural responses? Three EEG experiments will be designed to address the above questions.

Chapter II aims to answer the above first issue by demonstrating the first EEG experiment that tests the lexical representation hypothesis from Frank and Yang (2018). Chapter III aims to provide empirical results by describing the second EEG experiment which manipulates speech rate and rhythmicity. It provides an evidence about whether low-frequency entrainments are evoked responses or endogenous oscillations that are modulated by predictive processing. Chapter IV addresses the third issue by describing the third EEG experiment which manipulated several seman-

tic variables to see what semantic information can modulate low-frequency neural activities. Chapter V concludes the current findings and provides future directions.

## CHAPTER II

# Lexical Representation Can' t Fully Capture the Functional Role of Low-Frequency Neural Activity During Language Comprehension

### 2.1 Introduction

Human language is compositional; we can create infinite and novel phrases or sentences from a finite number of words. Importantly, this compositional ability is highly structured; words must be combined according to syntactic rules to yield well-formed and interpretable phrases and sentences. Previous studies have narrowed down the location and timing of neural indexes for compositional processing (see Pylkkänen and Brennan (2019) for review). For example, Bemis and Pylkkänen (2011) examined how humans process two-word combinatorial phrases (e.g. red boat) vs. non-combinatorial phrases (e.g. xkq boat) in Magnetoencephalography (MEG) recording and found that an increased activity of combinatorial phrases was elicited at 200-250 ms after the presentation of the second word at the left anterior temporal lobe (LATL), compared to non-combinatorial phrases. Furthermore, Neufeld et al. (2016) found a greater negativity in the similar time-window (184-256 ms) for the combinatorial phrases compared to the non-word condition by using the same experimental paradigm in electroencephalography (EEG) recording. In addition, several recent fMRI studies have shown greater activation for sentences vs. word

lists in brain regions such as inferior frontal gyrus (Pallier et al., 2011; Schell et al., 2017; Zaccarella et al., 2017), posterior superior temporal sulcus (Zaccarella et al., 2017), anterior temporal lobe (Humphries et al., 2006; Matchin et al., 2017), angular gyrus (Humphries et al., 2006; Matchin et al., 2017), and temporal parietal junction (Matchin et al., 2017), further supporting the role of these regions for compositional processing.

Although many studies have provided neural evidence for *when* and *where* the neural indexes are for compositional processing, it is not yet clear “how” neural circuits actually implement compositional processes. Examining language comprehension from neural oscillations indeed provides a good way to see how these mechanisms are implemented since oscillatory activities have the property of a repetitive or rhythmic pattern that can be observed from a single neuron or a group of neurons in the brain or central nervous system (Buzsáki and Draguhn, 2004). Neural oscillations can be characterized by frequency, phase, and amplitude, and they may synchronize with extraneous stimulation such as periodic acoustic or visual stimuli. The process of synchronization is also called *cortical entrainment*. Neural oscillations have been linked to many cognitive functions such as memory, attention, and language (e.g. Jensen et al., 2007; Lakatos et al., 2008; Meyer, 2018).

For language specifically, different frequency bands have also been linked to different stages of language comprehension and speech processing. As reviewed by Meyer (2018), especially for language comprehension, delta-band oscillations (0.5-3 Hz) have been associated with syntactic chunking, theta-band oscillations (4-8 Hz) have been related to memory retrieval, alpha-band oscillations (9-13 Hz) can be linked to the storage of syntactic categories, and higher-frequency bands such beta (14-30 Hz) and gamma (>30 Hz) appear to correlate with semantic prediction. However,



previous studies focused more on the explanation of speech processing in examining the relationship between oscillatory activity and low-level acoustics and phonology. It remains unclear how high-level information (i.e., syntax and semantics) modulate oscillatory activities. Detailed modulation for each frequency band during language comprehension remains elusive and more fine-grained definitions for the functional interpretation of each frequency band are necessary. As discussed in more detail below, a recent study has shown that the low-frequency delta oscillations appear to be entrained with abstract hierarchical structure (Ding et al., 2016; 2017). Martin and Doumas (2017) further provides a computational model and supports the structural explanation for the findings from Ding et al.. However, the lexical representation, a model proposed by Frank and Yang (2018), suggests that the results from Ding et al. are not necessarily entrained by hierarchical information, but can be explained by lexical information or part-of-speech only. To get a clearer functional significance of neural activity in the delta band, the current study tests the above two hypotheses by having a novel experimental condition and simulating results from Frank and Yang’s model with different word representations acquired from three language models.

### **2.1.1 Cortical tracking of abstract linguistic structure**

Speech signals are complex; their acoustic envelope and fine structure change quickly over time (Arnal et al., 2016). Previous studies have shown that the size of units correspond to certain neural frequencies. That is, larger acoustic units show entrainment with neural signals at slower frequencies. Namely, word and syllable boundaries are signaled at slow-wave frequencies (delta-theta), which is linked with the perception of these acoustic regularities as discrete events (Ghitza, 2011; Ghitza and Greenberg, 2009). Phonemes are signaled and perceived as discrete events at faster frequencies (gamma) (Giraud and Poeppel, 2012; Di Liberto et al., 2015).

Ding and colleagues (2016) argue that such entrainment patterns extend to more abstract properties of linguistic input as well. They used an isochronous speech paradigm with sentence stimuli composed from four one-syllable words in Mandarin Chinese. Each word lasted just 250 ms, so each sentence was exactly 1 second long. With this design, syllables and words were repeated at 4 Hz, two-word phrases at 2 Hz, and sentences repeated at 1 Hz. When native speakers of Mandarin listened to these stimuli during MEG recording, they observed spectral peaks from the neuromagnetic signals at 1, 2, and 4 Hz. The follow-up work by Ding and colleagues (2017) confirmed such entrainment can also be observed using an EEG recording. Ding et al. (2017) further shows that these peaks were observed in the so-called evoked power (i.e., the power that is phase-synchronized with speech stimuli) and also inter-trial phase coherence (i.e., whether the differences between phase angles and speech stimuli across trials are consistent), but not induced power (i.e., the power that is not phase-aligned with speech stimuli). Crucially, English speakers without Mandarin linguistic knowledge only show entrainment at the syllable rate (4 Hz), but not at the phrase (2 Hz) and sentence rate (1 Hz). This result was replicated cross-linguistically: English stimuli presented in the same paradigm to English-speaking listeners also elicited entrainment patterns at sentence and phrasal rates.

However, stimuli with structure elicit entrainment at phrase and sentence level might be confounded by the following. First, previous studies have shown that delta entrainment can be associated with prosody (Ghitza, 2013, Ghitza et al., 2012). Ding et al. (2016) controlled this potential factor by using isochronous speech recorded by a computer program to avoid prosodic cues. However, it is still impossible to rule out the possibility that the parser would generate internal prosody when parsing

the isochronous speech (Breen, 2014). Second, the transitional probability between words might play a role when a sentence is being unfolded and the upcoming elements become more predictable. Ding et al. (2016) also addressed this factor by having the stimuli which have either equal or non-equal transitional probabilities between words. Both equal and non-equal conditions show neural entrainment with abstract hierarchical structures at its corresponding frequencies. Third, it is possible that the delta entrainment at the phrase and sentence level correlates to the regularity of part-of-speech. Stimuli in Ding et al. (2016) show the regular repetition. For example, nouns occur at 2 Hz while verbs occur at 1 Hz. The consequence of entrainment might reflect solely the part-of-speech repetition (Frank and Yang, 2018). The current study focuses on whether the delta entrainment at the phrase and sentence level reflects merely part-of-speech repetition.

As mentioned above, whether these peaks were indeed modulated by structural information remains under debate. Two computational approaches were proposed to interpret the functional significance of these peaks. The first is the structural account, proposed by Martin and Doumas (2017). Their model adopts a time-based binding mechanism which allows for representations to be bound by using asynchrony of firing units and thus having different levels of representations without losing the information for each representation. In the model, time is used to carry out information and the model assumes independent values for variables. The independent value for variables would allow a particular value for a variable represented in a specific timing. In addition, this independence can allow the association between variables computed based on statistics without changing any core representation. The model has different layers representing different levels of a hierarchical structure. Take an adjective phrase “dry fur” for example. The model encodes semantic features for

each word at the lowest layer, word information such as [dry *adj*] and [fur *noun*] is encoded in the second layer and they fire at an asynchrony. Then the layer above encodes phrase information and will be activated after [dry *adj*] and [fur *noun*] fire. They conducted a simulation by using the English stimuli from Ding et al. (2016). The simulation results show that the grammatical condition (e.g. dry fur rub skin) elicits spectral peaks at 1 Hz, 2 Hz, and 4 Hz, consistent with the experimental results from (Ding et al., 2016). They further show that the word list condition (i.e. no syntactic relationship between words) shows the 4-Hz peak only. Similar to the grammatical condition, the jabberwocky condition, which preserved syntactic relationship between words without plausible semantic composition, shows the power increases at 1 Hz, 2 Hz, and 4 Hz, suggesting that the hierarchical structures were activated.

In addition, Martin and Doumas tested a three-layer recurrent neural network (RNN), which does not encode time-based binding representations between words. The RNN model was trained to predict the next word in a sentence and presented one word at a time for every sentence in each condition. The simulation from the RNN model shows that the power increases at 4 Hz in all conditions and the phrase condition shows the increased power at 2 Hz. However, the grammatical and jabberwocky conditions failed to show the increased power at 1 Hz and 2 Hz, although the RNN model was able to predict the next word accurately. These results suggest that delta entrainment encodes hierarchical information, rather than reflects statistical sequence-based prediction only.

Consistent with the model results from Martin and Doumas (2017), Kaufeld et al. (2020) further support the view that neural activities in the delta band reflect the combinatorial processes and high-level linguistic content (i.e. meaning and syntax)

beyond prosody by using naturalistic stimuli during an EEG recording. Their results reveal that the mutual information between neural signals and speech stimuli is higher in the sentence condition, compared to the jabberwocky and word list conditions.

The second approach to address the functional role of low-frequency activities is the lexical representation, proposed by Frank and Yang (2018). Instead of reflecting hierarchical structure, they suggest, such entrainment could arise from regularities at the lexical level that arise from the pattern of words repeated across the stimuli. They test this hypothesis using a series of simulations with a model that encodes words as vectors in a high-dimensional semantic space (a “word embedding”). Thus, the stimuli from Ding et al. (2016) are cast as sequences of high-dimensional numerical vectors (Mikolov et al., 2013). Such vectors capture semantic information (e.g. words with similar meanings will have more similar vectors) and also linguistic regularities like grammatical category of each word, such that two nouns tend to have more similar vectors than a noun and a verb. No further syntactic information was encoded in their model. The simulation results from their model show similar power spectra to the results from Ding et al. (2016), suggesting that those neural entrainment patterns may follow from the tracking of lexical or grammatical category sequence information and are not necessarily interpreted as the cortical entrainment for hierarchical structures. That is, verbs occur at the frequency of 1 Hz while nouns occur at a rate of 2 Hz and neural activities in the delta band could be modulated by these regularities. The simulation for both English and Chinese grammatical sentences elicits increased power at 1 Hz, 2 Hz, and 4 Hz. The Chinese VP elicits increased power at 2 Hz and 4 Hz, but not 1 Hz. The shuffled Chinese syllables show increased power at 4 Hz only.

The above two accounts provide different interpretations of the functional signifi-

cance for the entrainment of delta frequencies. A summary for these two accounts is given in Table 2.1. The model from Martin and Doumas (2017) shows how hierarchical information could be reflected by time-based binding mechanism. However, their results fail to show the difference between the grammatical condition and the jabberwocky condition, suggesting that the increased power at corresponding frequencies reflects purely hierarchical information in the model, but not semantic information. On the other hand, Frank and Yang's model suggests that the increased power at corresponding frequencies does not necessarily represent the processing of hierarchical information. Their model suggests that the increased power at corresponding frequencies might reflect lexical semantics or the part-of-speech information, without considering syntactic information. These two accounts both provide similar predictions for the grammatical sentences, phrases, and word list stimuli from (Ding et al., 2016, 2017). Thus, it is still unclear what high-level information drive neural activities in the delta band. In order to tease these two theories apart, the current experiment uses reversed phrases, where the semantic information and the frequency of part-of-speech are still preserved but the sequences are reversed, yielding an ungrammatical structure (see details in Materials section). The account offered by Frank and Yang (2018) predicts that such stimuli will show 1 and 2 Hz peaks since the frequencies of part-of-speech are still preserved. On the contrary, the account offered by Martin and Doumas (2017) predicts such stimuli will not show 1 and 2 Hz peaks since the structure is not allowed in Mandarin by reversing the phrase.

Table 2.1: Summary of two accounts and predictions for reversed phrases

Accounts	Major study	Critical simulation results	Predictions for reversed phrases
<b>Structural account</b>	Martin & Doumas: time-based encoding representations	Sentence: 1, 2, 4 Hz Phrase: 2, 4 Hz Word list: 4 Hz Jabberwocky: 1, 2, 4 Hz	<b>4 Hz</b>
<b>Lexical representation</b>	Frank & Yang: Lexical semantics and POS	Grammatical: 1, 2, 4 Hz Phrase: 2, 4 Hz Word list: 4 Hz	<b>1, 2, 4 Hz</b>

### 2.1.2 Supporting evidence for the delta oscillations in language comprehension

How neural activities in the delta band relate to the syntactic processing is further supported by multiple recent works (e.g. Bonhage et al., 2017, Meyer et al., 2016). Neural activities in the delta band have been associated with the processing of chunking continuous speech into phrases and sentences. Bonhage and colleagues (2017) conducted an EEG recording to examine whether a chunking-related modulation occurs during processing sentences, as compared to random word lists. During the experiment, participants either saw the grammatical German six-word sentence fragments (e.g. *Sie wurden gestern Abend an der* “they were yesterday evening at the”) or ungrammatical word list (e.g. *sie der an Abend gestern wurden* “they the at evening yesterday were”). Participants were asked to either memorize the stimuli for 6 seconds and recall whether a specific word is followed by another word in the trial, or decide whether a word is a real word or not. The results show that delta power increases during encoding sentence fragments, compared to random word lists, regardless of which task (i.e. working memory task or lexical decision task) is asked after the encoding stage. Their results suggest that the increased delta power reflects automatic linguistic chunking process and this process is effortless, automatic, and highly effective retention for sentences, but not for random word lists. In addition to increased power in delta-band oscillations, Meyer and colleagues (2016) show that

grouping bias of words in ambiguous sentences, regardless of prosodic cues, modulates the phase in the delta-band. In other words, parsers follow their internal grouping bias and chunk the words into meaningful phrases.

In addition to the processing of chunking, Meyer and Gumbert (2018) further show evidence for the role of delta-band oscillations with the processing of syntactic structure. Their results show that delta-band oscillations are phase-locked to the syntactic structure of speech stimuli and a correlation between the syntactic surprisal value and the phase was shown in the study, suggesting that the syntactic knowledge for the incoming word facilitates language processing and thus cause the alignment of neural excitability and syntactic information. However, as they noted, their results cannot fully rule out the possibility that the phase alignment in delta oscillations might merely reflect a transient prosodic/acoustic effect that is disguised in oscillations. That is, the synchronization they found might merely reflect the rhythmicity of the stimuli.

The above studies associated with the delta entrainment in syntactic processing are indeed in line with Ding et al. (2016), although they adopted very different methodologies to examine oscillations in delta bands. Specifically, consistent with Ding et al. (2016), Meyer et al. (2016) show that the delta oscillations are modulated by internal structure, but not prosody. However, as Meyer and Gumbert (2018) noted, the modulation of entrainment might result from the rhythmicity in stimuli. Previous studies have proposed that delta oscillations may correlate to two functions: they are driven by either acoustic features or syntactic information (Ghitza, 2017; Meyer et al., 2016). Under the acoustic account, delta is responsible for chunking speech into discrete units based on acoustic cues. Under the syntactic account, delta has the top-down process for applying internal syntactic knowledge to the speech



sequences. Although these two processes are highly correlated with each other, as shown in Meyer et al., 2016, these two processes may be disentangled since participants in that study show internal grouping bias and ignore prosodic boundaries. In addition, while previous work argue that delta-band oscillations are domain-general modulated function and are not necessarily interpreted as the modulation by syntactic processing (see Meyer, 2018 for discussion), the current study focuses on providing neural evidence to see to what extent neural activities in the delta band are driven by higher level syntactic information. Specifically, the study will further answer what kind of linguistic information (syntax or lexical sequences) would modulate neural activities in the delta band.

## 2.2 Experiment

The experiment aims to test whether neural activities in the delta band reflect lexical sequence or hierarchical information. If neural activities in the delta band can be merely modulated by lexical information, specifically, verb occurring at a frequency of 1 Hz and noun occurring at a frequency of 2 Hz, we would expect that the entrainment still occur at its corresponding frequencies even when the part-of-speech information is reversed (Frank and Yang, 2018). However, if neural activities in the delta band are not modulated by the lexical information only, increased power should not be observed in the reversed version of grammatical sentences since the hierarchical information is distorted and the stimuli are ungrammatical. This experiment also provides an opportunity to replicate results from Ding et al. (2016) using Mandarin stimuli with EEG techniques. It should be noted that the signals measured in the current study reflect low-frequency neural activities. One hypothesis is that the measurements reflect neural oscillations. However, the current experiment does not

allow us to conclude that the results can be directly mapped to neural oscillations.

### 2.2.1 Participants

Thirty-seven native speakers (22 females, 15 males) of Mandarin Chinese between the age of 19 and 52 (mean = 27.7) participated in the experiment. They were all right-handed and had normal hearing. They self-reported that they did not have any neurological disorders. They gave informed consent and were reimbursed for their time (15 USD/hour). Data from six participants were excluded from the analysis due to having many noisy trials (>40 %). Thus, data from 31 participants (18 female, 13 males) were included in the final analysis.

### 2.2.2 Materials

All experimental items are four-syllable Chinese sequences. Four experimental conditions are included in this experiment, as shown in Table 2.2. The first one is *Four-syllable sentence* (ABCD). We used 50 items adapted from Ding et al. (2016), with some modifications. The first two syllables constituted an NP, which is either an Adjective + Noun NP (e.g. *lao + niu* ‘old + cattle’ ) or a Noun + Noun NP (e.g. *shu + mu* ‘tree + wood (tree)’ ). The last two syllables constituted a VP (e.g. *chi + cao* ‘eat + grass’ or disyllabic verb *jiang jie* ‘to explain’). Six items from Ding et al. (2016)’s study were replaced or modified due to the following two reasons: (1) Stimuli that do not sound natural for native speakers from either Taiwan or mainland China were replaced since participants we recruited were either from mainland China or Taiwan; (2) Stimuli including disyllabic morphemes such as *heshang* ‘monk’ and *hudie* ‘butterfly’ were avoided since these NPs cannot be broken down further into Adjective+Noun or Noun+Noun.

Table 2.2: Stimuli design.

<b>Four-syllable sentence (ABCD)</b> 綿 羊 吃 草 mian yang chi cao Cotton sheep eat grass ‘Sheep eat grass.’	<b>Semantically-mismatched sequence</b> –give the structure as in the others (ABCD) 軍 孩 奔 草 jun hai ben cao Soldier child run grass
<b>Two-syllable phrase (ABAB)</b> 老 牛 青 草 lau niu qing cao Old cattle green grass	<b>Reversed phrase (BADC)</b> 羊 棉 草 吃 yang mian cao chi Sheep cotton grass eat

The second condition is *Semantically-mismatched sequence*. Following Ding et al. (2016)’s design, we randomly replaced each of the four syllables in the Four-syllable sentence condition independently with a new syllable from other sentences in the same condition while keeping the original positions of the syllables unchanged. After we had the sequences, we checked over all the items to make sure that these sequences do not sound meaningful or even familiar to native speakers. This is important as there are many syllables in Chinese that are completely different in meaning but share the same sounds.

The third condition is *Two-syllable phrase (ABAB)*. Fifty items were constructed by extracting the first two syllables from the items in the first condition, combining them two by two into NPNP sequences.

The fourth condition is *Reversed phrase (BADC)*. Crucially, this is the condition to examine whether neural activities in the delta band are entrained by merely lexical information. Same as condition 1, this condition also includes regular lexical sequences (i.e., noun at 2 Hz and verb at 1 Hz) but with reversed order, yielding an ungrammatical structure in Mandarin. Neither phrasal nor sentential neural encoding is expected in this condition if neural activities in the delta band reflect hierarchical processing.

All stimuli were recorded with text-to-speech from the website developed by iFLY-

Tek ([https://www.xfyun.cn/services/online\\_tts](https://www.xfyun.cn/services/online_tts)). Each syllable was recorded separately and the averaged duration of each recorded syllable is around 300 ms. Then each syllable was compressed to 240 ms with pitch preserved and a 10-ms silence gap was added after each syllable by using Praat (Boersma and Weenink, 2019). Thus, the whole duration of each syllable is 250 ms. Therefore, a four-syllable item was 1-second long and a trial that included ten four-syllable items was 10-second long. The power spectrum of the speech stimuli is shown in Figure 2.1. The power spectrum was computed using a fast Fourier transform based on the broadband envelope of the stimulus defined by the absolute value of the Hilbert transformation of the stimuli. The power spectrum was averaged over all 10-second trials for each condition. Only the syllable-level peak was observed in every condition.

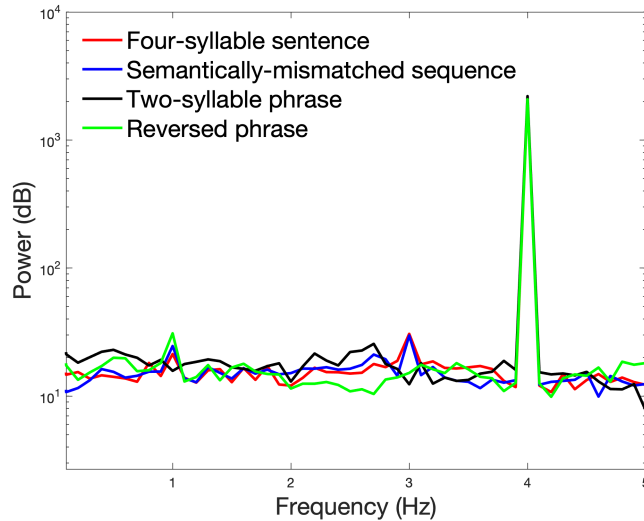


Figure 2.1: Power spectrum of speech stimuli

Eight blocks were included in this experiment. Each block included 20 plausible and 20 implausible trials. Thus, 320 trials were used in the whole experiment. The implausible trials serve not only to ensure the attentiveness of participants (as in Ding et al., 2016), but also to address critical research questions in the present study. A

plausible trial drew items from either Four-syllable sentence or Two-syllable phrase, while an implausible trial drew from either Semantically-mismatched sequence or Reversed phrase. A block either pairs items from Four-syllable sentence with those from Semantically-mismatched sequence, or items from Two-syllable phrase with those from Reversed phrase. The plausible trials and the implausible trials were intermixed and presented randomly in each block.

### **2.2.3 Procedure**

Participants sat comfortably in front of a computer screen in a quiet room. Prior to the main session, participants were fitted with an electrode cap. Electrodes were also affixed above and below the left eye. Electrolyte gel was applied to minimize impedance. The setup took approximately 30 minutes. Then the stimulus volume was set for each ear based on hearing level of each participant with the threshold of 45 dB determined using 1000 Hz tones (300 ms length). After setting up the volume, a pre-test presenting 120 1,000 Hz tones was conducted to ensure the data quality was sufficient to run the main session.

During the main session, participants were instructed to judge whether a trial includes plausible sentences/phrases or not by button press. After the button press, the next trial was played after a delay randomized between 800 - 1,400 ms (Ding et al., 2016). Stimuli were presented with Psychopy2 (v1.84.2; Peirce, 2007, 2009). Participants were also instructed not to frequently blink their eyes and avoid unnecessary body position adjustment when the stimuli were presented. Participants can take break between blocks. After the instructions, participants had 4 practice trials to familiarize with the whole procedure of the experiment. The order of blocks and the button press were counterbalanced across participants. The main experiment took about 1.5 hours to finish. After the main session, participants washed

their hair to remove the electrolyte gel and were debriefed about the goals of the experiment.

#### 2.2.4 EEG recording and data analysis

EEG data were recorded at 500 Hz from 61 active electrodes (actiCHamp, Brain-Products GMBH) in a 0.01-200 Hz band with online reference to an electrode placed on the left mastoid. Impedance were kept below 25 kOhms. FieldTrip software (Oostenveld et al., 2011) was used to analyze the data. Artifacts related to eye blinks were removed via Independent Component Analysis (Jung et al., 2000; Makeig et al., 1995) and remaining artifacts were removed manually following visual inspection. Following Ding et al. (2017), the first sentence from each trial was excluded to avoid potential EEG responses to sound onset. Data were filtered from 0.1-25 Hz, and re-referenced offline to a common average. For each condition, we computed evoked power (EP), induced power (IP), and inter-trial phase coherence (ITPC) from 0.5 to 10 Hz in increments of 0.111 Hz (i.e., 1/9-second trial) when removing the first sentence in a trial due to a possible transient effect (Ding et al., 2017). Evoked power, induced power and inter-trial phase coherence were computed based on the algorithms given in Ding et al. (2017), as listed and defined in (Equation 2.1-2.3). While Ding et al. (2016) demonstrated the total power recording from MEG, the current study follows the analysis from Ding et al. (2017). The total power can be divided into the evoked power and the induced power, which shows the clear picture of how stimuli are synchronized to the brain responses.

(1) **Evoked power** (EP) reflects the power of EEG responses that are phase-locked and time-locked to speech stimuli. The Discrete Fourier Transform (DFT) of the response in trial  $k$  is denoted as  $X_k(f)$  and  $X_k(f)$  is a complex-value Fourier coefficient. Thus, evoked power is the summation of complex-value Fourier coefficient

of trials averaged over the total number of trials  $K$ .

$$E(f) = \frac{1}{k} \left| \sum_k X_k(f) \right|^2 \quad (2.1)$$

(2) **Inter-trial phase coherence** (ITPC) reflects phase coherence across trials (Cohen, 2014). The summation of cosine and sine values of phase angle  $\theta_k$  of each complex-value Fourier coefficient is computed and then the square root of the summation is averaged over the total number of trials  $K^1$ . Higher ITPC indicates that phase angles are consistent across trials.

$$R(f) = \frac{\sqrt{(\sum_k (\cos\theta_k))^2 + (\sum_k (\sin\theta_k))^2}}{K} \quad (2.2)$$

(3) **Induced power** (IP) reflects the power of EEG responses that are time-locked but not phase-locked to speech stimuli. Induced power is computed from the difference between the complex-value Fourier coefficient and the mean over trials (denoted  $\langle X(f) \rangle$ ) from each trial  $k$ . Then the summation of difference from each trial is averaged over the total number of trials  $K$ .

$$I(f) = \frac{\sum_k |X_k(f) - \langle X(f) \rangle|^2}{K} \quad (2.3)$$

For statistical analysis, conditions were compared via one-way ANOVA for each measure at each frequency of interest. For evoked power, in order to remove the trend of  $1/f$  noise in power spectra, evoked power at each frequency was normalized by the neighboring frequency bins within  $\pm 0.5$  Hz via Equation 2.4 used in Ding et al. (2017).

---

<sup>1</sup>The original formula in Ding et al. (2017) did not take the square root.

#### (4) Normalized evoked power (EPn)

$$En(f) = \frac{E(f)}{\sum_w E(w)}, |w - f| < 0.5 \text{ Hz}, w \neq f \quad (2.4)$$

$w$  represents the neighboring frequency around the target frequency  $f$ .

#### 2.2.5 Predictions

As summarized in Table 2.3, for Four-syllable sentence, we would expect entrainment associated with syllable (4 Hz), phrase (2 Hz), and sentence (1 Hz) level should be observed. For the two-syllable phrase, we expect peaks at the syllable (4 Hz) and phrase level (2 Hz). As for the semantically-mismatch sequence, only the syllable-level peak is expected since the whole sequence is implausible. Crucially, for the reversed phrase, if neural activities in the delta band are modulated by lexical information only, we would expect the same results as the four-syllable sentence since the lexical items and the frequency of each syntactic category are exactly the same. However, if delta entrainment represents hierarchical information, we should expect that the reversed phrase could elicit the syllable-level peak only since the structural information is lost.

Table 2.3: Predictions of each condition

Conditions	Predictions
Sentence	1, 2, 4 Hz
Mismatched	4 Hz
Phrase	2, 4 Hz
Reversed	Structural: 4 Hz Lexical: 1, 2, 4 Hz

#### 2.2.6 Simulations

The current study also includes simulations to examine whether different word embeddings trained from different architectures would result in entrainment within the delta range. Since Frank and Yang (2018) have shown simulated results for



phrases and word list, here we simulated the critical comparison between the four-syllable sentence and the reversed phrase. Following Frank and Yang (2018), word vectors were extracted for conditions Four-syllable sentence and Reversed phrase. For the reversed phrases, we adopted Frank and Yang’s word vectors to create the word embeddings by swapping columns, without changing other parameters. The Matlab function for a Discrete Fourier Transform used by Frank and Yang (2018) was used here. Twelve simulated subjects and fifty randomized items modified from Ding et al. (2016) were tested in the model. The timing for each word is 250 ms. Given the assumption that word information would appear later after word onset, Word vectors were mapped to time between 0 and 250 ms by randomly sampling  $t$  from a uniform distribution (see Frank and Yang (2018) for details). Word vectors were then summed and a Discrete Fourier Transform was applied to obtain power spectrum for each simulated subject.

Second, in order to know whether word representations trained from different language models result in consistent outcomes above, different word embeddings from Wikipedia2vec (Yamada et al., 2020) and a Chinese pre-trained Bidirectional Encoder Representations from Transformers (BERT) (Cui et al., 2019) were encoded separately. Different from the word embeddings in gensim or fastText, which are trained by the conventional word-based skip-gram model, Wikipedia2vec was trained from the word-based skip-gram model, as well as the anchor context model and the link graph model, which learn embeddings by predicting the neighboring context from the given words and the link graphs on Wikipedia. Wikipedia2vec provides a better performance on the datasets of word analogy and word similarity. The second word embeddings were extracted from a pre-trained Chinese BERT with whole word masking (Cui et al., 2019). Different from the models from Frank and Yang (2018)

and Yamada et al. (2020), BERT is trained with an unsupervised learning and bidirectional approach, which means that the word vectors for the same word would be different, depending on the context. The Chinese BERT with whole word masking, different from the BERT developed by Google, takes the Chinese word segmentation into consideration before training. Thus, the model is trained from masking the whole word, instead of the fragment of a Chinese word. This model did have a good performance on various tasks, across the sentence-level and the document-level. The current study compares word vectors extracted from different models to see whether these models could elicit similar results in Frank and Yang (2018) for the four-syllable sentences and simulate how word vectors from these models modulate the delta oscillations for reversed phrases. These simulations serve as an important control to demonstrate whether lexical properties within the stimuli, when words are presented at the fixed rate as here, oscillate at 1 or 2 Hz.

## **2.3 Results**

### **2.3.1 Results of the simulations**

Figure 2.2 shows the simulated power spectra from 0-10 Hz. For Frank and Yang (2018)'s word2vec model, model results show that lexical properties of the stimuli oscillate at 1, 2, and 4 Hz in the reversed phrases. In addition, the results from the Wikipedia2vec and the Chinese BERT also show that lexical properties of the stimuli oscillate at 1, 2 and 4 Hz in both conditions.

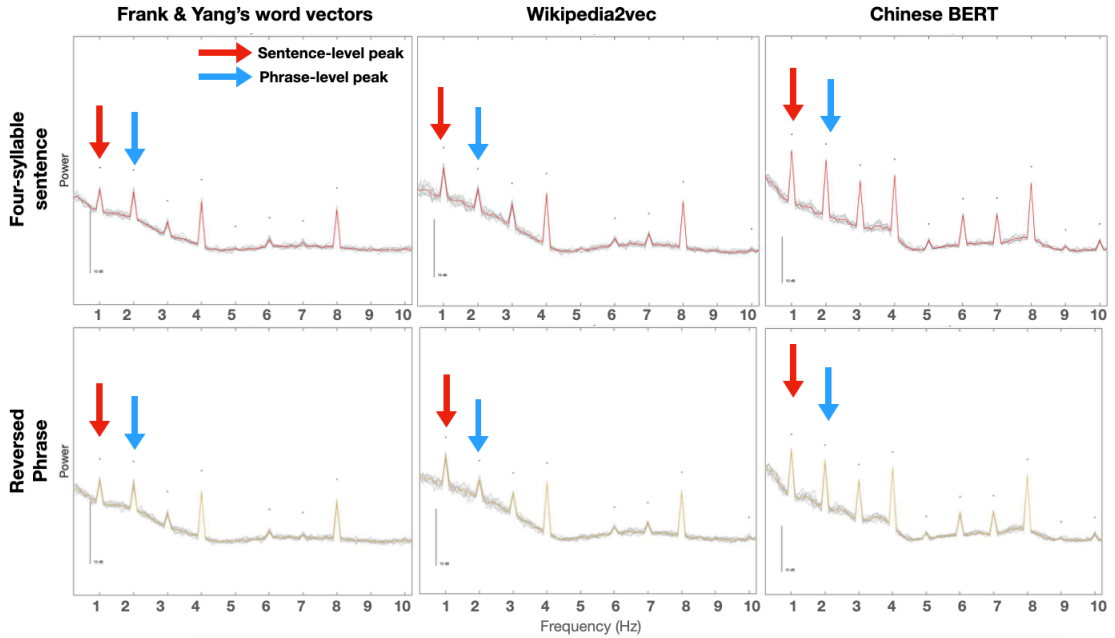


Figure 2.2: Simulated power spectra for three models

### 2.3.2 Results of the experiment

The results of evoked power are shown in Figure 2.3. The 4-Hz peak, which indicates the processing of syllables, can be found in all condition. The 2-Hz peak, which indicates the processing of phrasal level, is shown in the four-syllable sentence and two-syllable phrases. The 1-Hz peak, which indicates the processing of sentence level, is found only in the four-syllable sentences. Figure 2.4 shows the results of inter-trial phase coherence, which are consistent with the results of evoked power. Peaks at 4 Hz were found in all conditions. Peaks at 2 Hz were found in the four-syllable sentence and two-syllable phrase. Peak at 1 Hz was shown only in the four-syllable sentence. Figure 2.5 shows the results of induced power. Same as the results in Ding et al. (2017), no peaks were found in any conditions for this measure.

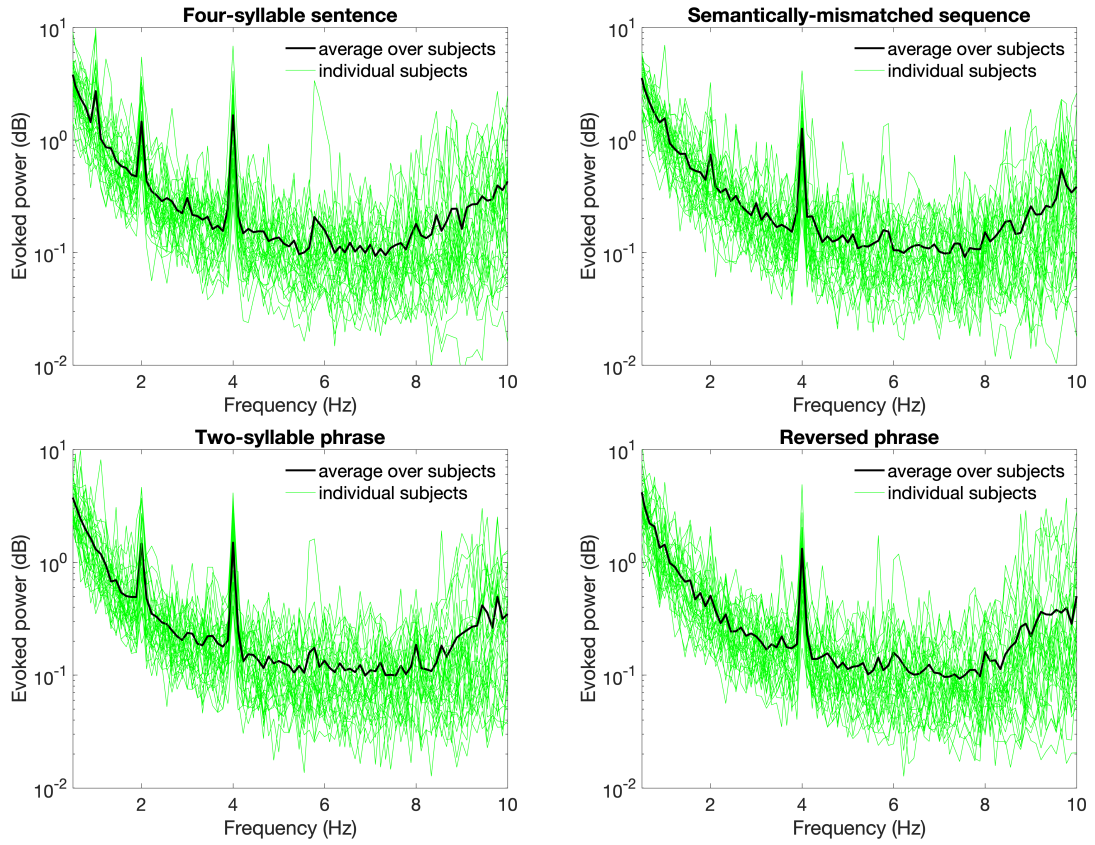


Figure 2.3: Evoked power for each condition

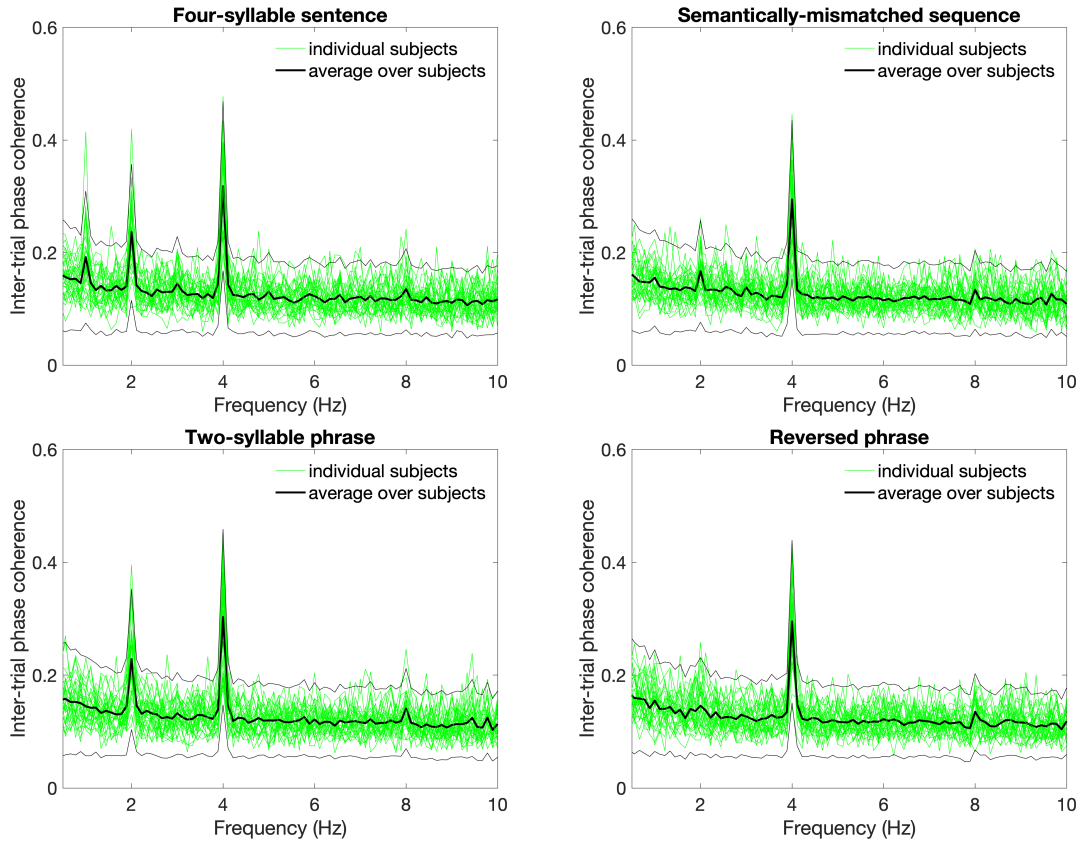


Figure 2.4: Inter-trial phase coherence for each condition

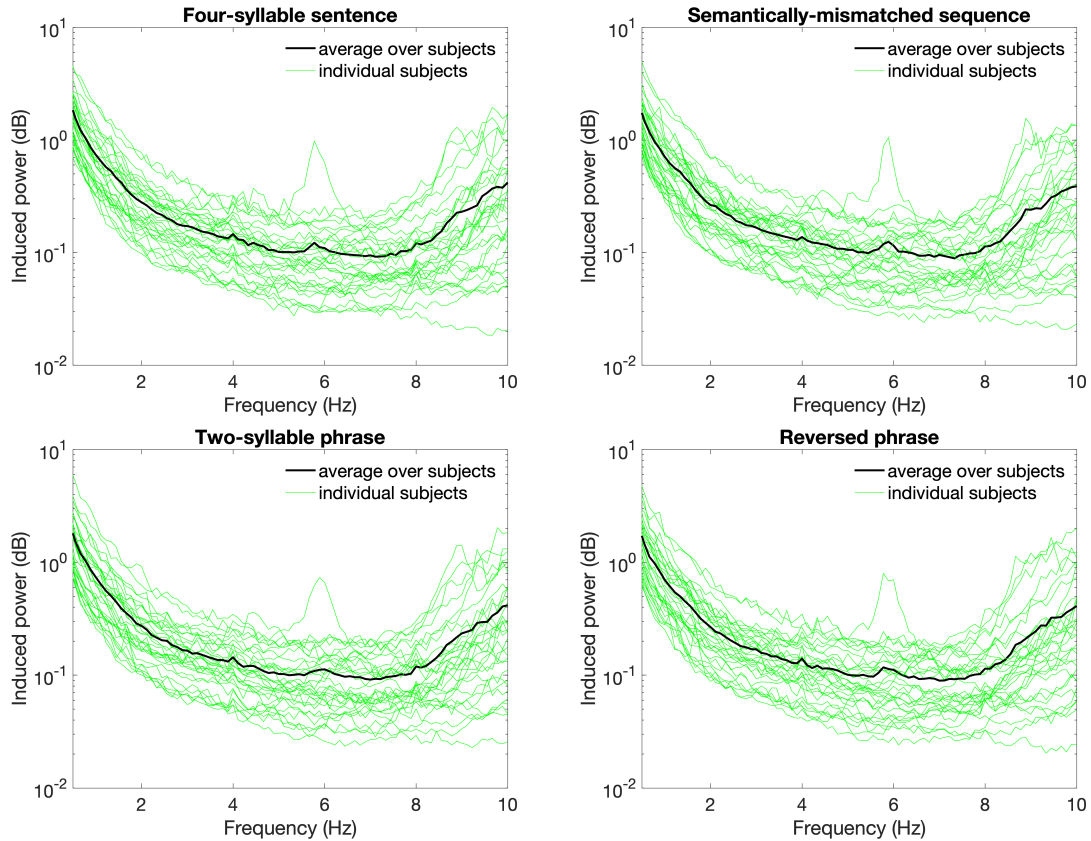


Figure 2.5: Induced power for each condition

Statistical comparison for each frequency of interest is shown in Figure 2.6. For the normalized evoked power, statistical significance was found at 1 Hz ( $p < 0.001$ ). Post-hoc pairwise Tukey's tests for the 1-Hz peak show significant difference in the comparison of the four-syllable sentence condition and the other three conditions ( $p < 0.001$ ) and there was no statistically significant difference for the comparison between the mismatched and the phrases ( $p = 0.67$ ), the comparison between the mismatched and the reversed ( $p = 0.99$ ), and the comparison between the phrases and the reversed ( $p = 0.6$ ). A statistically reliable difference was also found for the 2-Hz peak ( $p < 0.001$ ). Post-hoc pairwise Tukey's tests for the 2-Hz peak show significant difference in the comparison between the phrases and the reversed ( $p < 0.001$ ), the comparison between the phrases and mismatched ( $p < 0.001$ ), the

comparison between the sentence and the reversed ( $p < 0.001$ ), and the comparison between the sentence and the mismatched ( $p < 0.001$ ). No significant effect was found in the comparison between the sentence and the phrases ( $p = 0.98$ ) and the comparison between the mismatched and the reversed ( $p = 0.65$ ). No statistical significance was found at 4 Hz ( $p = 0.45$ ).

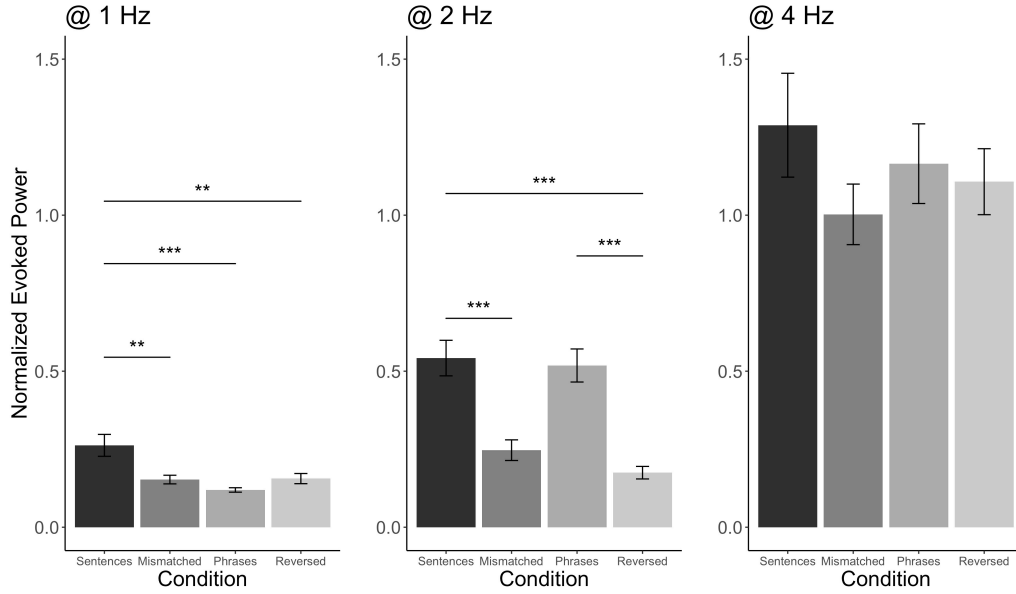


Figure 2.6: EPn at each frequency of interest

As for inter-trial phase coherence (see Figure 2.7), consistent with results of evoked power, a statistically reliable difference was also found at 1 Hz ( $p < 0.001$ ). Post-hoc pairwise Tukey's tests for the 1-Hz peak show significance in the comparison between the sentence and the other three conditions ( $p < 0.001$ ) and no significance in the comparison between the mismatched and the phrases ( $p = 0.8$ ), the mismatched and the reversed ( $p = 0.99$ ), and the phrases and the reversed ( $p = 0.85$ ). Statistical significance was found at 2 Hz ( $p < 0.001$ ). Post-hoc pairwise Tukey's tests for the 2-Hz peak show significant difference in the comparison between the phrases and the reversed ( $p < 0.001$ ), the comparison between the phrases and mismatched ( $p < 0.001$ ), the comparison between the sentence and the reversed ( $p < 0.001$ ), and the

comparison between the sentence and the mismatched ( $p < 0.001$ ). No significance was found in the comparison between the sentence and the phrases ( $p = 0.96$ ) and the comparison between the mismatched and the reversed ( $p = 0.45$ ). No significance was found in 4-Hz ( $p = 0.6$ ).

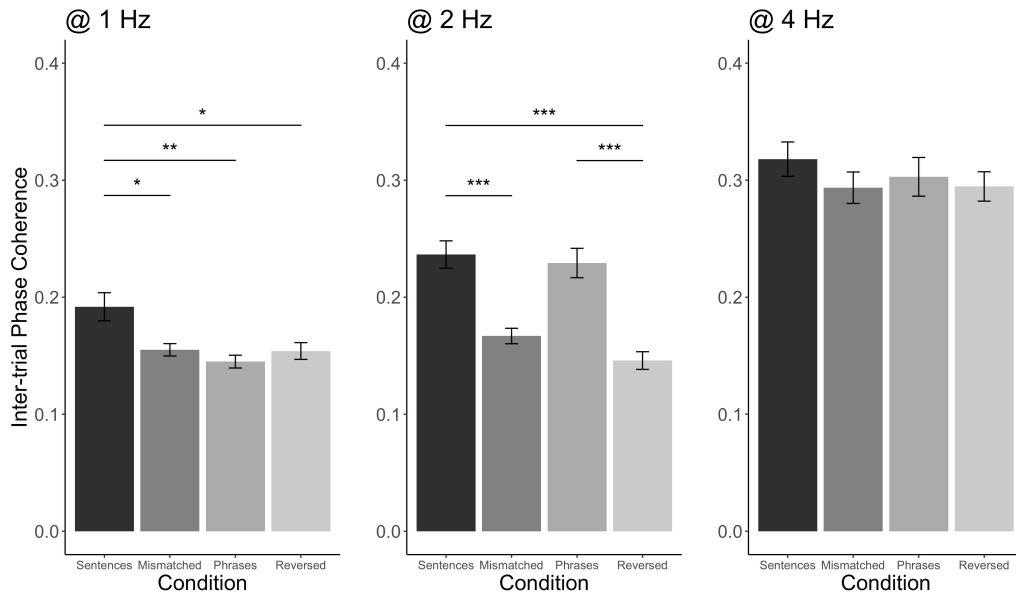


Figure 2.7: ITPC at each frequency of interest

For induced power (Figure 2.8), no significance was found in any of the target frequencies (1-Hz:  $p = 0.97$ , 2-Hz:  $p = 0.98$ , 4-Hz:  $p = 0.97$ ).



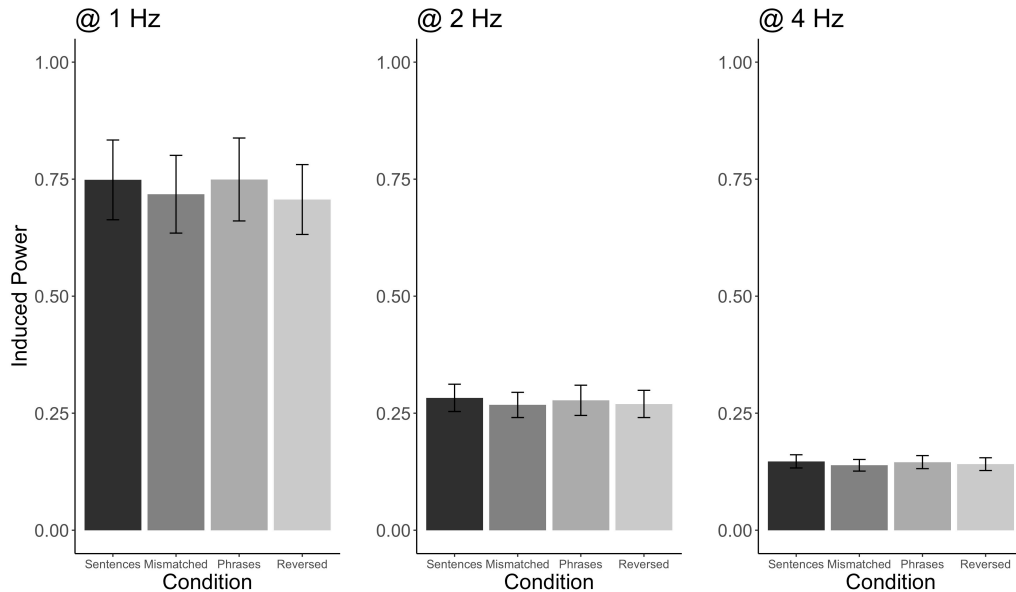


Figure 2.8: IP at each frequency of interest

## 2.4 Discussion

Neural activities in the delta band has been recently associated with the processing of abstract linguistic structure (Ding et al., 2016). However, previous research has not made clear what linguistic information modulate delta oscillations. Two accounts were proposed for the functional interpretation of this entrainment. The first one suggests that neural activities in the delta band might merely reflect word information or part-of-speech information of each word (Frank and Yang, 2018). The other structural account argues that this activity results from the syntactic structure encoding in time (Martin and Dumas, 2017), suggesting that the entrainment of delta frequencies is modulated greatly by hierarchical information. In order to tease apart these two theories, we use reversed phrases, which exhibit the same regular pattern of part-of-speech as the normal four-syllable sentences but result in ungrammatical sentences. In addition, the semantic information for each word and sentence is still preserved in this condition. If neural activities in the delta band merely re-

flect lexical sequence information, the reversed phrases should elicit peaks at 1, 2, and 4 Hz, same as the normal four-syllable sentences. However, if neural activities in the delta band reflect structural information, the reversed phrases should elicit syllable-level peak at 4 Hz only. Simulated power spectra by extracting word embeddings from three different models of lexical semantics were also provided. Word vectors from three models indicate that the lexical information could modulate 1-Hz and 2-Hz peaks. Inconsistent with the predictions from the three lexical models, our experiment results show that the reversed phrases elicit peaks at 4 Hz only, suggesting that neural activities in the delta band reflect the processing of hierarchical information, but not lexical information or part-of-speech information only.

Our results replicate the findings from different measurements shown in Ding et al. (2017). In Ding et al. (2016), they reported the total power, which computes the combination of evoked power and induced power. Ding et al. (2017) further separate total power into evoked power and induced power. In addition, the inter-trial phase coherence was also computed to see whether the phase angles across trials are consistent. In line with the finding of English in Ding et al. (2017), we do find consistent results for the evoked power and the inter-trial phase coherence. In addition, no significant peaks were found in the induced power.

The current study also successfully replicates several key results from Ding et al. (2016) by using EEG recording. The four-syllable sentences elicit peaks at 1, 2, and 4 Hz. The two-syllable phrases elicit peaks at 2 and 4 Hz, but not 1 Hz. Some more nuance trends are also observed in evoked power and inter-trial phase coherence. The peak value of the semantic-mismatched sequences and the reversed phrases are higher than the value of the two-syllable phrases condition at 1 Hz, although the difference is not statistically significant. On the contrary, the peak

value of semantically-mismatched sequences and the reversed phrases are lower than the value of the two-syllable phrases condition and the four-syllable condition at 2 Hz. The above observations might further suggest that delta entrainment at 1 Hz serve as an index of chunking by structure. In other words, the syntactic information at some degree are preserved in the semantically-mismatched condition and the reversed phrases. 1-Hz is an optimal timing for a parser to achieve the composition processes by integrating structural and semantic information. Thus, the value of those two conditions are higher than the value of the phrases condition at 1 Hz.

The current study also suggests that delta entrainment to speech may reflect endogenous neural responses but not merely the reflection of repetitive stimuli. Previous studies have argued that neural oscillations may reflect the repetitive pattern of external input. Whether neural oscillations reflect evoked responses or endogenous oscillatory activities remains under debate (Martorell et al., 2020; Zoefel et al., 2018). In our study, while the normal four-syllable sentences and the reversed phrases includes the same words, 1, 2 and 4 Hz peaks were observed in the normal four-syllable condition but not the reversed phrases, suggesting that neural oscillations do not just reflect the repetitive rhythms of external input, but involves the mechanism of using hierarchical knowledge.

One concern in the current study would be how the results from this unnatural experimental paradigm can map into a more naturalistic context and how we relate the current results to other relevant findings in neural activities in the delta band. This indeed requires future research to apply new methodologies or analysis to further attest the current results against other studies that adopts different methodologies (see Kaufeld et al., 2020). In addition, as Martorell et al. (2020) notes, it is unclear whether these findings could be elicited in different groups such as patient with

aphasia. Also, would similar peaks be elicited by using a different modality of stimuli presentation (i.e. visual vs. listening)? Another interesting future direction is to examine these low-frequency neural activities in language development (Getz et al., 2018). While it is much more clear how these low-frequency neural activities relate to language comprehension in adults, it is less clear when and how these patterns of oscillation are revealed in children (Maguire and Abel, 2013). If the peaks on delta could be an index of compositional processes, adopting the same experimental paradigm on children might provide insights on how these neural activities vary from children to adults.

## **2.5 Summary of the study**

The current study investigates whether neural activities in the delta band represent the processing of sequence-based lexical items only or that of hierarchical structure also. First, our results replicate results from Ding et al. (2016) by using EEG recording, confirming that cortical tracking of abstract hierarchical information can be detected robustly across different languages and brain-imaging techniques. Second, consistent with the English results from Ding et al. (2017), normal four-syllable sentences show sentence-level and phrasal-level peaks and two-syllable phrases show phrase-level peaks in evoked power and inter-trial phase coherence in Mandarin. Furthermore, our results of Reversed phrases are inconsistent with the lexical representation hypothesis. Only peaks at 4 Hz were elicited in the Reversed phrase condition, suggesting that delta oscillations are not modulated by part-of-speech or word information only but reflect structural information.

## CHAPTER III

# Testing Temporal Boundaries of Composition in Low-Frequency Neural Activity

### 3.1 Introduction

Successful speech comprehension relies on temporal integration of linguistic units at different time-scales; comprehenders combine phonemes into syllables, syllables into words, and words into phrases/sentences. In natural Speech, these units are quasi-rhythmic and overlapping, without any physical cues for boundaries, which means that speech comprehension is more than sound recognition. Many previous studies have adopted electroencephalography (EEG) and magnetoencephalography (MEG) techniques to examine the event-related potentials (ERP) or event-related fields (ERF) for time-locked events of speech stimuli reflected in the brain (e.g. Bendixen et al., 2009; Herrmann et al., 2013; Huotilainen et al., 1998). However, due to the irregularity and temporal dynamics in speech, the investigation of ERP/ERF has the limitation that the time-locked event to speech stimuli is not well suited to fully capture the temporal dynamics in continuous speech (Wöstmann et al., 2016). While some work approaches this challenge by fitting computational models to examine how neural activity maps to continuous speech (e.g. Lalor et al., 2009; Lalor and Foxe, 2009), it remains largely puzzling how linguistic units with different temporal properties are integrated in the brain.

Thus, in addition to probing how neural responses reflect speech stimuli in the time-domain, examining from the frequency-domain can serve as another way to overcome the above limitation by decomposing neural signals into phase and amplitude across different frequency bands (Shannon et al., 1995; Smith et al., 2002; Zeng et al., 2005). Moreover, as Wöstmann et al. (2016) points out, examining neural phase in cortical synchronization may offer precise timing of neural activity and how neural networks interact by knowing phase coupling across different frequencies (see also Sauseng and Klimesch, 2008; Tallon-Baudry et al., 1997).

Previous studies have shown that neural activity has the property of repetitive or rhythmic pattern that can be observed from a single neuron or a group of neurons in the brain or central nervous system; this is a neural oscillation (Schnitzler and Gross, 2005). Neural oscillations can be characterized by frequency, phase, and amplitude, and they may synchronize with extraneous stimulation such as periodic acoustic or visual stimuli. The process of synchronization is also called *entrainment* and neuronal populations entrain with different linguistic units. For example, previous studies have shown that word and phrases can be entrained with the frequencies at the delta band (0.5 - 3 Hz) (Ding et al., 2016) and syllables are synchronized with the theta band (4 - 8 Hz) (Ghitza, 2013; Giraud and Poeppel, 2012; Poeppel, 2003). However, given that language comprehension is sensitive to temporal properties of speech, how these entrainments interact with temporal dynamics remain unknown. Moreover, whether these entrainments reflect merely repetitive pattern of external punctual events or whether they are tie to functionally-specific bands are also under debate. The current study manipulates temporal properties of speech to investigate temporal dynamics of cortical tracking in the delta band.

### 3.1.1 Comprehension in low-frequency oscillations

Neural activities in the delta band have been associated with the processing of low-level acoustic information (Ding et al., 2014; Gross et al., 2013; Kayser et al., 2015) and high-level syntactic and semantic knowledge (Bonhage et al., 2017; Ding et al., 2016; Meyer et al., 2016). As Ding and Simon (2014) point out, delta oscillations might reflect onset tracking due to the sharpness of sound edges (Doelling et al., 2014), spectral-temporal feature tracking (Ding and Simon, 2012b), syllable parsing (Giraud and Poeppel, 2012), and sensory selection from noisy background (Ding and Simon, 2012a; Schroeder and Lakatos, 2009). These findings suggest that delta oscillations correlate with language comprehension significantly. However, language comprehension is sensitive to various factors such as temporal properties of speech. How temporal properties such as speech rate and rhythmicity interacting with high-level linguistic knowledge and thus resulting in neural tracking of speech remains unknown. In addition, given that language comprehension involves dynamic interaction between top-down and bottom-up processing, whether language comprehension requires neural responses phase-locked to speech envelope and how high-level linguistic knowledge guides language comprehension are still unclear. This section reviews previous work about the functionality of delta oscillations and the relationship between delta oscillations and language comprehension.

Delta oscillations have been linked to the decoding of low-level linguistic information such as extracting acoustic features or segmenting discrete units and also the integration of high-level information such as syntactic or semantic information (e.g. Bonhage et al., 2017; Bourguignon et al., 2013; Mai et al., 2016; Meyer et al., 2016). For high-level information processing specifically, neural activities in the delta band have been associated with the processing of abstract linguistic structure with dif-

ferent approaches (e.g. Ding et al., 2016; Kaufeld et al., 2020). For example, Ding et al. (2016) used a frequency-tagging paradigm, where native speakers of Mandarin listened to Mandarin four-syllable sentences devoid of prosodic contours that lasted 1 sec in total during MEG recording. Under these conditions, they observed that neural entrainment at 4 Hz reflect the processing of syllables, 2 Hz reflects the processing of phrases, and 1 Hz reflects the processing of sentences. Crucially, English speakers without Mandarin linguistic knowledge only show entrainment at the syllable rate (4 Hz), but not at the phrase (2 Hz) and sentence rate (1 Hz). Furthermore, Ding et al. (2017) conducted an English follow-up experiment using EEG and their results replicates the previous results from MEG. They observed neural tracking of syllable, phrase, and sentence-level peaks in evoked power (i.e. the power that is time-locked and phase-locked to speech stimuli) and inter-trial phase coherence (i.e. whether the differences between phase angles and speech stimuli across trials are consistent), but not induced power (i.e., the power that is not synchronized with speech stimuli). However, it is not yet clear from their studies whether neural activities in the delta band are indeed elicited by abstract structure information robustly or can be modulated only by lexical information of the stimuli (Martin and Doumas, 2017; Frank and Yang, 2018). Chapter II tested whether these patterns reflected entrainment to hierarchy or lexical properties of the stimulus. We offer evidence that delta entrainment correlates with abstract linguistic structure, rather than merely the reflection of lexical representation. A reversed phrase condition, which preserves the frequency of syntactic category but lost the structural information in the normal Mandarin sentence, was compared with the normal structure in Mandarin. Only a 4-Hz peak was observed in the reversed phrase, suggesting that structural information indeed plays a crucial role in the modulation of delta entrainment. A second open question is



whether neural activities in the delta band underlie hierarchical processing in natural speech, or whether the modulation seen here is specifically a function of entrainment due to highly rhythmic, artificial, isochronous speech. To answer this question, the current study manipulates temporal properties of speech to examine whether oscillations reflect arbitrary entrainment to external input or are tied to specific functional bands.

To probe the role of neural activities in the delta band in more natural speech, Kaufeld et al. (2020) further support the view that delta oscillations reflect the combinatorial processes and high-level linguistic content (i.e. meaning and syntax) beyond prosody by using naturalistic stimuli during an EEG recording. Four types of stimuli were tested in the study: (1) sentence condition which includes structure, lexical semantics, and prosody, (2) jabberwocky condition which includes structure and prosody, but not lexical semantics, (3) word-list condition which includes lexical semantics, but not prosody and structure, and (4) reversed speech as a control. By computing mutual information (MI) between neural signals and speech stimuli, they were able to quantify the relationship between different levels of structure (i.e., sentence, phrase, word) and the cortical tracking of speech. Their results reveal that higher MI values were shown at the phrase-level (0.8-1.1 Hz) and word-level (1.9-2.8 Hz) in the sentence condition, compared to the jabberwocky and word list conditions.

Other studies relate delta oscillations to the processing of chunking continuous speech into phrases and sentences (Bonhage et al., 2017; Meyer et al., 2016). For example, Meyer et al. (2016) show that grouping bias of words in ambiguous sentences, regardless of prosodic cues, modulates the phase in the delta band, suggesting a relationship between the internal grouping processes and delta entrainment. However, whether delta entrainment is sensitive to high-level linguistic content remains

under debate in part because delta oscillations have also been associated with lower-level aspects of acoustic processing, discussed below. Boucher et al. (2019) examines low-frequency oscillations when people listened to tones, nonsense sequences and normal speech with similar timing, pitch and energy contours. The peaks of inter-trial phase coherence in the delta range were observed in both nonsense sequences and normal speech, suggesting that delta oscillations associate with the sensory chunking of speech and the sequential properties of speech and indirectly link to the processing of the high-level content.

In terms of acoustic decoding, low-frequency neural oscillations have been associated with speech intelligibility. Previous research has identified that theta entrainment reflects syllable intelligibility in different experimental manipulations, given that the average syllable duration falls between 2.5 and 8 Hz (about 125-400 ms) (see Ding and Simon, 2014; Poeppel and Assaneo, 2020). For example, Peña and Melloni (2012) found that the theta power increased more in normal speech stimuli than reversed stimuli, suggesting the theta oscillations reflect some aspect of syllable tracking. Other studies manipulated the temporal structure of speech stimuli. Luo and Poeppel (2007) found that theta oscillations were phase-locked to the speech stimuli and this mechanism correlates with speech intelligibility. They manipulated the fine structure and speech envelope of speech stimuli and found that the reduction of intelligibility decreases the phase dissimilarity but not power dissimilarity in the theta range. Consistent with the above finding, Howard and Poeppel (2010) found that the speech stimuli from both the normal and time-reversed speech can be discriminated in theta-band phase, suggesting that the theta-band phase tracking pattern reflects early processing of acoustic properties in auditory cortex, but not the processing of lexicon, semantics and syntax.

While low-frequency oscillations have been linked to the processes of acoustic information, previous research suggests different functionalities for theta and delta bands. As Kayser et al. (2015) suggests, theta entrainment may relate to syllable parsing and speech segmentation directly while delta entrainment relates to the top-down processes that are sensitive to factors such as speech rate. Gross et al. (2013) found phase alignment between the speech envelope and the delta and theta ranges when participants listen to a normal story as compared to the backward condition, consistent with the hypothesis that the observed entrainment in the normal speech reflects a top-down mechanism. Ding et al. (2014) found that the delta-band synchronization is enhanced while the spectral resolution of stimuli is reduced, suggesting the possible role for delta entrainment as the top-down attention following an earlier proposal by Schroeder and Lakatos (2009). This reduction was considered as a “high-level entrainment of speech features” in Zoefel and VanRullen (2015), meaning that the entrainment can be observed even when the low-level features are removed. As Zoefel and VanRullen (2015) point out, this is distinct from entrainment that involves high-level processing is “the high-level modulations of phase entrainment”, meaning that oscillations are driven by the low-level information such as speech amplitude and further modulated by high-level processes such as prediction and attention. For example, Kerlin et al. (2010) had participants listened to two different speech streams simultaneously during EEG recording while attending to one speech stream only. Results showed greater enhancement for phase-locking response in the theta range in the attended speech, relative to the unattended speech.

Previous studies have shown that there is a dynamic interaction between bottom-up processing and top-down processing and it facilitates language comprehension (Davis and Johnsrude, 2007; Gwilliams et al., 2018; Gwilliams, 2020; Sohoglu et al.,

2012). From the above, multiple studies have provided some evidence for the functional significance of a specific frequency band. However, since neural responses track information from multiple time scales concurrently, cross-frequency coupling can demonstrate how different levels of processing influence each other. For example, Keitel et al. (2018) found that phase at the phrasal-level (0.6-1.3 Hz) is coupled to the power at beta range (13-30 Hz) in motor area, which has been associated with the predictive processing of rhythms or beats (Arnal et al., 2015). These results suggest that a top-down prediction at the phrasal level may involve the cross-frequency coupling with the beta range. In addition, Rimmele et al. (2019) found that the phase at 2 Hz is coupled with the amplitude at 4 Hz in MTG when German native speakers listened to real German and Turkish words during MEG recording with a frequency-tagging paradigm, but not the Turkish pseudo-words, which does not require the grouping of syllables into words. The above two studies suggest that we do not rely only on the bottom-up processing and provide evidence that the interaction between low-level and high-level information can be captured by cross-frequency coupling. Moreover, Park et al. (2015) found that stronger causal top-down effects at both delta and theta ranges in left auditory cortex when people listened to the intelligible speech, compared to the backward speech. In addition, they found that stronger top-down effects in the delta band, relative to the theta band, suggesting that a longer time window is preferred to extract high-level information. These findings suggest that there is an interaction between the high-level linguistic information from word to phrase level and the low-level acoustic processing and this interaction can be indexed by neural oscillations in the delta band.

Evidence further suggests that such oscillations are causally connected to speech processing. Specifically, speech intelligibility is also affected by whether oscillators

can track the amplitude of speech input well. Whether speech can be tracked well is largely affected by the speech rate. The optimal syllable rate for understandable speech is between 2 - 8 Hz (Poeppel and Assaneo, 2020). Comprehension decreases when the syllable rate falls outside of that frequency range (Ghitza, 2014; Ghitza and Greenberg, 2009). For example, Ahissar et al. (2001) examined the correlation between neural signals and comprehension of speech with four different time-compressed ratios (i.e., 0.2, 0.35, 0.5, 0.75). The results of comprehension significantly drop in the ratio of 0.2 and 0.35. They found that speech comprehension is correlated with the difference between the frequencies of the speech envelope and the cortical signals, and also the phase-locking values between the speech envelope and the neural signals. Their results suggest that speech comprehension requires the cortical responses phase to be locked to the speech envelope and that such phase-locking is only possible within a limited range. This phase-locking appears to be necessary for accurate segmentation of continuous speech into syllables. In addition, Nourski et al. (2009) found that power spectra of the low-frequency ECoG components matched the frequency of the stimulus envelope of sentences when time-compressed speech has a compressed ratio above 0.4. Moreover, a positive correlation between speech comprehension and the above frequency matching was observed in the study. However, they also observed that the synchronization between event-related band power and speech stimuli in the high-frequency range of the ECoG across all speech rates. Their results yield the question whether evoked responses should be modulated within certain time interval. Moreover, Assaneo and Poeppel (2018) recorded MEG responses when participants listened to trials included syllables with various speech rates. They found that brain responses can be synchronized with the speech envelope across different speech rates (i.e., 2.5 syl/s, 3.5 syl/s, 4.5

syl/s, 5.5 syl/s, 6.5 syl/s). From the above, it remains unclear how speech rates modulate low-frequency oscillations and whether there is an optimal time-scale for brain responses to achieve language comprehension. Moreover, while most studies focus on how low-level information processing reflects on low-frequency oscillations, it remains elusive how the integration between the high-level information and these low-level acoustic information modulates low-frequency oscillations during language comprehension.

### **3.1.2 Oscillation-based model vs. Evoked-response model**

The hypotheses mentioned earlier for speech perception may also characterize the functional interpretation of delta entrainment in sentence comprehension. Specifically, the role of syllable parsing can be generalized to explain the relationship between delta oscillations and high-level information processing. Under the syllable parsing hypothesis, parsers are able to segment continuous speech into syllable-level chunks since syllables can be matched well in speech envelope and this matching process occurs in the theta cycle (Ghitza, 2013; Giraud and Poeppel, 2012). The cortical oscillations within a theta cycle are thus viewed as the reflection of “packaging” acoustic features of speech signals (Ghitza, 2011). To be specific, since speech involves various linguistic units processed in multiple time frames simultaneously, the oscillation-based model TEMPO (Ghitza and Greenberg, 2009; Ghitza, 2011) proposed that syllables are matched in a theta cycle (about 200 ms) while finer phonetic features are matched within a beta cycle (about 50 ms). In TEMPO, if the oscillator is able to track the amount of information from the input, given the necessary decoding time, intelligibility can be achieved. This theory carries the prediction that intelligibility decreases if the speech rate falls outside of its corresponding frequency range (i.e., theta range for the syllable rate). In terms of sentence com-

prehension, the evidence reviewed so far suggests that delta oscillations are driven either by acoustic features or higher-level information and these two factors are highly interconnected (Meyer et al., 2016; Ghitza, 2017). The TEMPO model can be generalized to the word/phrasal level, meaning that the process for decoding phrases is guided by prosodic segmentation as long as neural activities in the delta band can be synchronized to the speech rhythm (Ghitza, 2017). Ghitza (2017) suggests that delta can be a acoustic marker for the segmentation by conducting two recall task of context-free digits. In their second experiment, syllable rate was manipulated so that the chunking rate of a phrase can be inside or outside the delta frequency band. The chunking rate was manipulated by inserting different silence gaps, but not changing the length of syllables. Four bands of chunking rates were examined (i.e., 2-2.5 Hz, 2.5-3 Hz, 3-3.5 Hz, and 3.5-4 Hz). Their results show that the error rate of behavioral responses increases when the chunking rate increases, supporting their hypothesis that a successful segmentation occurs only when the chunking rate falls in the delta range.

Other studies manipulating speech rates also provides supporting evidence for oscillatory models. Kösem et al. (2018) examined whether there is a causal role of neural oscillations in speech processing. Participants listened to Dutch sentences with either fast or slow speech rate from the beginning to the word before the target. The participants judged whether the target with the original speech rate contained a short or long vowel. The sustained oscillatory activities were observed from the carrier window to the target word. Also, the perception bias was observed while people tended to choose the long vowel in the fast preceding context and short vowel in the slow preceding context. These findings suggest that entrainment to the preceding context affect how parsers perceive the subsequent words. Most importantly,

the results further support oscillation-based models. That is, by varying the time frame of syllables, theta oscillations actively adapt to the speech rate and modify the expected duration of syllables in one cycle. This mechanism then guides how parsers perceive speech, leading different selections of vowels to form different meanings.

However, it is also possible to consider neural activities in the delta band as tracking the transient events in sentences. Lau and Liao (2018) point out that results shown in Ding et al. (2016) might reflect purely punctual events at critical timescales during structure-building processes, but not the sustained activity for maintaining the structure over time, proposed by Pallier et al. (2011). Alexandrou et al. (2020) also suggest that “cortical entrainment” in most studies are stimulus-driven responses or evoked responses, given that the stimuli in most studies are unnatural and the results are difficult to generalize to the naturalistic settings such as spontaneous speech. Moreover, Zoefel et al. (2018) also point out that the observed neural entrainments with repetitive stimuli may reflect merely the repetitive evoked responses (see also Capilla et al., 2011, Keitel et al., 2014). However, how to dissociate evoked responses and endogenous oscillatory activities is difficult. The oscillations and the repetitive evoked responses share multiple properties (e.g., the dominant frequency corresponds to the rhythmicity of stimuli) and sometimes the signals are the combination of both types (Doelling et al., 2019; Zoefel et al., 2018).

Previous studies manipulate jittered inter-stimulus intervals to test whether the elicited neural signals are endogenous oscillatory activities or evoked responses. The assumption is that the neural responses actively adapt the upcoming changes if the oscillatory activities are indeed a predictive process. Whereas for the evoked responses, the neural signals are expected to passively follow the timing of stimuli. Capilla et al. (2011) conducted two EEG experiments by using checkerboard stimuli



with different presentation rates and simulated models from both oscillation and evoked response accounts. Their results show that neural responses do not adapt changes depending on the preceding inter-stimulus intervals as they hypothesized and thus the neural responses are better explained by the evoked response. However, in contrast to Capilla et al. (2011), Kayser et al. (2015) manipulates jitter in speech by either expanding or shrinking the gaps between words but still maintaining a constant average inter-stimulus interval. In line with the account of oscillatory activity, they found that the mutual information between the delta range and the speech stimuli reduces when jitters increase, along with the lack of change in early auditory responses across conditions. These results suggest that delta entrainment can be affected by regularity of speech, rather than the transient evoked responses to the syllable onsets.

Another way to distinguish them is that brain oscillations can occur without actual external stimuli input. More relevantly, Zoefel et al. (2018) consider that findings in Ding et al. (2016) as the endogenous neural oscillations, rather than the repetitive evoked responses, since the speech stimuli only display syllable peaks at 4 Hz but neural responses are able to generate the internal rhythms at the phrase and sentence level by demonstrating phrase/sentence-level peaks at 1 Hz and 2 Hz. Instead of taking side from one of the above hypotheses, Zoefel and VanRullen (2015) consider results from most studies are a mixture of evoked responses and endogenous entrainment, given that entrainment can still be observed even when the low-level features are absent.

From the above, whether the regular pattern observed in neural signals represents the endogenous oscillations or repetitive evoked responses is still under debate. If we also consider how speech rate modulates neural activities, results from Ahissar et al.

(2001) and Nourski et al. (2009) would carry the prediction that neural signals can be evoked for each particular as long as the event occurs within a certain time threshold. However, part of results from Nourski et al. (2009) indicates that evoked signals can be observed even for very rapid compressed speech. The present study manipulates different speech rates to disentangle whether the observed neural activities represents oscillation-based model or evoked responses.

### **3.1.3 Speech rhythmicity in low-frequency oscillations**

Regularity can also modulate delta entrainment. Cravo et al. (2013) had participants look at a series of pictures that include the Gaussian noisy patches (standard) and the Gabor patches embedded in the Gaussian noise (target) with either regular or irregular stimulus-onset asynchrony between each standard during an EEG recording. They found that phase alignment of delta frequency and better behavioral performance were shown in the regular non-speech visual stimuli, compared to the irregular ones. Daily speech is quasi-rhythmic. Speech includes various information such as the acoustic features of syllables and the production and perception of speech are affected largely by the neighbouring contexts. Recent studies have shown that regularity in speech also plays an important role in modulating delta oscillations. To be specific, Kayser et al. (2015) manipulated the temporal structure of speech by expanding or shrinking the pauses between syllables. Crucially, this manipulation reduces entrainment in delta, but not theta, to the speech envelope. Their results suggest that delta entrainment reflects the top-down processes that are sensitive to speech regularities, rather than the bottom-up encoding processes of acoustic features.

However, as Ding and He (2016) point out, while the behavioral results show high accuracy for each condition with different percentage of jitters in Kayser et al. (2015),

how regularity in speech affect language comprehension remains unclear if the brain relies on the temporal regularities to predict upcoming sequences. Moreover, whether rhythm variations at the phrasal level are sufficiently regular for a flexible delta mechanism to drive a reliable segmentation process is yet to be tested (Ghitza, 2017). A recent study has shown that regularity in speech does not affect the behavioral results from a digit-recall task, as long as the rate for the target sequence falls in the delta range. Ghitza (2017) conducted a context-free digit recall task and manipulated the regularity of chunk size (regular chunk (e.g. 22222), and irregular chunk (e.g. 424)) and two prosodic cues (stimuli were recorded either consecutively as chunks or recorded individually and connected afterwards), The psychophysical data show that similar error patterns in both prosodic conditions, as long as the chunking rate falls in the delta range. Based on the above findings, delta oscillations should adjust rhythmicity in speech flexibly during language comprehension.

#### **3.1.4 Summary**

The current study aims to investigate how altering the temporal dynamics of isochronous linguistic events affects entrainment to low-frequency neural activities. Following Ding et al. (2016), the frequency-tagging paradigm is adopted in the experiment. Speech rate and regularity in speech are manipulated. The frequency-tagging paradigm, as Keitel et al. (2014) points out, can associate multiple concurrent events to multiple frequencies of interest. That is, as mentioned earlier, if a syllable lasts for 250 ms, then a two-syllable phrase lasts for 500 ms and a four-syllable sentence lasts for 1 sec. Thus, the syllable-level processing is tagged at 4 Hz, the phrasal-level processing at 2 Hz, and the sentence-level processing at 1 Hz. Two broad questions are addressed in the present study. The first issue is the relationship between speech comprehension and neural activities in the delta band. Four different syllable rates

(250 ms, 200ms, 100 ms, 75ms) were tested and the accuracy of comprehension for trials was computed. While the lowest compressed ratio is 0.3 for the 75ms/syl condition in the current study, we would expect that comprehension is significantly impaired in this condition, but not other conditions. Following Nourski et al. (2009), we should find high comprehension accuracy, along with synchronization of speech stimuli and cortical activities in the conditions of 250ms/syl, 200ms/syl, and 100ms/syl.

The second issue is whether neural activities in the delta band reflect a process of packaging information from the input or a process of punctual events. Ghitza (2017) focuses on examining acoustic-driven delta in the bottom-up segmentation by using context-free digits recall tasks and providing psychophysical evidence. The current study examines entrainment for the real sentences and aims to test predictions from the TEMPO model. The processing of real sentence would require both bottom-up segmentation and top-down parsing and results in the interaction between the acoustic-driven and context-driven delta. Since Ghitza (2017) have already shown that delta can be a acoustic-driven marker, the current study can show the product of the interaction between acoustic-driven and context-driven delta. We can indeed test whether TEMPO can be extended to explain the chunking processes at the phrasal level. Similar to the second experiment in Ghitza (2017), syllable rate was manipulated so that the chunking rate of a phrase can be either inside or outside the delta frequency band. Different from Ghitza (2017), the current study used time-compressed speech and controlled the length of syllables. Building on the results from Ghitza (2017), we should not find the phrasal-level peaks if the timing for phrasal composition falls outside of the delta range. However, if delta entrainment reflects evoked responses, we would expect entrainment shown in every

condition with a fixed speech rate. Also, based on Nourski et al. (2009), it is expected to find entrainment if the compressed ratio is above 0.4 and the sentences are comprehensible. In addition, we would not expect entrainment in the conditions with irregular speech rates, if the responses are hypothesized to actively adjust the upcoming changes under the oscillation-based models. The current study also tests how irregularity reflects on low-frequency oscillations. Three different syllable rhythmicity (regular, semi-regular, and irregular) were included. Based on Ghitza (2017), if delta entrainment can flexibly adapt regularity in speech, we expect entrainments to be observed in all conditions.

It should be noted that the signals measured in the current study reflect low-frequency neural activities. One hypothesis is that the measurements reflect neural oscillations. However, the current experiment does not allow us to conclude that the results can be directly mapped to neural oscillations.

## **3.2 Experiment**

### **3.2.1 Participants**

Forty-two native speakers (28 females, 14 males) of Mandarin Chinese between the age of 18 and 38 (mean = 24.6) participated in the experiment. They were all right-handed and had normal hearing. They self-reported that they did not have any neurological disorders. They gave informed consent and were reimbursed for their time (15 USD/hour). Data from twelve participants were excluded from the analysis due to having excessive noisy trials (>30 %). Thus, data from 30 participants (19 female, 11 males) were included in the final analysis.

### 3.2.2 Materials

Following the experimental paradigm in Ding et al. (2016), syllables used in the stimuli were recorded from the website developed by iFLYTEK ([https://www.xfyun.cn/services/online\\_tts](https://www.xfyun.cn/services/online_tts)). Then syllables were compressed to 240 ms by using Praat vocal toolkit (Corretge, 2020) and appended 10 ms silence to have the length of 250 ms in Praat (Boersma and Weenink, 2019). Table 3.1 shows the experimental conditions and the frequency of interest for each condition. Three different syllable rates were included: 250 ms, 200 ms, and 100 ms. Syllables with the rates of 200 ms and 100 ms were compressed from the length of 250 ms, using Praat vocal toolkit also (Corretge, 2020). Then syllables were combined into four-syllable sentences. Three different rhythmicities were tested. That is, the regular condition included the fixed syllable rate at 250 ms. The semi-regular condition included the fixed sequencing syllable rate: 300 200 300 200 ms, and thus the sentence length was kept at 1 second. The irregular condition included random syllable length (either 200 ms or 300 ms) and the sentence length was also 1 second. Thus, total of five conditions were included. Each condition had 60 trials in total and was presented in two separate blocks composed into 30 trials. Each trial was made up of 10 sentences. In each block, 30 trials included 20 grammatical trials and 10 semantically-improbable trials (Table 3.2). For the semantically-improbable trials, we randomly select syllables from the grammatical condition and assign them to the new four-syllable sequences with their original syntactic positions unchanged. The stimuli were the same as the conditions of the four-syllable sentence and the semantically-mismatched sequence in the Chapter II, listed in the Appendix A.1. The whole experiment included 10 blocks (300 trials in total). Only grammatical four-syllable sentences were analyzed.

In addition to the above conditions, we added a condition with 75 ms per syllable

during the middle of data collection. We found overall performed high accuracy in all conditions. Thus we decided to include the 75 ms/syl condition to see whether comprehension would be impaired under the ratio of 0.4. We report the intermediate results from 17 subjects for this condition.

Table 3.1: Experimental conditions and frequency of interest

Condition	<b>Regular:</b> 250 ms/syl (4Hz/syl)	<b>Regular:</b> 200 ms/syl (5Hz/syl)	<b>Regular:</b> 100 ms/syl (10Hz/syl)	<b>Regular:</b> 75 ms/syl (13 Hz/syl)
Frequency of interest	Sentence: 1 Hz Phrase: 2 Hz Syllable: 4 Hz	Sentence: 1.25 Hz Phrase: 2.5 Hz Syllable: 5 Hz	Sentence: 2.5 Hz Phrase: 5 Hz Syllable: 10 Hz	Sentence: 3.33 Hz Phrase: 6.66 Hz Syllable: 13.3 Hz
Condition	<b>Semi-regular:</b> 300 200 300 200 ms	<b>Irregular:</b> 300 200 300 200, 300 300 200 200...		
Frequency of interest	Sentence: 1 Hz Phrase: 2 Hz Syllable: 4 Hz	Sentence: 1 Hz Phrase: 2 Hz Syllable: 4 Hz		

Table 3.2: Example stimuli

<b>Grammatical four-syllable sentence</b>	<b>Semantically-implausible sequence</b>
綿 羊 吃 草 mian yang chi cao Cotton sheep eat grass 'Sheep eat grass.'	軍 孩 奔 草 jun hai ben cao Soldier child run grass

The power spectrum of the speech stimuli for different speech rates is shown in Figure 3.1. We used Hilbert transformation to obtain the analytic signal from each stimulus trial, and used the power of the analytic signal to compute the broad-band envelope. We then computed the power spectrum by using fast Fourier transform. The power spectrum was averaged over all trials for each condition. All the analyses were done in Matlab. Only the syllable-level peak was found in these power spectra (see Figure 3.1). Peaks appearing after the first peak are harmonic series. The power spectrum for different rhythmicity is shown in Figure 3.2. Note that there is a peak at 2 Hz in the semi-regular condition. It likely reflects the fact that the temporal sequence is always 200ms 300ms 200ms 300ms. Every two syllables forms

a rhythmicity at each 500 ms.

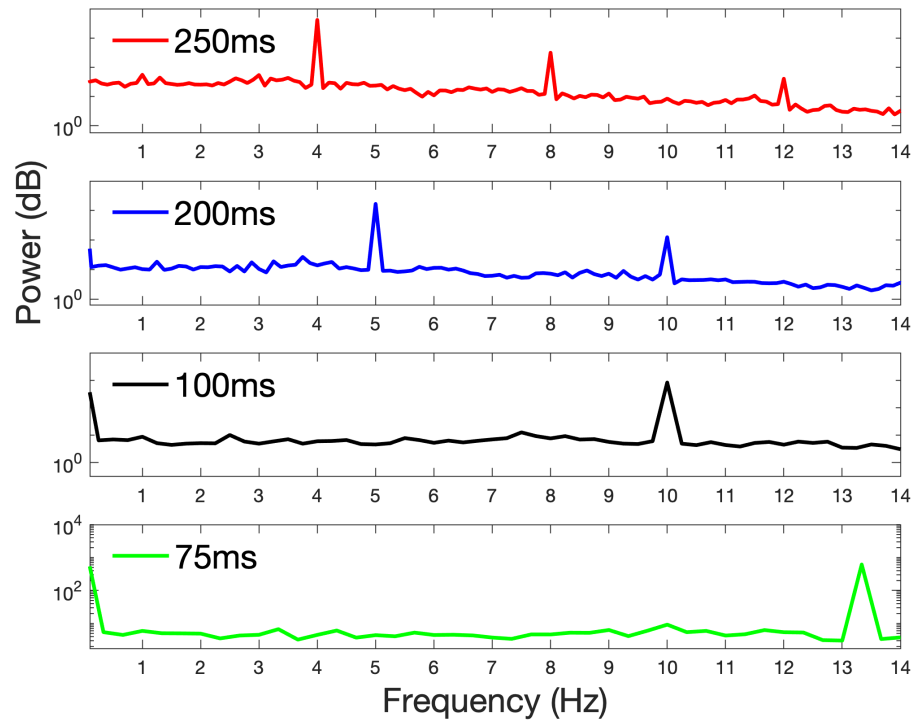


Figure 3.1: Power spectrum of speech stimuli for different speech rates

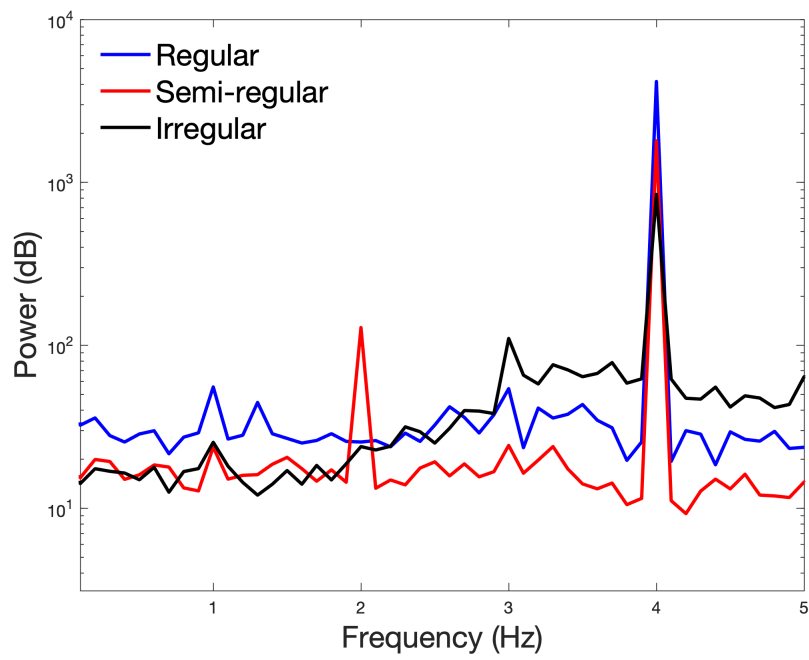


Figure 3.2: Power spectrum of speech stimuli for different rhythmicity



### 3.2.3 Procedure

Participants sat comfortably in front of a computer screen in a quiet room. Prior to the main session, participants were fitted with an electrode cap. Electrodes were also affixed above and below the left eye. Electrolyte gel was applied to minimize impedance. The setup took approximately 30 minutes. Then the stimulus volume was set for each ear based on hearing level of each participant with the threshold of 45 dB determined using 1000 Hz tones (300 ms length). After setting up the volume, a pre-test presenting 120 1,000 Hz tones was conducted to ensure the data quality was sufficient to run the main session.

During the main session, participants were instructed to judge whether a trial includes plausible sentences/phrases or not by button press. After the button press, the next trial was played after a delay randomized between 800 - 1,400 ms (Ding et al., 2016). Stimuli were presented with Psychopy2 (v1.84.2; Peirce, 2007, 2009). Participants were also instructed not to frequently blink their eyes and avoid unnecessary body position adjustment when the stimuli were presented. Participants could take a short break between blocks. After the instructions, participants had a practice session that included the training and testing phases to ensure that participants are used to different speech rates and familiar with the whole procedure of the experiment. The training phase included 22 trials in total with different speech rates and different answers controlled. Speech rates with 75 ms, 100 ms, 200 ms, 250ms, 300 ms, semi-regular 250 ms and irregular 250 ms were included. Four trials were included for each 75 ms, 100 ms, 200 ms, 300 ms condition. Two trials were included for each 250 ms, semi-regular 250 ms and irregular 250 ms condition to balance the condition of syllable length and shorten the training phase. Participants completed the training phase with the experimenter. The testing phase included 12 trials (2/per

condition). During the training and testing phases, participants got feedback on the screen for each trials. Comprehension in the testing phase was assessed and reported in the result section. The order of blocks was counterbalanced across participants. The main experiment took about 1 hour to finish. After the main session, participants washed their hair to remove the electrolyte gel and were debriefed about the goals of the experiment.

#### **3.2.4 EEG recording and data analysis**

EEG data were recorded at 500 Hz from 61 active electrodes (actiCHamp, Brain-Products GMBH) in a 0.01-200 Hz band with online reference to an electrode placed on the left mastoid. Due to the technical issues caused by the machine, impedance measurements were not available; electrodes with excessive noise were identified by visual inspection. FieldTrip software (Oostenveld et al., 2011) was used to analyze the data. Artifacts related to eye blinks were removed via Independent Component Analysis (Jung et al., 2000; Makeig et al., 1995) and remaining artifacts were removed manually following visual inspection. Following Ding et al. (2017), the first sentence from each trial was excluded to avoid potential EEG responses to sound onset. Data were filtered from 0.1-25 Hz, and re-referenced offline to a common average. Since each condition differs in length, they afford different spectral resolution. To facilitate comparison between conditions, evoked power and ITPC were computed from 0.5 to 20 Hz in increments of 0.1 Hz with Hanning tapers after zero padding with 10 seconds. Inter-trial phase coherence and evoked power were computed based on the equations given in Ding et al. (2017), as listed and defined in Equation (3.1-3.2). To remove the trend of  $1/f$  in the power spectra and test the statistical reliability, evoked power at each frequency was normalized by the neighboring frequency bins within  $\pm 0.5$  Hz via Equation 3.3 used in Ding et al. (2017). For statistical analysis,

conditions were compared via one-way ANOVA for each measure at each frequency of interest.

(1) **Inter-trial phase coherence** (ITPC) reflects phase coherence across trials (Cohen, 2014). The summation of cosine and sine values of phase angle  $\theta_k$  of each complex-value Fourier coefficient is computed and then the square root of the summation is averaged over the total number of trials  $K$ . Higher ITPC indicates that phase angles are consistent across trials.

$$R(f) = \frac{\sqrt{(\sum_k (\cos\theta_k))^2 + (\sum_k (\sin\theta_k))^2}}{K} \quad (3.1)$$

(2) **Evoked power** (EP) reflects the power of EEG responses that are phase-locked and time-locked to speech stimuli. The Discrete Fourier Transform (DFT) of the response in trial  $k$  is denoted as  $X_k(f)$  and  $X_k(f)$  is a complex-value Fourier coefficient. Thus, evoked power is the summation of complex-value Fourier coefficient of trials averaged over the total number of trials  $K$ .

$$E(f) = \frac{1}{K} \left| \sum_k X_k(f) \right|^2 \quad (3.2)$$

(3) **Normalized evoked power** (EPn)

$$En(f) = \frac{E(f)}{\sum_w E(w)}, |w - f| < 0.5 \text{ Hz}, w \neq f \quad (3.3)$$

$w$  represents the neighboring frequency around the target frequency  $f$ .

### 3.2.5 Predictions

The regular 250ms/syl condition serves as a control condition. As in Ding et al., 2016, 2017, and Chapter II, evoked power should be elicited at 1 Hz for sentence rate, 2 Hz for phrasal rate, and 4 Hz for syllable rate. For the regular 200ms/syl condition, the syllable rate falls in the theta range while the sentence rate and the

phrasal rate fall in the delta range. According to TEMPO and results from Ding et al. (2016), cortical tracking should be observed at its corresponding frequency. As for the regular 100ms/syl and 75ms/syl conditions, these are crucial conditions since the phrase-level composition falls in the theta range, but not the delta. In addition, the syllable rate falls out of the theta range in these two conditions. While the sentence-level composition falls in the delta range and phrase-level composition falls in the theta range, the functionalities of theta and delta entrainment can be compared in these two conditions. That is, if delta entrainment, but not theta, correlates with phrase/sentence-level comprehension, we should expect to see peaks in the delta range, but not theta. However, if delta entrainment does not specifically correlate with phrase/sentence-level comprehension, we might see the peaks related to the phrasal-level processing appearing in the theta range. Based on the TEMPO oscillation-based model, it is expected that neural responses might lose track of syllables for the 75ms/syl condition. However, it is unknown whether the failure of tracking syllables leads to the failure of tracking the higher information. As for the semi-regular and irregular conditions, based on the psychophysical data that different prosodic cues do not affect the error pattern in Ghitza (2017), we should expect that entrainments can be observed in both conditions if delta oscillations can flexibly adapt rhythmicity.

### **3.3 Results**

#### **3.3.1 Comprehension score**

The overall accuracy percentage in the practice session is 98%. The overall accuracy in the main session is 98%. Figure 3.3 shows the accuracy percentage for each condition. This indicates that even at very rapid rates, participants show a high degree of comprehension after just a few minutes of exposure (Ahissar et al., 2001;

Nourski et al., 2009).

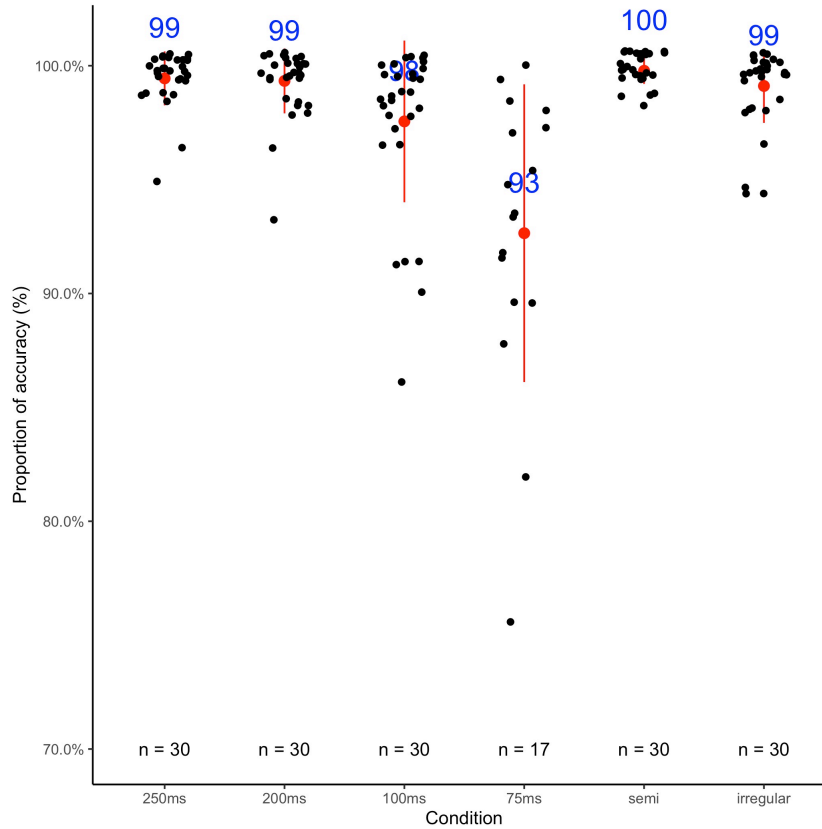


Figure 3.3: Accuracy percentage for each condition (Red dot: mean; Red line: +/-1 SD)

### 3.3.2 EEG results

The current experiment manipulates different speech rates and different rhythmicities to examine the functional significance of low-frequency neural activities. The following results first compare the conditions with different speech rates and then conditions with different rhythmicities.

First, we computed ITPC and EPn for each condition with different speech rates. ITPC measures the phase consistency of neural responses to stimuli across trials while EPn computes the normalized amplitude synchronized with speech stimuli. Figure 3.4 shows the ITPC and Figure 3.5 shows EPn computed from 0 to 14 Hz for each condition. Note that there is an edge effect shown in the results of EPn across

conditions. The edge effect might come from the algorithm that removes  $1/f$  trend.

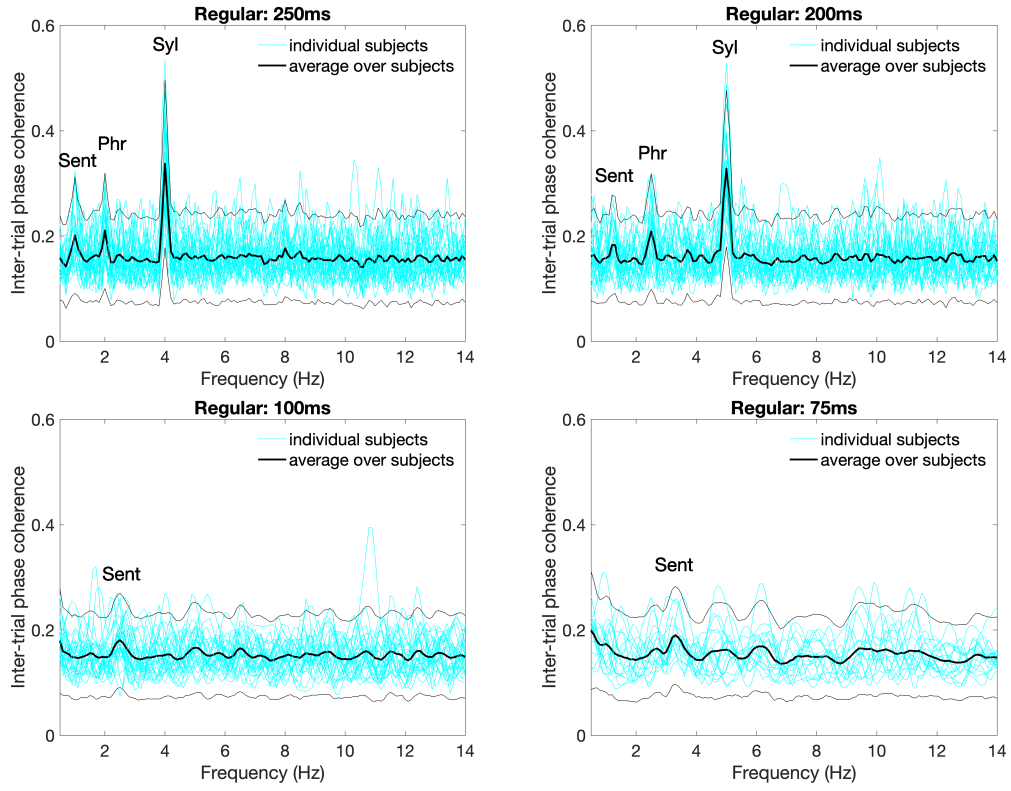


Figure 3.4: Inter-trial phase coherence for conditions with different speech rates. Regular:75ms shows the results from the subset of  $N = 17$ .

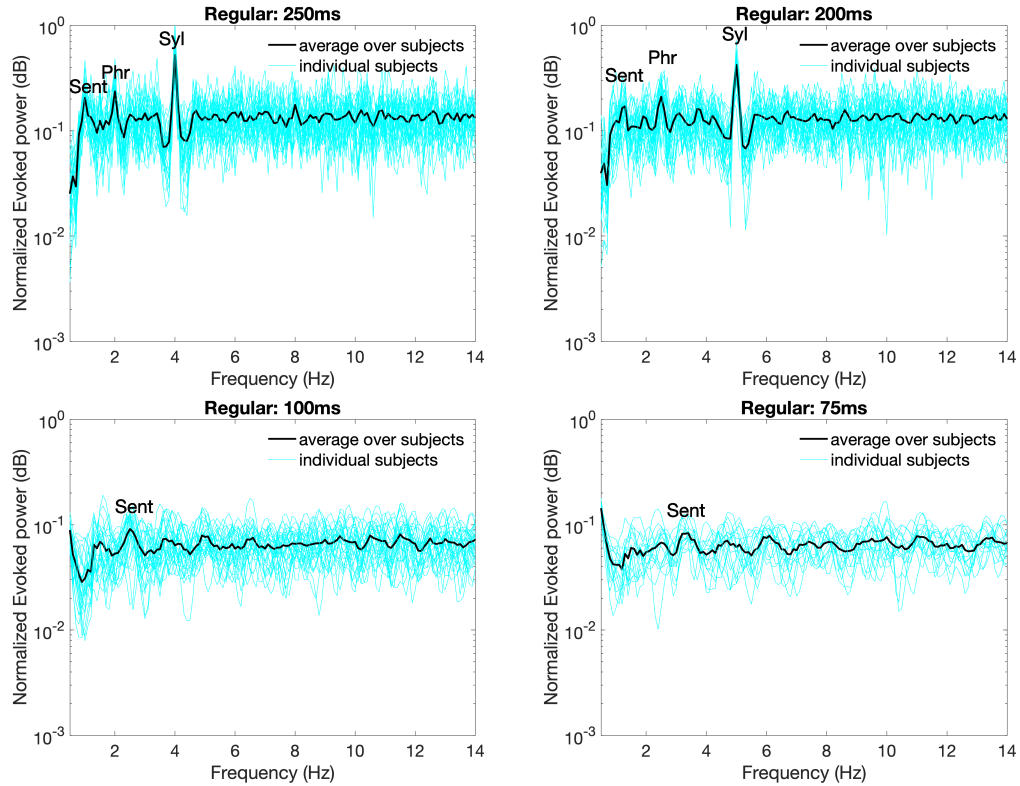


Figure 3.5: Normalized evoked power for conditions with different speech rates. Regular:75ms shows the results from the subset of  $N = 17$ .

Peak values for each frequency of interest are plotted in Figure 3.6. Peaks at 7 Hz in the 250/syl condition serves as a baseline activation. For ITPC, no significant difference in the amplitude of the sentence-level peak across the conditions ( $p = 0.25$ ,  $\eta^2 = 0.032$ ). For the phrasal-level, there is a statistical significance ( $p < 0.001$ ,  $\eta^2 = 0.18$ ). Post-hoc pairwise Tukey tests were also conducted for the phrase-level. There is no difference between 250ms/syl and 200ms/syl ( $p = 0.71$ ). A difference was found between 250ms/syl and 100ms/syl ( $p < 0.001$ ) and between 200ms/syl and 100ms/syl ( $p < 0.01$ ). As for the syllable-level, a statistically significant effect was found when comparing the speech rates ( $p = 0.001$ ,  $\eta^2 = 0.84$ ). The post-hoc pairwise Tukey tests show that there was no significant difference between 250ms/syl and 200ms/syl ( $p = 0.83$ ). A significant difference was found between the 250ms/syl and 100ms/syl

( $p < 0.001$ ) and 200ms/syl and 100ms/syl ( $p < 0.001$ ). Summarizing, syllable and phrasal peaks are significant lower for 100 ms than for 200 ms and 250 ms speech rates, while the sentence-level peak shows a small decline for 100 ms alone. The results suggest neural responses lose tracking of syllable if syllable rates fall outside of the theta band and also lose tracking of phrase if the rate of composing a phrases falls outside of the delta band.

For EPn, the one-way ANOVA test shows that there is a statistical reliable effect at the sentence-level ( $p < 0.001$ ,  $\eta^2 = 0.36$ ). Post-hoc pairwise Tukey tests show that there is no significant difference between 250ms/syl and 200ms/syl ( $p = 0.09$ ). A significance was found between 250ms/syl and 100ms/syl ( $p < 0.001$ ) and between 200ms/syl and 100ms/syl ( $p < 0.001$ ). As for the phrase-level, a significant effect was found ( $p < 0.001$ ,  $\eta^2 = 0.46$ ). Post-hoc pairwise Tukey tests show that no significant difference was found between the 250ms/syl and 200ms/syl ( $p = 0.37$ ). A significant difference was found between 200ms/syl and 100ms/syl ( $p < 0.001$ ) and between 250ms/syl and 100ms/syl ( $p < 0.001$ ). For the syllable-level, a significant effect was found ( $p < 0.001$ ,  $\eta^2 = 0.64$ ). Post-hoc pairwise Tukey test shows that a significant difference in all paired comparisons (250ms/syl vs. 200ms/syl:  $p < 0.05$ ; 250ms/syl vs. 100ms/syl:  $p < 0.001$ ; 200ms/syl vs. 100ms/syl:  $p < 0.001$ ).



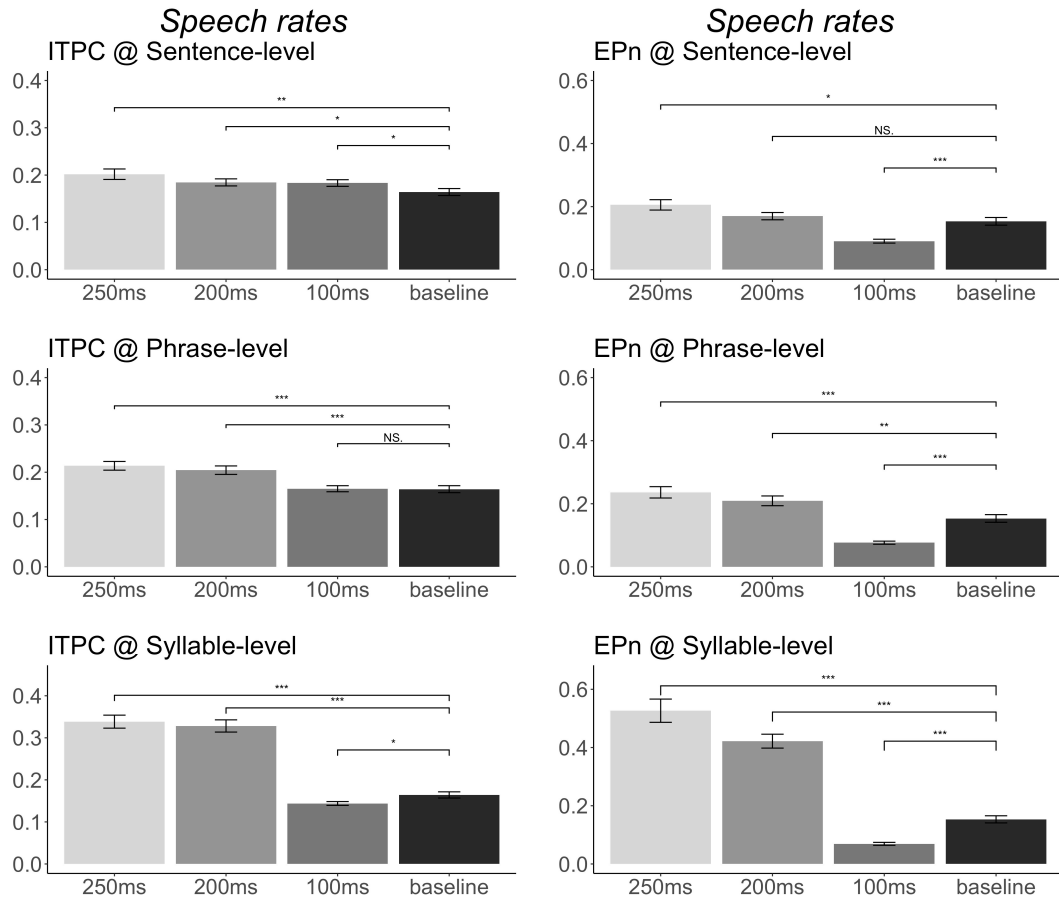


Figure 3.6: ITPC and EPn across conditions with different speech rates. Peaks at 7-Hz in the 250/syl condition serves as a baseline activation.

Second, we compare different regularities. Figure 3.7 shows the ITPC and Figure 3.8 shows EPn computed from 0 to 14 Hz for each condition.

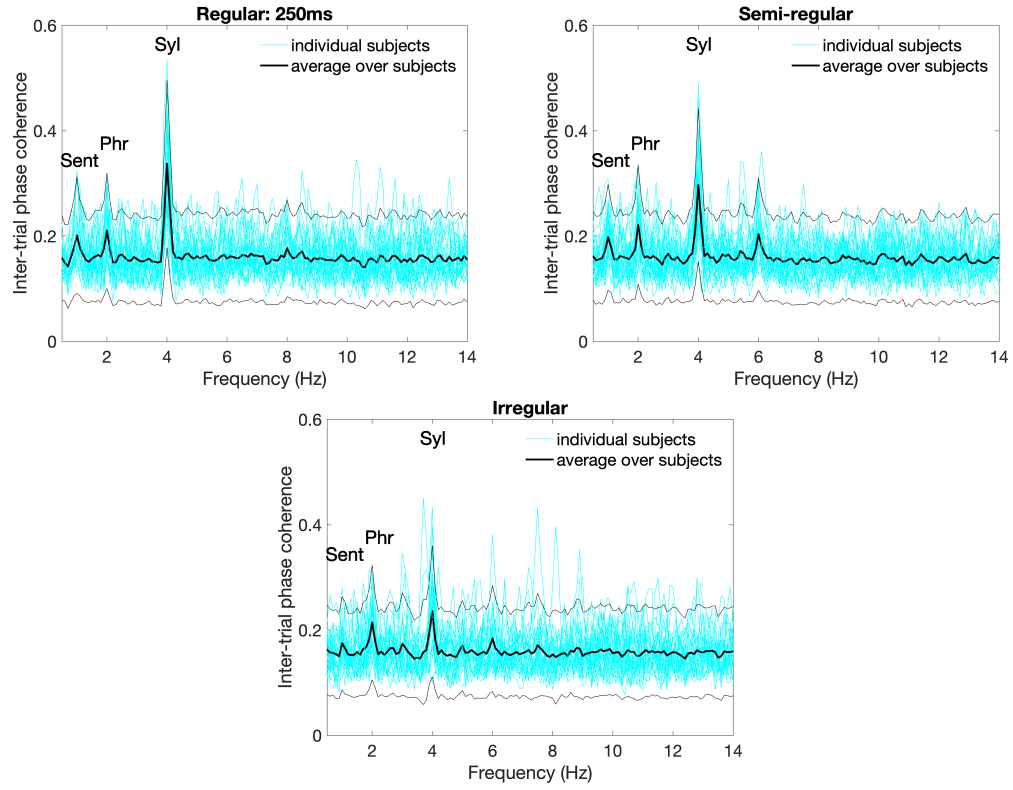


Figure 3.7: Inter-trial phase coherence for conditions with different rhythms.

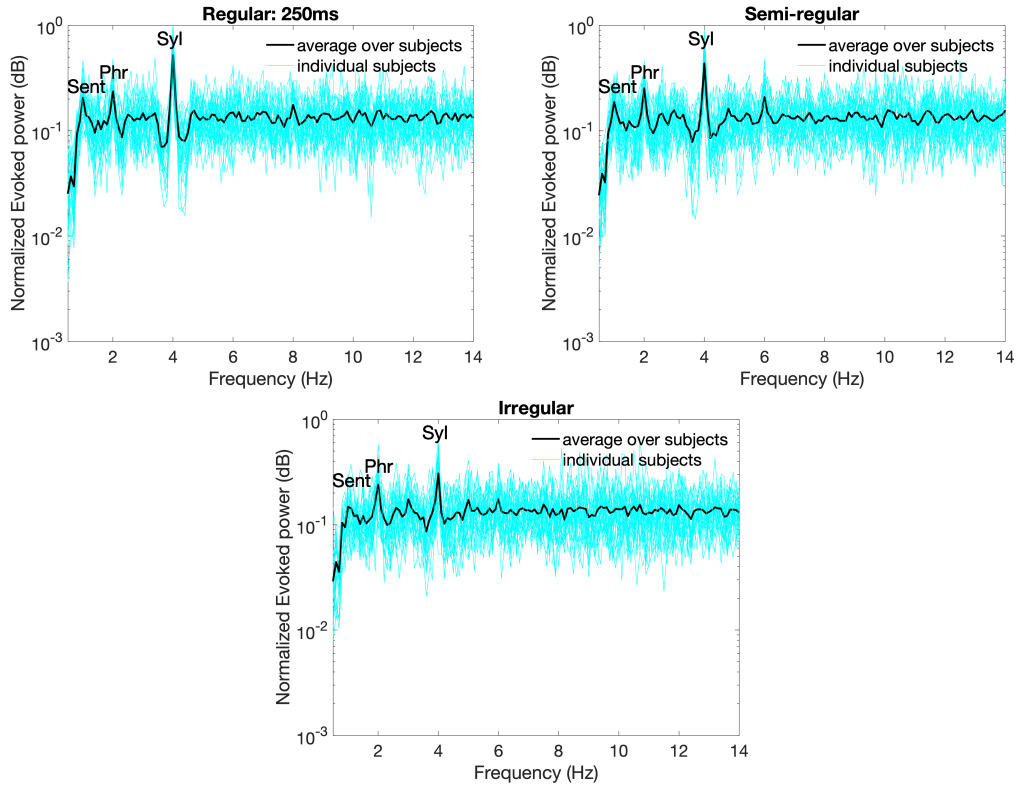


Figure 3.8: Evoked power for conditions with different rhythms.

Peak values for each frequency of interest are plotted in Figure 3.9. Peaks at 7 Hz in the irregular condition serves as a baseline activation. For ITPC, there is also no statistical reliable effect in the comparison of sentence-level peaks ( $p = 0.08$ ,  $\eta^2 = 0.06$ ). There is also no significant differences in the comparison of phrasal-level peaks ( $p = 0.87$ ,  $\eta^2 = 0.003$ ). As for the syllable-level, a significant difference was found ( $p < 0.001$ ,  $\eta^2 = 0.2$ ). Post-hoc pairwise Tukey tests show that no significant difference was found between the regular 250ms/syl and semi-regular condition ( $p < 0.05$ ). A significant difference was found between the regular 250ms/syl and irregular condition ( $p < 0.001$ ) and between the semi-regular and irregular conditions ( $p < 0.05$ ). Overall, the results suggest that delta entrainment can flexibly adjust speech rhythmicity, in line with the behavioral results from Ghitza (2017).

As for EPn, a significant difference was found in the comparison of sentence-level

EPn ( $p < 0.01$ ,  $\eta^2 = 0.1$ ). Post-hoc pairwise Tukey tests show that no significant difference was found between the regular 250ms/syl and the semi-regular condition ( $p = 0.6$ ) and between the semi-regular and irregular conditions ( $p = 0.1$ ). A significant difference was found between the regular 250ms/syl and the irregular condition ( $p < 0.05$ ). For the phrasal-level EPn, no significant difference was found ( $p = 0.85$ ,  $\eta^2 = 0.003$ ). As for the syllable-level EPn, a significant difference was found ( $p < 0.001$ ,  $\eta^2 = 0.19$ ). Post-hoc pairwise Tukey tests show that there is a significant difference between the regular 250ms/syl and the irregular condition ( $p < 0.001$ ) and between the semi-regular and irregular condition ( $p < 0.05$ ). There is no significant difference between the regular 250ms/syl and the semi-regular condition ( $p = 0.17$ ).

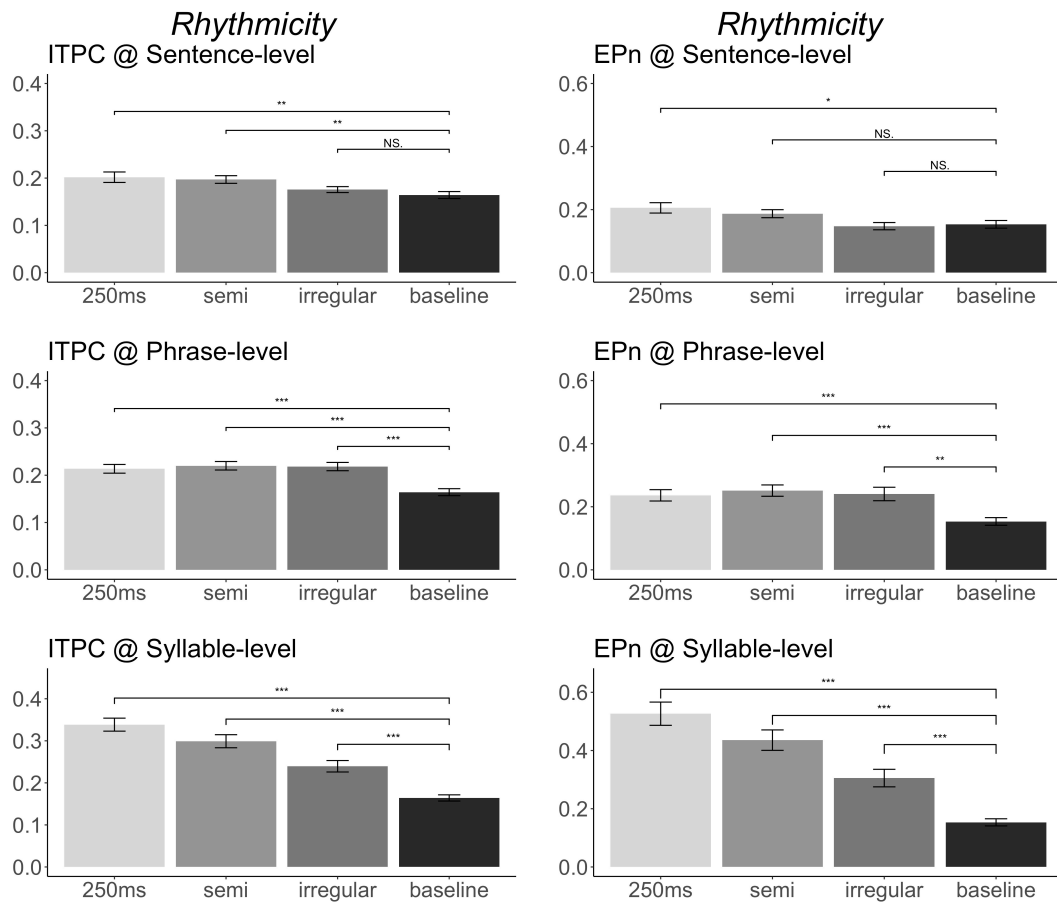


Figure 3.9: ITPC and EPn across conditions with different rhythmicities. Peaks at 7-Hz in the 250/syl condition serves as a baseline activation.

We added a new 75ms/syl condition during the middle of data collation. Here we compared different speech rates with the condition 75ms/syl included for a subset of  $N=17$  subjects. Peak values for ITPC and EPn are shown in Figure 3.10. Peaks at 7 Hz in the 75/syl condition serves as a baseline activation. Overall, the statistical results including 75ms/syl are in line with previous results. For ITPC, no significant difference was found between all conditions at the sentence-level. As for the phrasal-level, post-hoc pairwise Tukey tests show that no significant difference was found between 250ms/syl and 200ms/syl ( $p = 0.99$ ) and between 100ms/syl and 75ms/syl ( $p = 0.99$ ). A significant difference was shown in other paired comparisons (250ms/syl vs. 100ms/syl:  $p < 0.001$ ; 250ms/syl vs. 75ms/syl:  $p < 0.001$ ; 200ms/syl vs. 100ms/syl:  $p < 0.001$ ; 200ms/syl vs. 75ms/syl:  $p < 0.001$ ). As for the syllable-level, post-hoc pairwise Tukey tests show that no significant difference was found between 250ms/syl and 200ms/syl ( $p = 0.99$ ) and between 100ms/syl and 75ms/syl ( $p = 0.99$ ). A significant difference was shown in other paired comparisons (250ms/syl vs. 100ms/syl:  $p < 0.001$ ; 250ms/syl vs. 75ms/syl:  $p < 0.001$ ; 200ms/syl vs. 100ms/syl:  $p < 0.001$ ; 200ms/syl vs. 75ms/syl:  $p < 0.001$ ).

As for EPn, in the sentence-level peaks, post-hoc pairwise Tukey tests show that no significant difference was found between 250/ms and 200/ms ( $p = 0.7$ ) and between 100ms/syl and 75ms/syl ( $p = 0.99$ ). A significant difference was shown in other paired comparisons (250ms/syl vs. 100ms/syl:  $p < 0.001$ ; 250ms/syl vs. 75ms/syl:  $p < 0.001$ ; 200ms/syl vs. 100ms/syl:  $p < 0.001$ ; 200ms/syl vs. 75ms/syl:  $p < 0.001$ ). For the phrase-level, post-hoc pairwise Tukey tests show that no significant difference was found between 250ms/syl and 200ms/syl ( $p = 0.34$ ) and between 100ms/syl and 75ms/syl ( $p = 0.87$ ). A significant difference was shown in other paired comparisons (250ms/syl vs. 100ms/syl:  $p < 0.001$ ; 250ms/syl vs. 75ms/syl:  $p < 0.001$ ;

200ms/syl vs. 100ms/syl:  $p < 0.001$ ; 200ms/syl vs. 75ms/syl:  $p < 0.001$ ). As for the syllable-level, post-hoc pairwise Tukey tests show that no significant difference was found between 250ms/syl and 200ms/syl ( $p = 0.09$ ) and between 100ms/syl and 75ms/syl ( $p = 0.99$ ). A significant difference was shown in other paired comparisons (250ms/syl vs. 100ms/syl:  $p < 0.001$ ; 250ms/syl vs. 75ms/syl:  $p < 0.001$ ; 200ms/syl vs. 100ms/syl:  $p < 0.001$ ; 200ms/syl vs. 75ms/syl:  $p < 0.001$ ).

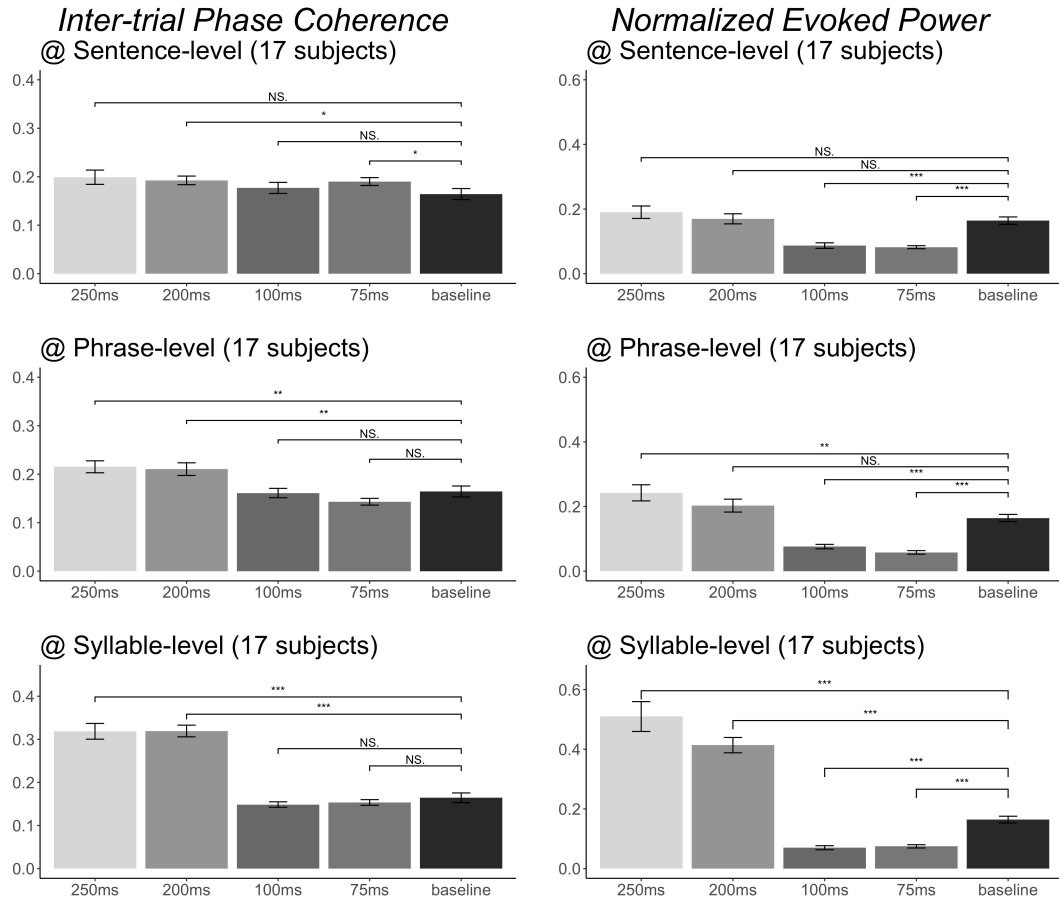


Figure 3.10: The comparison of inter-trial phase coherence and evoked power for conditions with different speech rates (17 subjects included). Peaks at 7-Hz in the 250/syl condition serves as a baseline activation.

### 3.4 Discussion

In this frequency-tagging experiment, we investigate temporal dynamics at different scales with variation of speech rates and regularities. Results have shown

that all conditions have high accuracy in comprehension. As for the results of oscillatory activities, as a control, we see the syllable-level peak at 4 Hz, phrasal-level peak at 2 Hz, and sentence-level peak at 1 Hz for both ITPC and EPn in 250ms/-syl condition. In 200ms/syl condition, we also see robust peaks occurring at 5 Hz for the syllable-level processing, 2.5 Hz for the phrasal-level processing, and 1.25 Hz for the sentence-level processing for ITPC. Robust peaks were also observed in EPn for the phrasal-level and sentence-level processing, but statistical weaker peak for the syllable-level processing, compared to the baseline. As for the 100ms/syl condition, only the sentence-level peak is not statistically different from the above two conditions in ITPC. Same results were also shown in the 75ms/syl condition. However, EPn results show that statistically weaker peaks in all levels of processing in 100ms/syl, compared to the 250ms/syl and 200ms/syl conditions. As for the variation of regularity, ITPC results show that a significant difference was found at the syllable-level processing, but not the phrasal-level and sentence-level processing, suggesting rhythmicity plays a role in syllable tracking, but not influence the higher-level processing. EPn results show that the irregular condition has weaker peak at the sentence-level and the syllable-level processing, but not the phrasal-level processing.

In the current study, high accuracy of comprehension was found among all conditions. Inconsistent with the previous findings (Ahissar et al., 2001; Nourski et al., 2009), high accuracy of comprehension was even found in 75ms/syl conditions (a compression ratio of 0.3). For the 250ms/syl and 200ms/syl conditions, as expected, syllables are tracked well when the frequency of syllable falls in the theta range. In addition, the phase consistency is robust in both the phrase and sentence levels. The 100ms/syl and 75ms/syl conditions show no statistically reliable effect at the

sentence-level, compared to the 250ms/syl and 200ms/syl conditions, suggesting a phase consistency at the sentence-level was also found in both conditions. Taken together the higher accuracy of comprehension and the robust phase consistency in these two conditions, the results provide supporting evidence for the relationship between delta entrainment and the processes of comprehension that involve high-level information.

Consistent with the previous studies (Ding et al., 2014; Kayser et al., 2015; Park et al., 2015), theta and delta entrainments play different role in language comprehension. In the present study, the preservation of delta entrainment and the reduction of theta entrainment suggest different functionalities at different time scales. Since the timing for the completion of a whole sentence in these two conditions falls in the delta range, parsers have time to extract semantic and syntax information and achieve a successful comprehension. On the other hand, the data also shows that this high-information processing relates to the delta range specifically, but not the theta range. In both fast conditions, the frequency of the phrase-level processing 5 Hz and 6.6 Hz respectively, which fall in the theta range. The phase consistency are significantly lower than the ones in the other two conditions. These results support the view from Howard and Poeppel (2010), that theta entrainment reflects the processes of directly encoding acoustic features and syllable parsing, but not processes relating to syntax and semantics.

The results also support the view that delta entrainment reflects a top-down predictive process. Previous fMRI studies have shown that activities from the high-order frontal regions modulate the activities from the lower-order temporal regions when speech is intelligible but distorted (Davis and Johnsrude, 2003; Giraud et al., 2004). However, as pointed out by Davis and Johnsrude (2007), it is difficult to dissociate



the bottom-up processing and the top-down processing and align the neural responses with these processes separately in the time domain since speech unfolds over time. The present study provides evidence from the frequency-domain and suggests that the observed sentence-level phase coherence with the absence of the syllable-level phase coherence indicates a top-down predictive process. From the 100ms/syl and 75ms/syl conditions, the syllables do not appear to lead to neural entrainment but there is still a robust phase consistency found at the sentence-level and the high accuracy of comprehension is shown. Even though losing track in syllables from the bottom-up processing, the top-down processing allows parser to access the high-level information in a longer time window and aid language comprehension. However, as Lewis (2020) points out, it is still unclear when the system knows that it is time to rely more on the top-down or bottom-up information to modulate low-frequency neural activities. The current results suggest that we rely more on top-down information when speech becomes fast. However, whether there is an ongoing change of weighting between the top-down and bottom-up information remains unknown. More fine-grained future studies are needed to test how low-frequency neural responses reflect the balance between the top-down and bottom-up processing.

The high accuracy may reflect the fact that comprehension was measured with an offline task. Participants had at least 3 seconds, up to 10 seconds, to process input before rendering a judgment and they could make the judgment having just processed a few phrases in the trial. In addition, parsers are possible to get used to the faster speech rates by listening multiple times. This would require more future work to rule out this possibility.

Inconsistent with evoked models, our results suggest that delta oscillations elicited from the frequency-tagging paradigm are not responses that merely reflect the rhythms

from the external speech or internally generated punctual operations. In line with the oscillation-based models, theta and delta oscillators are expected to flexibly adapt the speech rates so that each cycle of theta and delta is able to segment discrete syllables or words/phrases from connected speech. Our results have shown that entrainment was found among conditions with different speech regularities. Furthermore, the current results also provide neural evidence supporting the behavioral results from Ghitza (2017). If the chunking rate of a phrase falls in the delta range, comprehenders are able to achieve successful segmentation. This is shown in the 100ms/syl and 75ms/syl conditions, where display the sentence-level phase coherence only but not the phrase and syllable levels of phase coherence.

Consistent with Ghitza (2017), the results show that delta entrainment can flexibly adapt to rhythmicity in speech. As shown in Figure 3.9, no significant difference was found at the sentence and phrase levels, suggesting a robust phase consistency in all conditions with different regularities. This rhythmicity is not reflected on the syllable-level oscillations. As shown in Figure 3.9, when the syllable rhythm became more irregular, the phase consistency (ITPC) became lower gradually. These findings also support that the role for delta entrainment reflects the quality of speech perception Kayser et al. (2015).

### **3.5 Summary of the study**

The current study investigates the temporal dynamics of information integration reflected on low-frequency neural activities by manipulating temporal properties of stimuli. Different speech rates and regularities were manipulated and a frequency-tagging paradigm was used during an EEG recording. High accuracy of comprehension was shown among all conditions. Phase consistency was shown at the sentence-

level in all conditions with different speech rates, suggesting a top-down predictive role for neural activities in the delta band. Furthermore, inconsistent with evoked models, phase consistency across conditions with different regularities was observed and delta entrainment at the sentence-level across conditions with different speech rates were shown, suggesting that the observed neural oscillations reflect endogenous signals, as proposed by the oscillation-based models. Moreover, the results also support the view that delta entrainment reflects a flexible process as it can adapt to different speech rhythms of stimuli. Taken together, these results narrow down how speech rates and regularity can affect language comprehension and how these processes are reflected on low-frequency neural activities.

## CHAPTER IV

# Testing Semantic Compositionality in Low-Frequency Neural Activity

### 4.1 Introduction

Cortical oscillatory activity plays a critical role in transferring and conveying information for multiple fundamental cognitive functions such as memory, attention, and language (Meyer, 2018; Ward, 2003). For language particularly, a growing number of studies have associated different ranges of neural oscillatory activity with different levels of linguistic representation in speech and language processing (see review in Meyer, 2018). Specifically, delta oscillations (0.5-4 Hz) have been linked with the processing of abstract hierarchical structure (Ding et al., 2016), theta oscillations (4-8 Hz) have been associated with syllable processing (Giraud and Poeppel, 2012; Luo and Poeppel, 2007), and the high-frequency ranges such as beta and gamma have been linked with the processing of phonemic feature in speech (Di Liberto et al., 2015; Gross et al., 2013) and predictive coding framework in semantic processing (Lewis and Bastiaansen, 2015; Lewis et al., 2017).

Recent work has associated delta oscillations with the chunking processes of continuous speech and the reflection of hierarchical information processing (Bonhage et al., 2017; Meyer et al., 2016; Meyer and Gumbert, 2018). These studies look at oscillatory power and phase to linguistic stimuli that vary structure and semantics.

For example, Bonhage et al. (2017) report that delta power increases when participants encode German six-word sentence fragments, relative to random word lists, suggesting the role of delta oscillations in an automatic linguistic chunking process. In line with the above finding, Meyer et al. (2016) have shown phase modulation of the delta range in the processing of ambiguous German sentences, regardless of prosodic cues.

In addition to the above studies, there are findings with isochronous speech showing that neural activities in the delta band may entrain to syntactic structure (Ding et al., 2016; Ding et al., 2017). Ding et al. (2016) observed spectral peaks at 1 Hz, 2 Hz, and 4 Hz from a MEG recording when Mandarin native speakers listened to a continuous speech presented at a fixed rate of 4 syllables-per-second. Without physical cues for the two-word phrase and the four-word sentence in the speech stimuli, the neural signals shown at 1 Hz thus reflect the processing of sentence-level while the signals at 2 Hz reflect the phrasal-level processing. However, numerous linguistic processes correlate with hierarchical structure, including syntactic parsing, semantic composition, conceptual processing (Pylkkänen, 2019). It is not yet clear whether, and in what way semantic information engages in modulating neural activities in the delta band. Using isochronous speech, the current study focuses on a variety of factors associated with different aspects of semantic combinatorial processing in order to narrow down the kinds of representations that are associated with low-frequency entrainment to linguistic structure.

#### **4.1.1 Delta oscillations correlate with syntactic and semantic processing**

As mentioned above, neural activities in the delta band may reflect the processing of hierarchical information. However, what high-level information modulates neural activities in the delta band remains unclear. Prior work has shown that delta

oscillations can be associated with the processing of syntactic information. Meyer and Gumbert (2018) show evidence for the role of delta-band oscillations with the processing of syntactic structure. Electrophysiological signals were recorded when participants listened to sentences with morphological violations and then were asked to detect violations. They observed that delta-band oscillations are phase-locked to the syntactic structure of speech stimuli and a correlation between the syntactic surprisal value and the phase is shown in the study, suggesting that the syntactic knowledge for the incoming word facilitates language processing and thus cause the alignment of neural excitability and syntactic information.

In addition, delta oscillations have been associated with combinatorial processes which involve syntax and semantics. Kaufeld et al. (2020) recorded EEG signals when participants listened to naturalistic stimuli. Four types of stimuli were tested in the study: (1) sentence condition which includes structure, lexical semantics, and prosody, (2) jabberwocky condition which includes structure and prosody, but not lexical semantics, (3) word-list condition which includes lexical semantics, but not prosody and structure, and (4) reversed speech as a control. Their results reveal that the mutual information between neural signals and speech stimuli is higher at the phrasal (0.8-1.1 Hz) and word (1.9-2.8 Hz) timescales in the sentence condition, compared to the jabberwocky and word list conditions. The results suggest that combinatorial semantics plays a role in modulation of delta oscillations.

Moreover, delta oscillations have been linked to the reflection of word-level semantics. Frank and Yang (2018) simulated a model which encodes words and syntactic categories as vectors in a high-dimensional semantic space (a “word embedding”). Similar as findings from Ding et al. (2016), they found peaks at 1, 2, 4 Hz in the sentence condition, suggesting that neural activities in the delta band may follow

from the tracking of word-level semantics or the regularities of syntactic categories and are not necessarily interpreted as the cortical entrainment for hierarchical structures. Based on the results, neural activities in the delta band may correlate with lexical semantics of each word to some degree.

From the above, neural activities in the delta band may be modulated by syntactic information, combinatorial semantics, and lexical semantics. However, whether neural activities in the delta band can be modulated by semantic information is still under debate. Mai et al. (2016) conducted an EEG experiment using Mandarin nine-syllable sentences as stimuli. Three conditions were included: real-word (as shown in 1), pseudo-word (as shown in 2), and backward condition of real-word or pseudo-word. In order to prevent participants making prediction, the sentences in real-word condition were actually semantically-anomalous but syntactically-acceptable. For the pseudo-word condition, two adjacent syllables did not form a meaningful word (i.e. word list). They did not find increased delta power in the real-word condition as compared to the pseudo-word condition, inconsistent with the results from Ding et al. (2016). They only found that increased delta power in the real-word and the pseudo-word conditions, relative to backward condition. They suggest that the delta frequencies track the phonetic features and involve in phonological processing, but not semantic processing. However, inconsistent with the above finding, Mai and Wang (2019) re-analyzed the data from Mai et al. (2016) and had a different interpretation for delta oscillations. While Mai et al. (2016) computed normalized EEG power after the Hilbert transform, Mai and Wang (2019) adopted multivariate temporal response functions (mTRF) to directly map multiple bands of speech envelopes to EEG signals. They found that the entrainment in the delta band was larger in the pseudo-word condition than the real-word and backward condition. Additionally,

the absolute weighting of temporal properties for mTRF, which reflects the degrees of neural entrainment across time series, was greater for the theta band in the early time lags (0-160 ms), compared to the late time lags (160-300 ms). Contrast to the the results of the theta band, the delta shows no significant difference between the early and late time series, suggesting that delta oscillations involve both phonological and high-level processing.

(1) *wanglu xihuan jianjiang DE kongqi*  
 internet enjoy tough DE air

(2) *shu xi sheng yu shu fei DE shi shen*  
 book learn born universe tree fly DE vision body

Aligned with the above findings with non-entrainment approach, as mentioned already, other studies have shown that neural signals appear to be entrained to abstract linguistic structures when listening to isochronous speech (Ding et al., 2016; Ding et al., 2017). Martin and Doumas (2017) propose a model to account for the findings in Ding et al. (2016). The model adopts a time-based binding mechanism, which demonstrates the fully parsing steps and allows the representations to be bound by using asynchrony of firing units and thus having different levels of representations without losing the information for each representation. The model has different layers representing different levels of a hierarchical structure. Take an adjective phrase “dry fur” for example. The model encodes semantic features for each word at the lowest layer, word information such as [dry *adj*] and [fur *noun*] is encoded in the second layer and they fire at an asynchrony. Then the layer above encodes phrase information and will be activated after [dry *adj*] and [fur *noun*] fire. A simulation was conducted by using the English stimuli from Ding et al. (2016). The simulation results show



that the grammatical condition (e.g. dry fur rub skin) elicits spectral peaks at 1 Hz, 2 Hz, and 4 Hz, consistent with the experimental results from Ding et al. (2016). They further show that the word list condition (i.e. no syntactic relationship between words) shows the 4-Hz peak only. Similar to the grammatical condition, the jabberwocky condition (i.e., Syntactic relationship between words is preserved but no plausible semantic composition is possible.) shows the power increases at 1 Hz, 2 Hz, and 4 Hz, suggesting that the hierarchical structures were activated. Their results suggest that neural activities in the delta band correlate with syntactic and semantic composition. Moreover, in our Chapter II, by using isochronous speech, we observed the 4-Hz peak only when the speech stimuli are semantically-anomalous but syntactically-preserved, suggesting that semantic information indeed plays a role in delta oscillations to some degree.

From the above, studies from both non-entrainment and entrainment approaches support the role of neural activities in the delta band in hierarchical processing. However, this hierarchical processing can be related to either syntactic or semantic information. Even from the very basic two-word composition, prior work has shown that the combinatorial processes indeed involve a semantic component (Bemis and Pylkkänen, 2011; Bemis and Pylkkänen, 2013a; Bemis and Pylkkänen, 2013b). Bemis and Pylkkänen (2011) examined the processing of two-word simple phrases by comparing the neural activities between the phrase (e.g. red boat), the non-phrase (e.g. xkq boat and the list condition (e.g. cup boat) during an MEG recording. An increased activity was found in the noun at roughly 250 ms in the anterior temporal lobe (ATL) for the phrase condition. Moreover, as the studies with non-entrainment approach show (e.g. Kaufeld et al., 2020), results support either direction and thus it remains unclear what high-level information is reflected in the modulation of delta

oscillations.

#### 4.1.2 Semantic variables in semantic processing

Semantic information can affect hierarchical processing in various ways. Specifically, predictability, which indicates the predicted lexico-semantic activation, and plausibility, which indexes post-semantic result of semantic composition, are associated with the ERP component N400 in previous work (Nieuwland et al., 2019). Semantic similarity, which indexes lexical activation, is also a key factor examined in semantic processing. Semantic specificity, which is used to index conceptual combinatorics, has been associated with basic composition in the ATL. Taken together, the present study probes which, if any, of semantic variables might modulate delta oscillations. The following reviews the above semantic variables that may involve in hierarchical processing.

Low-frequency oscillations can be contributed by slow event-related responses such as the P300, N400, and P600 (Park et al., 2015, Zhou et al., 2016). For example, Park et al. (2015) found that the top-down information, which was computed from the difference between the transfer entropy (i.e., the directed statistical dependencies between two signals) in story and backward condition, modulate the phase of low-frequency oscillations in the left auditory cortex, consistent with the finding that the context-specific N400 observed in the left hemisphere (Federmeier, 2007). Additionally, Park et al. (2015) point out that the N400 originates from the delta and theta oscillations and they are all localized in the left hemisphere. Previous studies have demonstrated that various semantic variables modulate the N400, which is a negative-going ERP component occurring centro-parietally within roughly 250-500 ms post-stimulus and sensitive to semantic processing (Kutas and Hillyard, 1980, 1984). Specifically, the N400 can be modulated by word predictability, which

indicates the statistical association between words. Word predictability can be formalized by different methodologies. For example, it can be obtained from cloze probability and acquired from an offline fill-in-blank task. Words with high cloze probability means that a word is highly-predicted based on its sentence context. Kutas and Hillyard (1984) have shown that the amplitude of the N400 is negatively correlated to the cloze probability. That is, less-expected words show a greater N400 than highly-expected words. Word predictability can also be obtained from language models and may be characterized as surprisal (Hale, 2001), which indicates the amount of information conveyed by a word  $w$  in a context  $C$ , as formulated in (3). Words with high surprisal values indicate that words are less-predicted as they convey more information. Prior work such as Frank et al. (2015) has shown that word surprisal correlates with the amplitude size of the N400, providing supporting evidence for the relationship between word predictability and the N400.

$$(3) \quad \textit{surprisal}(w) = -\log_2(p(w|C))$$

Several studies have shown that word predictability modulates high-frequency neural activities in the beta and gamma ranges (Lewis and Bastiaansen, 2015; Molinaro et al., 2013; Wang et al., 2012a; Wang et al., 2012b). For example, Wang and colleagues (2012a) recorded brain’s magnetic fields when participants listened to sentences including either congruent ending or incongruent ending. Their results have shown a positive linear relationship between the N400m and the beta power (16-19 Hz). That is, decreased beta power was found in the sentences with incongruent ending and also a N400m component was shown in that condition, compared to the sentences with congruent ending. In addition, Wang et al. (2012b) further show that the increased gamma power (40-50 Hz) was found in the sentences that are both

congruent and highly predictable, relative to the sentences that are congruent but less predictable and the sentences that are incongruent and unpredictable. The results suggest the gamma power change relates to the predictability of the upcoming word. Contrary to the high-frequency neural activities, low-frequency neural activities are not sensitive to predictability. Ding et al. (2016) manipulated transitional probability between words. Words with equal transitional probability and varying transitional probability are included. They found that neural tracking of sentence is shown in both conditions, suggesting that low-frequency neural activities are not affected by predictability.

In addition to word-level processing, sentence plausibility, which indicates the consequence of semantic composition and world knowledge, is also considered as a critical factor for the modulation of the N400 (Nieuwland et al., 2019). The plausibility of sentence can be acquired by asking people to judge whether the sentence is plausible or not. Some studies have shown that less plausible words are difficult to be integrated in world knowledge and context as compared to highly-plausible words (Hagoort et al., 2004; van Berkum et al., 1999). Moreover, Nieuwland et al. (2019) found that the N400 is correlated to word predictability at roughly 200-500 ms post-stimulus with the peak occurring at about 330 ms. As for plausibility, the N400 is elicited at approximately 200-650 ms post-stimulus with the peak at about 350 ms. The association of the N400 and two different processing timings for word predictability and plausibility suggest that the N400 reflects prediction from beginning and integration afterwards, which supports the view of multiple processes occurring in the N400 time window. In other words, people tend to form predictions by pre-activating the upcoming word based on the semantic features or previous context and then integrate the word into context and world knowledge in the sentence-level.

A growing number of studies also show that plausibility relates to high-frequency neural activities (Hald et al., 2006; Hagoort et al., 2004). For example, Hagoort et al. (2004) relates gamma power changes to high integration demands. An increased gamma power was shown in the sentences that violates world-knowledge but still semantically congruent (e.g. The Dutch train are WHITE and very crowded), compared to the sentences that are semantically congruent (e.g. The Dutch train are YELLOW and very crowded) and the sentences that are semantically incongruent (e.g. The Dutch train are SOUR and very crowded). However, inconsistent with the above findings, Hald et al. (2006) found that an increased gamma power change was found in the semantically-correct sentences (e.g. The Dutch trains are YELLOW and blue.), relative to the semantically-incongruent sentences (e.g. The Dutch trains are SOUR and blue.). These studies show that sentence plausibility might play a role in modulating oscillations but it remains unclear whether these oscillations could be robustly modulated by sentence plausibility.

In addition, semantic similarity between words, which indicates lexical-level semantic relationship between words, can affect semantic processing. Semantic similarity can be computed from using latent semantic analysis, which extracts and represents meaning by using vectors to indicate semantic distance in a large corpus (Landauer and Dumais, 1997) or word2vec, which trains a neural network to predict words from a context (Mikolov et al., 2013). Previous studies have shown that semantic similarity can modulate semantic priming effects in naming tasks (Jones et al., 2006) and lexical decision tasks (Günther et al., 2016). Several studies have shown that the modulation of the N400 correlates to semantic similarity. For example, Ettinger et al. (2016) computed semantic distance between the critical word and its previous context for the stimuli in Federmeier and Kutas (1999) and found

the correlation between the N400 and the semantic distances. Frank and Willems (2017) also found that the N400 is affected by both surprisal model and semantic similarity model. Both models show the similar onset and offset of the effect and also the similar effect size. However, inconsistent with the above findings, Nieuwland et al. (2019) did not find that the semantic similarity affects the modulation of the N400.

A recent study also suggests that the semantic relatedness does not modulate low-frequency neural activities. Jin et al. (2020) tests the association between semantic similarity and chunking structure by using MEG recording. During the experiment, participants were asked to chunk a sequence of words into 2-word chunks and find whether there is an invalid chunk in a trial. Stimuli are either living things (L) or non-living things (N). Two conditions were included: the same-category condition consists of LL and NN as valid chunks while the different-category condition consists of NL and LN as valid chunks. The two conditions were presented in separate blocks. Both conditions was either presented in a regular order or in a random order. Three simulation models were built based on different hypotheses. The lexical property model was built by encoding word-level information only. The semantic relatedness model was built based on semantic similarity between words but no further information for syntactic structures. The rule-based chunking model assumes a constant change of neural responses within a chunk and predicts different neural trackings for different chunk rules. The MEG results show that the significant increased power at the chunk rate, which is consistent with the predictions from the semantic relatedness model and rule-based model, regardless of the alternating order or the random order of the stimuli presentation. Moreover, consistent with the simulation from the rule-based model, the phase difference between conditions is closer to 0 while

the semantic relatedness model predicts a 180 phase difference. Their results show that neural activities in the delta band track the chunking rules, rather than word properties.

Another approach looks into semantic processing by examining the neural basis of semantic composition and conceptual combination in simple phrases (Bemis and Pylkkänen, 2011; Westerlund and Pylkkänen, 2014; Zhang and Pylkkänen, 2015). For example, Westerlund and Pylkkänen (2014) conducted a MEG experiment to compare the conditions where the general nouns (e.g. boat) and the specific nouns (e.g. canoe) are modified by an adjective (e.g. blue). In addition, they included the conditions that block the composition by having consonant clusters (e.g. qktz) in the modifier position. They found that the general noun with the modifier elicited greater activity than the specific noun with the modifier at 221-264 ms. Also, they found that the specific noun elicited greater activity than general noun in the non-combinatorial conditions. Their results suggest that the left anterior temporal lobe (LATL) involves both the processing of composition and also the processing of the conceptual specificity. Crucially, they demonstrated that the conceptual specificity involves in the modulation of neural activities at very early processing stage. Based on the findings, neural activity can be modulated by conceptual specificity during semantic processing.

To sum up, semantic processing involves several different components such as pre-activating an upcoming word, composing two elements to form meaning, and integrating the word into sentences as sentence unfolds. How to integrate different semantic components to achieve successful sentence processing and how this process reflects on the neural oscillations remains unclear. The current study examines the above four factors (i.e., predictability, plausibility, semantic similarity, and concep-

tual specificity) to see how these factors modulate the neural activities in the delta band and provide a link between the slow-wave event-related response N400 and the low-frequency neural responses.

## 4.2 Experiment

The goal of the current experiment is to examine what aspects of semantic information modulate neural activities in the delta band. The current experiment tests the following semantic variables: (i) predictability, which denotes statistical association between words; (ii) plausibility, which represents the consequence of semantic composition and world knowledge; (iii) specificity, which reflects the outcome of composition itself; and (iv) semantic similarity, which represents the basic semantic relationship based on the co-occurrence of words within a context. It should be noted that the signals measured in the current study reflect low-frequency neural activities. One hypothesis is that the measurements reflect neural oscillations. However, the current experiment does not allow us to conclude that the results can be directly mapped to neural oscillations.

### 4.2.1 Participants

Twenty-one native speakers (10 females, 11 males) of Mandarin Chinese between the age of 18 and 31 (mean = 24) participated in the experiment. They were all right-handed and had normal hearing. They self-reported that they did not have any neurological disorders. They gave informed consent and were reimbursed for their time (15 USD/hour). Data from nine participants were excluded from the analysis due to having many noisy trials (>40 %) and one participant was excluded due to the recording error. Thus, data from 11 participants (6 female, 5 males) were included in the final analysis. Due to COVID-19, we report the intermediate results



here.

#### 4.2.2 Materials

Eight-syllable Chinese sentences were constructed to cross two variables: plausibility and specificity, as described in detail below. Then subsequently, the stimuli were assessed on additional two variables: predictability and semantic similarity. The sentence sequence is bisyllabic subject + bisyllabic transitive verb + bisyllabic modifier + bisyllabic object. Table 4.1 shows an example of the stimuli. The whole stimuli were shown in the Appendix B.1. Sixty sentence sets were created and fifty sentence sets were selected for the main experiment based on the norming results. During the norming process, participants were asked to judge whether the sentence is plausible or not. 24 native Mandarin speakers were recruited on-line for the norming test. During the main experiment, only four-syllable noun phrases (i.e. bisyllabic modifier + bisyllabic object) were used to elicit stronger neural signals.

Table 4.1: Example stimuli for the norming process.

+specific, +plausible	爸爸 摘了 新鲜 蕃茄 baba zhai-le xinxian fanqie Father pluck-ASP fresh tomato 'The father plucked fresh tomato.'
-specific, +plausible	爸爸 摘了 新鲜 蔬菜 baba zhai-le xinxian shucai Father pluck-ASP fresh vegetable 'The father plucked fresh vegetables.'
+specific, -plausible	爸爸 摘了 初生 蕃茄 baba zhai-le chusheng fanqie Father pluck-ASP new-born tomato 'The father plucked new-born tomato.'
-specific, -plausible	爸爸 摘了 初生 蔬菜 baba zhai-le chusheng shucai Father pluck-ASP new-born vegetable 'The father plucked new-born vegetables.'

Conceptual specificity was controlled by manipulating the object (e.g. tomato vs. vegetable). Specificity was defined relatively by using E-HowNet (<http://ehow.net>.

iis.sinica.edu.tw/index.php), which provides semantic relations between Mandarin words. An example is shown in 4.1. In E-HowNet, the items marked in blue (e.g. vegetable) is the main category, which includes several subcategories (marked in pink) such as celery, potato, onion, etc. Thus a non-specific noun would be “vegetable” and a specific noun would be one of the subcategories.



Figure 4.1: An example for word relations in E-HowNet.

Plausibility was defined by manipulating the modifier (e.g. fresh vs. newly-appointed). The implausible stimuli was created by replacing the modifier in the plausible sentence with an implausible modifier with a similar meaning based on E-HowNet. For example, the adjective *xin ren* ‘newly-appointed’ in the implausible item was created by replacing the second syllable from the adjective in the plausible item *xin xian* ‘fresh’. Then we also made sure that the implausible item (e.g. *xin ren fan qie* ‘newly-appointed tomato’) is impossible using a google search.

The results of plausibility (i.e., the offline judgment results) for each of the four conditions are shown in Figure 4.2. A two-way ANOVA shows that there is a statistically significant difference in plausibility ( $p < 0.001$ ) and in specificity ( $p < 0.05$ ). There is no significance in interaction effects ( $p = 0.61$ ). As expected, highly

plausible sentences were rated higher than less plausible sentences. Additionally, non-specific sentences were rated higher than specific sentences.

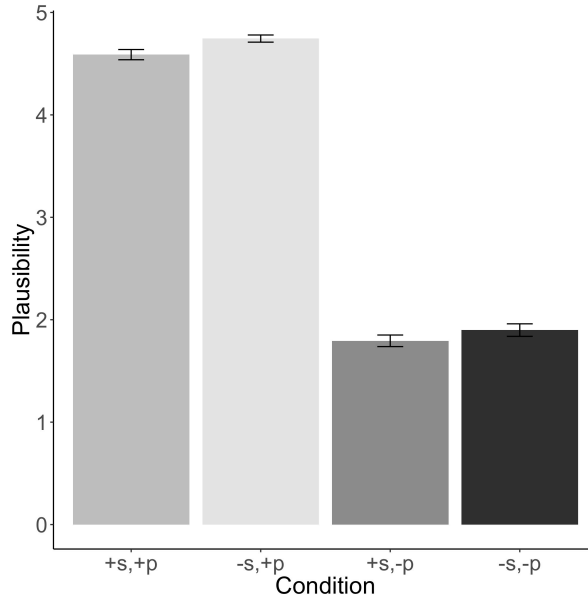


Figure 4.2: Results of plausibility and specificity

Predictability was computed quantified in two separate ways. The first is from Google Chinese BERT model (Devlin et al., 2019). We obtained the values by masking the last character of the stimuli, as shown in Algorithm 4.1. As BERT was trained bidirectional, we computed the value for the last syllable to avoid the bidirectional prediction. An example is shown in Table 4.2. Additionally, we computed word surprisal values from a Chinese GPT-2 model (Du, 2019), as shown in Algorithm 4.2. GPT-2 model was not trained for any specific tasks and it performs well in several language tasks such as generating text or answering questions. We obtained the values for the third syllable in the stimuli as it is the first syllable of the upcoming noun.

$$Predictability = P(\text{last word} | \text{BERT left context}) \quad (4.1)$$

Table 4.2: An example for extracting the value from Google BERT

Example stimuli	Predictability from BERT
xin-xian fan _____ : “qie”	0.94
xin-xian shu _____ : “cai”	0.16
xin-ren fan _____ : “qie”	0
xin-ren shu _____ : “cai”	0.85

$$Surprisal(w) = -\log_2(P(\text{third word}|\text{first two-word context})) \quad (4.2)$$

The results of Predictability are shown in Figure 4.3. There is a significant difference in plausibility ( $p < 0.001$ ). No significance was found in specificity ( $p = 0.07$ ) and interaction effect ( $p = 0.48$ ). Predictability was higher in the plausible sentences than the implausible sentences.

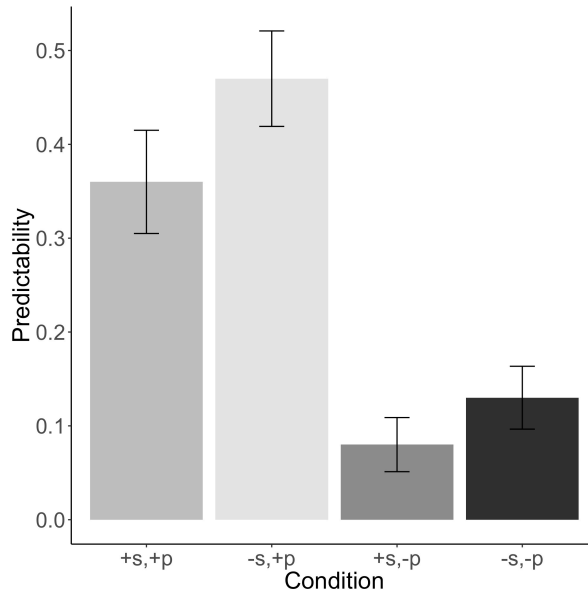


Figure 4.3: Results of predictability

The results of word surprisal are shown in Figure 4.4. A statistically reliable effect was found in plausibility ( $p < 0.001$ ). In addition, there is a significant difference in specificity ( $p < 0.05$ ). There is no interaction effect ( $p = 0.12$ ). As expected, surprisal values in the implausible sentences were higher than the plausible sentences.

In addition, surprisal values were higher in the specific items than the non-specific ones.

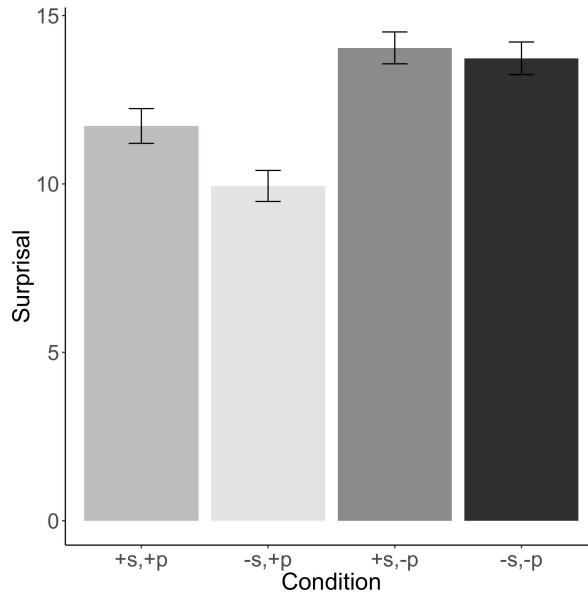


Figure 4.4: Results of word surprisal

Semantic similarity was obtained from computing cosine similarity by extracting word embeddings from Wikipedia2vec (Yamada et al., 2020). An example is shown in Table 4.3. The similarity between the adjective *xin xian* ‘fresh’ and the both noun *fan qie* ‘tomato’ and *shu cai* ‘vegetable’ are equally higher than the similarity between *xin ren* ‘newly-appointed’ and the both noun.

Table 4.3: An example for semantic similarity

Example stimuli	Semantic similarity
xin-xian fan-qie fresh tomato	0.41
xin-xian shu-cai fresh vegetable	0.43
xin-ren fan-qie newborn tomato	0.08
xin-ren shu-cai newborn vegetable	0.12

The results of semantic similarity are shown in Figure 4.5. A statistical significance was found in plausibility ( $p < 0.001$ ). No statistically significant difference

was found in specificity ( $p = 0.19$ ) and interaction effect ( $p = 0.35$ ). As expected, semantic similarity differed across the four conditions; the highly plausible conditions showed higher values for similarity than the less plausible conditions.

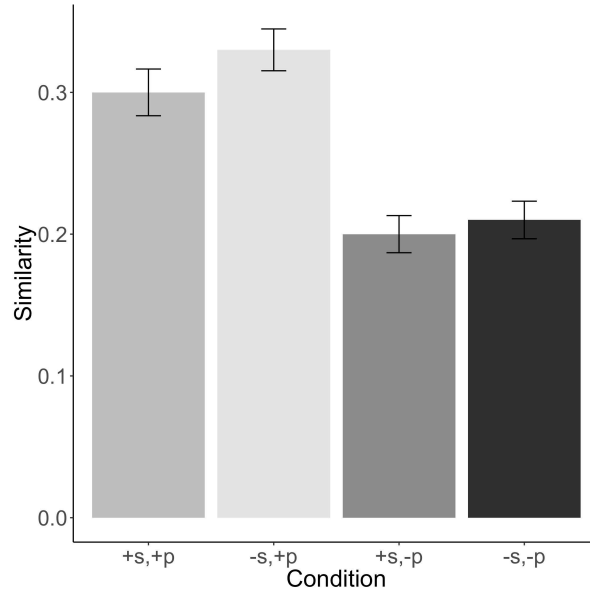


Figure 4.5: Results of semantic similarity

From the above, we see high correlation between predictability, plausibility, and semantic similarity, as they all show higher values for the plausible items. In addition, non-specific items have higher values than the specific items. As for predictability and word surprisal, the results extracted from two language models and different syllables also match for the stimuli.

All stimuli were recorded from the website developed by iFLYTEk ([https://www.xfyun.cn/services/online\\_tts](https://www.xfyun.cn/services/online_tts)). Each syllable was recorded separately, and the duration of each syllable was around 250 ms. Then each syllable was compressed to 240 ms and a 10-ms silence gap was added after each syllable by using Praat (Boersma and Weenink, 2019). Thus, the whole duration of each syllable is 250 ms. Therefore, a four-syllable item was 1-second long and a trial that included ten four-syllable items was 10-second long. Eight blocks were included in the current experiment.

Each block included 20 plausible and 20 implausible trials. Thus, 320 trials were used in the whole experiment.

#### **4.2.3 Procedure**

Participants sat comfortably in front of a computer screen in a quiet room. Prior to the main session, participants were fitted with an electrode cap. Electrodes were also affixed above and below the left eye. Electrolyte gel was applied to minimize impedance. The setup took approximately 30 minutes. Then the stimulus volume was set for each ear based on hearing level of each participant with the threshold of 45 dB determined using 1000 Hz tones (300 ms length). After setting up the volume, a pre-test presenting 120 1,000 Hz tones was conducted to ensure the data quality was sufficient to run the main session.

During the main session, participants was instructed to judge if they heard the sentence item appeared on the screen by button press. After the button press, the next trial was played after a delay randomized between 800 - 1,400 ms. Stimuli was presented with Psychopy2 (v1.84.2; Peirce, 2007, 2009). Participants was also instructed not to blink their eyes or move their body a lot when the stimuli were presented. Participants can take breaks between blocks. After the instructions, participants had 8 practice trials to familiarize with the whole procedure of the experiment. The plausible trials and the implausible trials were intermixed and presented randomly in each block. The order of blocks was counterbalanced across participants. The main experiment took about 1.5 hours to finish. After the main session, participants washed their hair to remove the electrolyte gel and was debriefed about the goals of the experiment.

#### 4.2.4 EEG recording and data analysis

EEG data was recorded at 500 Hz from 32 active electrodes (actiCHamp, Brain-Products GMBH) in a 0.01-200 Hz band with online reference to an electrode placed on the right mastoid. Impedance will be kept below 25 kOhms. FieldTrip software (Oostenveld et al., 2011) was used to analyze the data. Artifacts related to eye blinks were removed via Independent Component Analysis (Jung et al., 2000; Makeig et al., 1995) and remaining artifacts were removed manually following visual inspection. The first sentence from each trial was excluded to avoid potential EEG responses to sound onset (Ding et al., 2017). Data were filtered from 0.1-25 Hz, and re-referenced offline to a common average. For each condition, we computed evoked power and inter-trial phase coherence from 0.5 to 10 Hz in increments of 0.111 Hz since the whole trial after removing the first sentence is 9-second long and therefore the frequency resolution of the DFT is 0.111 Hz. Inter-trial phase coherence and evoked power were computed following Ding et al. (2017), as listed and defined in (Equation 4.3-4.4). To remove the trend of  $1/f$  in the power spectra and test the statistical reliability, evoked power was normalized via Equation 4.5 used in Ding et al. (2017). The evoked power at each frequency was normalized by the neighboring frequency bins within  $\pm 0.5$  Hz. For statistical analysis, conditions were compared via two-way ANOVA for each measure at each frequency of interest.

(2) **Inter-trial phase coherence (ITPC)** reflects phase coherence across trials. The summation of cosine and sine values of phase angle  $\theta_k$  of each complex-value Fourier coefficient is computed and then the square root of the summation is averaged over the total number of trials  $K$ . Higher ITPC indicates that phase angles are consistent across trials.



$$R(f) = \frac{\sqrt{(\sum_k(\cos\theta_k))^2 + (\sum_k(\sin\theta_k))^2}}{K} \quad (4.3)$$

(3) **Evoked power** (EP) reflects the power of EEG responses that are phase-locked and time-locked to speech stimuli. The Discrete Fourier Transform (DFT) of the response in trial  $k$  is denoted as  $X_k(f)$  and  $X_k(f)$  is a complex-value Fourier coefficient. Thus, evoked power is the summation of complex-value Fourier coefficient of trials averaged over the total number of trials  $K$ .

$$E(f) = \frac{1}{k} \left| \sum_k X_k(f) \right|^2 \quad (4.4)$$

(4) **Normalized evoked power** (EPn)

$$En(f) = \frac{E(f)}{\sum_w E(w)}, |w - f| < 0.5 \text{ Hz}, w \neq f \quad (4.5)$$

$w$  represents the neighboring frequency around the target frequency  $f$ .

#### 4.2.5 Predictions

As plausibility typically correlates with predictability and semantic similarity, if one of these three variables modulates delta oscillations, we should expect a main effect of plausibility based on the experimental design. As for specificity, if delta oscillations can be modulated by specificity, we should see a positive correlation between the EPn/ITPC and specificity.

### 4.3 Results

Figure 4.6 and Figure 4.8 show the ITPC and EPn results for each condition. From both ITPC and EPn, peaks are shown at 4 Hz, 2 Hz, and 1 Hz for every condition. The target frequency that tags the composition of the whole phrase is 1 Hz. The results of ITPC and EPn at 1 Hz are shown in Figure 4.7 and Figure 4.9. A two-way

ANOVA was conducted for these two measures. For ITPC, no statistical significance was found in plausibility ( $p = 0.79$ ), specificity ( $p = 0.15$ ), nor for their interaction ( $p = 0.33$ ). As for EPn, there is also no statistical significance in plausibility ( $p = 0.57$ ), specificity ( $p = 0.27$ ), nor for their interaction ( $p = 0.36$ ).

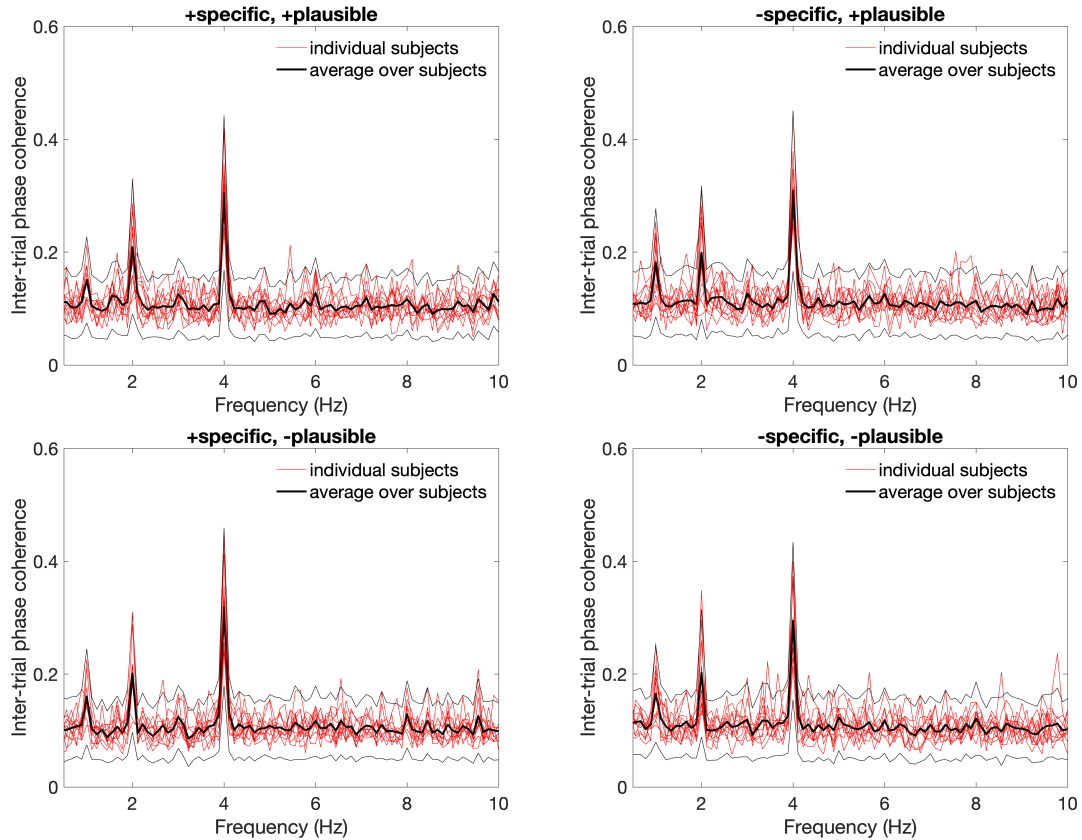


Figure 4.6: ITPC for each condition

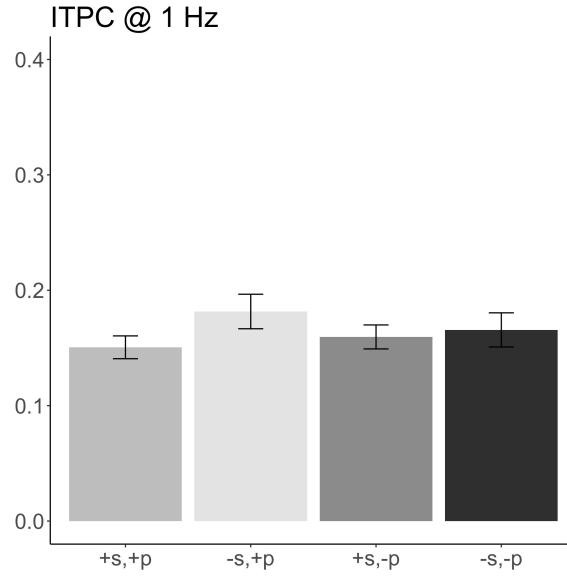


Figure 4.7: ITPC at 1 Hz

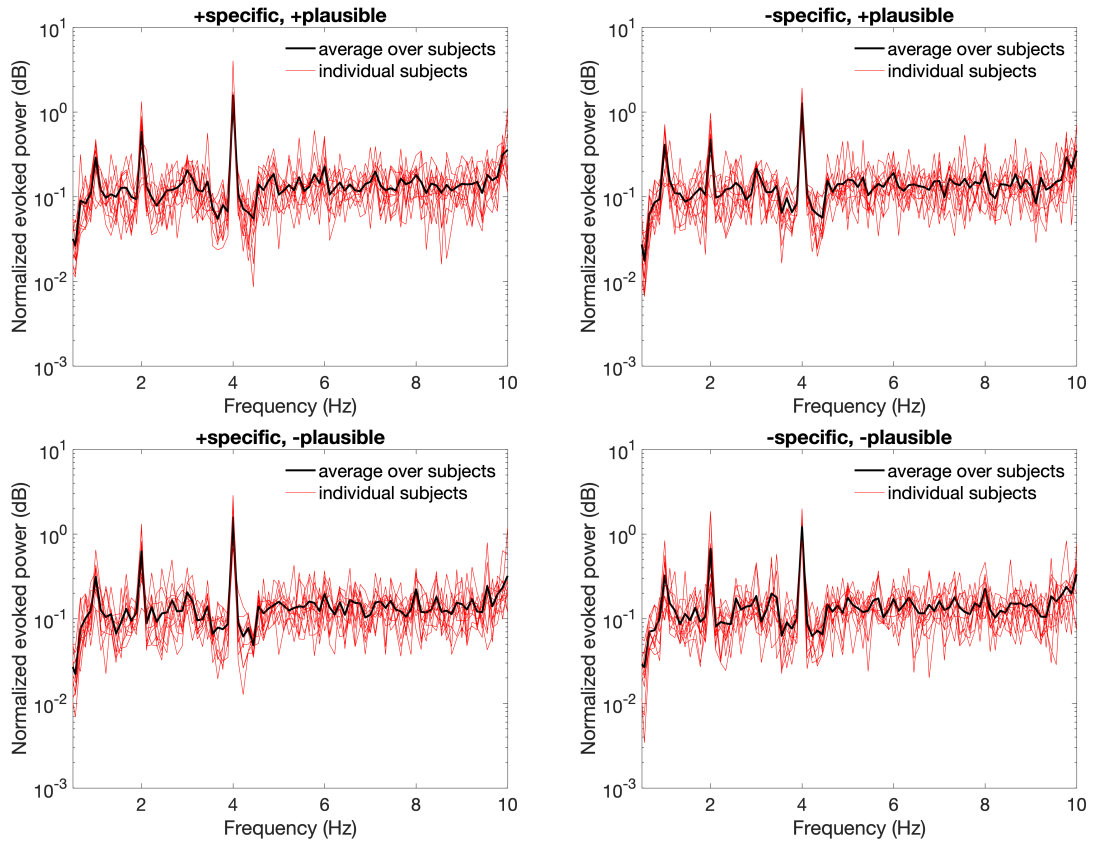


Figure 4.8: EPn for each condition

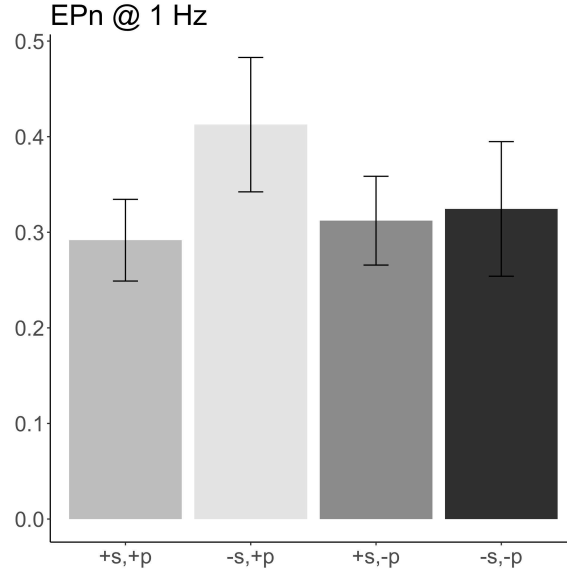


Figure 4.9: EPn at 1 Hz

#### 4.4 Discussion

The current experiment tests different semantic variables to examine whether neural activities in the delta band are driven by semantic processing. The results have shown that no statistical difference is found in ITPC and EPn between conditions. In addition, the results for plausibility, predictability and semantic similarity show that a significant difference in plausibility and no difference in specificity. As these three predictors are highly correlated, the ITPC and EPn do not show difference between conditions, suggesting that plausibility, predictability, and semantic similarity do not modulate neural activities in the delta band. Although there is no statistical reliable effect for the partial  $N = 11$  data, we observed a trend towards increased synchronization for non-specific, compared to the specific conditions.

While keeping the structure consistent across conditions, the results of ITPC and EPn show peaks at 1 Hz across all conditions and do not reflect the results of predictability, plausibility, and semantic similarity. We see no evidence that

delta oscillations are modulated by combinatorial semantics and lexical semantics. There is a trend showing that delta oscillations may be modulated by conceptual specificity. Inconsistent with the current findings, in the Chapter II, the absence of phrasal-level and sentence-level peaks in the semantically-mismatched condition and the appearance of 1, 2, and 4 Hz peaks in the normal four-syllable sentences suggest combinatorial semantics plays a role in modulating delta oscillations to some degree. If neural activities in the delta band could be modulated by pure syntactic information, we would expect that peaks at the phrasal and sentence level should be elicited in the semantically-mismatched condition of that earlier study, which preserved the syntactic information without semantics. This combinatorial processes may interact with temporal properties such as word length in the stimuli, as detailed below. Taken together, the current study suggest that delta oscillations reflect syntactic information more directly than semantic information.

Peaks at 1 Hz across all conditions do not reflect the pattern of predictability. Values for predictability were obtained from Google Chinese BERT and word surprisal was obtained from a Chinese GPT-2. Based on the current results, we see no evidence that neural activities in the delta band reflect predictability and we suggest that prediction that modulates neural activities in the delta band may come from syntactic prediction, but not semantic prediction. The current results are indeed consistent with the findings for transitional probability in Ding et al. (2016). They also observed sentence-level peaks in both equal and varying probability conditions. However, results in our Chapter III suggest that neural activities in the delta band reflect a top-down predictive processing. What parsers exactly predict and thus modulate neural activities in the delta band remain elusive. In addition, the processing between semantics and syntax does not have a clear cut. For example, the lexi-

cal category is syntactic information but it is realized via semantic interpretation. More fine-grained manipulations may be required in the future study. Moreover, predictability and word surprisal computed in the current study were obtained from the output layer of the language models. It is possible that low-frequency neural responses may reflect the internal or intermediate stage of predictability. One future direction would be obtaining predictability and word surprisal in the internal layers of the language models and then do FFT to examine whether predictability or word surprisal oscillate in the stimuli.

For plausibility, we also did not see that peaks at 1 Hz reflect the pattern of plausibility. However, the results in our Chapter II show peaks at 1, 2, 4 Hz in the normal four-syllable condition, but only 4-Hz peak in the semantically-mismatched condition. The above seems to suggest that semantics should play a role in modulating neural activities in the delta band. The inconsistency of results in the current study and the Chapter II might due to the fact that the length for a word/phrase should be taken into account. To be specific, in the Chapter II, each syllable represents one word and thus each word lasts only 250 ms, which falls in the theta band. However, in the current study, each word consist of two syllables, which falls in the delta band. Thus, parsers might have enough time to access the meaning of a word. The current results are still in line with our previous findings, suggesting that neural activities in the delta band correlate with comprehension above syllable-level processing.

For semantic similarity, we also see no evidence that peaks at 1-Hz from neural signals reflect the results of semantic similarity. In line with Nieuwland et al. (2019) and Jin et al. (2020), the current study shows that semantic relatedness does not modulate neural activities in the delta band. The current results are inconsistent with the simulation from Frank and Yang (2018), which suggests that neural activities in

the delta band may reflect the tracking of word information. The current results and findings from the Chapter II may suggest that neural activities in the delta band do not reflect semantics at word-level processing.

As for specificity, we see that peak values at 1 Hz in the non-specific conditions are higher than the values in the specific conditions, although this is not statistically significant. The intermediate results suggest that the compositionality reflect conceptual semantics might play a role in modulating neural activities in the delta band. However, there are many other ways to manipulate conceptual specificity. More future studies are needed for adopting various semantic stimuli to test whether conceptual semantics modulate the delta oscillations. In the current experiment, conceptual specificity was manipulated by the noun only. It would be interesting to create stimuli by manipulating specificity for the first and second noun such as “vegetable dish vs. tomato dish vs. vegetable soup vs. tomato soup” in Zhang and Pylkkänen (2015). This kind of stimuli may force participants to combine two words and yield a more direct connection between combinatorial processing and modulation of low-frequency neural activities.

#### **4.5 Summary of the study**

The current study tests different semantic variables to see whether neural activities in the delta band correlate with combinatorial semantics or lexical semantics. Predictability, plausibility, conceptual specificity, and semantic similarity were examined. 1-Hz peaks of neural signals were observed across all conditions. We see no evidence that the patterns of predictability, plausibility, and conceptual specificity reflect the results of 1-Hz peaks. However, there is a trend towards synchronization in non-specific condition. The overall results suggest that neural activities in the

delta band may reflect the processing of syntactic information more directly than the semantic processing.



## CHAPTER V

### Conclusion

#### 5.1 Summary of contributions

To summarize, this dissertation investigates what high-level linguistic information modulates neural activities in the delta band and how high-level information interact with temporal properties using a frequency-tagging paradigm. Three EEG experiments were conducted. In Chapter II, we tested whether neural activities in the delta band are modulated by hierarchical information or lexical information. By including reversed phrases that preserved syntactic regularities but lost structure in Mandarin, we only found a 4-Hz peak, but not peaks at 1 and 2 Hz in the reversed phrases, suggesting that neural activities in the delta band may be modulated by hierarchical information. In Chapter III, different speech rates and rhythmicities were manipulated to examine the interaction between high-level information and temporal properties. High accuracy of comprehension across all conditions and the sentence-level peaks observed in the faster conditions suggest that neural activities in the delta band may reflect a top-down predictive processing and this processing guides language comprehension when speech is fast. In addition, the results suggest that the observed neural signals represent endogenous oscillatory activities, not just a reflection of rhythms from external speech input. In Chapter IV, we tested

several semantic variables to see what semantic information may modulate neural activities in the delta band. the observed 1-Hz peaks across all conditions did not reflect the results of predictability, plausibility, and semantic similarity. There is a trend towards the synchronization of non-specific stimuli. The results from these three studies suggest that neural activities in the delta band reflect a predictive top-down processing and may be modulated more directly by syntactic information, not semantic information. Moreover, the results suggest that these low-frequency neural responses reflect endogenous oscillatory activities, not merely evoked responses.

## **5.2 Directions for future work**

Finally, we discuss some potential future directions to probe the functional interpretation of low-frequency oscillations, as well as some extensions related to the dissertation.

First, how neural patterns of composition represent in different groups of people is unknown. One future direction is to examine these low-frequency neural activities in language development (Getz et al., 2018). If low-frequency neural activities can be an indicator of linguistic chunking or compositional processes, can it be observed in children and what age can we observe the neural patterns? Adopting the same experimental paradigm on children might provide insights on how these neural activities vary from children to adults. Moreover, we can also compare children with or without developmental language disorders (Montgomery et al., 2018). How difficulty of structural chunking for children with certain language disorder reflects on low-frequency neural activities would be a potential future direction.

The other direction that extends the current work is to develop a unified oscillation-based model to explain how the structure-building processes and semantic composi-

tion associate with other cognitive faculties. Specifically, how memory architecture or attention interacts with language processing and how these processes integrate and are reflect by low-frequency neural activities are unclear. In particular, we can implement a cognitive architecture, adaptive control of thought - rational model (ACT-R, Anderson et al., 2004; Lewis and Vasishth, 2005; Lewis et al., 2006), to examine how memory mechanisms incorporated to language comprehension are reflected by neural signals.

Whether neural signals elicited from the frequency-tagging paradigm can be extended to explain language processing in a naturalistic setting remains unclear. One future direction is to develop a more naturalist approach to examine this low-frequency neural activity (e.g. Kaufeld et al., 2020). One related future direction would be examining this low-frequency neural activity in conversation. There is still a lack of understanding of how this low-frequency neural activity reflects in conversation (e.g. brain-to-brain entrainment while speaking and listening, Pérez et al., 2017) and how general cognitive processes such as retrieval of linguistic information from memory and planning are carried out by these neural activities during conversation.

Our third study includes several semantic variables but we see no evidence that these semantic variables play a role in modulating neural activities in the delta band. However, one possibility might be these semantic variables are not well-disentangled in the current study. Different methods of manipulating semantic variables can be further examined. For example, like the stimuli design in Zhang and Pykkänen (2015), we can manipulate the specificity of the first noun and also the second noun to have four conditions (e.g. vegetable dish vs. vegetable soup vs. tomato dish vs. tomato soup).

Our first study simulated brain responses from using word embeddings from dif-

ferent language models to do FFT. It is yet to be tested whether these low-frequency neural activities can be modulated by other factors. One future direction could be examining word surprisal extracted from large-scale language models and do FFT to see whether surprisal can also predict low-frequency neural responses. In addition, we can also use this method to examine different layers in large-scale language models to investigate which layer might elicit similar responses, which may provide an insight of the ongoing status and timing for modulating these low-frequency neural responses.

## APPENDICES

## APPENDIX A

### Supplementary material for Chapter II and III

Table A.1: Stimuli for Chapter II and III

Condition	Four-syllable sentence	Semantically-mismatched sequence	Two-syllable phrase	Reversed phrase
1	老牛耕地	行爭爬解	老牛青草	牛老地耕
2	樹木生長	廚生鼓跑	樹木太陽	木樹長生
3	冰雪融化	軍孩奔草	冰雪醫生	雪冰化融
4	嬰兒哭泣	小客融舞	嬰兒駿馬	兒嬰泣哭
5	獵犬奔跑	開鼠看馳	獵犬小偷	犬獵跑奔
6	老師講課	松士倒泳	老師觀眾	師老課講
7	綿羊吃草	英刀判覺	綿羊風爭	羊綿草吃
8	學生寫字	母師劃退	學生行人	生學字寫
9	觀眾鼓掌	朋船做山	觀眾烏龜	眾觀掌鼓
10	演員跳舞	公魚轉掌	演員電腦	員演舞跳
11	朋友請客	游司過架	朋友英雄	友朋客請
12	汽車轉彎	姐羊跌行	汽車外公	車汽彎轉
13	燈泡發光	剪泡打閉	燈泡開水	泡燈光發
14	農民種菜	輪水彈蛋	農民朋友	民農菜種
15	青草發芽	法馬告針	青草飛機	草青芽發
16	太陽落山	電友降走	太陽汽車	陽太山落
17	開水沸騰	學果戰客	開水老師	水開騰沸
18	導游講解	叛子種掉	導游軍隊	游導解講
19	叛徒告密	護徒落航	叛徒法官	徒叛密告
20	蘋果爛掉	飛兒開字	蘋果獅子	果蘋掉爛
21	廚師做飯	農犬發彎	廚師護士	師廚飯做
22	外公看報	醫民生光	外公導游	公外報看
23	小孩跌倒	嬰隊打報	小孩樹木	孩小倒跌
24	鴨子游泳	導民游倒	鴨子老鷹	子鴨泳游
25	美女彈琴	冰鷄下化	美女冰雪	女美琴彈
26	鯨魚噴水	老陽爬水	鯨魚演員	魚鯨水噴
27	小偷逃走	獅官吵落	小偷游客	偷小走逃
28	松鼠睡覺	鯨鷹救長	松鼠鯨魚	鼠松覺睡
29	英雄救火	獵公吃密	英雄剪刀	雄英火救
30	獅子打架	演游發菜	獅子學生	子獅架打
31	電腦開機	烏草講船	電腦小孩	腦電機開
32	輪船起航	外偷請門	輪船獵犬	船輪航起
33	士兵戰鬥	癩兵看飯	士兵松鼠	兵士鬥戰
34	老鷹飛翔	美師講落	老鷹士兵	鷹老翔飛
35	風爭墜落	駿員噴琴	風爭廚師	爭風落墜
36	公司倒閉	祖牛奔架	公司叛徒	司公閉倒
37	姐妹吵架	小機下火	姐妹綿羊	妹姐架吵
38	駿馬奔馳	燈生耕病	駿馬蘋果	馬駿馳奔
39	母鷄下蛋	太眾生騰	母鷄輪船	鷄母蛋下
40	飛機降落	汽龜逃街	飛機祖父	機飛落降
41	游客爬山	觀父寫地	游客姐妹	客游山爬
42	漁民划船	漁子飛案	漁民公司	民漁船劃
43	軍隊撤退	老雪哭課	軍隊漁民	隊軍退撤
44	剪刀生鏽	鴨雄爛泣	剪刀農民	刀剪鏽生
45	祖父下棋	青木起翔	祖父燈泡	父祖棋下
46	醫生看病	風人沸棋	醫生老牛	生醫病看
47	護士打針	老腦睡機	護士鴨子	士護針打
48	烏龜爬行	士車跳鏽	烏龜母鷄	龜烏行爬
49	行人過街	樹女墜芽	行人嬰兒	人行街過
50	法官判案	綿妹撤山	法官美女	官法案判

## APPENDIX B

### Supplementary material for Chapter IV



Table B.1: Stimuli for Chapter IV

Condition	specific, plausible	non-specific, plausible	specific, implausible	non-specific, implausible
1	新鮮蕃茄	新鮮蔬菜	新任番茄	新任蔬菜
2	黑暗石窟	黑暗空間	暗自石窟	暗自空間
3	奇怪藥味	奇怪氣味	奇遇藥味	奇遇氣味
4	完整密碼	完整符號	周到密碼	周到符號
5	罐裝咖啡	罐裝飲料	纏足咖啡	纏足飲料
6	國內競走	國內比賽	之內競走	之內比賽
7	地方喜訊	地方新聞	地步喜訊	地步新聞
8	徹底緝毒	徹底搜索	無疑緝毒	無疑搜索
9	高級海鮮	高級菜肴	跳高海鮮	跳高菜肴
10	精密核彈	精密武器	中肯核彈	中肯武器
11	實際法案	實際建議	結實法案	結實建議
12	明確錄音	明確記錄	明知錄音	明知記錄
13	重要鋪路	重要建設	重聽鋪路	重聽建設
14	優秀才女	優秀人才	優惠才女	優惠人才
15	管理網路	管理系統	褲管網路	褲管系統
16	健康餅乾	健康食品	健談餅乾	健談食品
17	暢銷小說	暢銷書籍	暢飲小說	暢飲書籍
18	貴重水晶	貴重珠寶	貴庚水晶	貴庚珠寶
19	專業銀行	專業組織	面熟銀行	面熟組織
20	勞工糾紛	勞工事務	工整糾紛	工整事務
21	新興柔道	新興武術	繁盛柔道	繁盛武術
22	大型會議	大型聚會	重型會議	重型聚會
23	嚴重癌症	嚴重疾病	威嚴癌症	威嚴疾病
24	個人成績	個人能力	私怨成績	私怨能力
25	多元課程	多元教育	多汁課程	多汁教育
26	生活惡習	生活習慣	生路惡習	生路習慣
27	木制鋼琴	木制樂器	木屑鋼琴	木屑樂器
28	學術論文	學術文章	學乖論文	學乖文章
29	熱門電玩	熱門遊戲	趁熱電玩	趁熱遊戲
30	臨時班會	臨時會議	時針班會	時針會議
31	知名外商	知名公司	著稱外商	著稱公司
32	創意食譜	創意書籍	重創食譜	重創書籍
33	周邊道路	周邊設備	周報道路	周報設備
34	綜合筆試	綜合審查	無疑筆試	無疑審查
35	嚴格口試	嚴格考驗	嚴斥口試	嚴斥考驗
36	年終獎金	年終待遇	年輪獎金	年輪待遇
37	众多居民	众多人民	充沖居民	充沖人民
38	合法釣魚	合法捕捉	清白釣魚	清白捕捉
39	有利證據	有利訊息	上風證據	上風訊息
40	基礎數學	基礎科目	基調數學	基調科目
41	全面辦案	全面厘清	全票辦案	全票厘清
42	安全巡邏	安全檢查	安息巡邏	安息檢查
43	年度預算	年度方案	年僅預算	年僅方案
44	高級牛肉	高級食物	高樓牛肉	高樓食物
45	優良政黨	優良團體	優劣政黨	優劣團體
46	有效簽證	有效公文	靈驗簽證	靈驗公文
47	改革草案	改革構想	改天草案	改天構想
48	精密平台	精密設施	精壯平台	精壯設施
49	龐大家產	龐大財富	遠大家產	遠大財富
50	巨大資助	巨大幫助	高大資助	高大幫助

## BIBLIOGRAPHY

## BIBLIOGRAPHY

- Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., and Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *PNAS*, 98(23):13367–13372.
- Alexandrou, A. M., Saarinen, T., Kujala, J., and Salmelin, R. (2020). Cortical entrainment: what we can learn from studying naturalistic speech perception. *Language, Cognition and Neuroscience*, 35(6):681–693.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111(4):1036–1060.
- Arnal, L. H., Doelling, K. B., and Poeppel, D. (2015). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*, 25(9):3077–3085.
- Arnal, L. H., Poeppel, D., and Giraud, A.-L. (2016). A neurophysiological perspective on speech processing in “the neurobiology of language. In Hickok, G. and Small, S. L., editors, *Neurobiology of Language*, pages 463–478. Academic Press, New York.
- Assaneo, M. F. and Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, 4(2):eaao3842.
- Bastiaansen, M., Magyari, L., and Hagoort, P. (2010). Syntactic unification operations are reflected in oscillatory dynamics during on-line sentence comprehension. *Journal of Cognitive Neuroscience*, 22(7):1333–1347.
- Bemis, D. K. and Pykkänen, L. (2011). Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Cognitive Neuroscience*, 31(8):2801–2814.
- Bemis, D. K. and Pykkänen, L. (2013a). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex*, 23(8):Cerebral Cortex.
- Bemis, D. K. and Pykkänen, L. (2013b). Flexible composition: Meg evidence for the deployment of basic combinatorial linguistic mechanisms in response to task demands. *PLoS ONE*, 8(9):e73949.
- Bendixen, A., Schröger, E., and Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, 29(26):8447–8451.
- Benítez-Burraco, A. and Murphy, E. (2019). Why brain oscillations are improving our understanding of language. *Frontiers in Behavioral Neuroscience*, 13:190.
- Boersma, P. and Weenink, D. (2019). Praat: doing phonetics by computer [computer program]. version 6.1.03, retrieved 1 september 2019. <http://www.praat.org>.
- Bonhage, C. E., Meyer, L., Gruber, T., Friederici, A. D., and Mueller, J. L. (2017). Oscillatory eeg dynamics underlying automatic chunking during sentence processing. *NeuroImage*, 152:647–657.
- Boucher, V. J., Gilbert, A. C., and Jemel, B. (2019). The role of low-frequency neural oscillations in speech processing: Revisiting delta entrainment. *Journal of Cognitive Neuroscience*, 31(8):1205–1215.

- Bourguignon, M., Tiège, X. D., de Beeck, M. O., Ligot, N., Paquier, P., Bogaert, P. V., Goldman, S., Hari, R., and Jousmäki, V. (2013). The pace of prosodic phrasing couples the listener’s cortex to the reader’s voice. *Hum Brain Mapping*, 34(2):314–326.
- Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass*, 8(2):37–50.
- Brennan, J., Nir, Y., Hasson, U., Malach, R., J. Heeger, D., and Pykkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, 120(2):163–173.
- Brennan, J. R. and Pykkänen, L. (2016). Meg evidence for incremental sentence composition in the anterior temporal lobe. *Cognitive Science*, 41(S6):1515–1531.
- Buzsáki, G. and Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304(5679):1926–1929.
- Capilla, A., Pazo-Alvarez, P., Darriba, A., Campo, P., and Gross, J. (2011). Steady-state visual evoked potentials can be explained by temporal superposition of transient event-related responses. *PLoS ONE*, 6(1):e14543.
- Cohen, M. X. (2014). *Analyzing Neural Time Series Data: Theory and Practice*. The MIT Press, Cambridge, MA.
- Corrette, R. (2020). Praat vocal toolkit. <http://www.praatvocaltoolkit.com>.
- Cravo, A. M., Rohenkohl, G., Wyart, V., and Nobre, A. C. (2013). Temporal expectation enhances contrast sensitivity by phase entrainment of low-frequency oscillations in visual cortex. *Journal of Neuroscience*, 33(9):4002–4010.
- Cui, Y., Che, W., Liu, T., Qin, B., Yang, Z., Wang, S., and Hu, G. (2019). Pre-training with whole word masking for chinese bert. *arXiv preprint arXiv:1906.08101*.
- Davis, M. H. and Johnsrude, I. S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 23(8):3423–3431.
- Davis, M. H. and Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing Research*, 229:132–147.
- Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv*, page 1810.04805v2.
- Di Liberto, G. M., O’ Sullivan, J. A., and Lalor, E. C. (2015). Low-frequency cortical entrainment to speech reflects phoneme-level processing. *Current Biology*, 25(19):2457–2465.
- Ding, N., Chatterjee, M., and Simon, J. Z. (2014). Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage*, 88:41–46.
- Ding, N. and He, H. (2016). Rhythm of silence. *Trends in Cognitive Sciences*, 20(2):82–84.
- Ding, N., Melloni, L., Yang, A., Wang, Y., Zhang, W., and Poeppel, D. (2017). Characterizing neural entrainment to hierarchical linguistic units using electroencephalography (eeg). *Frontiers in human neuroscience*, 11:481.
- Ding, N., Melloni, L., Zhang, H., Tian, X., and Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, 19:158–164.
- Ding, N. and Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *PNAS*, 109(29):11854–11859.

- Ding, N. and Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, 107(1):78–89.
- Ding, N. and Simon, J. Z. (2014). Cortical entrainment to continuous speech: functional roles and interpretations. *Frontiers in Human Neuroscience*, 8:311.
- Doelling, K. B., Arnal, L. H., Ghitza, O., and Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage*, 85, Part 2(15):761–768.
- Doelling, K. B., Assaneo, M. F., Bevilacqua, D., Pesaran, B., and Poeppel, D. (2019). An oscillator model better predicts cortical entrainment to music. *PNAS*, 116(20):10113–10121.
- Du, Z. (2019). Gpt2-chinese: Tools for training gpt2 model in chinese language. <https://github.com/Morizeyao/GPT2-Chinese>.
- Ettinger, A., Feldman, N. H., Resnik, P., and Phillips, C. (2016). Modeling n400 amplitude using vector space models of word representation. In Papafragou, A., Grodner, D., Mirman, D., and Trueswell, J., editors, *Proceedings of the 38th annual conference of the Cognitive Science Society*, page 1445–1450. Cognitive Science Society, Austin, TX.
- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, 44(4):491–505.
- Federmeier, K. D. and Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4):469–495.
- Frank, S. L., Otten, L. J., Galli, G., and Vigliocco, G. (2015). The erp response to the amount of information conveyed by words in sentences. *Brain and Language*, 140:1–11.
- Frank, S. L. and Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience*, 32(9):1192–1203.
- Frank, S. L. and Yang, J. (2018). Lexical representation explains cortical entrainment during speech comprehension. *PLoS ONE*, 13(5):e0197304.
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., and Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *PNAS*, 103(7):2458–2463.
- Friederici, A. D., Meyer, M., and von Cramon, D. (2000). Auditory language comprehension: An event-related fmri study on the processing of syntactic and lexical information. *Brain and Language*, 74(2):465–477.
- Getz, H., Ding, N., Newport, E. L., and Poeppel, D. (2018). Cortical tracking of constituent structure in language acquisition. *Cognition*, 181:135–140.
- Ghitza, O. (2011). Linking speech perception and neurophysiology: speech decoding guided by cascaded oscillators locked to the input rhythm. *Frontiers in Psychology*, 2:130.
- Ghitza, O. (2013). The theta-syllable: a unit of speech information defined by cortical function. *Frontiers in Psychology*, 4:138.
- Ghitza, O. (2014). Behavioral evidence for the role of cortical oscillations in determining auditory channel capacity for speech. *Frontiers in Psychology*, 5:652.
- Ghitza, O. (2017). Acoustic-driven delta rhythms as prosodic markers. *Language, Cognition and Neuroscience*, 32(5):545–561.

- Ghitza, O., Giraud, A.-L., and Poeppel, D. (2012). Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Frontiers in Human Neuroscience*, 6:340.
- Ghitza, O. and Greenberg, S. (2009). On the possible role of brain rhythms in speech perception: Intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*, 66(1-2):113–126.
- Giraud, A., Kell, C., Thierfelder, C., Sterzer, P., Russ, M., Preibisch, C., and Kleinschmidt, A. (2004). Contributions of sensory input, auditory search and verbal comprehension to cortical activity during speech processing. *Cerebral Cortex*, 14(3):247–255.
- Giraud, A.-L. and Poeppel, D. (2012). Cortical oscillations and speech processing: emerging computational principles and operations. *Nature Neuroscience*, 15:511–517.
- Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., and Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology*, 11(12):e1001752.
- Gwilliams, L. (2020). Hierarchical oscillators in speech comprehension: a commentary on meyer, sun, and martin (2019). *Language, Cognition and Neuroscience*, 35(9):1114–1118.
- Gwilliams, L., Linzen, T., Poeppel, D., and Marantz, A. (2018). In spoken word recognition, the future predicts the past. *Journal of Neuroscience*, 38(35):7585–7599.
- Günther, F., Dudschig, C., and Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology*, 69(4):626–653.
- Hagoort, P. (2019). The neurobiology of language beyond single-word processing. *Science*, 366(6461):55–58.
- Hagoort, P., Hald, L., Bastiaansen, M., and Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, 304(5669):438–441.
- Hald, L. A., Bastiaansen, M., and Hagoort, P. (2006). Eeg theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, 96(1):90–105.
- Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*. NAACL’ 01, Stroudsburg, PA, USA.
- Herrmann, B., Henry, M. J., and Obleser, J. (2013). Frequency-specific adaptation in human auditory cortex depends on the spectral variance in the acoustic stimulation. *Journal of Neurophysiology*, 109(8):2086–2096.
- Howard, M. F. and Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, 104(5):2500–2511.
- Humphries, C., Binder, J. R., Medler, D. A., and Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18(4):665–679.
- Huotilainen, M., Winkler, I., Alho, K., Escera, C., Virtanen, J., Ilmoniemi, R. J., Jääskeläinen, I. P., Pekkonen, E., and Näätänen, R. (1998). Combined mapping of human auditory eeg and meg responses. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 108(4):370–379.
- Jensen, O., Kaiser, J., and Lachaux, J.-P. (2007). Human gamma-frequency oscillations associated with attention and memory. *Trends in Neurosciences*, 30(7):317–324.

- Jin, P., Lu, Y., and Ding, N. (2020). Low-frequency neural activity reflects rule-based chunking during speech listening. *eLife*, 9:e55613.
- Jones, M. N., Kintsch, W., and Mewhort, D. J. K. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, 55(4):534–552.
- Jung, T., Makeig, S., Humphries, C., Lee, T., McKeown, M. J., Iragui, V., and Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2):163–178.
- Kaufeld, G., Bosker, H. R., ten Oever, S., Alday, P. M., Meyer, A. S., and Martin, A. E. (2020). Linguistic structure and meaning organize neural oscillations into a content-specific hierarchy. *Journal of Neuroscience*, 40(49):9467–9475.
- Kayser, S. J., Ince, R. A., Gross, J., and Kayser, C. (2015). Irregular speech rate dissociates auditory cortical entrainment, evoked responses, and frontal alpha. *The Journal of Neuroscience*, 35(44):14691–14701.
- Keitel, A., Gross, J., and Kayser, C. (2018). Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features. *PLoS Biology*, 16(3):e2004473.
- Keitel, C., Quigley, C., and Ruhnau, P. (2014). Stimulus-driven brain oscillations in the alpha range: Entrainment of intrinsic rhythms or frequency-following response? *Journal of Neuroscience*, 30(31):10137–10140.
- Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a “cocktail party”. *Journal of Neuroscience*, 30(2):620–628.
- Kutas, M. and Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207(4427):203–205.
- Kutas, M. and Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307:161–163.
- Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., and Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, 28:2867–2875.
- Kösem, A. and van Wassenhove, V. (2017). Distinct contributions of low- and high-frequency neural oscillations to speech comprehension. *Language, Cognition and Neuroscience*, 32(5):536–544.
- Lakatos, P., Karmos, G., Mehta, A. D., Ulbert, I., and Schroeder, C. E. (2008). Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science*, 320(5872):110–113.
- Lalor, E. C. and Foxe, J. J. (2009). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *European Journal of Neuroscience*, 31(1):189–193.
- Lalor, E. C., Power, A. J., Reilly, R. B., and Foxe, J. J. (2009). Resolving precise temporal processing properties of the auditory system using continuous stimuli. *Journal of Neurophysiology*, 102(1):349–359.
- Landauer, T. K. and Dumais, S. T. (1997). A solution to plato’s problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240.
- Lau, E. (2018). Chapter five - neural indices of structured sentence representation: State of the art. *Psychology of Learning and Motivation*, 68:117–142.
- Lau, E. and Liao, C.-H. (2018). Linguistic structure across time: Erp responses to coordinated and uncoordinated noun phrases. *Language, Cognition and Neuroscience*, 33(5):633–647.

- Lewis, A. G. (2020). Balancing exogenous and endogenous cortical rhythms for speech and language requires a lot of entraining: a commentary on meyer, sun & martin (2020). *Language, Cognition and Neuroscience*, 35(9):1133–1137.
- Lewis, A. G. and Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, 68:155–168.
- Lewis, A. G., Schoffelen, J.-M., Christian Hoffmann, M. B., and Schriefers, H. (2017). Discourse-level semantic coherence influences beta oscillatory dynamics and the n400 during sentence comprehension. *Language, Cognition and Neuroscience*, 32(5):601–617.
- Lewis, A. G., Schoffelen, J.-M., Schriefers, H., and Bastiaansen, M. (2016). A predictive coding perspective on beta oscillations during sentence-level language comprehension. *Frontiers in Human Neuroscience*, 10:85.
- Lewis, R. L. and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3):375–419.
- Lewis, R. L., Vasishth, S., and Dyke, J. A. V. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, 10(10):447–454.
- Luo, H. and Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, 54(6):1001–1010.
- Maguire, M. J. and Abel, A. D. (2013). What changes in neural oscillations can reveal about developmental cognitive neuroscience: Language development as a case in point. *Developmental Cognitive Neuroscience*, 6:125–136.
- Mai, G., Minett, J. W., and Wang, W. S. Y. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage*, 133:516–528.
- Mai, G. and Wang, W. S. Y. (2019). Delta and theta neural entrainment during phonological and semantic processing in speech perception. *BioRxiv*.
- Makeig, S., Bell, A. J., Jung, T.-P., and Sejnowski, T. J. (1995). Independent component analysis of electroencephalographic data. In Touretzky, D. S., Mozer, M. C., and Hasselmo, M. E., editors, *Advances in Neural Information Processing Systems 8*, pages 145–151. MIT Press, Cambridge, MA.
- Martin, A. E. (2020). A compositional neural architecture for language. *Journal of Cognitive Neuroscience*, 32(8):1407–1427.
- Martin, A. E. and Dumas, L. A. A. (2017). A mechanism for the cortical computation of hierarchical linguistic structure. *PLoS Biology*, 15(3):e2000663.
- Martorell, J., Morucci, P., Mancini, S., and Molinaro, N. (2020). Sentence processing: How words generate syntactic structures in the brain. *PsyArXiv*.
- Matchin, W., Hammerly, C., and Lau, E. (2017). The role of the ifg and psts in syntactic prediction: Evidence from a parametric study of hierarchical structure in fmri. *Cortex*, 88:106–123.
- Menenti, L., Petersson, K. M., Scheeringa, R., and Hagoort, P. (2009). When elephants fly: Differential sensitivity of right and left inferior frontal gyri to discourse and world knowledge. *Journal of Cognitive Neuroscience*, 21(12):2358–2368.
- Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms. *European Journal of Neuroscience*, 48(7):2609–2621.
- Meyer, L. and Gumbert, M. (2018). Synchronization of electrophysiological responses with speech benefits syntactic information processing. *Journal of Cognitive Neuroscience*, 30(8):1066–1074.



- Meyer, L., Henry, M. J., Gaston, P., Schmuck, N., and Friederici, A. D. (2016). Linguistic bias modulates interpretation of speech via neural delta-band oscillations. *Cerebral Cortex*, 27(9):4293–4302.
- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint*, page 1301.3781v3.
- Molinaro, N., Barraza, P., and Carreiras, M. (2013). Long-range neural synchronization supports fast and efficient reading: Eeg correlates of processing expected words in sentences. *NeuroImage*, 72(15):120–132.
- Montgomery, J. W., Evans, J. L., Fargo, J. D., Schwartz, S., and Gillam, R. B. (2018). Structural relationship between cognitive processing and syntactic sentence comprehension in children with and without developmental language disorder. *Journal of Speech, Language, and Hearing Research*, 61(12):2950–2976.
- Neufeld, C., Kramer, S. E., Lapinskaya, N., Heffner, C. C., Malko, A., and Lau, E. F. (2016). The electrophysiology of basic phrase building. *PLoS ONE*, 11(10):e0158446.
- Nieuwland, M. S., Barr, D. J., Bartolozzi, F., Busch-Moreno, S., Darley, E., Donaldson, D. I., Ferguson, H. J., Fu, X., Heyselaar, E., Huettig, F., Husband, E. M., Ito, A., Kazanina, N., Kogan, V., Kohút, Z., Kulakova, E., Mézière, D., Politzer-Ahles, S., Rousselet, G., Rueschemeyer, S.-A., Segaert, K., Tuomainen, J., and Wolfsturn, S. V. G. Z. (2019). Dissociable effects of prediction and integration during language comprehension: Evidence from a large-scale study using brain potentials. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*.
- Nourski, K. V., Reale, R. A., Oya, H., Kawasaki, H., Kovach, C. K., Chen, H., Matthew A. Howard, I., and Brugge, J. F. (2009). Temporal envelope of time-compressed speech represented in the human auditory cortex. *Journal of Neuroscience*, 29(49):15564–15574.
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.-M. (2011). Fieldtrip: Open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011:156869.
- Pallier, C., Devauchelle, A.-D., and Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *PNAS*, 108(6):2522–2527.
- Park, H., Ince, R. A. A., Schyns, P. G., Thut, G., and Gross, J. (2015). Frontal top-down signals increase coupling of auditory low-frequency oscillations to continuous speech in human listeners. *Current Biology*, 25(12):1649–1653.
- Pearce, J. W. (2007). Psychopy—psychophysics software in python. *Journal of Neuroscience Methods*, 162(1-2):8–13.
- Pearce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics*, 2:10.
- Petersson, K.-M., Folia, V., and Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, 120(2):83–95.
- Peña, M. and Melloni, L. (2012). Brain oscillations during spoken sentence processing. *Journal of Cognitive Neuroscience*, 24(5):1149–1164.
- Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech Communication*, 41(1):245–255.
- Poeppel, D. and Assaneo, M. F. (2020). Speech rhythms and their neural foundations. *Nature Reviews Neuroscience*, 21:322–334.

- Poeppl, D. and Embick, D. (2005). Defining the relation between linguistics and neuroscience. In Cutler, A., editor, *Twenty-First Century Psycho-linguistics: Four Cornerstones*, page 103–120. Lawrence Erlbaum, Hillsdale.
- Pylkkänen, L. and Brennan, J. (2019). Composition: The neurobiology of syntactic and semantic structure building. In Gazzaniga, M. S., Mangun, G. R., and Poeppel, D., editors, *The Cognitive Neurosciences*. The MIT Press, New York.
- Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, 366(6461):62–66.
- Pérez, A., Carreiras, M., and Duñabeitia, J. A. (2017). Brain-to-brain entrainment: Eeg interbrain synchronization while speaking and listening. *Scientific Reports*, 7.
- Rimmele, J., Sun, Y., Michalareas, G., O., G., and Poeppel, D. (2019). Dynamics of functional networks for syllable and word-level processing. *bioRxiv*.
- Sauseng, P. and Klimesch, W. (2008). What does phase information of oscillatory brain activity tell us about cognitive processes? *Neuroscience and Biobehavioral Review*, 32(5):1001–1013.
- Schell, M., Zaccarella, E., and Friederici, A. (2017). Differential cortical contribution of syntax and semantics: An fmri study on two-word phrasal processing. *Cortex*, 96:105–120.
- Schnitzler, A. and Gross, J. (2005). Normal and pathological oscillatory communication in the brain. *Nature Reviews Neuroscience* volume, 6:285–296.
- Schroeder, C. E. and Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, 32(1):9–18.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270(5234):303–304.
- Smith, Z. M., Delgutte, B., and Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416:87–90.
- Sohoglu, E., Peelle, J. E., Carlyon, R. P., and Davis, M. H. (2012). Predictive top-down integration of prior knowledge during speech perception. *Journal of Neuroscience*, 32(25):8443–8453.
- Tallon-Baudry, C., Bertrand, O., Delpuech, C., and Pernier, J. (1997). Oscillatory  $\beta$ -band (30–70 hz) activity induced by a visual search task in humans. *Journal of Neuroscience*, 17(2):722–734.
- van Berkum, J. J. A., Hagoort, P., and Brown, C. M. (1999). Semantic integration in sentences and discourse: Evidence from the n400. *Journal of Cognitive Neuroscience*, 11(6):657–671.
- Wang, L., Jensen, O., van den Brink, D., Weder, N., Schoffelen, J.-M., Magyari, L., Hagoort, P., and Bastiaansen, M. (2012a). Beta oscillations relate to the n400m during language comprehension. *Human Brain Mapping*, 33:2898–2912.
- Wang, L., Zhu, Z., and Bastiaansen, M. (2012b). Integration or predictability? a further specification of the functional role of gamma oscillations in language comprehension. *Frontiers in Psychology*, 3:187.
- Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, 7(12):553–559.
- Westerlund, M. and Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, 57:59–70.
- Wöstmann, M., Fiedler, L., and Obleser, J. (2016). Tracking the signal, cracking the code: speech and speech comprehension in non-invasive human electrophysiology. *Language, Cognition and Neuroscience*, 32(7):855–869.

- Yamada, I., Asai, A., Sakuma, J., Shindo, H., Takeda, H., Takefuji, Y., and Matsumoto, Y. (2020). Wikipedia2vec: An efficient toolkit for learning and visualizing the embeddings of words and entities from wikipedia. *arXiv preprint 1812.06280v3*.
- Zaccarella, E., Meyer, L., Makuuchi, M., and Friederici, A. D. (2017). Building by syntax: The neural basis of minimal linguistic structures. *Cerebral Cortex*, 27(1):411–421.
- Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargava, A., Wei, C., and Cao, K. (2005). Speech recognition with amplitude and frequency modulations. *PNAS*, 102(7):2293–2298.
- Zhang, L. and Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An meg study. *NeuroImage*, 111(1):228–240.
- Zhou, H., Melloni, L., Poeppel, D., and Ding, N. (2016). Interpretations of frequency domain analyses of neural entrainment: Periodicity, fundamental frequency, and harmonics. *Frontiers in Human Neuroscience*, 10:274.
- Zoefel, B., ten Oever, S., and Sack, A. T. (2018). The involvement of endogenous neural oscillations in the processing of rhythmic input: More than a regular repetition of evoked neural responses. *Frontiers in Neuroscience*, 12:95.
- Zoefel, B. and VanRullen, R. (2015). The role of high-level processes for oscillatory phase entrainment to speech sound. *Frontiers in Human Neuroscience*, 9:651.