Doctoral Dissertation

Investigation on the neural oscillatory mechanisms of speech production and comprehension

ZHAO BIN

Supervisor: Unoki Masashi

Graduate School of Advanced Science and Technology

Japan Advanced Institute of Science and Technology

Degree of Information Science

March, 2022

Abstract

Speech communication is a trait of human species. It is the most natural and convenient means of interpersonal interaction and information acquisition. Normally, we can quickly and accurately understand the meaning of spoken language based on our acquired linguistic knowledge. Also, we can precisely control articulatory movements via appropriate speech planning, motor programming and auditory feedback. These seemingly quite flexible and easy operations, however, require extremely complex brain network coordination. With the mutual development and continuous interactions, there is a tight link formed between the speech comprehension and production systems, called the 'speech chain'. Understanding the principles governing the neural dynamics related to the speech chain remains to be an unmet goal within neuroscience. Recent brain imaging advances have contributed abundant of anatomical and functional cortical localizations and neural pathways. However, it is not sufficient for us to form a comprehensive understanding of the organizational principles underlying the local computation and long-range communication across multi-scale brain networks. Our study aims to unravel the speech production and comprehension mystery by probing into the neural oscillatory mechanisms and reconstructing the temporal-spatial-spectral brain network dynamics in a series of listening and speaking tasks. Technically, we utilized a multi-modal data acquisition system, which includes a high-density (128-channel) electroencephalography (EEG) recorder, an eye movement tracking system, and an electret condenser microphone. Experimentally, we designed a word listening task and a sentence oral reading task to explore speech processing at different linguistic scales and along successive functional stages. Materially, we take advantage of distinctive Chinese properties (e.g., uniformity and rhythmicity) to address some linguistic controversies that have not been settled in other languages. Computationally, we performed EEG artifact reduction, source reconstruction, effective connectivity estimation, and cross-frequency coupling (CFC) evaluation, etc. These developed methods enable us to investigate some interesting topics, such as the motor theory of speech perception, the syntactic and semantic effects in sentence phrase building, and the interactive nature of the speech perception-production loops. By incorporating prevailing theoretical and computational frameworks with our results in the brain network dynamics and CFC patterns, we proposed a neurofunctional model of speech production and comprehension (SPAC) to explain the dynamic, hierarchical and interactive organization of speech functions. The SPAC model (1) considers the active nature of the speech functions with top-down regulations from higher cognition; (2) extends the framework up to the sentence level with the consideration of syntactic and semantic effects on low-level sensorimotor processing. (3) complements earlier anatomical and functional models with spatiotemporal brain network dynamics. (4) indicates bidirectional information flows in the dorsal stream for speech production and ventral stream for speech comprehension. (5) explains with CFC mechanisms for the bridging of linguistic form diversities and representational hierarchies, as well as the bidirectional interactions between the bottom-up sensory input and top-down cognitive regulations. The SPAC model is supposed to forward our understanding of the speech chain from a more comprehensive perspective.

Key words: Speech production and comprehension, cross-frequency coupling, brain network dynamics, EEG source reconstruction, SPAÇ model.

Declaration

It is hereby declared that

1. The thesis submitted is my own original work while completing degree at Japan Advanced Institute of Science and Technology.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.
5. This dissertation was prepared according to the curriculum for the Collaborative Education Program organized by Japan Advanced Institute of Science and Technology and Tianjin University, China.

Student's Full Name & Signature:

Student Name

Student ID

Ethics Statement

The studies involving human participants were reviewed and approved by the Research Ethics Committee of Tianjin University, China.

The experimental procedures have been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki).

The participants provided their written informed consent to participate in this study, with the understanding that anonymized data could be used for research and publications.

Acknowledgments

First and foremost, I have to thank my research supervisors (Prof. Jianwu Dang and Prof. Masashi Unoki), my researcher advisors (Dr. Gaoyan Zhang, Prof. Kiyoshi Honda, Prof. Longbiao Wang, and Prof. Hirokazu Tanaka), and my minor research supervisors (Prof. Tzyy-Ping Jung, Prof. Masato Akagi, Prof. Bo Hong, and Prof. Atsuo Yoshitaka). Without their assistance and dedicated involvement in every step throughout the process, this paper would have never been accomplished and I would never be who I am today. Their moral character and life attitude are what I am always looking up to.

I would like to thank my research teams in His-lab at JAIST and in Tianjin Key Laboratory of Cognitive Computing and Applications at Tianjin University. I spent most of my research period and had my best memories there. I would also like to thank the research team in Swartz center for computational neuroscience (SCCN) at the University of California San Diego (UCSD), from where I step on the road of neuroscientific exploration. The inspiration and instruction that I received from all my teachers and lab mates surely have a life-long influence on me.

I would also like to express my deepest gratitude to my host family in San Diego (my Mom Kim Phillips, Dad Jeff Phillips, and sister Trinity) and my life mentor in Japan (Prof. Shungo Kawanishi and Kotona Motoyama). They gave me the warmth of family while I was overseas and made me feel the unconditional love without boundaries.

Finally, none of this could have happened without my family. They are the source of love, the source of encouragement, and the source of motivation. They give me the greatest courage, holding me up and leading me forward. Hereby thanks to everyone and everything that come into my life.

Table of Contents

Chapter 1:

Chapter 2:

Chapter 3:

Abbreviations

Lists

AAC - amplitude-amplitude coupling

ATL – anterior temporal lobe

CFC - cross-frequency coupling

dDTF - direct directed transfer function

ERP - event-related potential

ERSP - event-related spectral perturbation

HSFC - hierarchical state feedback control

IC - independent component

ICA - independent component analysis

IPL – inferior parietal lobe

ISPC - intersite phase clustering

GC - granger causality

M1 - primary motor area

PAC - phase-amplitude coupling

PC - parietal cortex

PFC – prefrontal cortex

pITG - posterior inferior temporal gyrus

pMTG - posterior middle temporal gyrus

PPC - phase-phase coupling

S1 - primary somatosensory area

S2 - secondary somatosensory area.

SFC - state feedback control

SMA – supplementary motor area

SMC – sensorimotor cortex

SPAC - speech production and comprehension

Spt - Sylvian fissure at the parieto-temporal boundary

STG – superior temporal gyrus

STS - superior temporal sulcus

VC – visual cortex

vPMC - ventral premotor cortex

Chapter1

Introduction

Brain research of speech functions roots back to the lesions studies in the 19th century and even earlier. Recent advances in neuroimaging techniques provided us with wider opportunities to investigate finer speech operations in healthy subjects. This chapter introduces the history of these neurolinguistic studies, technological advancement, and the current issues that remain to be overcome. It leads to the motivation and originality of the current study. At the end, the structures of the following chapters will be outlined.

1.1 Background

Since the first discovery that language functions are directly related to brain tissue, people have been interested in understanding the neural basis of language. For many centuries, there are many diseases and/or types of injury to the human brain that lead to language disturbances, and these are important for us to understand the basic neurobiology and develop treatments for conditions that affect language [1]. During the 19th century, brain research was mainly observational and autopsy driven [2, 3], which offered fundamental knowledge about the loci of speech perception and production. The 20th century witnessed the advent of brain imaging techniques such as electroencephalography (EEG) [4], magnetoencephalography (MEG) [5], and magnet resonance imaging (MRI), which can be used in vivo to image cognitive functions in the brain (fMRI) as well as gray matter anatomy and white matter fiber tracts (diffusion-weighted MRI) [6]. These new methodologies has lead to a considerable increase in brain-based language studies and greatly improved our understanding of the cognitive function of language. Below is a historic review of these backgrounds.

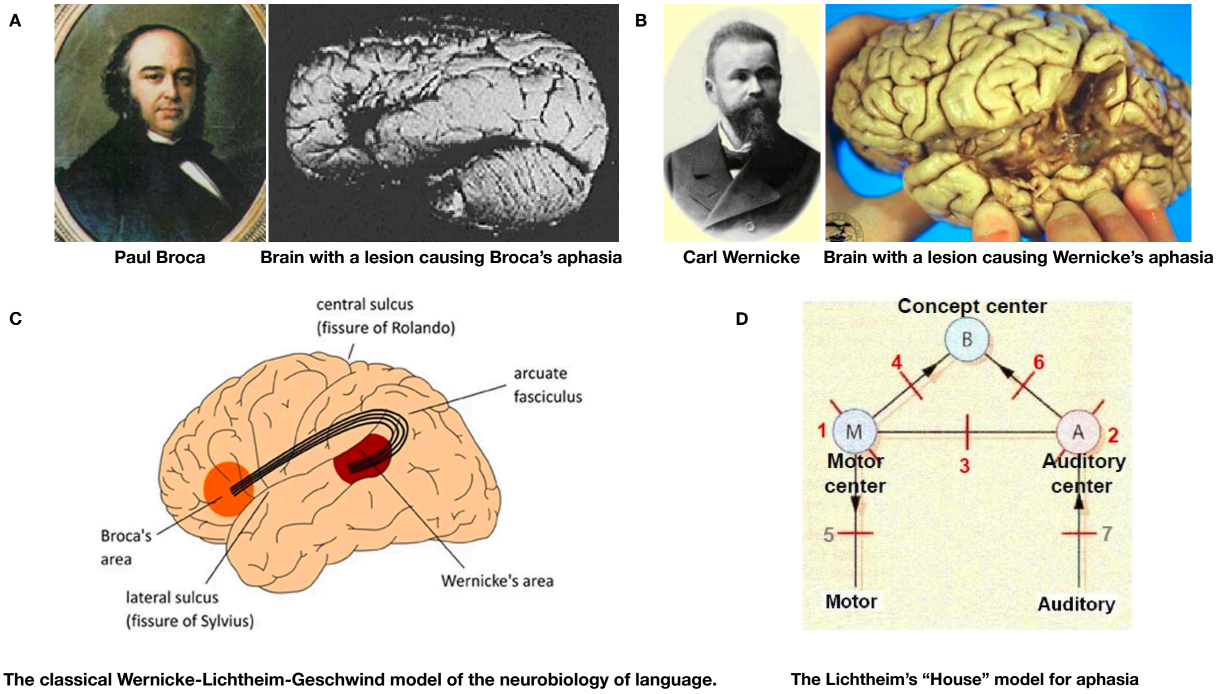
1.1.1 Lesion studies

The biological basis of human thought has been an important focus of attention in medicine, with particular interest in the brain basis of language sparked by the famous patients of Pierre Paul Broca (1861) in the mid 19th century [2]. In postmortem inspection, Broca found that some neuroanotomical damage in the left posterior inferior frontal gyrus could cause severe difficulty in fluid speaking. This region is now known as Broca's area. Broca' work not only established the principle of localization, but also introduced the notion of hemispheric differences.

The era of brain localization for language blossomed after this with the famous doctoral dissertation of Wernicke (1874) Based on deficit-lesion correlations of postmortem dissections [3] Wernicke localized recognition of sound images of words to the posterior superior temporal lobe (now known as Wernicke's area) and representations of motor images of words to the inferior frontal lobe (the Broca's area). He hypothesized that words as supramodal linguistic units emerge from an interaction between distant temporal and frontal areas, thus reconciled the principle of localized functions with a connectionist framework [7].

Lichtheim (1885) later translated Wernicke's ideas into illustrative diagrams, formulating a three-component model of language, in which Broca's and Wernicke's areas are interconnected via a hypothetical (not anatomically localized) 'concept center' involved in semantic processing. This model became the standard reference for clinicians to predict aphasic syndromes from lesions of either a center or a connection.

Norman Geschwind recapitulated and added to the language 'center' models that preceded him and presented a reconceptualized 'connectionist' view of the brain mechanisms of language, and culminated with the classic Wernicke-Lichtheim-Geschwind model [8].



Paul Broca **Brain with a lesion causing Broca's aphasia** **Carl Wernicke** **Brain with a lesion causing Wernicke's aphasia**

The classical Wernicke-Lichtheim-Geschwind model of the neurobiology of language. **The Lichtheim's "House" model for aphasia**

Figure 1.1: Early lesion studies and models for aphasia. (A) Paul Broca and the brain with a lesion causing Broca's aphasia. (B) Carl Wernicke and the brain with a lesion causing Wernicke's aphasia. (C) The classical Wernicke-Lichtheim-Geschwind model of the neurobiology of language. Broca's area and Wernicke's area are connected by a large bundle of nerve fibres called the arcuate fasciculus. (D) The Lichtheim's 'House' model for aphasia. A represents the center for sound images; M is the center for movement images; B is the center for concepts; a is the acoustic pathway and m is the motor pathway. 1–7 represent various forms of aphasia, depending on the destruction of a center or pathway.

Figure 1.1 depicts these early lesion studies and the models of speech processing based on aphasia. By their very nature, lesion analysis studies tend to relate single foci of brain injury to single psychological or linguistic phenomena. The classic models derived from these lesions studies saw the language system as highly specialised and clearly defined, which has been quite successful in giving insights into the neurobiology of language. However, it is obsolete and no longer fit for purpose [9].

1.1.2 Cognitive neuroscience

In the past decade of the 20th century, the advent of high-resolution functional brain imaging has dramatically changed the conduct of studies of brain and behavior [10]. These applications allow us to examine healthy subjects when they perform particular cognitive tasks [11]. Below I'll brief introduce the major techniques applied in cognitive studies, their advantages and limitations.

1.1.2.1 Brain imaging techniques

The commonly applied neuroanatomical techniques include:

(1) X-ray Computed Tomography (CT), which is constructed based on the differences in x-ray absorption.

(2) Structural Magnetic Resonance Imaging (sMRI), which is constructed based on the differences in the radio frequency signal of excited hydrogen atoms as they return to their equilibrium states.

(3) Tractography, which images brain pathways (tracts) using diffusion tensor imaging (DTI) or diffusion spectrum imaging (DSI) based on the diffusion of water molecules in the brain.

The neurophysiological techniques include:

(1) Single- or multi-unit spiking train recording from single or several neurons directly in the brain.

(2) Local Field Potential (LFP), which records cortical electrical activity from tens of thousands of neurons with microelectrodes in the brain.

(3) Intracranial EEG (iEEG), which records cortical electrical activity from patch of neurons with macroelectrodes in the brain.

(4) Electrocorticogram (ECoG), which records cortical electrical activity from several millions of neurons with macroelectrodes on the cortical surface.

(5) Electroencephalogram (EEG), which records cortical electrical activity from multiple cortical areas with extracranial sensors. (6) Magnetoencephalogram (MEG), which records cortical magnetic activity from extracranial sensors.

The functional brain imaging techniques include:

(1) Functional Magnetic Resonance Imaging (fMRI), which reflect tomographic imaging of brain activity from the Blood Oxygen Level Dependent (BOLD) signals.

(2) Positron Emission Tomography (PET), which reflect tomographic imaging of brain activity based on the emitted gamma rays from radioactive tracers.

There are also perturbation techniques that can be used to test causalities, which include:

(1) Transcranial Magnetic Stimulation (TMS), which is a noninvasive technique that uses magnetic stimulation to activate a focal brain region or produce a 'virtual lesion' by disrupting ongoing activities.

(2) Transcranial Direct Current Stimulation (tDCS), which is a noninvasive technique that uses electrical stimulation to activate or inhibit focal brain regions.

1.1.2.2 Technical advances and limitations

The brain imaging techniques introduced above hold advantages over either temporal/spatial precision, or operational practicability, thus can be selectively adopted in the investigation of different perspectives of brain activities. Figure 1.2 compares the temporal and spatial resolution for different brain imaging techniques. Among them, the PET and fMRI are called active recording techniques considering they interference with normal brain metabolism. These active methods have pretty good spatial resolution but poor temporal resolution. In contrast, EEG and MEG do not alter the brain activities but record the changes in the local electric or magnetic field at the surface of the scalp. Thus they are called passive recording techniques and are extremely superior in temporal resolution. However, they face major challenge in spatial localization, particularly in the 'inverse problem' when trying to identify the source of the underlying signal. Fortunately, now EEG and MEG are equipped with more electrodes, thus butter dipole fitting. Besides, better mathematical tools could also be used for combating noise in measurement, resulting in more accurate cortical source localization. With these techniques, more widely distributed cortical regions and dense fiber tracts have been identified [6], and critical temporal windows have been recognized

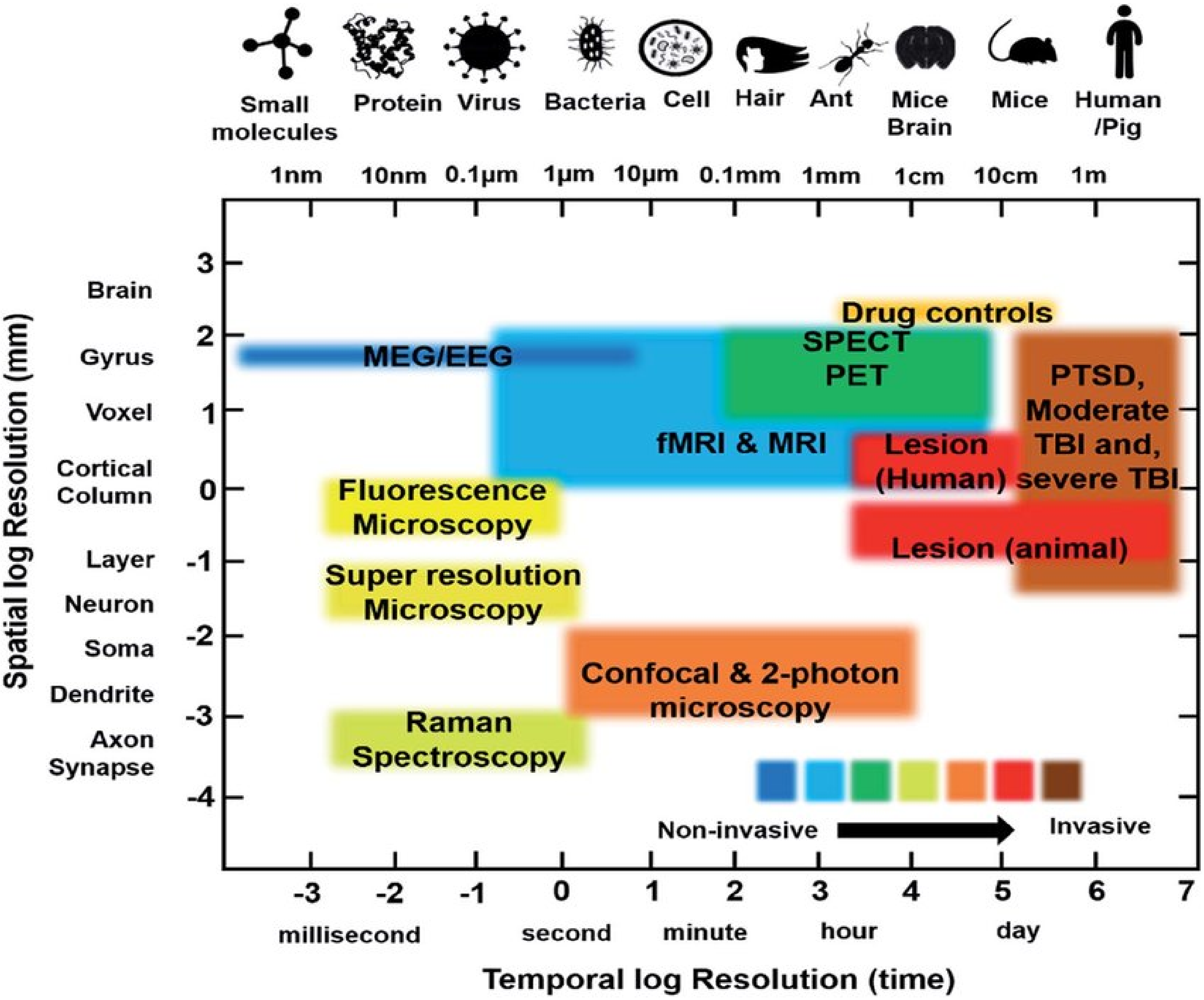


Figure 1.2: Comparison of temporal and spatial resolution for different brain imaging techniques. Adapted from [13] Fig. 9.

during different stages of language processing [12]. However, the sum of projections does not equal to the object itself. Analogously, separate revelation of the involved brain regions or the temporal response windows could not produce a brain dynamic view. Therefore, more precise recording techniques that balanced in both spatial and temporal resolution, more developed source reconstruction methods and connectivity estimation algorithms, as well as more profound knowledge of the intrinsic neural oscillatory mechanisms are well-appreciated in the essential revelation of the neural activities associated with speech functions.

1.2 Current issues

Despite the fact that earlier findings have made immense contribution to people's basic understanding of speech processing in the brain and are still of great referencing importance for future studies. However, the deeper revelation of the neural mechanisms underlying speech production and comprehension still remains pretty vague. The difficulties come from various sources and some prominent issues are briefly listed below:

1.2.1 Theoretical issues

(1) Brain as an inherently dynamic system, not only responds to external stimuli passively and forward information sequentially for a final cognition [14]. Instead, it is actively engaged in constant anticipation, comparison and error correction for dynamic reconfiguration of response strategies [15]. However, conventional studies tend to treat the brain as a stimulus-driven device, and lead the mechanisms of the bottom-up and top-down interactions obscure.

(2) Speech as a highly evolved cognitive function, is not supposed to solely rely on language-specific areas, such as the traditional peri-Sylvian regions, the lexical representation regions, and the motor control regions. Instead, other cognitive functions such as attention, prediction, memory, and emotion are also involved. However, these higher-order functions have been less considered in the framework of speech processing [16].

(3) Sentence as a hierarchical structure, compose multi-scale linguistic units, such as phonemes, lexicon, and clauses. These structures were thought to be serially processed but met challenges in explaining some contradictory facts, such as injured phonetic discrimination ability but preserved lexical comprehension ability. This requires us to reconsider the organizational principle of multi-scale linguistic processing in parallel. Promising mechanisms (e.g., via cross-frequency coupling) for such explanation are yet to be explored.

1.2.2 Technological issues

(1) The trade-off between temporal prcision and spatial resolution has long-been a dillemma in the cognitive research field. The spatially-precise brain-imaging techniques (e.g., fMRI and PET) place more emphasize on functional localization whereas. However, they lack temporal description of language processing dynamics. In contrast, electro-physiological studies (e.g., EEG and MEG) are more temporally precise. Yet, they are not good at revealing brain network organization by using traditional analysis methods (e.g., ERP analysis and time-frequency analysis). Even though merging these techniques might be a solution, economic cost is another concern.

(2) Even though spatial and temporal investigations are critical windows for mechanical insight, yet too much relying on these aspects may lead to the ignorance of the neural functional basis, namely neural oscillations. The organizational principles of oscillation in different frequency bands are much less clear than the functional regions in human brain map, which remains a big to be made up.

(3) In cognitive research, collecting relatively clean brain activities is difficult enough itself, letting along introducing large facial muscle movement during speech production tasks. This problem has impeded progress in the investigation of speech production mechanisms, especially at the sentence level.

In summary, the complexity of the speech chain operation and brain network organization require not only sufficient temporal and spatial resolution for neural dynamic tracking, but also a deeper understanding of the underlying neural oscillatory mechanisms that govern the spatiotemporal dynamics. Moreover, advanced neuro-computational algorithms, such as artifact reduction, source localization and connectivity estimation are necessary for acquiring clean electrophysiological data and reconstructing the brain network dynamics. Finally, and most importantly, it requires an updated view of us to trait the brain as an active and holistic system and analyze properly with consideration of its higher cognitive functions.

1.3 Research goals

This study aims to reveal the neural oscillatory mechanisms involved in speech production and comprehension. Specifically, we would like to know (1) how linguistic structures (e.g., lexicons, clauses, and sentences) are represented and processed on the cortex? (2) What are the contributions of different oscillatory bands in organizing these hierarchical linguistic structures? (3) Whether there is any interaction between the speech production and comprehension systems? Are they functioning independently or collaborate as a whole? To answer these questions, there are several steps we need to take as listed below:

1.3.1 Experimental design

To understand the speech functioning in both comprehension and production tasks, we will design two experiments. Experiment 1 is a passive listening task with word items as auditory materials. These word items appear as synonym pairs with similar semantic meanings but different phonetic features (articulatory gestures). Experiment 2 is a sentence oral reading task with syntactically consistence sentences as visual stimuli. The sentences consist of 8 disyllable words that could be further parsed into 3 clauses that can be tested for semantic and syntactic effects. The two experiments are designed to investigate linguistic processing at multiple level (e.g., phonemic, lexical, syntactic), with various modalities (e.g., auditory, visual, and motor) and in different functional stages (e.g., perception, comprehension, and production) of the speech functions.

1.3.2 Brain network reconstruction

This study aims to balance the trade-off between the spatial and temporal resolutions by utilizing a high-density (128 channel) EEG equipmenent, in combination with recently developed computational algorithm for artifact reduction, source reconstruction, and effective connectivity estimation, etc. The reconstructed brain network dynamics would help us to identify the involved cortical regions and critical time windows that are more functionally specialized in different stages of speech processing. The estimated effective connectivity and information flow across the brain networks would provide us with a clearer image of the inter-regional relationship and the possible functional transformation between different speech modules.

1.3.3 Neural mechanical explanation

Another aim of this study is to investigate the neural oscillatory mechanisms underlying the spatial-temporal brain network dynamics. The neural oscillations with different temporal periods and spatial coverage are suitable for entraining with linguistic units at multi-levels of linguistic units. And the cross-frequency couplings between different frequency bands are promising to relate the linguistic hierarchy with the multi-scale cortical networks. Thus it would be for the current investigation in validating the neural operational mechanisms underlying the spatial-temporal dynamics.

1.3.4 Neurofunctional model building

Finnaly, we aim to build a neurofunctional model for speech production and comprehension (SPAC) based on the spatial-temporal brain network dynamics and neural oscillatory mechanisms revealed from above analyses. The SPAC model aims to extend previous models of speech processing from the lexical level to the sentence level. The cross-frequency coupling patterns would also be involved in the SPAC model to deeply reveal the neural oscillatory mechanisms underlying the speech production-perception loops. In addition, The SPAC model considers the active nature of the speech functions with top-down regulations from higher-level of cognitive control (e.g., with syntactic and semantic constraints) and the bidirectional interactions between the bottom-up sensory input and top-down cognitive regulations.

1.4 Originality

The originality of the current study may include the theoretical, technological, and methodological aspects as explained blow:

1.4.1 Theoretical originality

Our study considers the brain as an active and interactive complex system. Rather than tracking the bottom-up or feedforward information flow in speech comprehension and production, we considers the top-down or feedback influence of the internal mental state and linguistic knowledge on the sensorimotor processes. In this respect, we incorporated the concept of dual-stream representation and the feedforward and feedback loop from the hierarchical state feedback control (HSFC) model into our SPAC model and introduced the cross-frequency coupling mechanisms to explain the multi-scale organization and bidirectional interactions.

1.4.2 Technological originality

This study introduced multi-model data acquisition system for simultaneously recording the cognitive (EEG) and behavioral (eye movement and speech) response patterns during speech production and comprehension tasks. The adoption of EEG technique is in the consideration its high-temporal-resolution and its deeper revelation of the neural oscillatory properties, which are suited to capture the fast, dynamic, and temporally sequenced speech events. The combination of eye movement and speech data could not only give us a clue to investigate the behavioral patterns of speech processing, but also helps to segment the cognitive stages and relate the behavioral patterns with brain dynamics for a functional association.

1.4.3 Methodological originality

This study adopted experimental stimuli at multiple scales of the linguistic hierarchy, ranging from phonemes, lexicons to the whole sentences. Such stimuli selection paradigm would enable us to test the staged processing (e.g., phonemic, lexical, syntactic, etc.) and the serial or parallel relationship between different thread of processing streams. In addition, we associated the hierarchical linguistic structure with the hierarchical neural oscillatory system to explain the multi-scale linguistic processing over distributed cortical regions and in various temporal windows, as well as the interaction across the linguistic and oscillatory hierarchy.

This study adopted advanced computational methods, such as the artifact reduction space analysis, the independent component analysis methods to reduce the noisy artifact induced in EEG signals by orofacial muscle movement. Besides, we performed effective connectivity estimation and examined the cross-frequency coupling patterns in various forms. These are recently developed methods that could help us to incorporate the temporal dynamics, anatomical localization, and oscillatory patterns within a single framework.

This study adopted Chinese language as the speech materials, which has many specific properties suitable for address some intricate problems that are not easily solved in other languages. For example, the uniformity the rhythmicity of the Chinese character composition could be used to test the rhythmic chunking and the oscillatory correspondence in sentence parsing.

1.5 Dissertation outline

Figure 1.3 shows the research framework of the current study. Our research includes two modules, speech comprehension and speech production. For each module, we designed an experiment to discusses several topics involved. In the literature review, I will introduce the theories and models related to our study. In our experiment, we will use multi-modal data

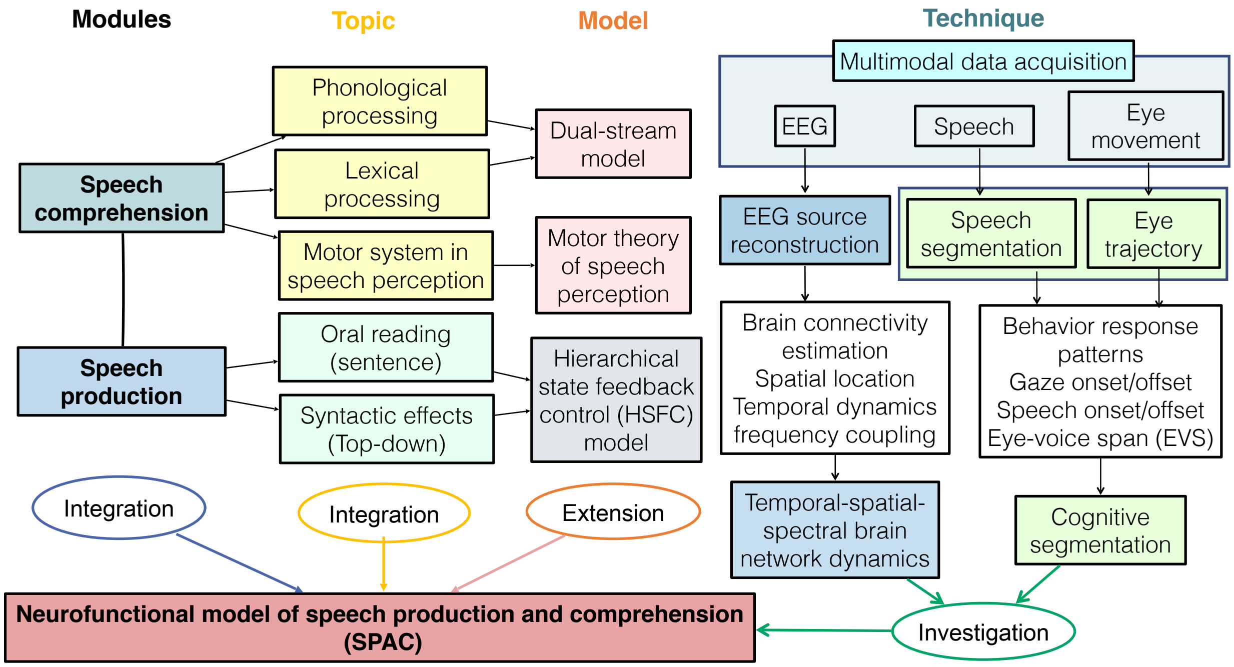


Figure 1.3: Research framework of the current study.

acquisition system of EEG, eye movement, and speech signals, and and reconstruct the brain network dynamics in different cognitive stages. Finally, we will proposed the neurofunctional model of speech production and comprehension.

Chapter 2 firstly lays the knowledge foundation of human cortical structure and the speech system, which includes the cortical circuit hierarchy, cortical oscillatory system, brain connectivity organization. Then it focuses on the current understanding of the neural mechanisms related to speech functions, which includes the cortical organization of different speech functions and the classic models for functional transformation in speech comprehension and production. I will also indicate the weaknesses of these models that I would like to overcome in our studies.

Chapter 3 explains the experimental design and the EEG analysis methods performed in the current study. The experimental design includes our methodology, the experimental procedures, the multi-modal data acquisition procedures. The data analysis methods include artifact reduction pipelines in EEG data reprocessing, time-frequency analysis, source reconstruction, brain connectivity estimation, cross-frequency analysis and statistical analysis.

Chapter 4 elaborates on experiment 1 - speech perception and comprehension at the phonological and lexical levels. The issues of interest we would like to address with this experiment includes: (1) speech perception-production interaction at the phonological scale, and (2) lexical representation and processing and the neural oscillatory mechanisms involved.

Chapter 5 elaborates on experiment 2 - speech production during oral reading at the sentence level. The topic of interest is the hierarchical structure of sentence organization and its neural correspondence. A key question we would like to address is whether the syntactic structure and semantic association from our internal knowledge could top-down influence the behavior of visual reading and speech production process.

Chapter 6 discusses the insights reveal from current analyses, which includes the cortical representation, temporal procedures, frequency characteristics and neural oscillatory mechanisms of speech processing along the dual-stream paths. Finally, I will propose a neuro-functional model of speech production and comprehension (SPAC) that we have constructed based on the temporal-spatial-spectral brain network dynamics and neural cross-frequency coupling patterns.

Chapter 7 summarizes the dissertation with the insights reveal from the current study, its contributions and future plans.

Chapter2

Literature review

This chapter lays the foundation of our current understanding regarding the neural basis of brain structures and speech functions. The involved aspects include: (1) The cortical organization of fractal structure and hierarchical loops. (2) The neural oscillatory system, neural synchronization and desynchronization, and the cross-frequency coupling mechanism. (3) The neuronal substrates and models for speech comprehension. (4) The neuronal substrates and models for speech production. These knowledge comes from neural investigation with different techniques and emphasizes the spatial, temporal and frequency aspects differentially.

2.1 Human cortical organization

The human brain, with its exquisite complexity, is considered to be a fractal object with its self-similar patterns in both neuroanatomical architecture (spatial aspect) and neurophysiological time-series (temporal aspect) [17]. The dynamic operation underlying this fractal structure is subserved by a system of neural oscillations at multiple scales (frequency aspect). This section will briefly introduces the fractal organization of the human brain.

2.1.1 Spatial: Fractal structure and hierarchical circuits

One of the fundamental discoveries of neuroscience is that cortical regions are formed of many functionally specialized areas that are organized into hierarchical networks [18]. From a single neuron, to a cortical column, and to an interconnected cortical patches, various levels of neuronal structure in the brain share tree-like branching characteristics with recurrent feedforward and feedback projections. Along the evolution course, the basic circuits have been remained with parallel neuronal circuits superimposed on the existing wiring, forming a fractal structure with multilevel, self-similar loops. Upon the hierarchically organized circuits, links between different layers are also formed, generating a complex interconnected organization among the parallel loops [18].

The fractal nature of the human brain gives it many excellent properties to deal with complex circumstances: (1) Robust - fractal systems could maintain their coherence and hold together well, even under tough circumstances. They are balanced in this respect. (2) Connectivity - the self-similar and self-organizing fractal structures are very good at connecting and reorganizing, both internally and also to other surrounding systems. (3) Plasticity - activity in a small branch could quickly spread over the whole system. This enables the brain to quickly integrate internal information and timely respond to external stimuli. One the other hand, the neuronal branches within each local part may be trimmed without affecting the overall structure. (4) Economical - the fast response property could also help the brain to minimize wring cost while support dynamic complexity. It is cost-effective. (5) Small-worldness - the dense local interconnections and random long-range interconnections of the cortical circuits is a crucial aspect of efficient brain organization that support both segregated (specialized) and distributed (integrated) information processing.

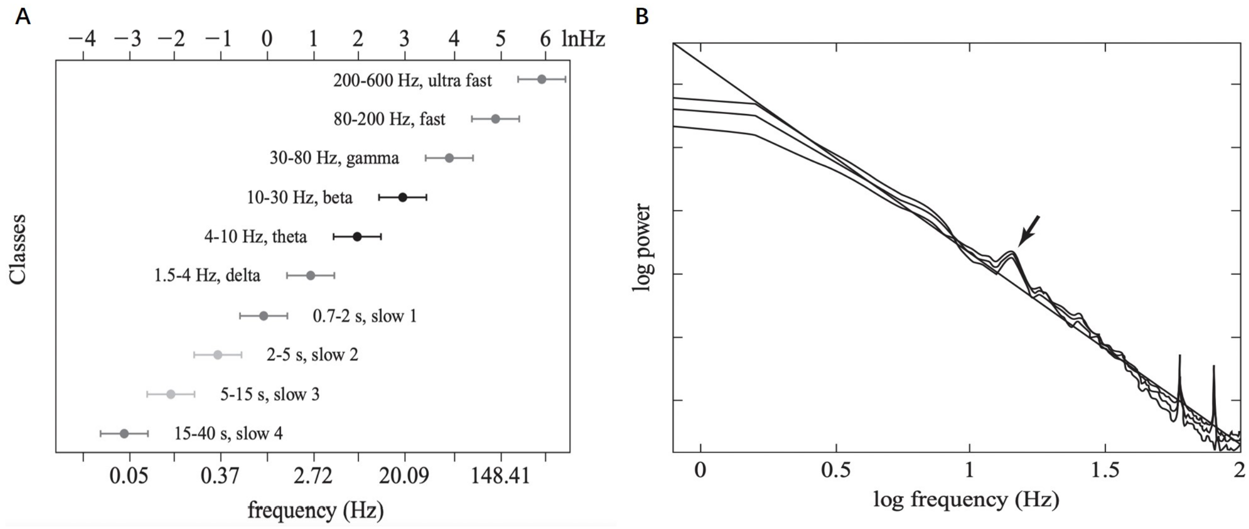


Figure 2.1: Hierarchical oscillatory system. (A) A hierarchical system with multiple oscillators in the cerebral cortex. (B) Illustration of the 1/f power law of spontaneous cortical oscillations. Adapted from [18] Fig. 5.1 and Fig. 5.2

2.1.2 Temporal-spectral: multi-scale rhythmic system

The organization of the brain is fractal not just in the structure but also in the processes. Contributing to the dynamic complexity and flexibility of the fractal brain is a system of oscillatory rhythms, ranging from ultraslow oscillations less than 0.05 Hz to very fast oscillations over hundreds of hertz [18]. This wide range of scale enables spatial coherence extending from centimeters to micrometers away, and temporal coherence from millisecond to minutes long [19]. Figure. 2.1 (A) shows a hierarchical system with multiple oscillators resonating in multiple frequency bands [20]. It could be found that these discrete oscillation bands form a linear progression of the frequency classes on a natural logarithmic scale, showing a geometric progression in the hierarchical system of the cerebral cortex.

One reason for the brain to use so many neural oscillators instead of a uniform rhythm might be attributed to the economical compromise in the brain wiring between size and speed. For information transfer across brain networks, the effective connectivity and path length determines the possible routes, and the oscillatory cycle length limits how far information get transferred in one step. When complex events involving a large pool of neuronal assembles occurs, the arrival time of distant neuronal charges must be coordinated precisely to exert an impact on their postsynaptic target in time. These separate processing inevitably require neuronal networks operating at multiple spatiotemporal scales [18]. For the various neural oscillators, each oscillatory cycle corresponds to a temporal processing window which determines the transferring period and the neuronal pool involved. Following this speculation, different frequencies are suitable for different types of connections and different levels of computation. Fast oscillations, therefore, favor local decisions, whereas the involvement of distant neuronal groups for obtaining a global consensus may requires slow oscillations with longer time span.

Another factor affecting the effect scope of different oscillators is the amount of energy they could afford to dissipate. Because the amplitude of oscillations is proportional to the number of synchronously active neural elements, slowly oscillating cell assemblies comprise more neurons than fast oscillating cells [21]. Figure. 2.1 (B) shows the power spectrum of various neural oscillators with a 1/f statistics. This power-law scaling reflects the fact that higher frequencies are more locally focused while lower frequencies tend to decay more slowly and extend over larger distances in time and space. In short, the self-similarity and hierarchical organization of cortical circuits, the logarithmically progressing frequency bands and the power-law scaling behavior of neural oscillations give the brain its fractal nature and its flexibility in multi-scale communication and adaptive reorganization according to demands, either externally or internally. It also gives us the insight to explore the brain mechanisms of speech functions from the perspective of neural oscillations.

2.1.3 Brain connectivity

Brain connectivity refers to a pattern of anatomical links formed by synapses or fiber pathways (structural connectivity), or statistical dependencies measured as correlation, covariance, spectral coherence or phase-locking, etc (functional connectivity) or causal interactions measured as information flow or causality (effective connectivity).

The structural/anatomical connectivity refers to a physical network linking sets of neurons or neuronal elements, as well as their associated biophysical attributes such as synaptic strength or effectiveness. It can be studied in vivo using invasive axonal labeling techniques or noninvasive MRI-based diffusion weighted/tensor imaging (DWI/DTI) methods.

Functional connectivity is more of a statistical concept. It captures deviations from statistical independence between spatially distributed neuronal units without explicit reference to specific directional effects or to an underlying structural model. Unlike structural connectivity, functional connectivity is highly time-dependent and fluctuate on multiple time scales, from tens to hundreds of milliseconds.

Effective connectivity aims to identify the directionally specific (although not necessarily causal) effective connectivity between neuronal structures. It may be viewed as the union of structural and functional connectivity and can be inferred through perturbation examination or through time series analysis. Popular effective connectivity methods include dynamic causal modeling, structural equation modeling, transfer entropy, and Granger-causal methods.

Brain connectivity can be described at several levels of scale [22]. These levels include: (1) individual synaptic connections that link individual neurons at the microscale. (2) networks connecting networks of columns and minicolumns at the mesoscale, as well as (3) brain regions that are interconnected by fiber pathways at the macroscale. EEG investigation of brain networks tends to focus on the functional and effective connectivities and fall under the macroscale categories. Over the past several decades, considerable empirical research from multiple species, theories, and mathematical models points to oscillatory synchronization as being a key mechanism by which neural populations transmit information and form larger networks [23], giving support to the current study to use neural oscillation for the reconstruction of brain network activities.

2.2 Cortical oscillatory system

Cortical oscillations reflect cyclical variations in excitation and inhibition of a population of neurons [24, 25]. They are commonly recorded at a variety of different scales, ranging from single neuron spikes to synchronized population activity [26]. Despite the wide range of spatiotemporal scales, some common frequency bands have been revealed across species, tasks and signal types [27]. It has also been found that these rhythmic activities are functionally related to the timing and direction of neuronal activates and correlate with task demands (e.g., perceptual, cognitive, motor, linguistic, social, emotional, mnemonic, and other functional processes) [28] .

2.2.1 Oscillatory hierarchy

Cortical oscillations in different frequency ranges have different spatiotemporal coverage constrained by the maximal conduction delays [29]. Commonly observed and studied are the delta (1–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–30 Hz), and gamma (>30 Hz) oscillations as well as finer subdivisions within each band. For these neural oscillators, each oscillatory cycle corresponds to a temporal processing window that determines the transferring period and the neuronal pool involved [30]. This suggests that different frequencies favor different types of connections and different levels of computation.

Delta oscillations are a predominant slow wave that are broad spread during deep sleep, and are associated with general functions such as learning, motivation and reward [31, 32].

Theta oscillations are prominent in the hippocampus and frefrontal cortex during exploratory movements [33], and is associated with the encoding of working memory, and the retrieval of episodic and spatial memory [34, 35].

Alpha oscillations are most dominant under conditions of relative mental inactivity. It is blocked or desynchronized by sensory inputs, attention and mental effort [36]. The archetypal alpha desynchronization are in the occipital lobe and the frontal eye fields during eye opening, eye movement and visual imaginary. Moreover, alpha desynchronization has also been found in the sensory-motor regions (also called mu rhythm) [37] and auditory regions (also called tau rhythm) [38] in respond to motor and auditory activities.

Beta oscillations tend to synchronize in sensorimotor regions during stable postures [29] but desynchronize at the initiation of a movement [39], then resynchronize after the movement termination [40]. Thus, beta rhythm is associated with preparation and inhibitory control in the motor system [41].

Gamma oscillations are considered as a fundamental process that subserves the elemental operation of cortical computation. It could be induced by various stimuli or tasks, thus is related to multiple cognitive capacities [42].

2.2.2 Oscillatory synchronization

Oscillatory synchronization is considered as a core mechanism for integrating spatially distributed neural populations into a coherent functioning network [43]. Specifically, for effective communication between two groups of neurons, the sending group must precisely time its output so that the signal arrives when the receiving group is most excitable. Otherwise, the communication channel would be closed. This is known as the oscillatory gating hypothesis [44] or communication-through-coherence (CTC) [23]. It also suggests the predictive ability of neural oscillations [45]. There are two chief mechanisms for generating oscillatory synchronization, namely the amplitude modulation (stimulus-evoked response) and Phase modulation (phase-resetting) [46].

The amplitude modulation assumes that the stimulus drives the response of additive neuronal population, such as the visual- [47] or auditory-evoked potentials [48].

The phase modulation assumes that sensory stimulus induces phase resetting of ongoing rhythms without increasing the oscillatory amplitude, such as the somatosensory [48].

2.2.3 Oscillation functions

Neuronal oscillations have many properties that are suitable for them to organize cortical activities: (1) Oscillations can affect local computation by modulating both the output spike timing and the input synaptic sensitivity [23]. (2) Oscillations can mediate long-range communication by synchronizing the phase relevant neuronal populations [23]. (3) Oscillations can selectively enhance or suppress information transformation by adjusting the entrainment of external input and internal expectation or attention [49]. (4) Oscillations can form perceptual binding by packaging activities in different frequency bands via cross-frequency coupling (CFC) [50, 51]. (5) Oscillations can also achieve predictive coding considering its predictive nature in rhythmicity [52].

2.2.4 Cross-frequency coupling

Cross-frequency coupling is a phenomena that used to describe the statistical relationships between two oscillatory frequencies. CFC can take different forms as it associates with certain types of neural oscillators and inter-site communication needs, and can be used to estimate local organization (when measured at a single electrode) and long-range connectivity (when measured from different electrodes) [53, 54]. Currently, the mostly studied CFC forms include: (1) phase–amplitude coupling (PAC) [55], in which the phase of slow oscillation (SO) modulates the amplitude of fast oscillation (FO). PAC has been verified in physiological, computational and theoretical simulations [56, 57, 58] and associated with multi-item representation, long-distance communication, sensory stimulus parsing, etc [54]. (2) amplitude–amplitude coupling (AAC) [35], in which the amplitude of SO modulates the amplitude of FO. AAC occurs on top of PAC when the SO shape is asymmetric, either positive or negative [56]. (3) phase–phase coupling (PPC) [23] or phase synchronization, emphasize the phasic relationship either in the same or different frequency bands. In a synchronized network, PPC may not happen exactly at the same time, but with a certain time delay, which can be used to speculate the information flow directions [59]. These various forms of CFC are helpful for linking across frequency bands, distributed regions and oscillatory patterns, thus worth to be further explored for the discovery of neural organizational principles.

2.3 Neural mechanisms of speech comprehension

This section introduces the neural mechanisms of speech comprehension, including its cortical organization and the neuroanatomical model, specifically, the dual-stream model.

2.3.1 Cortical organization of speech comprehension

The cortical organization of speech comprehension involves multiple levels of computation and representation for distinctive features, such as phonemic segments, syllabic units, semantic items, and syntactic structures [60]. The earliest processing of **acoustic codes** is supposed to be conducted in the Heschl's gyrus, supratemporal plane, and ventro-lateral portions of the superior temporal gyrus (STG) and superior temporal sulcus (STS) [61, 62]. The **Phonological processing** showed increased activity in the supramarginal gyri, the left precentral sulcus, and the inferior frontal cortex [63]. The **semantic meaning** involved broad cortical regions for the representation of specific sensorimotor features [64, 14]. lexical-level representation is supposed to be associated in the posterior middle and inferior temporal regions (pMTG/pITG). Besides, there is a more focally organized semantic 'hub' in the anterior temporal region (ATL) that is related to syntactic-level semantic association [65]. Studies on the functional neuroanatomy of **syntactic processes** demonstrate involvement of the inferior frontal cortex (IFC) and the anterior portion of the temporal cortex (ATL) [66].

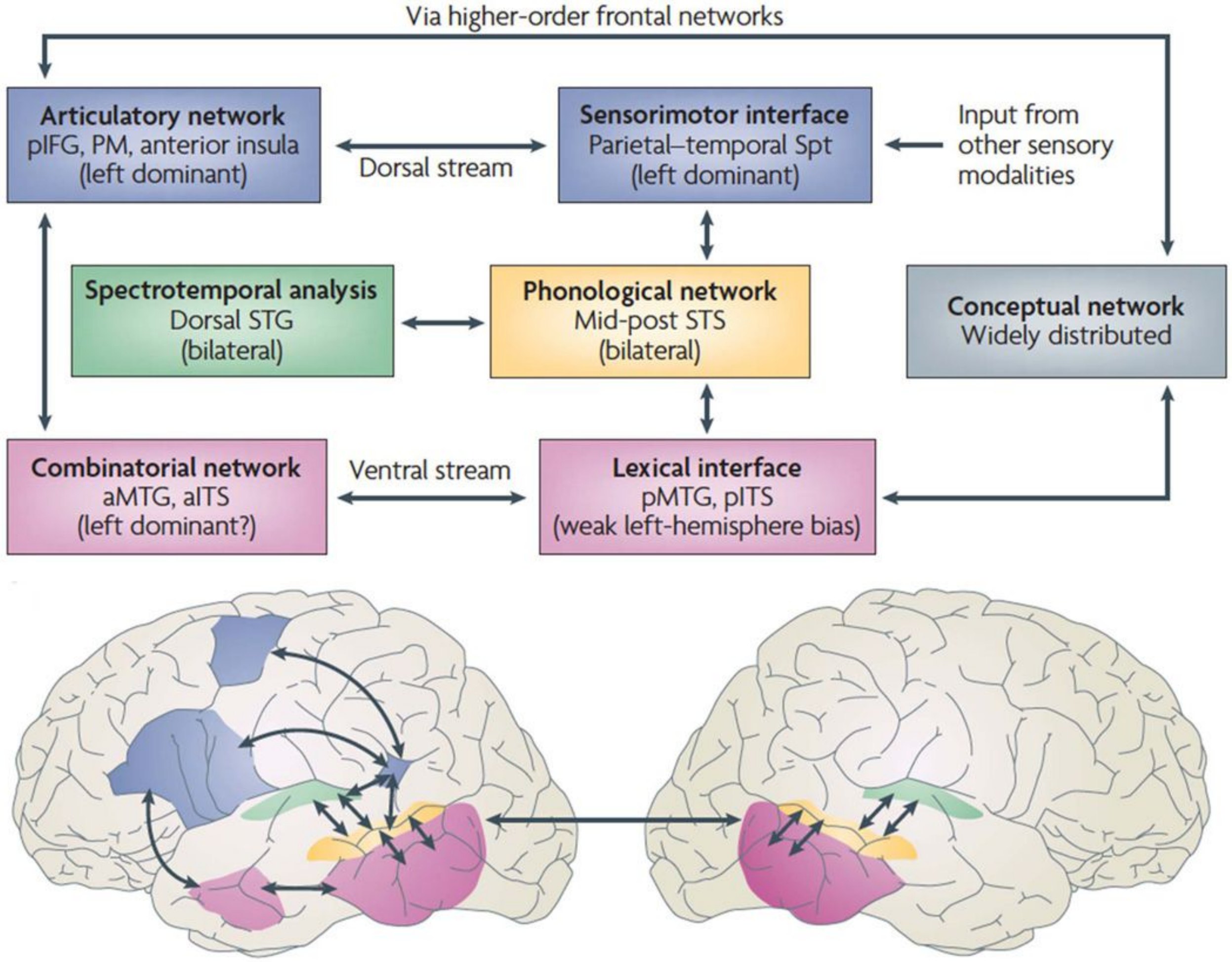


Figure 2.2: The dual stream model. IFG, inferior frontal gyrus; ITS, inferior temporal sulcus; MTG, middle temporal gyrus; PM, premotor; Spt, Sylvian parietal-temporal; STG, superior temporal gyrus; STS, superior temporal sulcus. Adapted from [68] Fig. 1.

2.3.2 The dual-stream model of speech perception

Currently, the most prevailing model for understanding the functional anatomy of speech perception and comprehension is the dual-stream model proposed by Hickok and Poeppel [67, 60, 68]. This functional anatomic framework is based on the idea that speech codes must interface with at least a semantic-conceptual system and a motor–articulatory system.

2.3.2.1 The functional anatomy of the DSM

Figure 2.2 is an illustration of the dual-stream functional anatomy. It posits that early stages of speech processing involves some form of spectrotemporal analysis and phonological-level processing, which occurs bilaterally in auditory regions on the dorsal STG (spectrotemporal analysis; green) and STS (phonological access/representation; yellow). The information flow then diverges into two broad streams: (1) a ventral stream supports speech comprehension (lexical access and combinatorial processes; pink) with the pMTG/pITL serve as an interface between sound-based representations of speech and widely distributed conceptual representations, and the ATL serves as the combinatorial network for sentence-level processing (syntactic and semantic integration processes). (2) a dorsal stream supports sensory-motor integration (speech motor programming and execution; blue), which involves structures at the parietal-temporal junction (Spt) as a sensorimotor interface, and the Broca's region and a more dorsal premotor site, correspond to portions of the articulatory network. Moreover, they also proposed a conceptual network (gray box) that is assumed to be widely distributed throughout cortex.

2.3.2.2 The bi-directioality of the DSM

The dual-stream model also propose some degree of bi-directionality in both the dorsal and ventral pathways. In the ventral stream, the pMTG/pITL networks mediate the relation between sound and meaning both for perception and production. In the dorsal stream, the Spt regions mediate the relation between auditory and motor representations. These two sensory-motor interfaces forms the basis of feedforward and feedback loops at the psycholinguistic and motor control levels [69].

2.3.2.3 The perception-production interaction in the DSM

The dual-stream model, a network of regions that include auditory areas in the superior temporal sulcus, motor areas in the left inferior frontal gyrus (parts of Broca's area), a more dorsal left premotor site, and an area in the left planum temporale region (referred to as area Spt) has been identified as a sensory-motor integration system that transforms between sensory and motor speech representations [67, 60, 68]. Recent progress has revealed why this sensory-motor link is so critical - motor acts aim to hit sensory targets, which is supposed to be the internal representations of the sound patterns. And the speech production is heavily rely on the semantic comprehension of the articulatory targets. The involvement of the production and comprehension systems, on the one hand, depends on the extent to which that task involves mapping between auditory and motor systems. On the other hand, depends on the strategies employed by individual subjects. A task which ostensibly involves only comprehension (say, passive sentence listening in a functional activation experiment) will primarily drive bilateral auditory, ventral stream areas, but may additionally recruit dorsal stream mechanisms if the subject uses articulatory re-mapping as an aid in task performance [67].

2.3.3 The neurocognitive model of sentence comprehension

In a neurocognitive model of sentence comprehension propose by Friederici [66], the temporal characteristics of auditory processing, segmental and suprasegmental phonological processing (prosody or pitch) as well as syntactic and semantic information are characterised as three phases. A schematic view of the model is given in Figure 2.3. In phase 1 (100–300 ms), the initial syntactic structure is formed on the basis of information about the word category. In phase 2 (300–500 ms), lexical-semantic and morphosyntactic processes take place with the goal of thematic role assignment. In phase 3 (500–1000 ms), the different (syntactic and semantic) types of information are integrated. In terms of the interactions between the syntactic and semantic processes, there are two alternative views among psycholinguistic discussions. The serial or syntax-first model holds that syntax is processed autonomously prior to semantic information [70]. In contrast, the interactive or constraint-satisfaction models claims that all types of information interact at each stage of language comprehension [71,

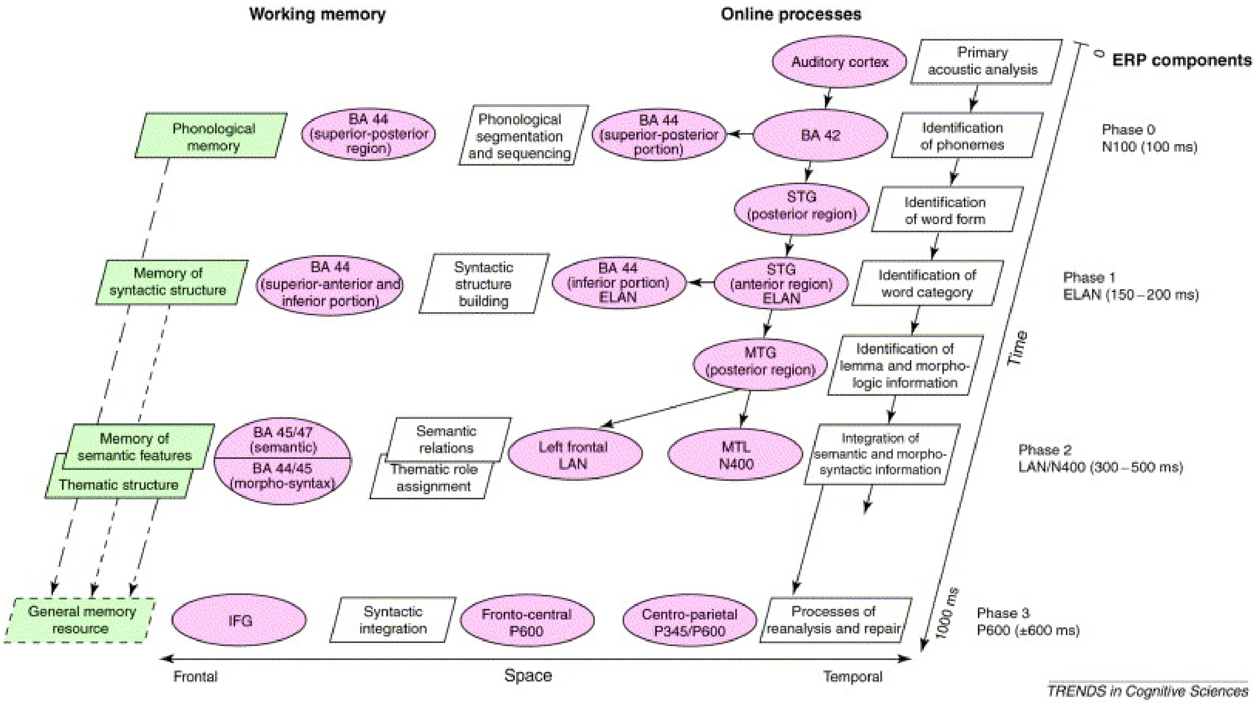


Figure 2.3: Neurocognitive model of auditory sentence processing. Adapted from [66] Fig. 1

72]. The Friederici model introduced here is compatible with both syntax-first models and interactive models. Although the building of the syntactic-phrase structure is autonomous and precedes semantic processes in the early-time windows, these processes also interact in the late-time window. From this perspective, both autonomous processing and interactive processing, hold in principle, but describe different processing phases during language comprehension.

2.4 Neural mechanisms of speech production

This section introduces the neural mechanisms of speech production, including the cortical organization of speech production and the neuroanatomical model of speech production (feedforward and feedback control models).

2.4.1 Cortical organization of speech production

The cortical organization of speech production has much overlapping with that for speech comprehension, in which the **phonological processing for speech production** has been implicated in the left posterior superior temporal regions, also known as the area Spt [67]. The **sensory-motor integration network for the vocal tract** involved tight connections between the the pars opercularis (BA 44) and the Spt regions [73]. The **motor system for speech production** involves the prefrontal lobe (PFC), the ventral premotor cortex (vPMC), the supplemantary motor cortex (SMA) and the primary motor regions (M1). Besides, some subcortical structures are also involved, such as the basal ganglia and cerebellum [74, 75]. In addition, the **auditory feedback** would recruit the posterior superior temporal gyrus (STG) and the superior parietal temporal area (Spt) [73]

2.4.2 Feedforward model of speech production

2.4.2.1 The WEAVER++ model

The WEAVER++ (Word-form Encoding by Activation and VERification) model is a typical example of feedforward lexical access in speech production [76]. An illustration of the WEAVER++ model is given in Figure 2.4, in which word generation proceeds through several processing stages: (1) Conceptual preparation activates the lexical concept via different forms of verbalization. During this stage, not only proximal and semantic, but also distal, pragmatic, context as well as environmental facts need to be considered. (2) Lexical selection selects function words based on syntactic grounds for further grammatical encoding. (3) Morphophonological encoding and syllabification prepare the appropriate articulatory gestures for the selected word in its prosodic context. (4) Phonetic encoding computes the phonological gesture of the articulatory target in its abstract forms. (5) Articulation process controls the execution of abstract gestural scores as well as a muscular machinery that controls lungs, larynx and vocal tract. (6) Self-monitoring detect errors, dysfluencies, or other problems of delivery in our own overt speech. In general, the WEAVER++ model is serially organized which may be theoretically reasonable but face problems in error correction and prediction.

2.4.3 State feedback control model

Recent speech production models, by adapting the concept of state motor control, have begun to investigate the computational advantages of using a set of internal feedforward prediction and feedback conrol, known as the state feedback control (SFC) models [77, 78]. These SFC models capture the forward or causal relationship between the input and the output of the system by introducing an efferent copy of the motor command to anticipate and cancel the sensory effects of a given movement. This structure allows prediction and feedback delays, thus could maintain stability and be of computational use in motor learning.

2.4.3.1 The SFC model

An illustration of the state feedback control (SFC) model with putative neural substrate has been shown in Figure 2.5 and Figure 2.6. In the anatomical structure, the primary motor cortex (M1) generates neuromuscular controls applied to brainstem and spinal cord and enable fine motor movements in the vocal tract [79]. After articulation, the auditory and somatosensory feedback first reaches the primary auditory (A1) and somatosensory (S1) cortices, then transferred to the higher order auditory (Spt/PT) and somatosensory (S2/PV) cortices, respectively. The premotor cortices, with bidirectional connection to both the motor cortex and the higher order auditory (Spt/PT) and somatosensory (S2/PV) cortices is ideal for an intermediary role: it receives both the motor command copy of from M1 and sensory feedback from Spt and S2, and computes the errors in between for correction in following speech motor programming. Conform to the general idea of the state feedback control models, the SFC model includes an internal forward model for the prediction of the sensory consequences of the motor command. For the state feedback, it incorporates two-levels of

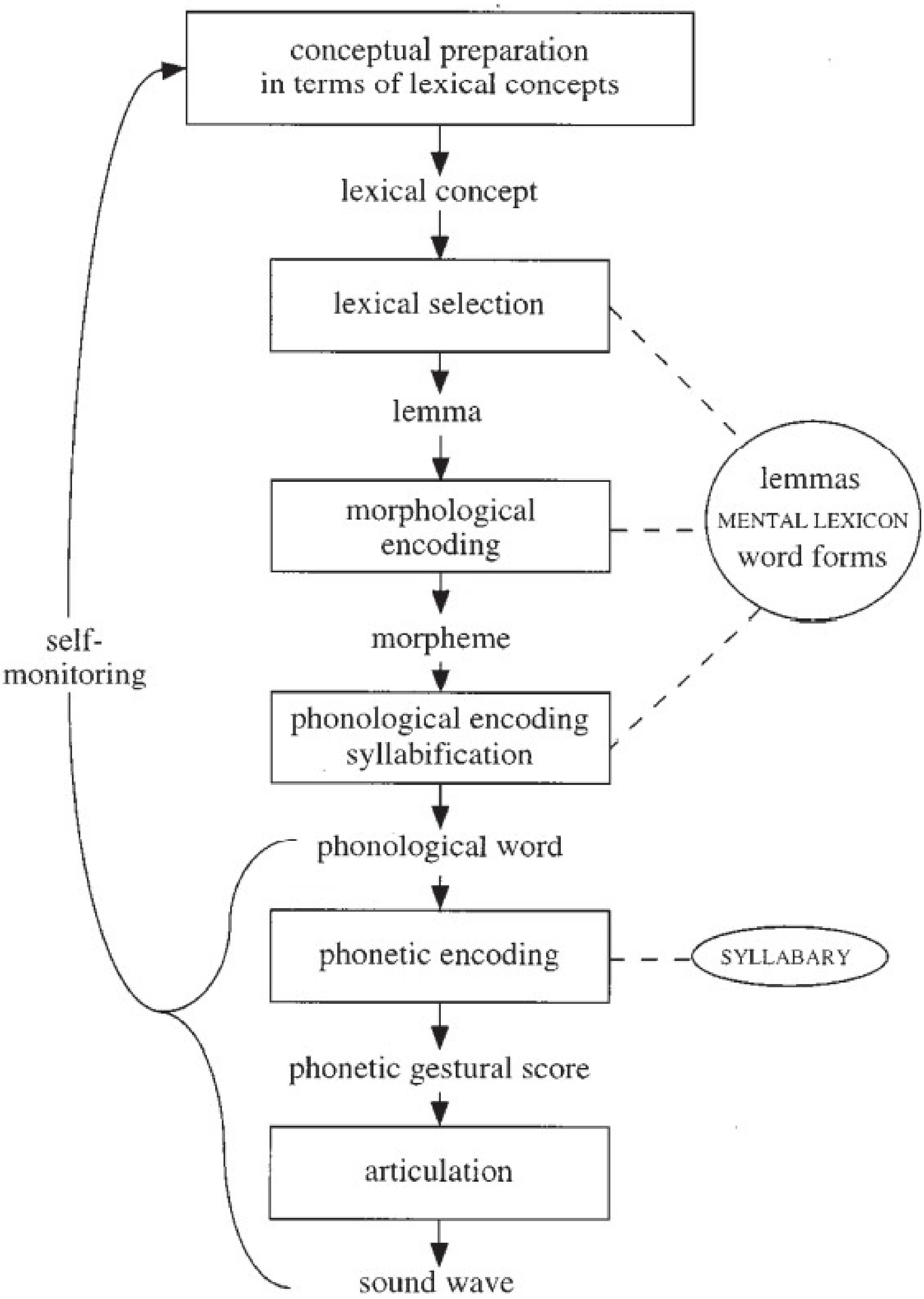


Figure 2.4: Illustration of the The WEAVER++ model. Adapted from [76] Fig. 1

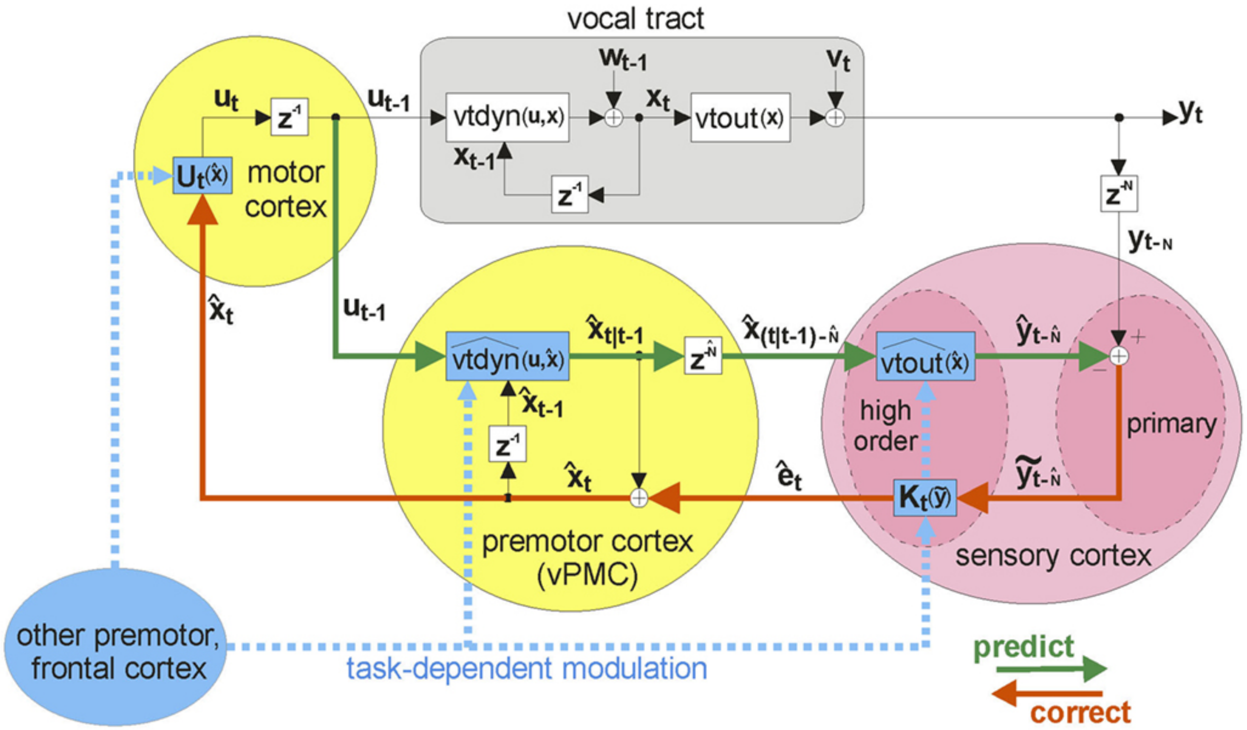


Figure 2.5: Illustration of the state feedback control (SFC) model with putative neural substrate. Adapted from [74] Figure. 6

control, namely a lexical–conceptual level and a phonological level. This framework laid the foundation for speech production at the lexical levels, however, the sentence level of organization remains to be complemented.

2.4.3.2 The HSFC model

An extension of the SFC model with multiple levels of control was proposed as the hierarchical state feedback control model (HSFC) of speech production [80]. Its organization spans the phonetic–phonological, morphological and phrase-level units and included a conceptual system in its organization (Figure 2.7). In essence, the HSFC model still focuses at two levels: the lower-level articulatory control and the higher-level linguistic processing. Specifically, the lower-level of feedback control involves a sensory-motor loop that includes sensory targets coded primarily in somatosensory cortex, motor programs coded in lower primary motor cortex (M1), and a cerebellar circuit mediating the relation between the two. It codes

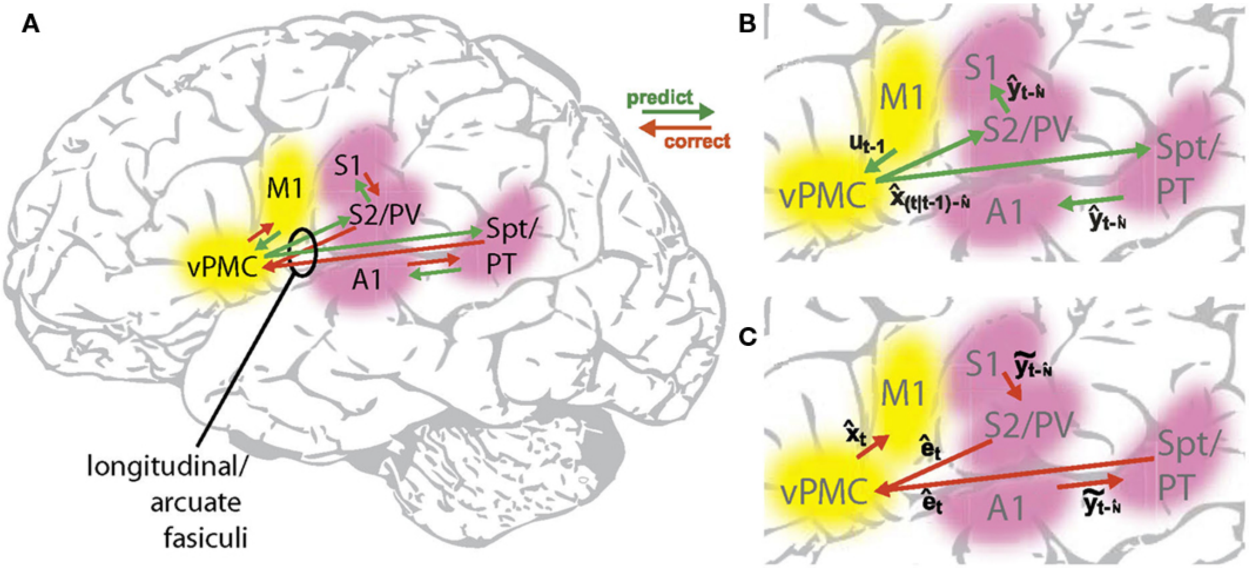


Figure 2.6: Cortical substrate of SFC model. (A) Anatomical locations of candidate cortical areas and white matter tracts comprising network of the core SFC model. (B) Cortical connections in the prediction (green) direction. Adapted from [74] Figure. 7

speech information at the level of articulatory feature clusters that roughly corresponds to phonemes. Whereas the higher-level of speech feedback control involves a sensory-motor loop that includes sensory targets in auditory cortex, motor programs coded in the Brodmann area (BA) 44 portion of Broca's area and lower BA6, and area Spt, which computes a coordinate transform between the sensory and motor areas, and it codes predominately at the syllable level. Unfortunately, the conceptual system is still a black box without further expansion in its component and relationship with sentence processing.

2.5 Summary

Although the dual-stream models and the HSFC model may have provided us with a general framework of neural anatomy and functional organization of speech comprehension and production. One common problems of these two models is that they focused their scope at the lexical level and left the sentence construction underexplored. Surely, simplicity is one consideration for choosing the lexical unit as a starting point for our essential understanding

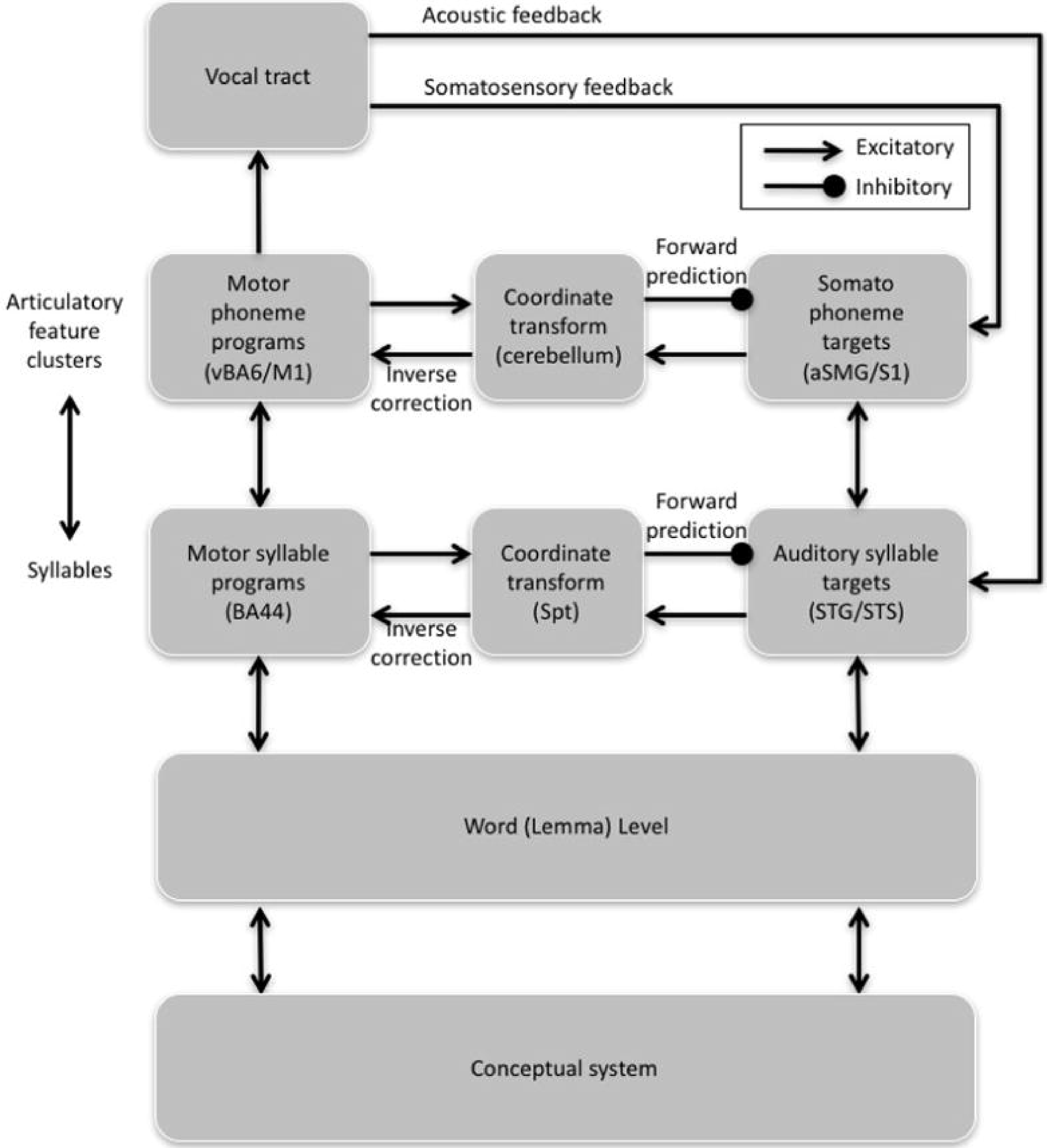


Figure 2.7: Illustration of the hierarchical state feedback control (HSFC) model. BA, Brodmann area; M1, primary motor cortex; S1, primary somatosensory area; aSMG, anterior supramarginal gyrus; STG, superior temporal gyrus; STS, superior temporal sulcus; vBA6, ventral BA6. Adapted from [80] Figure. 4

of language processing. Other obstacles impeding the progress in sentence-related neurophysiological researches may also include the artifact contamination of orofacial movement in the electrophysiological signals [81, 82] and the further complexity introduced by the syntactic structure and semantic associations in the linguistic hierarchy [83], which we aims to overcome in this study.

Another weakness of the these two models is that these they only explained from an 'observational' perspective instead of an 'mechanical' perspective. In other words, they did not trace back to the neural oscillatory mechanism that underlying the spatiotemporal brain dynamic organization. Thus, in our study, we will try to make up this gap by probing into the neural oscillatory details and integrating the temporal-spatial-spectral dynamics of functional transformations among specialized language modules in a systematic framework [84, 43].

Chapter3

Experiments and Methods

In our research, we conducted two experiments for speech comprehension and production, respectively. This chapter introduces generally the experimental design, equipment and procedures for collecting the multi-modal data and the data analysis methods. Chapter 4 and Chapter 5 will elaborate these two experiments in more details.

3.1 Experimental design

Two experiments were designed with different levels (phonological, lexical, sentential) of speech material and speech functioning tasks (listening and speaking) for different purposes.

Experiment 1 - The speech comprehension experiment is a passive listening task with word items as auditory materials. These word items appear as synonym pairs with similar semantic meanings but different phonetic features (articulatory gestures).Ones that require stronger lip-rounding gesture (Lip-R) for articulating while the others do not (Lip-N). The Lip-R vs. Lip-N contrasts could be used to test whether the speech motor regions for lip control would participate in and facilitate speech perception during the passive listening task. This design allows us to investigate not only the lexical semantic also the phonetic representation mechanisms. According to the motor theory of speech perception, speech motor system with particular articulatory features will also contribute to the perception of the corresponding acoustic stimuli. With our design, we could test whether the differentiated phonetic features could influence the perception of the synonym pairs by examining their brain network dynamics. Besides, it could also give us a clue regarding whether the speech production system participates in the speech perception process, and how it works.

Experiment 2 - The speech production experiment is a sentence oral reading task with syntactically consistence sentences as visual stimuli. The sentences consist of 8 disyllable words that could be further parsed into 3 clauses. This design has a two-fold consideration. At the lexical level, the grouping of the disyllable words are not visually cued but require mental resources of linguistic semantic knowledge, thus could be used in our study to test whether the subjects would retrieve actively mental lexicon in the visual processing and speech production of the word sequences. At the sentence level, the syntactic structure could also explored in the oral reading strategy by investigating the eye movement trajectory and speech segmentation and the structural pattern of the eye-voice span (EVS). The semantic and syntactic effects, if exist, could be an experimental evidence of the participation of comprehension system during the speech production process.

3.1.1 Experimental equipment

In the speech comprehension tasks (Experiment 1), EEG is the major equipment for collecting brain signals. Whereas in the speech production task (Experiment 2), EEG, eye movement, and speech signals will be simultaneously recorded. Below is an explanation of the multi-modal data acquisition system.

The experiments were carried out in a soundproof, electromagnetically shielded laboratory room. Video monitoring was applied to ensure that subjects were engaged in the task and did not make large body movements.

EEG data acquisition

Scalp EEG signals were recorded with a 128-channel Quik-Cap-SynAmps 2/RT (Neuroscan, USA) placed in accordance with the extended 10-5 system [85]. The horizontal electrooculogram (EOG) was measured with two electrodes at the outer canthi, and the vertical

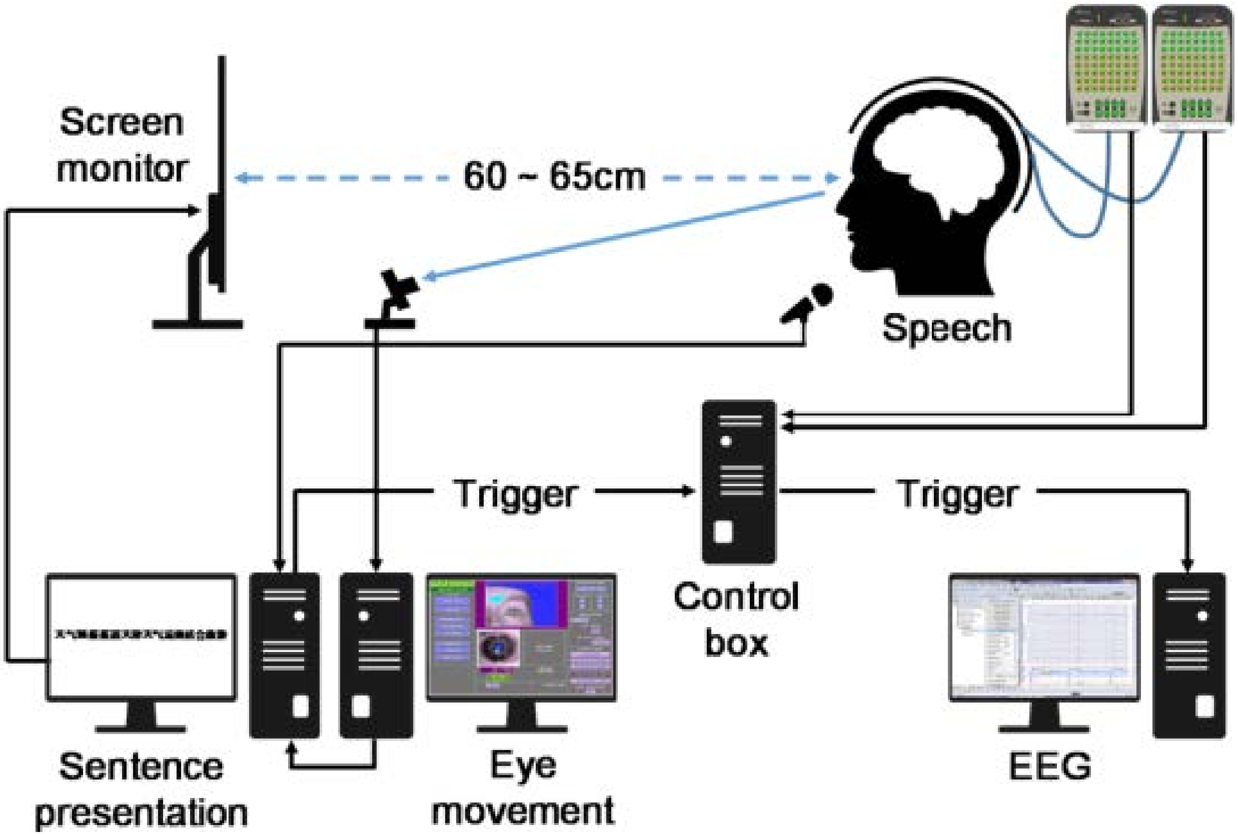


Figure 3.1: Schematic diagram of multimodal acquisition for EEG, eye movement and speech data.

EOG was measured with four electrodes placed above and below both eyes. The sampling rate was 1000 Hz, and the channel impedance was maintained below 5 kΩ throughout the acquisition.

Eye movement tracking

The eye movements were recorded via a monocular pupil tracking system (Eyelink 1000, SR Research Ltd., Mississauga, Canada) at a sampling rate of 1000 Hz. To avoid large movements, desktop support was installed for the participants to rest their foreheads on. Prior to the start of each trial, a three-point calibration was conducted to ensure accurate eye-tracking (Gaze accuracy deviation $< 0.50^{\text{o}}$). The gaze onset and offset of each word were extracted based on the location and duration of the participant's eye movements. The microsaccades and regression were eliminated in the segmentation of eye movement.

Speech signal recording

The speech signals were recorded using an electret condenser microphone (SONY ECM MS957) at a rate of 44100 Hz. The articulation of the whole sentence was typically finished within 4500 ms. The speech onset and offset of each word were segmented based on speech spectral analysis.

3.1.2 Experimental paradigm

3.1.2.1 Paradigm for Experiment 1 (speech comprehension)

In the speech comprehension task, E-prime 2.0 (PST, USA) was used for stimulus presentation and behavioral response collection. The participants were first given 32 trials for practice. Then, 320 experimental stimuli (160 disyllable words and 160 white-noise segments) were pseudo-randomly presented (no more than 3 items of the same category were in direct succession) across 8 blocks (40 trials per block). Figure 3.2 gives a general view of the procedures for each trial. At the beginning, a fixation cross appeared first in the screen center for a random duration around 400 ms. Then, a 900-ms auditory stimulus (44,100

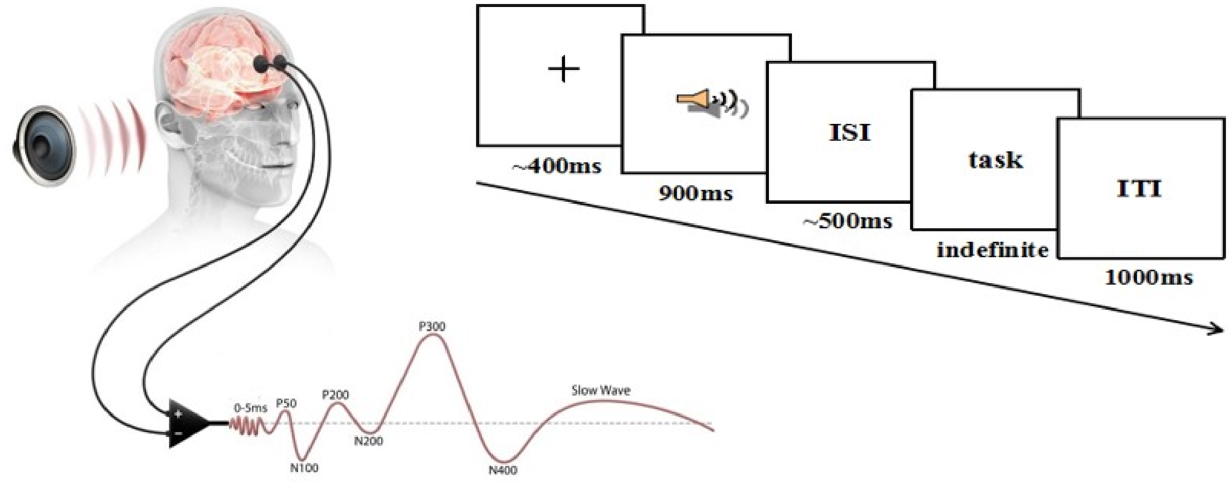


Figure 3.2: Experimental paradigm. An illustration of the experimental procedure for each trial. Orderly, a 400-ms fixation cross, a 900-ms auditory stimulus, a 500-ms inter-stimulus interval (ISI), a quick categorization judgment and a 1000-ms inter-trial interval (ITI). Adapted from [86] Fig. 2

Hz, 16-bit depth) was presented through headphones. After a 500-ms inter-stimulus interval (ISI), participants were asked to make a quick categorization judgment by clicking key “1” for all the disyllable (two-character) words and key “2” for the white-noise segments. In the end, a 1000-ms inter-trial interval (ITI) was arranged before the next trial. The entire experiment was limited within 30 min to prevent a learning effect.

3.1.2.2 Paradigm for Experiment 2 (speech production)

In the speech production task of sentence oral reading, the sentence presentation and behavioral (eye movement and speech) response collection were conducted with Psychtoolbox-3 (www.psychtoolbox.org) running under MATLAB R2018b (MathWorks). Figure 3.3 gives a general view of the procedures for each trial. At the beginning of each experimental trial, a three-point calibration was conducted optionally when eye-tracking got lost or shifted. After 2000 ms of a brain resting period, a cross appeared in the screen center for about 1000ms to attract subject’s attention and fixation. Once a sentence was presented on the screen as a single line, the subjects were asked to scan from left to right and recite it in a natural speed.

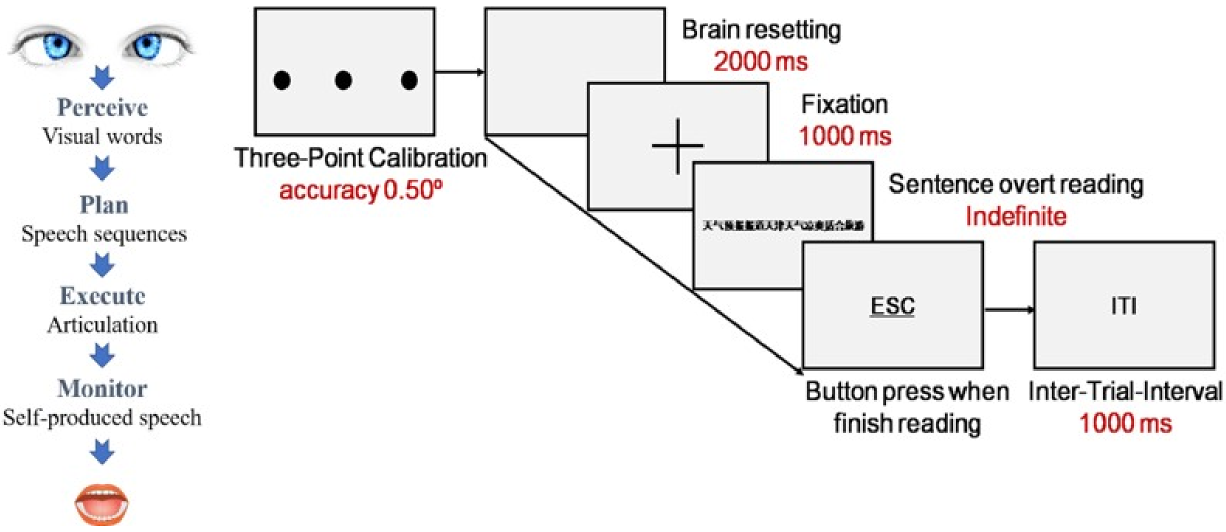


Figure 3.3: Experimental paradigm. An illustration of the experimental procedure for each trial. Orderly, a three-point calibration, a 2000-ms brain resetting period, a 1000-ms fixation cross, an indefinite period for sentence oral reading, a button 'ESC' press and a 1000-ms inter-trial interval (ITI).

Specifically, whenever the subject's gazing point fell in one of the 8 word fields, a trigger with the corresponding number (1-8) would be sent to EEG signals. It worth noting that the sentences were presented on the screen with no delimiter between the characters. That means when the subjects read aloud the sentence, there were no obvious cues for parsing out the sentence structure unless the participant resorted to the internal memories of their semantic and syntactic knowledge. When the reading is finished, the subjects press the 'ESC' button and skip to the next sentence after a 1000 ms inter-trial-interval (ITI). Each of the 60 sentences was repeated three times for a total of 180 trials in random order. The whole procedure varied from 52 to 88 minutes individually, with EEG, eye movement and speech data recorded in sync during the whole-range.

3.2 Analysis methods

3.2.1 EEG Data preprocessing

Since EEG signals collected at the scalp are a mixture of both brain source activities and artifacts from various sources, some preprocessing steps are needed to reduce unwanted noises before calculating behaviorally relevant EEG measures.

3.2.1.1 Re-sampling

EEG signals are usually recorded at relatively high sampling rate (e.g., 1000 Hz in our experiment). During offline analysis, we down-sampled the signals to lower frequencies (250 Hz to help with data compression and better ICA decomposition by cutting off unnecessary high-frequency information.

3.2.1.2 Filtering

Low-pass filtering at 60 Hz is performed before down-sampling to help reduce the high-frequency noise as well.

High-pass filtering at 1-Hz pass-band edge is applied in our practice to help remove the baseline drift. Usually, lower cutoff frequencies (0.01-0.1 HZ) are recommended, however, we choose 1 Hz as recommended for the following reasons:

1. Low-frequency (< 1 Hz) EEG could be contaminated by seating etc, which adds spatiotemporal non-stationarity.

2. EEG signal follows the 1/f power law. That means lower frequency components tend to have larger influences on later ICA decomposition results. Thus cutting off the leftmost values helps to focus on other frequencies of interests.

3. In time-frequnency analysis, a decent frequency precision often require several cycles of sliding window length (in EEGLAB parameters, 3 cycles). So the lowest frequency that

can be derived is also restricted by the available temporal window length.

3.2.1.3 Artifact rejection

Remove bad channels. Bad channels ($5\% \pm 3\%$) with more than 5 seconds of flat lines or having the correlation coefficient with surrounding channels below 0.8 were removed. This is a critical step particularly for later average referencing. Because including channels that are noisy into the average step will introduce noise to all the other channels.

Artifact subspace reconstruction. To removes transient and large-amplitude artifacts and reconstruct relatively clean data from the remaining components, we further applied the artifact subspace reconstruction (ASR) method that comes from the EEGLAB plugni clean-rawdata [87]. The ASR procedure consists three main steps: (1) automatically select clean portions of EEG by calculating channel-wise root-mean-square (RMS) values on each 1-s window, z-score the values, identifies clean windows with z-scored values between -3.5 and 5.5, and concatenates the clean windows to obtain the calibration data X_c; (2) determine the rejection criteria in the principal component (PC) space of the calibration data $Y_c = V_c^T \cdot X_c$ and select the threshold of rejection criteria with the formula $T_i = \mu_i + k \cdot \sigma_i$, where T_i is the threshold for each component i, μ_i and σ_i are the mean and standard deviation of RMS values across all 0.5–s windows of $Y - c$, and k is a user-defined cutoff parameter, which is set as 20 in our case as suggested by [88]; (3) apply a linear operator $R = VM(M \circ U)^{\dagger} V^T$ to EEG segments X along a sliding window with window size of 0.5 s and a step size of 0.25 s, such that $X_{clean} = RX$. Here $M = V^T \bar{M}$ is the projected matrix square root of the geometric median covariance of X_c, such that $\bar{M}\bar{M}^T = Cov(X_c)$. U is a threshold operator where $U_i = 0$ if $\sigma_i > T_i$, otherwise $U_i = 1$. Note that ASR is not simply rejecting PCs but rejecting projected PCs defined by $V^T \bar{M}$.

Remove line noise. To remove the 50-Hz line noise, we applied CleanLine, an EEGLAB plugin that uses shifting window to adaptively estimate sine wave amplitude to subtract without making a hole in the background EEG spectrum.

3.2.1.4 Channel Interpolation

After channel removal and artifact rejection, the missing channels were spatially reconstructed from the activity of neighboring channels using a Gaussian spline function. This could help to minimize potential bias in the next average referencing stage.

3.2.1.5 Reference

After interpolating the removed channels, the EEG data were average-referenced. Compared to other reference methods (e.g., mastoid reference), average reference is independent of reference locations and is based on the assumption that when EEG is generated on the cortex with dipolar current distribution without external sources, hence the scalp topography should sum to zero.

3.2.1.6 Epoch

To focus on peri-event brain dynamics while not being disturbed by data points from period of non-interest, data epoch were implemented around the stimulus onset with a pretrial segment as baseline and a post-trial segment that include the event-related activities. In the case of the speech comprehension task, the period of -100–1000 ms were epoched, and in the case of speech production tasks, -500–4500 ms segments were epoched.

3.2.1.7 Baseline correction

Baseline correction that subtract the mean of the baseline period (-100-0 ms for the speech comprehension task, and -500–0 ns for the speech production task), is selectively conducted based on the later analyses methods applied. One consideration is that mean value of such short baseline can easily fluctuate by background rhythm that changes the DC levels of all the channels independently.

3.2.2 Time-domain analyses

3.2.2.1 Event-related potentials, ERP

ERP analysis is a predominant method in traditional EEG studies and is still commonly practice today [89]. The mathematical computation of the ERP is imply by averaging the voltage at each time point over trials. The logic behind assumes that evoked potentials are stable while the noise is random. Thus averaging trials would cancels out unwanted noises. The ERP algorithm is pretty simple and has high temporal precision and accuracy. However, only those brain activities that are coherent both in time and phase could be remained in the ERP results, whereas those quasi-rhythmic brain activities not exactly synchronized in both time and phase tend to be averaged out through phase cancellation [89]. Beside, the neurophysiological mechanisms that produce ERPs are less well understood, leaving it difficult to link ERP results with physiological relevance. So in our analyses, ERPs are only used to identify critical temporal windows and compare with previous studies instead of interpreting its physical meaning and neural mechanisms.

3.2.2.2 ERP images – 2-D ERP

The ERP image is a 2-D extension of the ERP results with each single trial stacked vertically and color coded to show it amplitude changes. ERP images can be used to link trial-varying task parameters or behaviors to the time-domain EEG signal by sorting the EEG trials according to values of the aligning event, such as the reaction time or the phase of a frequency-band-specific signal at a certain time point [90]. In our analysis, ERP-images are used to reveal single trial dynamics that buried in the traditional ERP averaging.

3.2.3 Time-frequency-domain analyses

3.2.3.1 Fourier transform and Morlet wavelets

One way to visualize the non-phased-locked signaled buried in the average ERP results is to transform the EEG time series to the frequency domain by using Fourier transform. However, the Fourier-transform-based frequency representation could not capture the dynamic frequency changes in the EEG signals. For this reason, we used the Morlet wavelet, simply a sine wave windowed with a Gaussian, as formulated below:

$$GaussWin = ae^{-t^2/(2s^2)} \tag{3.1}$$

where a represents the amplitude of the Gaussian, t is the time, and s is the standard deviation or the width of the Gaussian.

For the sine wave, we use its complex form, and by multiplying it with the Gaussian window, we created the complex Morlet wavelet with the following equation:

$$cmv = Ae^{-t^2/2s^2}e^{i2\pi ft} \tag{3.2}$$

$$A = \frac{1}{(s\sqrt{\pi})^{1/2}} \tag{3.3}$$

where f is the peak frequency of the wavelet. A is a frequency band-specific scaling factor. s is the standard deviation of the Gaussian. After the Morlet wavelets convolution, the real value represents the bandpass-filtered signal. The magnitude of the vector is related to the similarity or overlap between the kernel and the EEG signal. And the angle of that vector represents the phase angle at the peak frequency and the center time point of the wavelet.

Event-related spectral perturbation analysis

By using the complex morlet wavelet, we could extract broadband time-frequency activities for different frequencies. However, these frequencies conform to the 1/f power scaling law, meaning the frequency spectrum of data tends to show decreasing power at increasing frequencies. To convert the time-frequency power to a comparable scale that is amenable to qualitative visual inspection and quantitative statistical analysis, we normalized the frequency-specific post-stimulus power spectra against their respective mean baseline spectra using decibel conversion:

$$dB_{tf} = 10 \log 10 \left(\frac{activity_{tf}}{baseline_{f}} \right) \tag{3.4}$$

in which t and f are time and frequency points. The *baseline* without t subscription indicates the mean across the baseline period. The decibel (dB) conversion is robust to many limitations of the raw power, such as the 1/f power scaling and subject- or electrode-specific idiosyncratic characteristics. After the baseline normalization, task-related changes in the post-stimulus period can also be disentangled from background activity, thus known as event-related spectral perturbation (ERSP) [46].

3.2.4 Source reconstruction analyses

The main purpose of source reconstruction is to identify the cortical sources that contribute to the EEG signals recorded at the scalp. To reconstruct the cortical sources, two steps are involved: (1) separation of the cortical independent component. (2) localization of the physical coordinates. We will approaches these two steps using independent component analysis (ICA) and boundary element method (BEM) as described below.

3.2.4.1 Independent component analysis (ICA)

The scalp-recorded EEG signals are a mixture of multiple brain sources with severe contamination by eye movements, blinks, muscle, heart and line noise, etc. This causes a serious problem for EEG interpretation and analysis. A preferable treatment to this issue is to apply ICA to multichannel EEG recordings and remove a wide variety of artifacts from EEG records by eliminating the contributions of artifactual sources onto the scalp sensors. ICA works by finding a matrix W and a vector w, so that the elements $u = [u_1, ...u_N]^T$ of the linear transformation $u = Wx + w$ of the random vector $u = [x_1, ...x_N]^T$ are statistically independent. In contrast with decorrelation techniques such as PCA which ensures that $< u_i u_j >= 0, \forall ij$, ICA impose the much stronger criterion that the multivariate probability density function (p.d.f) of u factorizes:

$$f_u(u) = \prod_{i=1}^{N} f_{u_i}(u_i) \tag{3.5}$$

Finding such a factorization involves making the mutual information between the u_i goes to zero: $I(u_i, u_j) = 0, \forall ij$. Mutual information is a measure which depends on all higher-order statistics of the u_i while decorrelation only take account of 2nd-order statistics.

3.2.4.2 IC selection

The function of ICA analysis is double folded. It could be used to find brain sources as well as remove artifacts embedded in the data (muscle, eye blinks, or eye movements). The main criteria to determine if a component is cognitively related or an artifact include: (1) scalp topography, (2) time series , (3) power spectrum, (4) dipole location, (5) ERP image (6) residual variance (RV), (7) Percent data variance accounted for. Below are some prominent features of typical artifacts and brain components that we use as the criteria for classifying components:

* Eye blinks/saccade (Electro-Oculogram, EOG): Scalp map with a strong far-frontal

projection, smoothly decreasing EEG spectrum. Vertical eye movement components will contain blinks in the data. Horizontal eye movement components will look like step functions.

* Muscle activities (electromyography, EMG): Power concentrated in higher frequencies (>20 Hz). Source can still be dipolar, but will be located outside the skull.

* Heart beat (electrocardiography, ECG or EKG): Clear QRS complex at about 1 Hz. Near linear gradient scalp topography. No peaks in power spectrum.

* Line noise: 50-Hz or 60-Hz peak in the power spectrum.

* Channel noise: Very focal scalp topography. Typically have a 1/f spectrum.

* Brain components: dipolar scalp topography, low (<15 %) residual variance, Power spectrum peaks usually between 5 and 30 Hz, most often 10 Hz.

In EEGLAB, the ICLabel plugin is an automated independent component classifier [91], which has been trained on thousands of labeled ICs by experts and hundreds of thousands of unlabeled ICs and is used in our practice for IC selection.

3.2.4.3 Source localization

The necessity for source localization is caused by volume conduction. The cortical generated electric or magnetic need to go through biological tissue to reach the measurement sensors. The reconstruction of the cortical sources from the scalp-recorded signals is known as the inverse problem [92]. To start with the inverse process, the electrical conduction properties of the human head (the volume conductor) have to be modeled.

Model setting and co-registration

A head model defines different regions of brain conductivities. In our analysis, we used the boundary element model (BEM) with three 3-D surface (skin, skull, cortex) extracted from the MNI (Montreal Neurological Institute) canonical template brain. The 128 electrode locations are then co-registered to the template electrode locations of the MNI average head model. Another model to choose is the source model, which defines the type of source to use. It could be a single dipole with optimized position (DIPFIT), or a distributed model with one dipole per brain voxel (Loreta).

Dipple fitting estimates activity at a single point or a small number of points in the brain that explain the maximal amount of topographical variance. There are two steps required to create equivalent dipole models for independent components: (1) Coarse-grained grid scanning involves scanning possible positions in a coarse 3-D grid to determine an acceptable starting point for fitting equivalent dipoles to each component. (2) Non-linear interactive fitting involves running an optimization algorithm to find the best position for each equivalent dipole. Once the dipole location, orientation, and magnitude are estimated, a weight for each electrode can be determined such that the weighted sum of activity at all electrodes is the estimate of the dipole activity.

Distributed source imaging estimates the magnitudes of thousands of dipoles all over the brain with fixed locations and orientations. LORETA and minimum-norm estimators are two commonly used nonadaptive inverse-source imaging techniques. Nonadaptive distributed source models can be computed quickly, however, the weights for each electrode are not fine-tuned to the statistical properties of the data.

3.2.5 Brain connectivity estimation

The brain is a highly dynamic system with complex network organizations. These are different methods to estimated the brain connectivity based on different assumptions.

3.2.5.1 Phase-based connectivity

Phase-based connectivity analyses counts on the distribution of phase angle differences between two electrodes or independent components. They rely on precise temporal relationships with the assumption that coupled neural populations are synchronized in their phase. However, they do not provide compelling evidence for directionality. There are several terms that are used in the literature to describe phase-based connectivity, including phase-locking value/statistic/factor, phase synchronization, Phase correlation, and phase coherence. Mathematically, intersite phase clustering (ISPC) is a more concise description of the method without any assumptions or interpretations regarding putative neurophysiological mechanisms. The formula for ISPC is given below:

$$ISPC_f = |n^{-1} \sum_{t=1}^{n} e^{i((\phi_{xt} - \phi_{yt})}| \tag{3.6}$$

where n is the total number of time points. ϕ_{xt} and ϕ_{yt} are the phases of the oscillation in region x and y at frequency f. Typically the angle of ISPC (also called the "preferred angle") is ignored. However, when the angle of the mean vector is extracted, it can be used as supportive evidence for directionality of connectivity, although insufficient for directionality or causality.

3.2.5.2 Power-based connectivity

Power-based connectivity analyses involve correlating time-frequency power between two electrodes across time or over trials based on correlation coefficients. A commonly practiced measure of power-based connectivity is the Pearson correlation coefficient, which is defined in the formula below:

$$r = \frac{\sum_{t=1}^{n} (x_t - \bar{x})(y_t - \bar{y})}{\sqrt{\sum_{t=1}^{n} (x_t - \bar{x})^2 (y_t - \bar{y})^2}} \tag{3.7}$$

The Pearson correlation coefficient requires that the data are normally distributed. Violations of this assumption will introduce a bias into the correlation coefficient. However, trial-averaged baseline-corrected power (decibel or percentage change) is often normally distributed and therefore can be tested using Pearson correlations.

3.2.5.3 Granger causality

Granger causality (GC) is an established measure for inferring directed functional connectivity between stochastic variables based on reduction of prediction error of a putative effect predicted by a past observation [93, 94]. The GC value is calculated as the logarithm of the ratio of the error variances between a univariate model in which current values of X are predicted only from past values of X, and a bivariate (multivariate) model in which current values of X are predicted both from past values of X and from past values of Y:

$$GC = ln(\frac{var(e_x)}{var(e_{xy})}) \tag{3.8}$$

In the Granger prediction from Y to X, if Y does not contribute to X, the error variances would be the same for the uni- and bivariate models, resulting in zero GC values.

3.2.6 Cross-frequency coupling analyses

3.2.6.1 Phase-amplitude coupling (PAC)

The quantification of the PAC values used in this study is based on the modulation index [95], which test the correlation between the instantaneous phase of a SO and the instantaneous amplitude of a FO by way of clustered vectors. represents a set of vectors as one vector. The average length of the vectors with the phase angles of a SO and the amplitudes of a FO over time is considered as the measure of the raw PAC values [95, 59]:

$$PAC = |N^{-1}\sum_{n=1}^{N} A_{f_a}(n)e^{i\phi_{f_p}(n)}| \tag{3.9}$$

in which n represents the time point, N is the total number of time points, $A_{f_a}(n)$ is the amplitude of the FO at time point n, $\phi_{f_p}(n)$ is the phase angle (in radians) of the SO at time point n, and i is the imaginary operator.

To perform statistical analysis on the PAC values and reduce the confounds due to different reference or baseline correction scheme [59], we applied non-parametric permutation testing with the null hypothesis of no temporal relationship between the phases and amplitudes. It implies that the shuffled high-frequency amplitude and low-frequency phase time series will not affect the corresponding PAC values. For this reason, we temporally shifted the amplitude time series by a random temporal offset without changing the phase angle time series and then computed the PAC values according to equation 3.9. This random shifting procedure was repeated for one thousand times to generate a distribution of raw PAC values that are expected under the null hypothesis. After that, the original PAC value was compared to the distribution of the null hypothesized PAC values by subtracting the mean and dividing by the standard deviation, resulting in a standardized Z-value of PAC (PACz). Different from the raw PAC value, the PACz values are independent of the scale of the original data, thus could be used for conditional comparisons, both at the individual level and the group level.

3.2.6.2 Phase-phase coupling (PPC)

The PPC, or phase synchronization/locking, is commonly established by demonstrating a consistent phase relationship over time, trials, or participants [96]. Besides, the phase differences could also be used to speculate the information flow directions, though not causal relationships [59]. Thus in our analysis, the PPC values are calculated based on both the consistency and the value of the phases differences using the following equation:

$$ISPC_f|N^{-1}\sum_{n=1}^{N} e^{i(\phi_{x_n}-\phi_{y_n})} \tag{3.10}$$

$$PPC = |ISPC_f| * angle(ISPC_f) \tag{3.11}$$

where $ISPC_f$ is the inter-site phase clustering, N is the total number of time points. ϕ_{xn} and ϕ_{yn} are the phases of oscillation x and oscillation y at frequency f. The PPC value is calculate as the product of the magnitude of $ISPC_f$ and the angle of the $ISPC_f$. In addition, considering that the oral reading pace varies across subjects and averaging over the whole range would mess up the phase lagging consistency, we selected the EVS of the first word for calculating the PPC values where the consistency among subjects was the highest.

3.2.6.3 Amplitude-amplitude coupling (AAC)

The AAC measures the covariation of spectral power across frequency bands and cortical sites [56]. In our analysis, the AAC values were calculated based on the ERSP results in which the normalized power changes were cross-correlated between cortical sites:

$$AAC = \frac{\sum_{n=1}^{N}(f_1x_n - f_1\bar{x})(f_2y_n - f_2\bar{y})}{\sqrt{\sum_{n=1}^{N}(f_1x_n - f_1\bar{x})^2(f_2y_n - f_2\bar{y})^2}} \tag{3.12}$$

where f_1x_n (f_2y_n) represents the ERSP value of x (y) at frequency f_1 (f_2) and time point n, with $\bar{x}$ ($\bar{y}$) as the average ERSP values of x (y) over all time points N.

3.2.7 Statistical analyses

In cognitive electrophysiology, statistical analysis could be performed at two levels according to different goals and different interpretations. The within-subjects analyses calculates on the trials of each subject and provide information regarding the cross-trial variability of an effect relative to the magnitude of the effect. The Group-level analyses calculates across subjects and could provide information regarding the consistency or generality of effects among groups.

3.2.7.1 Within-subject statistical analyses

The within- subjects statistics, also called level-1 analysis, consider the trial to be the unit for analysis, and usually applied in situations: (1) with a small number of subjects. (2) when testing for the robustness of an effect. (3) when the “raw” analysis values are either uninterpretable or uncomparable across subjects [59].

3.2.7.2 Group-level analyses

Group-level statistics, also called level-2 analysis, consider the subject to be the unit for analysis. This is a more commonly used approach in cognitive electrophysiology. In practice, the trials within each subject are first average to reduce noise. Then group-level analyses are performed across subjects to test the generality.

3.2.7.3 Correction for multiple comparisons

Correcting for multiple comparisons is often necessary with electrophysiology data because of the large number of tests (over electrodes, time points, frequency bands, and so on) [59]. Different multiple comparison methods are suitable for correcting based on different goal of the statistical threshold.

Bonferroni correction assumes that the statistical tests are independent of each other and involves dividing the p-value by the number of statistical comparisons. It is perhaps the most common way to practice multiple comparison in cognitive science. However, it may be inappropriate as time-frequency results in EEG analyses are usually autocorrelated [59].

False discovery rate correction is based on both the distribution of p-values and also on the number of tests performed. It works by controlling for the probability of type I errors within a distribution of p-values. It is also inappropriate in some cases as the FDR-defined critical p-value becomes exponentially smaller with linear increases in the number of p-values above 0.05, whereas the critical p-value increases only very slightly with linear increases in the number of p-values below 0.05 [59].

Nonparametric permutation testing assumes no assumption regarding the population distribution. It is widely applicable to: (1) nonnormally distributed data; (2) exploratory data-driven analyses with multiple comparisons correction. The procedures of nonparametric permutation testing is listed below:

(1) Creating a Null-Hypothesis Distribution. This is down by condition swapping labels either for comparing conditions (discrete tests) or for correlating continuous variables (con-

tinuous tests). After hundreds or thousands of swapping, it will create a distribution of test statistic values. Under the null hypothesis, the test statistic value would be expected to be zero-centered.

(2) Determining Statistical Significance. After obtaining the distribution of statistical test values expected under the null hypothesis, the next step is to compute a p-value associated with the observed test statistic. This can be down by either counting the number of extreme null-hypothesis statistical test values divided by the total number of tests, or by z-transforming the observed test statistic value to standard deviation units of the null-hypothesis distribution.

Chapter4

Investigation on the Speech Comprehension Mechanisms

This Chapter will focus on the first experiment for the investigation of speech perception and comprehension. The issues of interest we would like to address with this experiment includes: (1) whether and how the speech production system could facilitate speech perception; and (2) what is the neural oscillatory mechanisms underlying such interactions. To answer the first question, we will adopt auditory stimuli of synonym contrasts with articulatory differences - whether the articulation requires lip rounding or not (Lip-R vs. Lip-N). Then we will look into the lip motor control regions to compare their response differences to these two stimulus types. To address the second issue, we will construct frequency-specific brain networks and inspect the time-varying network organization across different cortical regions. Based on these results, we will discuss their implication and association with the motor theory of speech perception, as well as the neural oscillatory application in further neurolinguistic studies.

4.1 Introduction

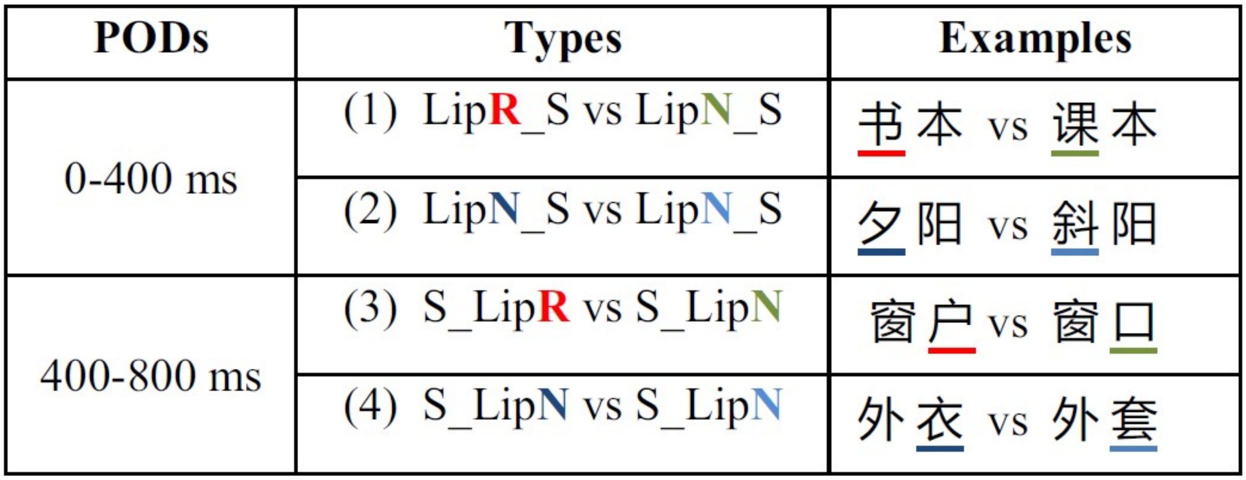
Neurobiological and psycholinguistic studies have long postulated that knowledge about articulatory features of individual phonemes has an important role in speech perception and comprehension [97, 98, 99]. One of the most intriguing and highly cited theories is the motor theory of speech perception [100, 101], which claims that the listener perceives speech by simulating ”intended articulatory gestures” of the speaker. This perception-production circuit was demonstrated by some neurobiological evidence showing that passively listening to phonemes and syllables tends to activate the motor and premotor cortex [102]. Interestingly, these activation were somatotopically organized according to the articulatory effector that recruited in the production of these phonemes [103]. As reported in an fMRI study [98], distinct motor regions in the left precentral gyrus governing articulatory movements of the lips were differentially activated when subjects listened to the lip-related phonemes. Researchers also used repetitive transcranial magnetic stimulation (TMS) to temporarily disrupt the lip representation area in the left primary motor cortex, and found that the TMS-induced disruption impaired the categorical perception of phonemes and syllables that involved lip movement in their articulation [104, 105, 106, 107]. However, in these studies, a critical component that has been ignored is the semantic interference [107]. It has been found that action-related verbs could also elicit brain activities in the somatotopic motor regions [108]. This is known as the theory of embodied semantics [109]. Lack of investigation on the semantic interference may obscure the results. Besides, the above-mentioned studies is unsatisfactory in temporal resolution, thus could not give us a clue about when does the motor contribution was involved.

To disentangle phonological motor effects from semantic ones, our study employed sets of Chinese synonyms that have near-identical or synonymous meanings but recruit different articulatory gestures. The adoption of synonymous is for minimizing the semantic influence. In addition, considering that verbs with action meaning tend to activate the frontal motor regions and cause confusions [110], all the synonyms used in this experiment were nouns that referring to relative static objects. For comparison with previous studies, we distinguished the synonyms by the recruited shape of the lips. In each pair of synonymous, one item requires lip rounding gestures (e.g., the syllable includes a labial consonant like /b/, /p/ or a vowel like /u/), and the other item require no lip rounding gestures (e.g., the syllable includes a consonant like /k/, /g/ or a vowel like /i/). We also designed control groups are have synonymous pairs, yet neither of the two items required lip gestures. To find a temporal clue for the articulatory motor effect, we utilized high resolution electroencephalograph (EEG) to detect instantaneous brain responses to the time-varying acoustic signals. The event-related potential (ERP) analysis was performed to compare group differences and detect critical temporal windows. Then EEG source reconstruction technique was employed to trace back to the sensorimotor cortex to find out whether the lip motor regions respond differently to speech sounds with different lip gestures.

4.2 Subjects and materials

Subjects. Twenty-two healthy native Mandarin (standard Chinese) speakers (12 male and 10 female participants) with the mean age of 22.3 years (range 20–24 years) were recruited from students at Tianjin University. All the participants were right-handed [111] with normal hearing and normal or corrected-to-normal vision, and reported no diagnosed history of psychiatric disorders or neurological deficits.

Materials. The auditory stimuli included four types of synonymous pairs (20 pairs or 40 items for each type) and 160 white noise segments (hissing sounds with the same frequency of the word stimuli). The synonyms were all two-character Chinese words, lasting for 800 ms (400 ms per character), and they were evaluated on their familiarity, concreteness, object-relatedness and emotional features by another 20 native Mandarin speakers, which is to make sure that each pair of the synonyms have no significant differences in the above possible influential factors ($ps > 0.05$). Figure 4.1 gives an illustration of the four types of synonyms. For type (1), the first characters of the two items in this pair are different, one of them require lip gestures (LipR) and the other one require no lip gestures (LipN). The second characters of this pair are the same. Type (2) is a contrast group for Type (1), in



PODs	Types	Examples
0-400 ms	(1) LipR_S vs LipN_S	书本 vs 课本
	(2) LipN_S vs LipN_S	夕阳 vs 斜阳
400-800 ms	(3) S_LipR vs S_LipN	窗户 vs 窗口
	(4) S_LipN vs S_LipN	外衣 vs 外套

Figure 4.1: Illustration of the four types of synonyms.

which the first characters of the pairs are different, yet neither of them need lip gestures (LipN). Their second characters are also the same. For type (3), the first characters are the same, while the second ones differ in their lip gesture requirement (the former) or not (the latter). Type (4) is a control group of type (3), in which the first characters are the same and neither of their second characters need lip gestures. The third column of Figure 4.1 gives one example of the synonyms for each type in Chinese. As the periods of difference (PODs) for type (1) and (2) are on the first character (0-400 ms), and for type (3) and (4) are on the second character (400-800 ms), we analyzed the Type (1) vs Type (2) contrast and Type (3) vs Type (4) contrast in these two PODs respectively.

4.3 Analysis results

4.3.1 ERP analysis results

After the preprocessing steps for noise reduction of the EEG data, ERP analysis was conducted on three electrodes in the precentral (FC3), central (C3) and postcentral (CP3) based on previous localization experiments for the motor-lip representation [104]. Figure 4.2 compares the ERP waves of the four synonym types (row1-4) and noise on the electrodes of

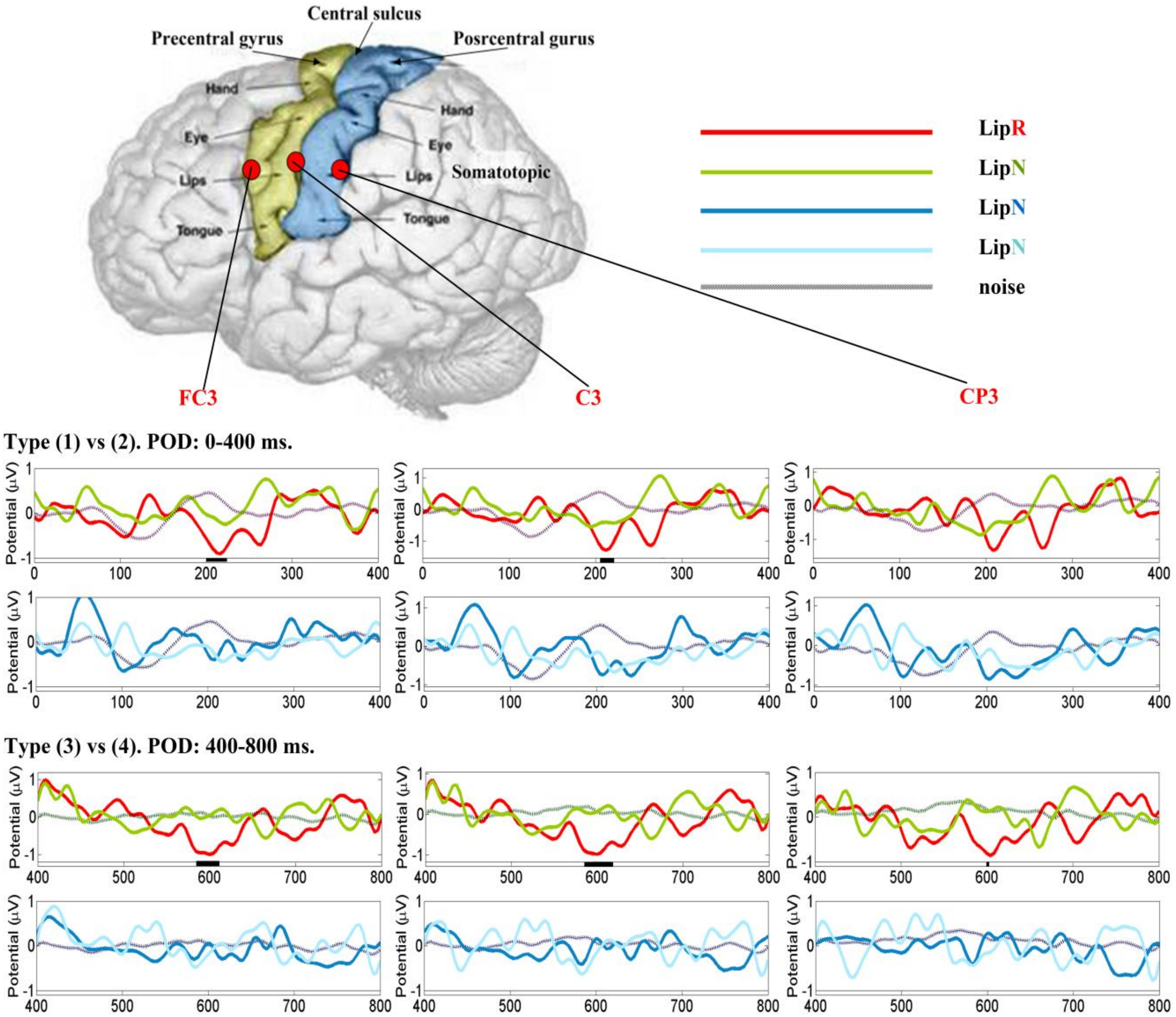


Figure 4.2: ERP waveform comparison of the four types of synonyms and noise at FC3, C3 and CP3 electrode sites. The four rows of the ERP plots correspond to the four types of synonyms in order. The three columns of the ERP plots correspond to the three electrode sites at the precentral, central and postcentral lip-related regions. Adapted from [112] Figure. 1

FC3, C3 and CP3 (column 1-3) in their corresponding PODs (0-400 ms for type (1) and (2); 400-800 ms for type (3) and (4)). In general, ERP wave of the noise showed less fluctuation than that in the word cases, especially during the 400-800 ms range. In the 0-400 ms period, noise stimuli elicited a negative peak shortly after 100 ms (N1 component) and a positive peak around 200 ms (P2 component) in all three electrode sites, as shown in the illustration of type (1) and (2). This could be explained by the electrocortical mapping from the near auditory-related cortical areas, namely the superior temporal gyrus, Broca's area, and Wernicke's area, where the acoustic-phonetic and phonological information is processed to discriminate noises from speech signals. After that, the process of noises was exempted from the higher-level analysis of semantic meaning in N400 and the later periods. In terms of the word conditions, ERP results of type (1) showed that compared to $LipN_S$, $LipR_S$ elicited a more significant negative peak ($p < 0.05$) around 200 ms at the electrodes of FC3 and C3, as marked with black lines on the time axis. This negative peak was not significant at the PC3 electrode. For type (2), no significant ($P > 0.05$) ERP difference was found between the two $LipN_S$ conditions during the whole POD. In the 400-800 ms range, a large negative peak was also found in the S_LipR condition of type (3) 200 ms after the onset of the second character (600 ms post-onset of the two-character word). This negative peak lasted for the longest duration at FC3 and C3 electrodes, while it was quite transient at CP3 electrode. In type (4), the two S_LipN conditions again failed to show significant differences at 200 ms after onset of the test word. These results revealed that in passive speech perception tasks, words with lip-related features elicited stronger activation in the precentral lip motor regions than their synonyms lacking lip-related gestures with a 200-ms latency. In short, at about 200 ms post-onset, the event-related potential of LipR and LipN showed a significant polarity reversal near the precentral lip motor areas.

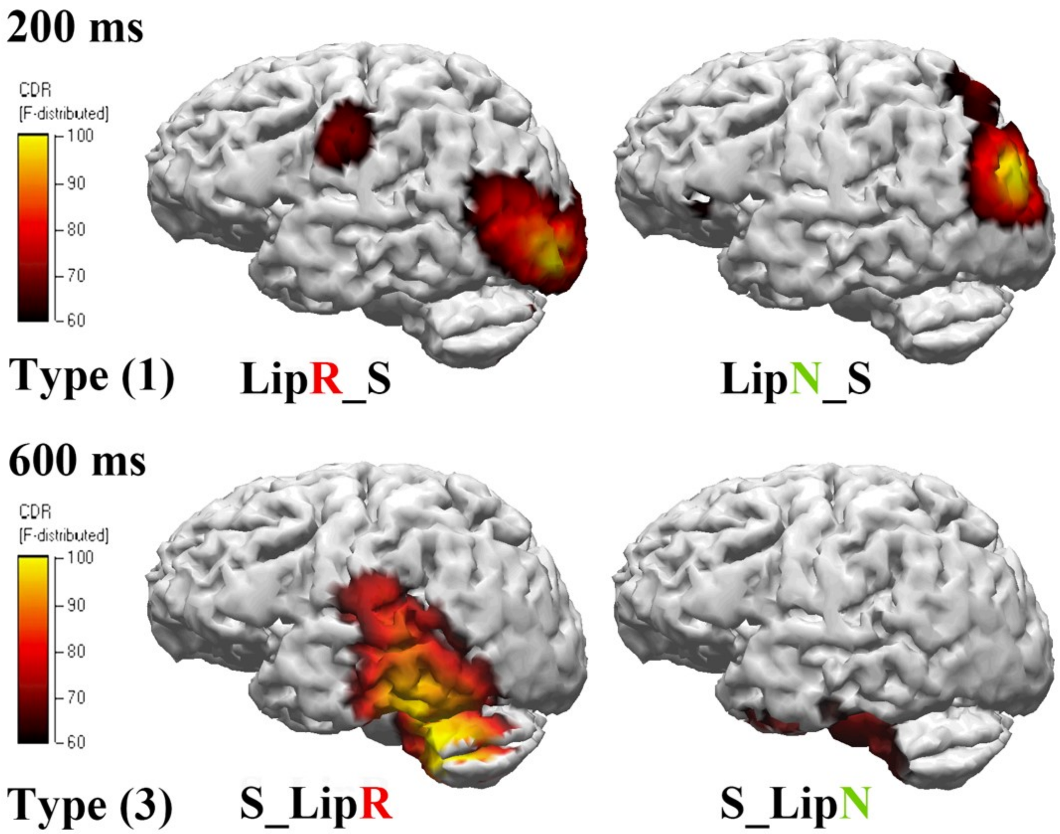


Figure 4.3: CDR results on the cortical maps for type (1) and (3) at 200 ms and 600 ms post-stimulus onset, respectively. The intensity of the maximum activation was normalized as 100%, and the cortical regions with an activation intensity over 60% was highlighted referring to the color scale on the left.. Adapted from [112] Figure. 2

4.3.2 EEG source reconstruction results

To inspect on the cortical sources of activation differences, we performed the current density reconstruction (CDR) on the constraint of standardized Low Resolution Electromagnetic Tomography, which is capable of exhibiting cerebral dynamic sources on a 3D cortical map and providing the activation extent at each time point. The CDR results of type (1) and (3) were inspected during the whole range of their PODs, but special attention was given to the time periods when significant ERP differences (p ¡ 0.05) were found between the LipR-LipN contrasts. As shown in Figure 4.3, the intensity of the maximum activation was normalized as 100%, where the cortical regions with an activation intensity over 60% were highlighted (referring to the color scale on the left). As expected, in type (1), at 200 ms after the onset of the first character, the inferior precentral and central gyrus, covering the lip-related motor and premotor areas were activated significantly by the LipR-S stimuli, which was absent in the lipN-S condition. Similarly, in type (3), when the second character was displayed for 200 ms (600 ms after onset), S-LipR also induced obvious stimulation near or over the lip somatotopic regions, which was also failed to be detected in the S-LipN condition. These findings are consistent with the ERP outcomes, providing neuro-experimental evidence for an involvement of the articulatory motor control in the lip-related regions in response to speech sounds involving lip-related articulatory information. In short, during speech perception and comprehension of synonyms, the items that recruit lip rounding gestures for its articulation elicited a stronger neural response in the precentral lip motor cortex around 200 ms post-onset compared to the items that involves little lip movement. This result suggests an automatic motor-somatotopic association when perceiving spoken words with specific articulatory features, even when the semantics are carefully controlled. Basically, our findings are in agreement with the motor theory of speech perception in the sense of the articulatory facilitation to comprehension. However, it's still arguable whether the motor circuits play a causal contribution to speech perception, which requires a step further investigation in future research.

4.3.3 Frequency-specific brain network reconstruction

To reveal the frequency-specific brain network dynamics, we used multivariate autoregressive modeling (MVAR) methods in the reconstruction of time-varying brain connectivity. Considering that the neural oscillatory activities in distinct frequency bands are associated with different brain states, we decomposed the network activities in the theta (4-8 Hz), alpha (8-15 Hz), beta (15-30 Hz) and gamma (30-45 Hz) bands separately. For each frequency band, the connectivity strength differences between word and noise conditions (word-noise) were depicted in Figure 4.4. As shown, network connections in the high gamma band were mainly distributed in the temporal lobe along the ventral path, which is conventionally correlated with the processing of the speech fine structure at the phonetic scale [113]. Theta rhythms, closely correlates with the acoustic envelope of naturalistic speech at the syllabic rate [114] was also found with dominant network activity in the ventral path. This theta-gamma coupling is also in line with its spatial function in the temporal lobe for the transformation from phonetic to syllable and lexical perception [115, 116]. Beta band,in close relation with sensorimotor behavior [117, 118], appeared with stronger connectivity along the dorsal stream. Alpha oscillation showed up with a wide connection along both ventral and dorsal streams, consistent with its mediation function for pathway direction and selective attention [119].

Figure 4.5 unfolded the brain network dynamics of word processing onto a time-frequency grid. Notice that functional network along the ventral stream showed up early in the gamma the theta bands for the primary auditory processing. Alpha oscillation also showed an early involvement for pathway direction. Over time, network connection in the gamma and theta bands flowed from STG to MTG and ITG in the posterior parts for higher cognition. Whereas dorsal connections in the beta band appeared much later near the end of the auditory stimuli as articulatory simulation for speech confirmation.

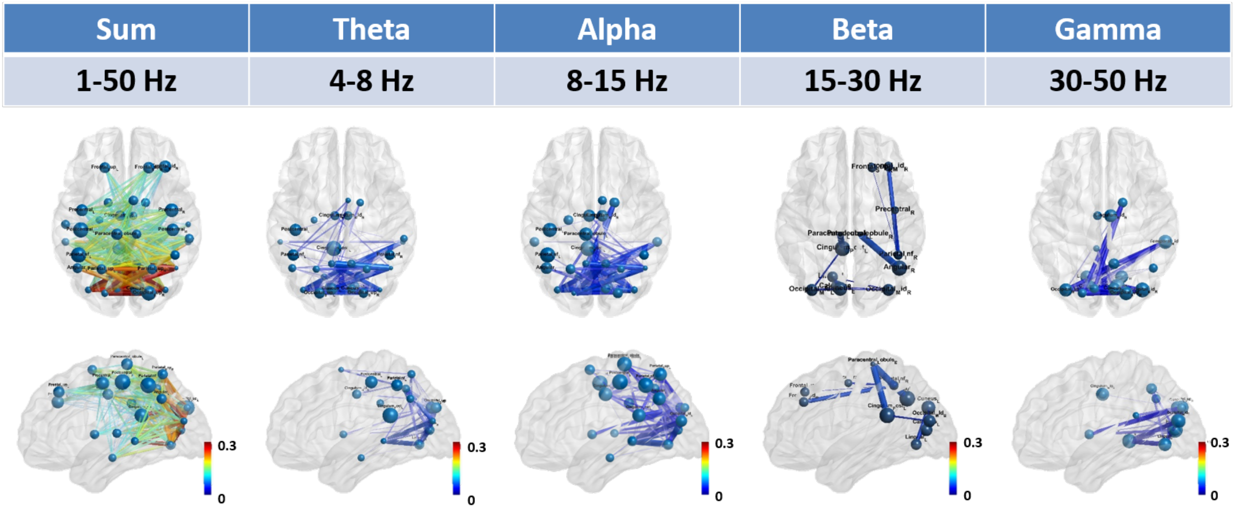


Figure 4.4: Brain connectives of word processing in the Theta, Alpha, Beta, and Gamma bands. Adapted from [120] Fig. 3

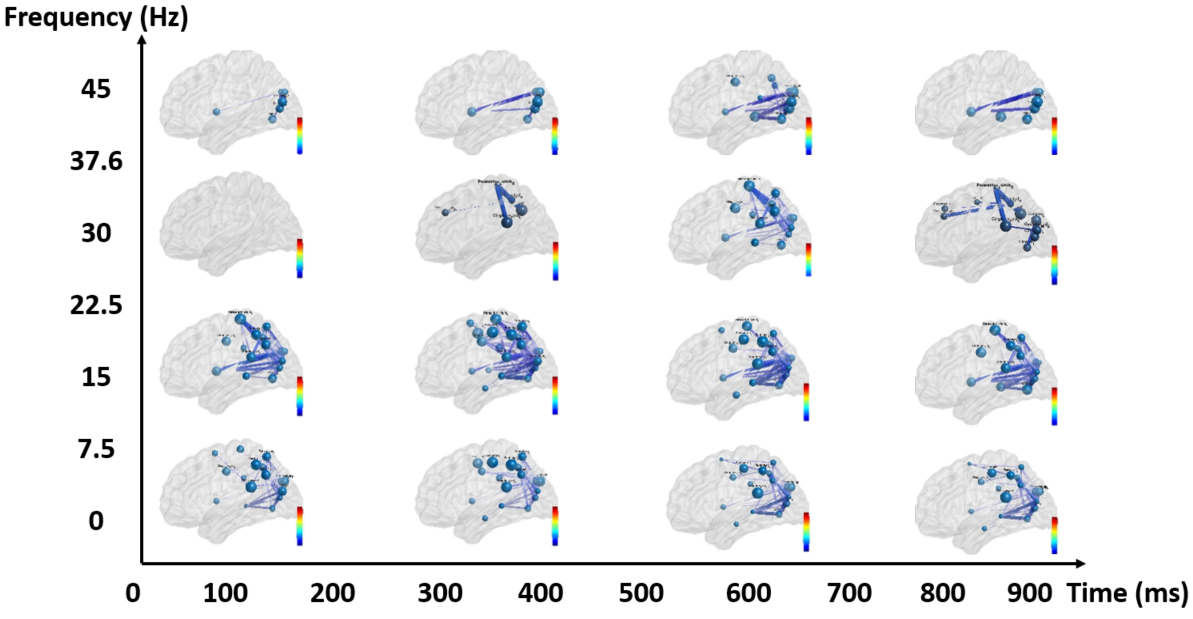


Figure 4.5: Temporal-spatial-spectral brain network dynamics of word processing. Adapted from [120] Fig. 4

4.4 Discussion

Speech perception and production are inextricably integrated. However, Evidence for the participation of one system in another among the perception-production loop is unevenly provided. For example, there is unequivocal evidence for an important influence of sensory systems on speech production [3, 121, 122]. Behaviorally, it is well-established that auditory input can produce rapid and automatic effects on speech production. For example, delayed auditory feedback of one's own voice disrupts speech fluency [121]. Other forms of altered speech feedback have similar effects: shifting the pitch or first formant (frequency band of speech) in the auditory feedback of a speaker results in rapid compensatory modulation of speech output [122], indicating that auditory feedback is important in maintaining articulatory tuning. As in our oral reading task for speech production (Experiment 3), the participation of the auditory system in the STG/STS has also been identified (Figure 5.2). Although not every subject recruited the auditory region for the oral reading task, some even intentionally suppress the auditory feedback for a smooth articulation. There is no doubt that speech perception system plays a role in the production process. More recently, research on motor control has revealed why this sensory-motor link is so critical [74, 80]. Motor acts aim to hit sensory targets. In the speech domain, the targets are not external objects but internal representations of the sound pattern (phonological form) of a word. During speech programming, the auditory information is compared with a prediction derived from the efferent copy of the motor output, with the resulting prediction error used to keep an internal state tracking of the vocal tract [74].

On the other hand, evidence for the role of motor speech system in speech recognition is relatively rare. This idea dates back to the 1960s when the motor theory of speech perception was proposed by Liberman and colleagues: listener perceives speech by simulating the ”intended articulatory gestures” of the speaker [100]. The motivation for this proposal was the observation that there is not a one-to-one correspondence between acoustic patterns and perceived phonemes. Therefore, the objects of speech perception are the intended phonetic gestures of the speaker (represented in the brain as invariant motor commands), rather than auditory transforms of the acoustic speech signal. Therefore, perceiving speech necessarily involves speech motor system recruitment [100]. Unfortunately, empirical work aimed at testing this hypothesis failed to support it by the 1990s. However, the discovery of mirror neurons in the frontal lobe of macaque monkeys revitalized the theory[123, 124, 103], at least among neuroscientists [125]. Mirror neurons respond both during action execution and action perception [123, 124, 103], a response pattern that had led to the hypothesis that 'we understand action because the motor representation of that action is activated in our brain' [123]. There is good evidence that motor-related systems and processes can play a role in at least some speech perception tasks. For example, stimulating motor lip areas produces a small bias to hear partially ambiguous speech sounds as a sound formed by lip closure (e.g., /b/) whereas stimulation of motor tongue areas bias perception toward sounds with prominent tongue movements (e.g., /t/) [126]. Following this logic, in our investigation of the perception-production interactions, we employed disyllabic synonyms with different extent of the lip involvement in a listening task. And we found that during the listening task of Lip-rounding sounds, the motor regions of lip motor control show stronger activation in contrast to sounds that does not involve lip rounding gestures. This results provided a supportive neural evidence for the motor theory of speech perception, showing that passively listening to phonemes and syllables tends to activate the motor and premotor cortex [102].

4.5 Summary

Experiment 1 investigated the neural mechanism of speech perception and comprehension. From LipR-LipN comparasion, we found active participation of the dorsal motor regions in the speech perception and comprehension process. This provides a neural experimental evidence for the motor theory of speech perception. Our frequency-specific analysis further identified that gamma the theta bands mainly functions in the ventral stream for phonological and lexical level processing, respectively. Beta band,in close relation with sensorimotor behavior, also participated and show stronger connectivity along the dorsal stream. Alpha oscillation showed up with a wide connection along both ventral and dorsal streams, consistent with its mediation function for pathway direction and selective attention. These oscillatory characteristics are consistent with their rhythmic properties in terms of temporal duration and spatial distribution and also consistent with the functionalities of the ventral stream for speech comprehension and the dorsal stream for speech production.

Chapter5

Investigation on the Speech Production Mechanisms

This Chapter will focus on the second experiment for the investigation of speech production during an oral sentence reading task. The topic of interest is the hierarchical structure of sentence organization and its underlying neural oscillatory mechanisms. A key question we would like to address is whether and how the syntactic structure and semantic association from our internal knowledge could top-down influence the behavior of visual reading and speech production process. I will analyze from the spatiotemporal brain network dynamics and introduce the cross-frequency coupling mechanism to address the organization of linguistic hierarchy and the bidirectional interactions between the top-down and bottom-up processes.

5.1 Introduction

Sentence oral reading requires cortical processing for top-down control over the semantic comprehension and speech production systems [127]. This top-down control involves hierarchical organization of the orthographic, phonological, semantic and syntactic information for sentence representation [83], as well as goal directing, motor planning, and concrete motor execution for speech production [128, 129]. These tasks require the activation of distinctive functional networks [130] and the transient modulation of effective connectivity strengths between areas [131] in both bottom-up and top-down directions [132]. To date, there still lacks a comprehensive framework that describes the involved spatiotemporal brain network dynamics, meanwhile explains its neural operational mechanisms. This absence greatly limited our overall understanding of the speech functions as an interactive chain, and left many long-standing disputes unresolved [133]. One hot topic around speech production and comprehension concerns whether and how our internal knowledge of the syntactic structure and semantic association could actively affect the strategy of sentence processing [45]. Conventional cognitive studies tended to treat the brain as a passive, stimulus-driven device [**RN3532**], by which sensory information is feed forward progressively to higher association cortices for cognition and motor actions [134]. Contrarily, accumulating anatomical, psychophysical, and physiological evidences are converging to a more active view of brain functions, where top-down influences such as expectation or goal orientation play a critical role in parallel with the bottom-up process [135, 131, 136].

Researchers aiming to address the controversy approached either behaviorally or neurophysiologically [133]. For the behavioral measurement, a common practice is to record the eye movement trajectory and speech signals during oral reading [137, 138, 139]. The latencies in-between the gaze onset and articulatory onset of each word in the sentence, known as the eye-voice span (EVS), could be used to evaluate the speech planning strategies in certain contexts [137]. It turned out that the incrementality of the EVS varies from a lexicon, a phrase, to a clause or even an entire sentence [138], which correspond to different speech planning strategies. Specifically, the lexical incrementality suggests a serial word encoding with a passive strategy, whereas the structural incrementality implies an organized sentence unfolding strategy with an active involvement of cognitive resources [139]. Further studies verified that these EVS variations depend on the accessibility of the contextual information and the availability of prior knowledge [137]. On the other hand, neurophysiological researchers are more interested in the cortical tracking of the visual and auditory stimuli at multiple scales [57, 140, 141, 142]. An interesting phenomena, called cortical entrainment, has been consistently identified [143, 144, 145, 146, 147, 148], which refers to the alignment of rhythmic neural activity with an external periodic or quasiperiodic stimulus [144, 133]. In speech processing, these quasiperiodic elements can be fluctuations in the amplitude envelope associated with syllables [149, 150] or linguistic representations generated in the mind of the listener, such as syntactic phrase boundaries [151, 152]. Correspondingly, neural oscillations in different frequency bands have been attributed with distinctive entrainment properties [151, 56, 153, 54, 154]. For instance, gamma (>30 Hz) oscillations with fast rhythmic cycles are suitable for entraining with fine-grained visual details and sub-phonetic speech structures [155, 156]. Theta (4–8 Hz) and delta (1–4 Hz) oscillations have longer duration that roughly corresponds to speech envelopes, thus are suitable for entraining with syllabic, phrasal and clausal structures [114, 157]. Moreover, alpha (8–13 Hz) oscillations, ubiquitous during resting state, usually get suppressed with sensory inputs along the processing path, thus were associated with sensory-gating process [158]. Beta (15–30 Hz) oscillations, usually in the form of suppression are commonly observed in the sensorimotor regions during motor planning and execution in speech production [159, 160].

several recent studies have provided evidence for oscillatory models that entail a more active role of cortical tracking in speech comprehension [161, 162, 163, 151, 149, 146, 150]. And it has been recently proposed that cortical tracking consists of both “entrainment proper” (phase-locking to acoustic periodicities of the signal) and “intrinsic synchronicities” reflecting the endogenous generation of linguistic structure and predictions [164].

Given these oscillatory features and corresponding cortical localization, however, is not sufficient for us to have a clear image of the cortical network dynamics for the syntactic and semantic influence from higher-cognition to sensorimotor regions. Besides, how different levels of linguistic units in the sentence hierarchy could be cortically entrained and integrated from distributed areas into a coherent concept remains obscure. The serial/parallel (visual, semantic, syntactic, motor, and auditory) processes of each word in a sentence require the activation of different functional networks and the transient modulation of the effective connectivity strength between areas. How is this dynamic coordination of networks accomplished? in a recurrent, concatenated and overlapping manner.

local high frequency neural activity typical of motor or perceptual cortices is modulated by low frequency neural oscillations typical of prefrontal cortex, which might be achieved via cross-frequency coupling (Canolty and Knight, 2010).

To address the issues of multi-scale integration and dynamics network construction, we introduced a mechanism that is biologically plausible and receives growing research attention, namely cross-frequency coupling (CFC) [56, 165, 27, 54]. CFC refers to a statistical relationship between neural oscillations in two different frequency bands. CFC can take different forms as it associates the phase, frequency, and amplitude of coupled oscillations [53]. Most commonly studied forms of CFC are phase-amplitude coupling (PAC) [55], phase-phase coupling (PPC) [23], and amplitude-amplitude coupling (AAC) [35]. Specific forms of coupling can be a signature of the underlying network physiology and associated with different putative functions, including neural representations of environment items [166], parsing of sensory stimuli with complex temporal structures [24], communication over distinct areas [131], and internal modulation of neural processing based on temporal predictions [45, 68]. For example, theta-gamma coupling is considered as a fundamental language processing mechanism that groups the sub-phonetic speech structures carried by gamma oscillations into the theta temporal windows for lexical interpretation [167, 24]. Alpha-gamma coupling is associated with a primary sensory gating mechanism where the gamma-modulated sensory input is selectively suppressed via alpha inhibition if it is irrelevant or unattended [168, 169]. Considering that oral sentence processing is characterized by an intrinsic quasi-rhythmic temporal structure at multiple linguistic levels [53], there ought to be a mental correspondence in the hierarchical cortical representation, as well as a higher cognitive control over the routing of information across scales from sensory to motor areas [23]. Thus, a neural oscillatory system with multiple frequency bands and various forms of cross-frequency coupling are well-suited for this purpose [23, 57].

Taking all the above mentioned, we designed an oral reading task with simultaneous recording of the eye movement, speech and electroencephalography (EEG) signals. The behavioral tracking would be used for the examination of the EVS strategies and for the segmentation of different cognitive stages. The adoption of the EEG technique is in view of its high temporal resolution and deeper revelation of the neural oscillatory mechanisms compared to functional magnetic resonance imaging (fMRI), positron emission tomography (PET), etc [84, 43]. Thankfully, recent improvements in EEG artifact reduction and source reconstruction algorithms have turned out to be successful in reducing the contamination of orofacial muscle movement and approximating the identification of cortical locations, which provided us with great opportunity to reconstruct the time-varying network organization with sufficient spatiotemporal satisfaction [92]. Furthermore, considering that different forms of CFC may reflect different aspects of oscillatory processing, we calculated the PAC, PPC and AAC values for the examination of regional cross-frequency nesting, inter-regional phase synchronization, and excitatory/inhibitory relations, respectively. These CFC patterns with inter-regional connections and information flow directions would be used to construct a neurofunctional model for the speech production and comprehension (SPAC) process. The current prevailing dual-stream model [60, 67, 68] and hierarchical state feedback control (HSFC) model [127] for speech processing have laid the foundation of a dual-stream hierarchical architecture with feedforward and feedback loops at the phonological and lexical levels. Our aims with the SPAC model is to (1) extend to the sentence level with the consideration of syntactic and semantic effects. (2) complement with spatiotemporal brain network dynamics beyond the neural pathways. (3) explain with the cross-frequency coupling mechanisms for the bridging of linguistic form diversities and representational hierarchies, as well as the bidirectional interactions between the bottom-up sensory input and top-down cognitive regulations.

5.2 Subjects and materials

Subjects. Twenty Mandarin speakers (12 male, 8 female. Age: 22.5 years, SD = 4.4) participated in this study. All the participants were right-handed [170], with normal or corrected-to-normal vision, normal hearing and speaking abilities, and no psychological disorders. Ethical approval for this experiment was obtained from the local Research Ethics Committee. The participants all signed the informed consent for this experimental participation.

Materials. The text materials consisted of 60 syntactically consistent sentences. Each sentence is composed of eight disyllabic words, corresponding to sixteen Chinese characters in total. An example sentence with English translation underneath is given in Figure 5.1 (A). During oral reading, the whole sentence was presented on the screen with the sixteen characters evenly spaced in a single line.

5.3 Analysis results

5.3.1 Behavioral analysis results

Figure 5.1 (A) illustrates the hierarchical structure for one example out of the sixty syntactically consistent sentences. Each sentence comprised sixteen Chinese characters (sixteen syllables). Every two characters constitute one semantic unit (one disyllabic word). The resulting eight disyllabic word items could be further segmented into three clauses (underlined with black dot lines) with the Stanford parser and the Chinese Treebank [171].

Figure. 5.1 (B) demonstrates the eye movement trajectory (blue trace) and speech spectrum (grey background) for one typical trial of the example sentence. The statistical analyses of the eye movement and speech signals were performed across trials and subjects. The averaged duration for processing the whole sentence (from the gaze onset of the first word to the speech offset of the last word) was 4412.8 ms (SD = 559.3 ms). As one can see from

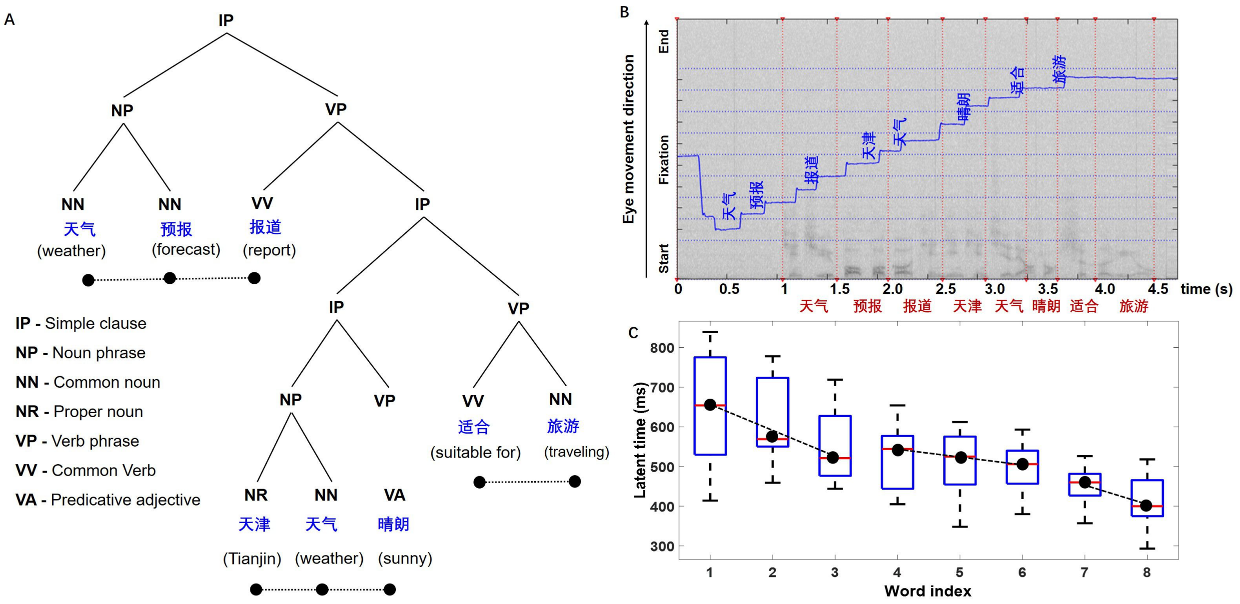


Figure 5.1: Sentence syntactic structure and behavioral results of eye movement and speech segmentation. (A) The syntactic structure of an example sentence. (B) Eye movement trajectory and speech spectrum for one typical trial of sentence oral reading. The x-axis represents the time from the onset of sentence presentation (0 s) to the offset of sentence articulation (4.5 s). The y-axis indicates eye movement from the left to the right. The gaze onsets of the eight words are marked with blue characters adjacent to the eye movement trajectory (blue step-wise traces) and segmented with horizontal blue lines as gaze boundaries. The speech onsets of the eight words are marked with red characters under the speech spectrum (grey background) and segmented with vertical red lines as phonological lexical boundaries. (C) The averaged eye-voice span (EVS) between the gaze onsets and speech onsets of the eight word items. The x-axis represents the word indexes. The y-axis indicates the time latency (ms). The box plot shows the averaged latencies and quartiles across trials and participants.

the eye movement trace, the initial visual fixation started from the screen center. After the onset of sentence presentation (at 0 ms), the participant gaze-shifted from the fixation point to the start of the sentence at around 369 ms. The scanning of the word sequence then proceeded in a step-wise manner, where the vertical lines represent the saccade from one character to another and the horizontal lines represent the fixation within one character field (The microsaccade, nystagmus and drift, etc were excluded from the current analysis). In some cases, the two characters of a disyllabic word were processed at once, implying a word grouping tendency. This phenomenon suggests that the subject may have retrieved his/her mental lexicon to be able to group-process the two characters as one unit. Subsequently, we outlined the boundaries of the eight disyllabic words by defining the falling of the eye sight into the visual field of a word as its gaze onset and the moving out of the eye sight from the word field to another as its gaze offset, as divided by the horizontal blue dotted lines and marked with blue characters in Figure 5.1 (B). Similarly, we segmented the speech boundaries between the eight disyllabic words, as divided by the vertical red dotted lines and marked with red characters. According to the time delay from the gaze onset and speech onset of each word, the articulatory initiation of each word is about two-word-length behind the corresponding gaze onset. This duration is almost long enough to precode a whole clause (maximum 3 words) during the speech planning period. These observation, in general, suggest a pre-reading phenomenon where the visual and articulatory processes interleave and a semantic effect where the subjects actively retrieve lexical meanings prior to articulation for optimizing the oral reading process.

In Figure 5.1 (C), we calculated the averaged EVS in between the gaze onset and speech onset of the eight words across subjects. The eye-voice span is indicative of the involved energy resources for semantic comprehension and speech motor programming, where longer EVS is taken as a more effortful process [137]. From the EVS plot, a general time-latency reduction tendency ($\beta = -32.193$, $p < .05$, adjusted $R^2 = 0.9570$) was noticeable along the whole word sequence (which is not shown in the figure). Further regression analyses indicated that a syntactically consistent pattern (piece-wise linear regression as three segments) could better fit the averaged EVSs (adjusted $R^2 = 0.9863$), as illustrated with the black 'dot-line-dot', than the simple linear regression. The structural incremantality suggests a syntactic involvement, which might be explained by active speaker control and the availability of context. Speakers normally prefer to maximize fluency by planning their utterance more extensively before speaking. Besides that, the repetitive syntactic patterns in all the sentence materials as well as the initial impression of the whole sentence on the screen may also contribute to the syntactic priming effect and facilitate later oral reading processes [139].

5.3.2 EEG source reconstruction results

To trace back to the cortical sources and have a broad view of the event-related spectral perturbation in different frequency bands, we first inspected the ERSP results of 16 IC clusters that were mainly involved in the oral reading process, as illustrated in Figure. 5.2. For each cluster, the ECD positions of the ICs were projected onto a standard MNI template (sagittal view, black background). Next to each fMRI projection on the left side, the averaged ERSP result of the IC cluster is plotted as a time-by-frequency plane (time range: -500–4500 ms, frequency range: 3–60 Hz, log-spectral power scale: -2–2 dB). In each ERSP plot, the positive (red) regions correspond to power increase, which indicate event-related synchronization (ERS) in that cortical area. In contrast, the negative (blue) regions correspond to power decrease, which indicate event-related desynchronization (ERD). Because macroscopic EEG recording could capture neural oscillations from both excitatory and inhibitory neurons, the event-related activation in different frequency bands may show up as either ERS and ERD according to the task demands and the oscillatory properties of corresponding brain origins. Generally speaking, delta, theta and gamma oscillations tend to synchronize (ERS) in response to cognitive demands [172], whereas alpha and beta oscillations tend to desynchronize (ERD) in response to sensorimotor activities [173]. To closely inspect the frequency-specific fluctuations in the ERSP results, the visual cortical cluster (VC) was found to be slightly left-biased with strong power increase in the 3–8 Hz range (delta and theta ERS) and power decrease in the 9–13 Hz range (alpha ERD). In the ventral stream, cortical areas presented strong delta, theta (3–8 Hz) and gamma (30-60 Hz) ERS, which include the bihemispheric posterior middle/inferior temporal gyrus (pMTG/pITG), the left anterior temporal lobe (ATL), and the right superior temporal gyrus/sulcus (STG/STS). Notice that in these ventral areas, delta and theta ERS occurred as early as the sentence onset, whereas gamma ERS occurred around 1000 ms, following the articulatory onset. In the dorsal stream including the left parietal cortex (PC), right

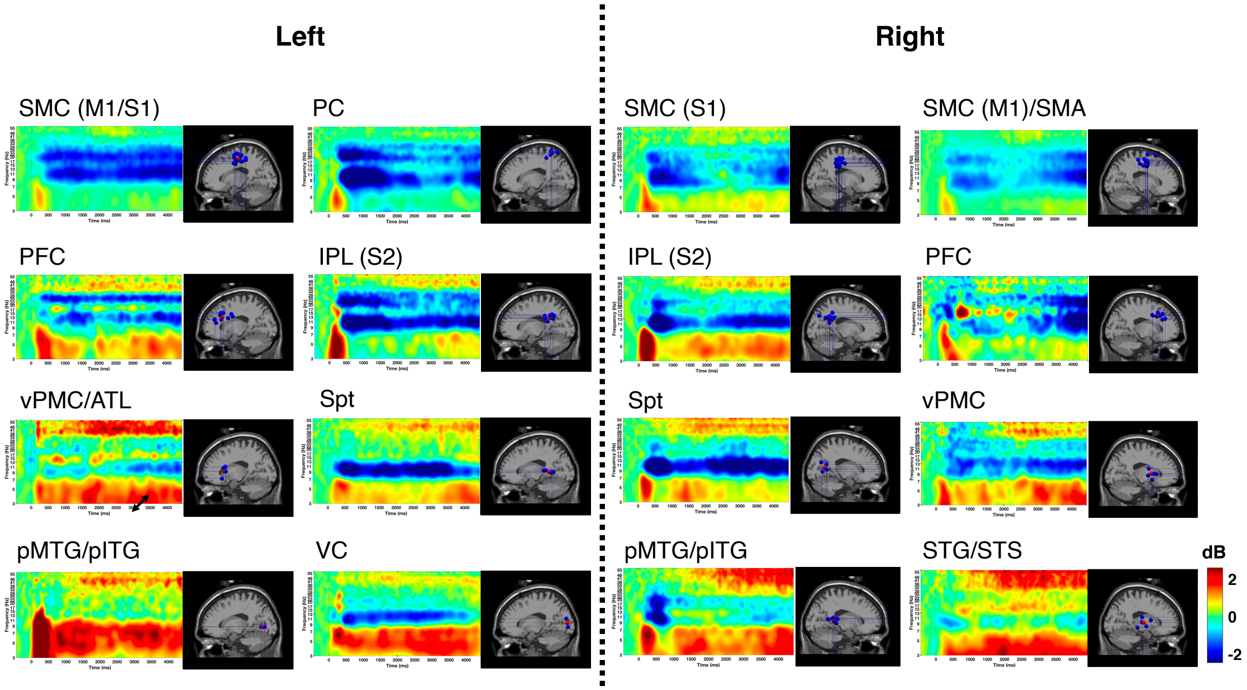


Figure 5.2: Independent component clustering and event-related spectral perturbation (ERSP) of cortical sources. For each cortical location, the clustered independent components (blue dots) and the cluster center (red dot) are mapped onto an fMRI template (sagittal view, black background). The corresponding ERSP plot for each IC cluster is shown by the left side, which is calculated by averaging the spectral perturbation of all the components within the cluster. The x-axis of the ERSP plot represents time (-500–4500 ms), and the y-axis frequency bands (δ (3–4 Hz), θ (4–8 Hz), α (8–15 Hz), β (15–30 Hz), γ (30–50 Hz)). The power scale of the ERSP results (log-scaled, -2–2 dB) is illustrated at the bottom right. The positive (red) regions in the ERSP plots indicate power increase, which imply event-related synchronization (ERS) in that cortical region. In contrast, the negative (blue) regions indicate power decrease, which imply event-related desynchronization (ERD). Abbreviations: VC – visual cortex, pMTG - posterior middle temporal gyrus, pITG - posterior inferior temporal gyrus, STG – superior temporal gyrus, STS - superior temporal sulcus, vPMC - ventral premotor cortex, ATL – anterior temporal lobe, Spt - Sylvian fissure at the parieto-temporal boundary, PFC – prefrontal cortex, IPL – inferior parietal lobe, SMC – sensorimotor cortex, SMA – supplementary motor area, PC - parietal cortex, M1 - primary motor area, S1 - primary somatosensory area, S2 - secondary somatosensory area.

supplementary area (SMA), and the bilateral sensorimotor cortex (SMC, which could be further divided into primary motor area (M1) and primary somatosensory area (S1)), the concurrent alpha (8–13 Hz) and beta (15–25 Hz) ERD are the most significant features, also known as μ rhythm [37]. Cortical areas in between the ventral and dorsal roots include the bilateral Sylvian fissure at the parieto-temporal boundary (Spt), the vental premotor cortex (vPMC), and the inferior parietal lobe (IPL). In these middle areas, broad range activities such as delta-theta-gamma ERS and alpha-beta ERD coexisted.

From a functional localization perspective, these cortical components are generally symmetric with only a few bihemispheric differences. Specifically, the ATL is more active in the left hemisphere, whereas the STG/STS activation is more obvious in the right hemisphere. Functionally, the ATL is speculated to serve as an interface between posterior lexical–semantic systems and frontal systems for structuring lexical meaning into sentence-level context [174, 175]. The STG/STS is thought to carry out the acoustic–phonetic processing for auditory feedback in speech production tasks [176]. Thus, the bilateral differences of these regions, under the framework of the dual-stream model [60], can be well-explained by the left-hemispheric dominance for the sentence-level semantic association in the ATL, and by the left-biased speech motor programming that sends inhibitory copies of motor commands to the ipsilateral auditory regions for prediction, which in turn suppress the activation of the left STG/STS during auditory feedback. Other than these left-right asymmetry, the pMTG/pITG regions as the sound-meaning interface for relating perceived sensory (auditory or visual) information into lexical representation were found to be bilaterally involved, which is consistent with the proposal in the dual-stream model [60]. The Spt region has been identified in the dual-stream model as a sensory-motor interface for coordinate transform from auditory or visual modalities into motor representations [60, 177]. The IPL region has been proposed in the state feedback control (SFC) model [178] as a somatosensory-motor interface where the motor efferent copy from the motor network is compared with the somatosensory feedback for future correction. Besides, another region for sensory-motor integration is the vPMC, which was found to connect with the Spt and IPL via the arcuate fasciculus (AF) and superior longitudinal fasciculus (SLF) respectively [179]. The left SMC covers the M1 and S1 that were clustered together due to their cortical proximity and oscillatory similarity. The right hemisphere was also found with the SMA response together with the M1 activities. The SMA is important for temporal (sequence) organization and initiation of movements. Moreover, The the prefrontal-parietal (PFC-PC) regions has long been suggested to play an important role in cognitive control, in the ability to orchestrate thought and action in accordance with internal goals [180, 181, 182]. In summary, our source reconstruction and IC cluster results identified bihemispheric involvement of cortical components in the oral reading task. We generally agree with the dual-stream model [178, 60] regarding the ventral and dorsal organization for speech comprehension and production and the left-dominance in higher-level semantic processing and speech motor programming. However, we emphasise a broader right-hemispheric participation and a higher-level cognition in the PFC-PC network over the ventral and dorsal streams.

5.3.3 Cross-frequency coupling results

5.3.3.1 Phase-amplitude coupling (PAC)

For each IC cluster illustrated in Figure 5.2, we calculated the PAC values between low frequency (1-15 Hz) for phase and high frequency (5-60 Hz) for amplitude (power) and z-transformed them into PACz values after permutation testing. Figure 5.3 A shows the PACz values of eight cortical regions in the left hemisphere. The PACz values for the right hemisphere are given in supplementary materials. The red regions with ovethreshld PACz values ($PACz > 1.65$, $p < 0.05$) were considered as significant PAC phenomena and marked on the cortical map (Figure 5.3 B) with certain PAC attributes. Totally, six PAC categories with clearer neurofunctional meaning were considered, namely the alpha-gamma coupling, alpha-beta coupling, theta-gamma coupling, theta-alpha coupling, delta-alpha coupling, and

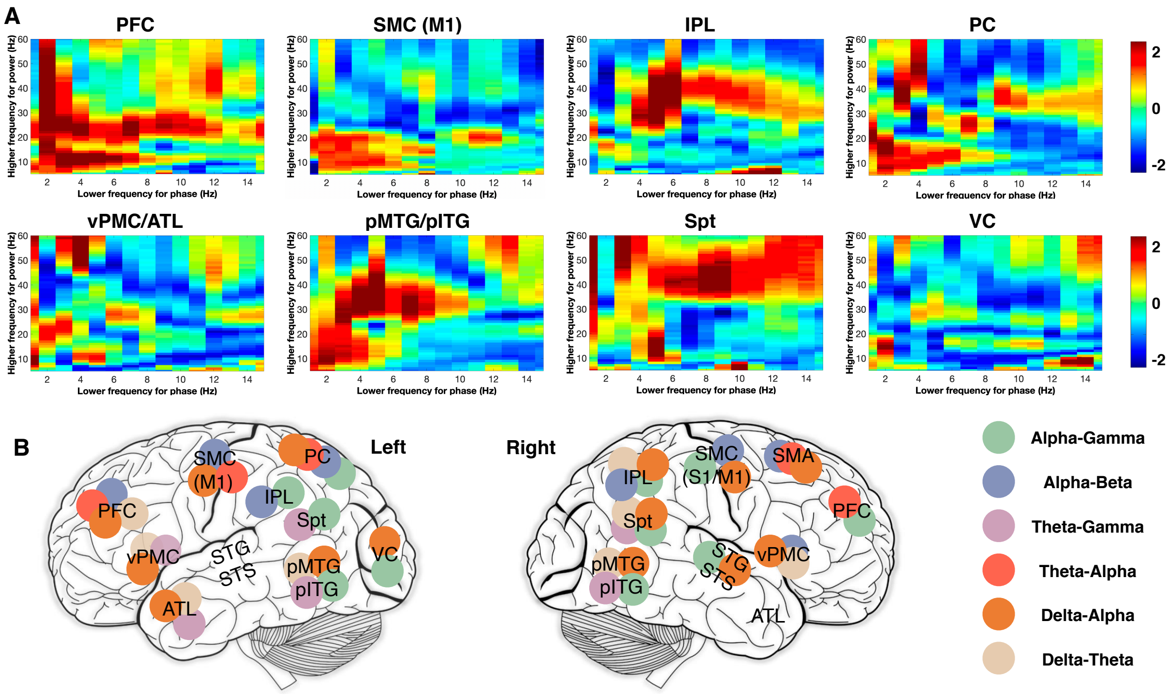


Figure 5.3: Phase-amplitude coupling (PAC) results of bihemispheric cortical regions. (A) Illustration of the normalized PACz values for eight cortical regions in the left hemisphere. The x-axis represents lower frequency for phase (1–15 Hz), and the y-axis represents higher frequency for power (5–60 Hz). The color-coded PACz values is calculated after permutation test and z-normalization from the original PAC values. (B) Cortical localization with five categories of significant PACz values ($PACz > 1.65, p < 0.05$) in both hemispheres.

delta-theta coupling. From the cortical PAC distribution in the left hemisphere, it can be noticed that alpha-gamma coupling (green dots) were mainly located in the VC and adjacent pMTG/pITG, Spt, IPL, and PC. Alpha-beta coupling (blue dots) were mainly located in the dorsal areas, e.g., PFC, SMC, IPL, and PC. Delta-theta coupling (yellow dots) and theta-gamma coupling (pink dots) were mainly located along the ventral areas, e.g., vPMC, ATL, and pMTG/pITG. Theta-alpha coupling (red dots) were mainly in the PFC and PC areas. Finally, delta-alpha coupling (orange dots) has a wide distribution over the dual-stream paths. The right hemisphere presented some similarities with the left part, except some additional delta-alpha and alpha-gamma couplings in the STG/STS, Spt, IPL and SMC (S1) areas. The account for these right dominance, or left absence, might also related to the cancellation of the auditory feedback (in the STG/STS, Spt) and somatosensory feedback (in the IPL and SMC (S1)) by the predictive motor commands that were mainly organized in the left hemisphere.

5.3.3.2 Phase-phase coupling (PPC)

Figure 5.4 shows an example of the intersite phase clustering (ISPC) results between cortical sites along the ventral stream. Figure 5.4 A, B, and C illustrate the three tested cortical pairs that are supposedly involved in the visual processing (VC), lexical representation (pMTG/pITG), semantic association (ATL) and high-level (syntactic) integration (PFC). The arrows indicate the information flow between the cortical pairs, which is derived from the ISPC results beneath. Figure 5.4 a1, b1, and c1 illustrate the signals of these regions filtered at 4 Hz. The selection of 4 Hz for the PPC testing is in consideration that theta band generally serves as the long-range communication channel for distributed cortical regions. In addition, the 4-Hz significance in the PAC has been verified in our PAC results for both ventral and dorsal connections (as illustrated in Figure 5.3). From the signal plots in Figure 5.4 a1, b1, and c1, it could be noticed that the major fluctuations occur between 0 ms and 1000 ms, which overlaps with the EVS of the first word. This dominance of fluctuations in

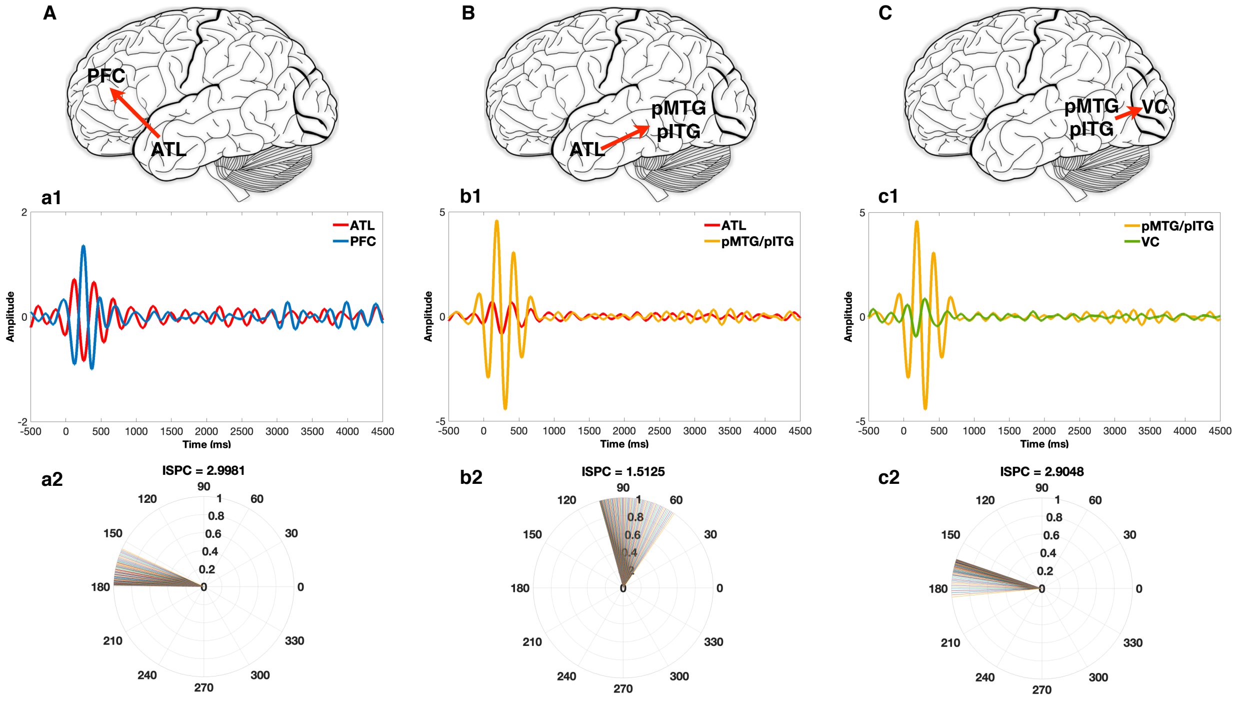


Figure 5.4: Intersite phase clustering (ISPC) between cortical sites along the ventral stream. (A) Illustration of the path linking the ATL and the PFC. (a1) Illustration of the filtered signals of the ATL (red) and the PFC (blue) at 4 Hz. (a2) Distribution of the phase differences between the ATL and the PFC at each time point (shown as unit-length vectors with a certain angle in the poplarplot). The ISPC value on top of the poplarplot is defined as the average angle of dark lines. Positive ISPC value implies phase leading of the first component (ATL) over the other one (PFC), and vice versa. (B) Illustration of the path linking the ATL and the pMTG/pITG. (b1) Illustration of the filtered signals of the ATL (red) and the pMTG/pITG (yellow) at 4 Hz. (b2) Distribution of the phase differences between the ATL and the pMTG/pITG over the first EVS. (C) Illustration of the path linking the pMTG/pITG and the VC. (c1) Illustration of the filtered signals of the pMTG/pITG (yellow) and the VC (green) at 4 Hz. (c2) Distribution of the phase differences between the pMTG/pITG and the VC over the first EVS. The time ranges for calculating the ISPC span the first EVS from the gaze onset (369 ms) to the speech onset (1099 ms) of the first word.

the early stage might be explained by neuroadaptation [23], namely the neuronal responses decrease over time to constant stimuli, and/or by the time response differences across subjects that dampen the signals over time. In view of above reasons, we calculated the ISPC in the time range of 369-1099 ms, which covers the EVS of the first word from gaze onset to speech onset. The clustered phase differences over time are illustrated in Figure 5.4 a2, b2, and c2. From the sign (positive or negative) of the ISPC values on top of each polar diagram, we could speculate the information flow direction based on the judging of phase leading or phase lagging. The absolute value of the ISPC may also provide information regarding the time difference for signal routing between each cortical pair. From the ventral ISPC results, the 4 Hz signal is supposed to originate from the ATL, propagating one way up to the PFC, and another way down to the pMTG/pITG, and furter to the VC. Hence, the ATL as an intermediate semantic association area could both top-down regulate the visual recognition process in the VC and bottom-up transfer the low-level information into PFC for appropriate action via higher-lever integration.

Figure 5.5 shows the ISPC results between cortical sites along the dorsal stream. Figure 5.5 A, B, and C illustrate the tested cortical pairs that are supposedly to be involved in cognitive control (PFC), motor execution (SMC), speech motor programming (vPMC) and sensory-motor interface (Spt). Figure 5.5 a1, b1, and c1 illustrate the signals of these regions filtered at 4 Hz. Figure 5.5 a2, b2, and c2 show the phase differences between these cortical pairs over the EVS of the first word. From the signs of the ISPC values on top of the polar diagrams, it is speculated that the PFC projects commands onto the SMC, which sends efferent copy to the vPMC, and then tansformed in the Spt regions. This estimated cortical information flow is consistent with the neural path proposed in the HSFC model [127], which in turn validated the effectiveness of using the neural oscillatory characteristics to speculate the information flow along the neural path. Therefore, besides the 4-Hz information flow tested in the above selected regions, we also investigated the whole cortical connections in the theta, alpha, beta and gamma bands, as illustrated in Figure 5.7 which we will explain

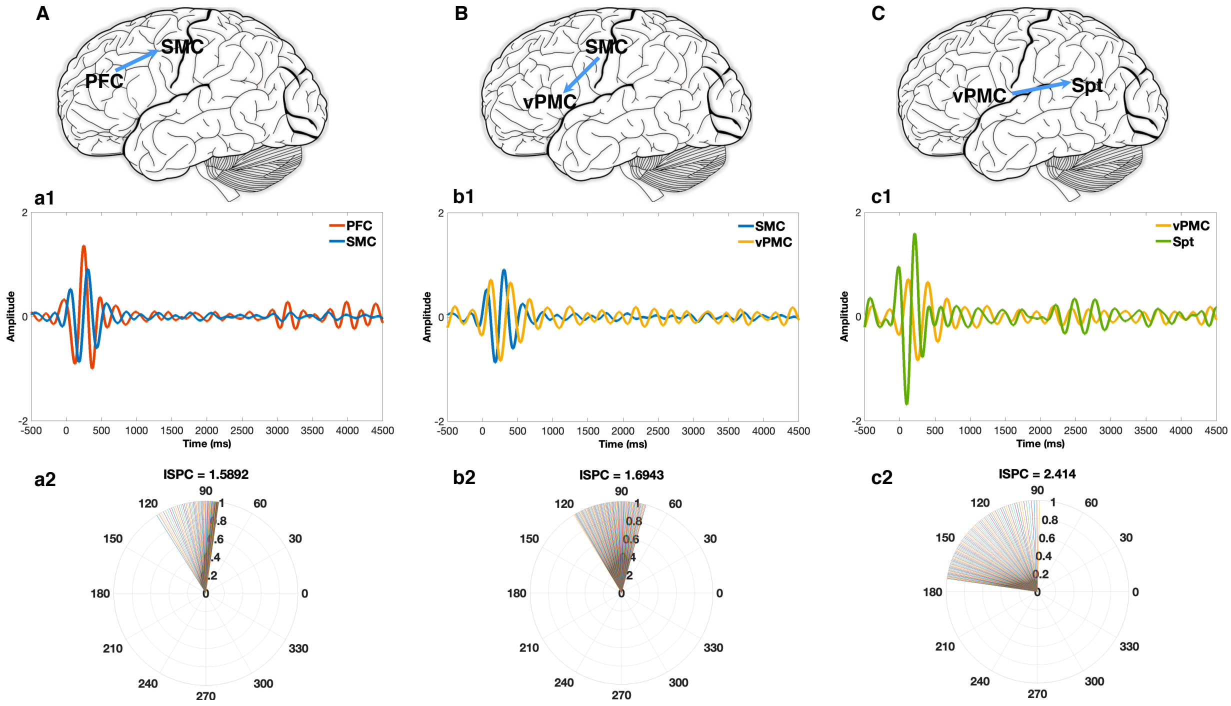


Figure 5.5: Intersite phase clustering (ISPC) between cortical sites along the dorsal stream. (A) Illustration of the path linking the PFC and the SMC. (a1) Illustration of the filtered signals of the PFC (orange) and the SMC (blue) at 4 Hz. (a2) Distribution of the phase differences between the PFC and the SMC over the first EVS (369–1099 ms). (B) Illustration of the path linking the the SMC and the vPMC. (b1) Illustration of the filtered signals of the SMC (blue) and the vPMC (yellow) at 4 Hz. (b2) Distribution of the phase differences between the SMC and the vPMC over the first EVS. (C) Illustration of the path linking the the vPMC and the Spt . (c1) Illustration of the filtered signals of the vPMC (yellow) and the Spt (green) at 4 Hz. (c2) Distribution of the phase differences between the vPMC and the Spt over the first EVS.

later.

5.3.3.3 Amplitude-amplitude coupling (AAC)

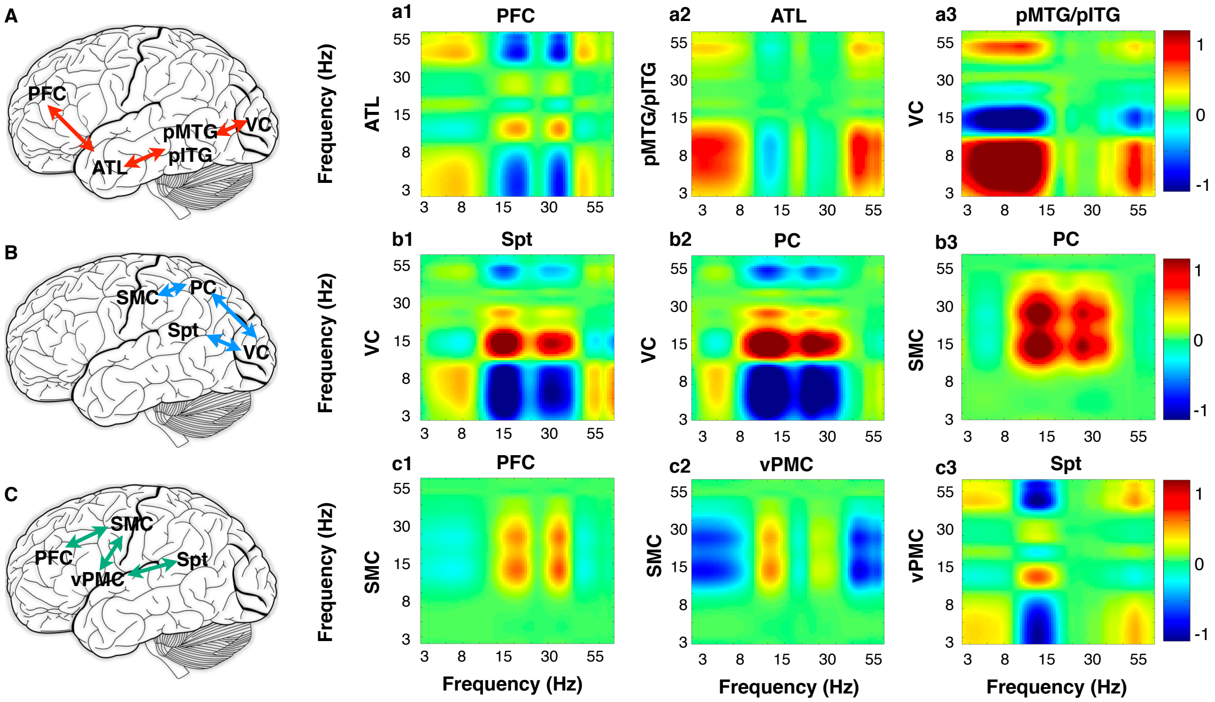


Figure 5.6: Amplitude-amplitude coupling (AAC) results between cortical pairs along the dual-stream paths. (A) Illustration of the connections between the VC, pMTG/pITG, ATL and PFC along the ventral stream. (a1) Normalized AAC values between the ATL and the PFC. (a2) Normalized AAC values between the pMTG/pITG and the ATL. (a3) Normalized AAC values between the VC and the pMTG/pITG. (B) Illustration of the connections between the VC with the Spt and PC, and between the PC with the SMC along the dorsal stream. (b1) Normalized AAC values between the VC and the Spt. (b2) Normalized AAC values between the VC and the PC. (b3) Normalized AAC values between the SMC and the PC. (C) Illustration of the the connections between the PFC, SMC, vPMC and Spt along the dorsal stream. (c1) Normalized AAC values between the SMC and the PFC. (c2) Normalized AAC values between the SMC and the vPMC. (c3) Normalized AAC values between the vPMC and the Spt.The y-axis and the x-axis of each AAC plot represent the frequency range (3-60 Hz) for the first and second cortical component, respectively. The color-coded AAC values in each plot indicate positive (red) or negative (blue) correlation between these cortical pairs in corresponding frequency bands.

In addition to the regional nesting patterns revealed in the PAC analysis and the inter-regional phase relationship revealed in the PPC analyses, we also performed the amplitude-amplitude coupling (AAC) to examine the excitatory (positive AAC) or inhibitory (negative

AAC) relations between cortical pairs. It is important to note that AAC itself is not indicative of the information flow directions as the PPC did, nor could it identify all CFC patterns as the PAC revealed. This is because the CFC is not necessarily accompanied by the increase or decrease of oscillatory magnitude, rather some synchronization may be formed via time coding or phase resetting without changing the spiking rates [46, 183]. In Figure 5.6, we illustrate the normalized AAC results of some prominent cortical regions with their bidirectional connections shown in Figure 5.6 A-C and their inter-regional AAC values shown in Figure 5.6 a1-c3. It could be noticed that theta band synchronization (shown as the red block at the left bottom) and theta-gamma AAC is mainly presented in the ventral stream, especially between the VC with the pMTG/pITG in (a3), and the pMTG/pITG with the ATL in (a2). In contrast, theta-alpha AAC between the pMTG/pITG and the VC (a3) is shown as negative values, indicating opposite oscillatory modulation patterns between the theta and alpha rhythms. Similar negative relationships could be found between the ATL and PFC in the theta-alpha and alpha-gamma pairs, reflecting the inhibitory nature of alpha oscillation in sensory processing tasks. In the dorsal path, the most significant AAC features are the alpha-beta cross coupling between SMC and PC in (b3), the alpha oscillation in the VC synchronizing with alpha and beta activities in the Spt and PC (b2 and b1), and the theta oscillation in the VC negatively correlating with the Spt and PC (b2 and b1). For top-down organizing the articulatory process, the PFC positively correlate the SMC in the alpha and beta bands (c1). The SMC send a negative copy of the motor commands in the alpha-beta bands into the theta-gamma bands of the vPMC (c2). Then the copy of the theta-gamma codes in the vPMC are transfered in the Spt into alpha activity negatively and theta-gamma activity positively (c2). Overall, the AAC results echo with the PAC and PPC outcomes and demonstrated the excitatory nature of the theta-gamma activities and the suppressive nature of the alpha-beta activities, as well as their inverse relationship in modulating the speech comprehension and production processes in different functional stages.

5.3.4 Effective connectivity results

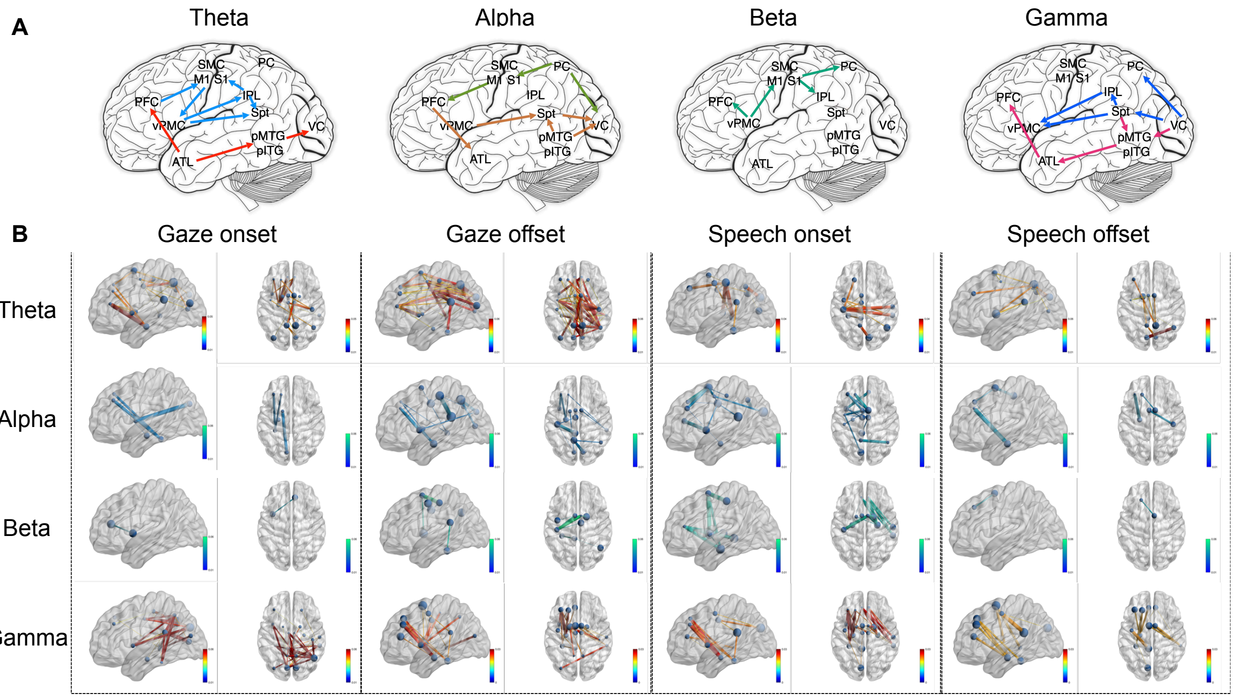


Figure 5.7: Frequency-specific brain connectivity and network dynamics. (A) Frequency-resolved cortical connections along the dual-stream based on the PPC results. The ventral and dorsal paths are distinguished with warm and cold colors, respectively. The arrows in each cortical map indicate the information flow directions in the corresponding frequency band. (B) Frequency-specific brain network dynamics during the first EVS based on the effective connectivity results. The theta and gamma activities are connected with warm colors whereas the alpha and beta activities are connected with cold colors. In each of the four stages, namely the gaze onset, gaze offset, speech onset, and speech offset, the sagittal and axial views of the cortex are shown for illustration.

Apart from the CFC analyses, we performed an effective connectivity analysis across the whole brain network using multi-variant and frequency-specific Granger causality methods [184] for comparison with the CFC-based connectivity (as shown in Figure 5.7 (A)) on the one hand, and for enriching the temporal dynamics of cortical networks (as shown in Figure 5.7 (B)) on the other hand. From the brain network dynamics in Figure 5.7 (B), we could observe gamma band activities at the gaze onset in the VC for visual processing. Theta oscillation also shows early participation in the PFC, ATL and PC at the same time. Meanwhile, alpha oscillation bridged the PFC, vPMC, ATL, and the VC, serving as a mediation for the ventral path. During the period between the gaze offset and the speech onset, gamma activity flowed from the VC to the ATL and connected with the PFC, supposedly for semantic association and syntactic integration. Beta activities in the PFC, vPMC, and SMC remain inactive until the speech onset for articulation. During speech programming, slow oscillations in the theta and alpha bands firstly strengthened around the PC multisensory integration center and the PFC control networks (in the gaze-onset period), then concentrated in the motor-related areas in support of the articulatory process (in the speech-onset period). This information flow followed the dorsal path but in a top-down direction from the higher-cognitive control to the lower-sensorimotor regions. Towards the speech offset, brain activities gradually attenuated, leaving some weak PFC-PC connections for the remaining cognitive control. In summary, the frequency-dependent network dynamics reflected two significant connectivity features reemerged. One is the organizational hierarchy for the multi-level processing in different frequency bands. Another is the bidirectional information flows for the top-down and bottom-up interactions along both the dorsal and ventral paths. In line with the effective connectivity results in Figure 5.7 (B), the CFC-based neural connections in Figure 5.7 (A) also reflected frequency-specific functional hierarchies. In the theta band, the ATL serves as the intermediary in the ventral stream, top-down regulating the pMTG/pITG and the VC while transferring the associated information up to the PFC. In the dorsal stream, the PFC organizes the speech production process at a high-level by regulating the motor commands in the SMC, which downwards to the vPMC, then linking with the Spt and IPL for feedback control. In the alpha band, two regulatory center, the PFC and PC, bridges the connectivity path, one direction to the VC where visual processing is conducted, another direction to the SMC where speech motor processing is executed. In the beta band, activities are focused in the dorsal stream, concentrating on speech motor cooperation. Finally, in the gamma band, activities start from with the visual inputs in the VC, one way projecting ventrally along the temporal lobe and end up in the PFC, another way towards the dorsal path, possibly for multi-sensory integration in the PC, sensory-motor transform for phonological processing in the Spt, and auditory/somatosensory feedback in the vPMC via the projections from the Spt and IPL. These four frequency-dependent brain networks dissociated the sensory (visual, auditory, and somatosensory) processing in the gamma band, the speech motor organization in the beta band, the selective attentional regulation in the alpha band, and the top-down predictive coding and sensory-motor integration in the theta band. However, these four networks are not functioning in separations. But exquisitely interact both forward and backward along the dual streams at different levels.

5.4 Discussion

In this study, we adopted both the behavioral and neurophysiological methods to investigate the oral sentence reading process, with particular interest in the top-down syntactic and semantic effects and the underlying neural oscillatory mechanisms. In addition to conventional measurement on the eye-voice span and cortical entrainment, We introduced the CFC mechanisms in the attempt to bridge the multi-modal and multi-scale hierarchies, as well as address the 'passive verses active' controversy. Below, we will expand further discussion on the behavioral evidence for the syntactic and semantic effects and the cross-frequency coupling mechanisms involved.

Behavioral evidence for the syntactic and semantic effects

In our oral reading task, participants were asked to read aloud the word sequences, whereas understanding the meaning of the sentence is not mandatory. However, it turned out that subjects autonomously initiated semantic comprehension and developed an anticipation for the structural progression. The verification of such semantic and syntactic involvement can be testified twofold from our design of the sentence composition. First, the sixteen characters in each sentence were organized in a way that every two characters form a disyllabic word, yet these characters were presented evenly on the screen without any visual segmentation cues. The second level is where the eight disyllabic words were further grouped into three clauses and the syntactic structures emerged (Figure 5.1 (A)). Our hypothesis is that if oral sentence reading is a pure bottom-up process in a feed-forward direction, then undifferentiated treatment to the sixteen characters wold be expected. However, as demonstrated behaviorally (Figure 5.1), the eye movement trajectory (Figure 5.1 (B)) appeared with word-grouping effect instead of lexical-wise incrementality. Beside, the EVS decreased structurally rather than keeping constant. Evidently, the brain is continuously employing linguistic knowledge to interpret the accumulated information, anticipate the upcoming inputs, and adjust the ongoing sensorimotor behaviors. This phenomenon is consistent with the sensorimotor integration models of movement control [185], which proposed that neural processing of timing information in the motor system is facilitated by an internal forward model to generate predictive codes that simulate the associations between motor commands and expected sensory inputs. In case of externally cued movements, the predictive codes enable the motor system to estimate the timing of upcoming sensory stimuli to plan and execute faster responses with higher temporal precision [186]. An additional contributor to the predictive coding might be related to the syntactic or structural priming effect [**RN322**, 187, 188]. It refers to the tendency to optimize the sentence reading strategy due to the prior experience of structurally similarity ("prime") sentences [189]. Presumably, the structurally congruent sentences in our materials primed the subjects with an internal syntactic representation with relatively fixed (i.e. predictable) timing intervals, thus leads to the development of a predictive code with faster responses [190].

Cross-frequency coupling mechanisms in speech production and comprehension

Cross-frequency coupling (CFC) in neuronal coding and communication has emerged as a topic of growing interest [191]. The coupling across multiple frequency bands has been found to be an ideal mechanism for cognitive flexibility, allowing for dynamic reconfiguration of neural ensembles/circuit at various spatiotemporal scales [192]. This is particularly suitable for organizing sentence structures with tree-like hierarchy and dealing with frequent transformation at different levels and stages during speech production and comprehension. The involved cortical oscillations span a broad range of frequencies that originate from various sources and are associated with distinct cognitive functionalities [18]. Fast gamma oscillations occur spontaneously at low amplitudes and increase transiently in response to sensory stimuli [193]. Their role in the processing of sensory information has been associated with binding low-level stimulus features within local cortical circuits [194]. Meanwhile, the predominantly bottom-up-directed gamma operations are temporally coordinated by slower rhythms in a top-down direction [195]. The slower rhythm could be alpha oscillation, which are cortically widespread and regionally attenuated by a diverse range of sensory stimuli or mental activities [36]. For example, alpha desynchronization (ERD) is commonly seen in the occipital alpha oscillation during eye opening, and in the frontal eye fields that control eye movements [196]. Alpha ERD has also been identified in the sensorimotor system prior and during movements (also called 'mu' rhythm) [37] and in the auditory regions in response to acoustic stimuli (also called 'tau' rhythm) [38]. In theory, alpha oscillation stems from rhythmic fluctuation of inhibitory neurons, thus its ERD reflects a state of comparatively high excitability [28]. Nevertheless, alpha ERD are not simply a result of sensory engagement but reflect internal mental processing with top-down control, such as attention and expectation [197, 198, 169]. Beside, alph ERD has also been linked with semantic long-term memory performance [199], where the search and retrieval processes are reflected in the thalamo-cortical feedback loops via alpha-gamma coupling [16]. Therefore, alpha-gamma coupling could both serve for initial sensory gating and sensory recognition via long-term memory retrieval [200].

Another slow rhythm that could modulate gamma oscillation is the theta rhythm. Theta responds in different and opposite ways with alpha [199, 166]. Instead of desynchronization, theta synchronize (ERS) to encode input information, which is often found along the temporal lobe via hippocampo-cortical feedback loops, and goes to the prefrontal cortex as short-term or working memory (WM) [36]. WM is clearly important for language comprehension, such as keeping track of many levels of sentences or using prior interpretations to correctly predict subsequent items, etc. [201]. Theta cycles that close to syllabic-scale envelops are well-suited to group gamma-carried segments into word units in the WM. These word units could be further grouped into phrases or clauses, which is assumed to be conducted via delta-theta coupling [113, 149]. Indeed, delta oscillation is often found in the ATL where syntactic structure is supposed to be formed [202] and in the PFC-PC network where cognitive control related to syntactic structure is modulated [203]. Shortly, delta-theta-gamma oscillations forms a hierarchy of CFC that could parse sentence structure at clausal, lexical and phonological levels [113].

In addition to cortical oscillations related to sensory comprehension as above mentioned, beta oscillations have been recognized in the sensorimotor system during motor preparation and execution [41]. Beta synchronization are particularly pronounced during steady contractions but become desynchronized by voluntary movements [75, 204, 205, 206]. Beta ERD in motor execution engage not just motor structures, such as the motor and premotor cortex, the basal ganglia, the cerebellum and even the peripheral effectors [75], but also return from muscle to the central nervous system via feedback afferent pathways to the somatosensory cortex [29]. The beta coherence between M1 and S1 suggests a senorimotor role of beta oscillation that allows sensory feedbacks to be interpreted in the context of the motor command that produced it [207, 29]. Several studies have shown that beta ERD can change depending on the expectancy of a forthcoming event, reflecting an anticipatory tendency [208, 209]. Moreover, beta ERD is always concurrent with alpha ERD but with different functionalities. The alpha ERD is supposed to originate from distributed sensory networks that contributes to establishing predictive codes during the planning phase of movement [82], while the beta ERD is supposed to originate from somatotopically organized motor control areas and sustain both before and after the onset of movement [173]. Meanwhile, beta ERD is also biased by delta oscillation in the PFC based on the linguistic context and task goals [210, 211]. In sum, the CFC mechanisms provide an oscillatory framework that could organize the dual-stream hierarchy for ventral semantic comprehension and dorsal motor execution at multiple linguistic scales.

5.5 Summary

Experiment 2 investigated the spatiotemporal dynamics and neural oscillatory mechanisms of oral sentence reading. It was found that during oral sentence reading, syntactic and semantic knowledge could top-down influence the visual perception and speech production from the bottom-up. The effective communication between higher-cognitive and lower-sensorimotor cortices is supposed to be mechanistically implemented by a system of neural oscillatory hierarchy via the cross-frequency coupling in various forms. In this framework, slow delta rhythm in the prefrontal-parietal control network with larger spatiotemporal coverage coordinate the whole brain network with an underlying syntactic structure. It couples with alpha oscillation broad-wide for the regulation of excitation or inhibition in specific regions. Alpha-gamma coupling then serves selective sensory (visual, auditory, somatosensory) perception, and feature extraction from long-term memory retrieval. The identified sensory code could be further coded as WM via theta-gamma and theta-alpha couplings in the ATL. Delta also couples with theta oscillation in the ATL for the transformation from WM encoding to WM maintenance, which forms a syntactic structure and in turn develops an anticipatory code that could be used for both anticipatory perception of sensory input and predictive coding of motor commands. Delta is also supposed to modulate beta via delta-alpha-beta chain couplings to forward the motor program into motor command. The sensory feedback after motor execution reenters the ascending path via alpha-gamma coupling, thus forming a complete loop, namely sensory perception - semantic comprehension - higher-cognitive regulation - motor execution - sensory feedback, nested via fast-slow-fast oscillatory couplings. This framework could be used to coordinate the hierarchical circuits that we propose in the SPAC model from an oscillatory perspective and explain the underlying neural mechanisms as we will elaborate in the the next chapter.

Chapter6

An interactive speech production and comprehension (SPAC) model

In this Chapter, we will propose a neurofunctional model of speech production and comprehension (SPAC) based on the spatiotemporal brain network dynamics and the neural oscillatory (cross-frequency coupling) patterns revealed from Experiment 1 and Experiment 2. The SPAC model is organized as interconnected hierarchical neural circuits, corresponding to the hierarchical linguistic units in the sentence structure. And the functions of the SPAC model is based on a hierarchical neural oscillatory system, in which multiple frequency bands are nested together in various CFC forms. The SPAC incorporates the spatial-temporal-spectral aspects of brain network activities into a principled framework, and is supposed to better our comprehensive understanding of the speech production and comprehension processes as well as their interactions.

6.1 The functional structure of the SPAC model

Figure 6.1 illustrates the neurofunctional model for speech production and comprehension (SPAC) with the functional anatomy of cortical regions shown in Figure 6.1 A and the schematic diagram of information flow shown in Figure 6.1 B. In the functional anatomy, cortical regions shaded with different colors corresponds to the different PAC patterns, con-

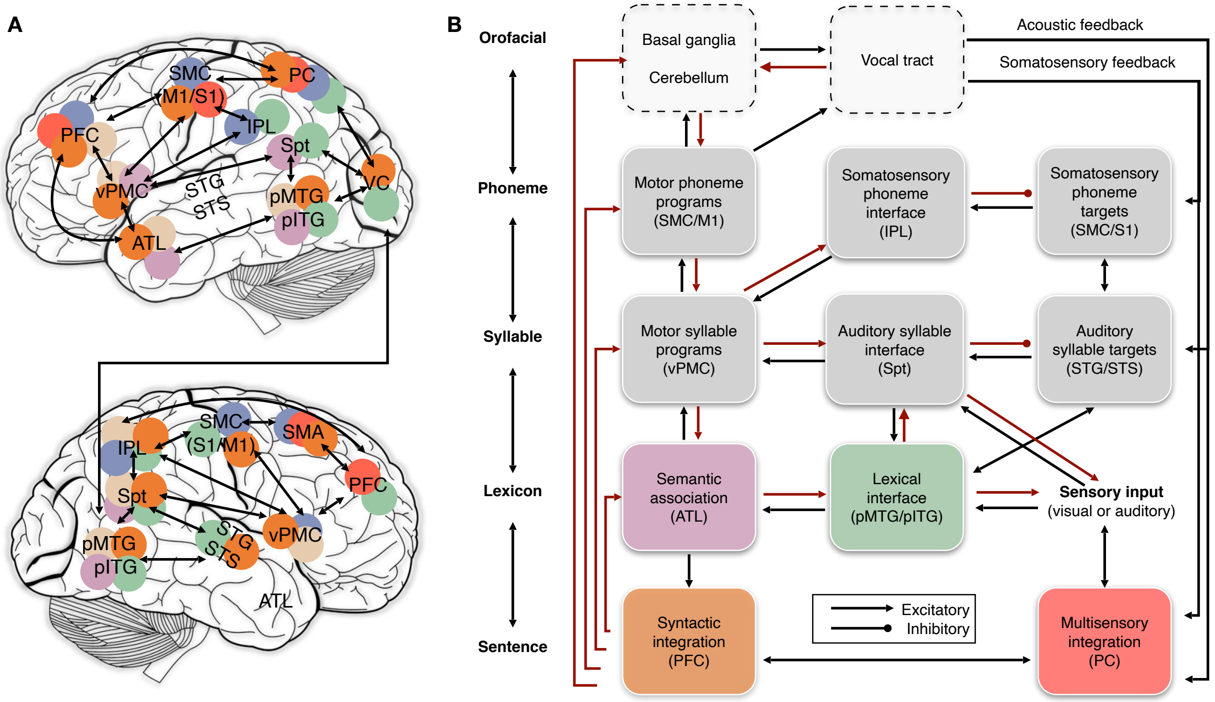


Figure 6.1: Neurofunctional model for speech production and comprehension (SPAC). (A) Functional anatomy of the SPAC model. Regions shaded with different colors corresponds to the different CFC patterns, consistent with those in Figure 5.3. (B) Schematic diagram of the SPAC mocel. The bottom-up (feedforward) and top-down (feedback) information flows are indicated by black and red arrows, respectively. The triangular and dot arrowheads represent excitatory and inhibitory effects on the downward regions.

sistent with those in Figure 5.3. Distributed cortical regions with shared CFC patterns were linked based on directions indicated by the PPC results (Figure 5.4). The excitatory and inhibitory inter-regional relations were illustrated with triangular and dot arrowheads respectively based on the AAC results in Figure 5.6. In addition, the bottom-up (feedforward) and top-down (feedback) information flows are indicated by black and red arrows respectively in the schematic diagram.

6.1.1 Sentence-level organization

The diagram is structured at five levels. From the bottom to the top covers the sentential level, the lexical level, the syllabic level, the phonemic level and the orofacial level. The model starts from the sensory input, which is the visual cortex where the character forms of the word sequences are perceived in our oral reading task. But we also included the auditory modality to generalize the SPAC model to other speech tasks involving production and comprehension, such as auditory repetition. The sensory information then flows in two directions. One flow goes along the ventral stream to the sensory-meaning interface in the pMTG/pITG for lexical interpretation, which then heads for semantic association and syntactic integration in the ATL, and forward to the frontal lobe (the PFC and vPMC) where the extracted sensory features could be transferred into appropriate actions. Another flow goes towards the Spt, the polysensory coordinate interface for phonological transformation between the sensory and motor forms [212, 73]. This transformation could also be achieved via the access to the lexical representation center (pMTG/pITG) first, then to the phonological center (Spt). The trandformed phonological form then forward to the speech programming center in the vPMC. The motor cortex (M1) in the SMC execute the speech motor plan by descending the signals to the vocal tract with the regulation of basal ganglia and cerebellum for finer tuning. On the other hand, the motor command is also supposed to be copied to the vPMC for forward prediction. This internal forward prediction in the SFC and HSFC models occurred at two level. At the auditory syllabic level, the motor features are transformed through the Spt interface to the STG/STS to compare with the auditory feedback. At the somatosensory phonemic level, it is transformed through the IPL interface to the somatosensory regions (S1) for sensorimotor integration. These error corrections are in turn sent back to the vPMC for further programming of the motor commands. Besides these two feedforward and feedback loops, we added another circuits for feedforward and feedback control at the sentence level based on the syntactic structure and semantic associations. This circuit involves the internal forward predictions from the vPMC to the semantic association center in the ATL, and compared in the lexical center in the pMTG/pITG with the sensory input. Moreover, this internal prediction and error correction are supposed to be regulated by the higher-level control in the PFC-PC cognitive network.

6.1.2 bidirectional organization

In addition to extending the SPAC model to the sentence level, another prominent feature of is the CFC explanation for the underlying neural communication mechanisms. As shown in the cortical connectivity maps on the left of Figure 6.1, we draw out the bidirectional arrows to represent the bottom-up and top-down communications between higher cognitive and lower sensorimotor regions. The forward and backward flows are supposed to be carried in different frequency bands depending on the specific communicators as differentiated by different color lines. The dual-stream model [68] also proposes some degree of bi-directionality in both the dorsal and ventral pathways, however, the details are not elaborated nor the neural oscillatory mechanisms explained. Conceptually, the effect of top-down control over the bottom-up process could be considered as a modification in the communication structure between brain areas [23], which may play out in a number of ways based on our results, such as the syntactic and semantic effect via theta-gamma coupling, the selective attention via alpha-gamma coupling, and the speech programming via theta-beta coupling, etc. Another important point to mention is the relationship between the ventral semantic comprehension and the dorsal speech production. Although we emphasized in last paragraph the sequential order of comprehension followed by production, in fact the reversed order may apply in terms of priority. Essentially, in the oral reading process, articulation is the primary purpose and comprehension the concomitant assistance. This idea has its evolutionary root of survival selection, in which survival depends on an appropriate action taken based on correct interpretation of the sensory input. In this view of motor primacy, the purpose of the brain is to guide action, whereas perception and cognition arise to adjust and optimize the actively ongoing process of the action selection [27]. Thus, our experimental support for the top-down regulation may be less of comprehension and anticipation, but more of the early preparation for a rhythmic and natural oral reading process [48]. In sum, the SPAC model levels up the speech production scheme to the sentence level and incorporates the organizational hierarchy at the syntactic, lexical and syllable levels for psycholinguistic representation as well as the motor and orofacial circuits for speech motor control (as hierarchically listed by the left side of the diagram).

6.2 Hierarchical circuits in the SPAC model

From a functional organization perspective, the SPAC model involves (1) a higher-level cognitive control circuit in the PFC and PC network for multisensory integration and internal regulation. (2) a semantic comprehension and prediction circuit in the ventral stream, which includes a sensory-semantic interface in the pMTG/pITG and a syntactic or compositional semantic center in the ATL; (3) a phonological-motor circuit in the dorsal stream, which includes the vPMC, Spt, and STG/STS for internal motor programming and auditory feedback at the syllabic level. (4) a somatosensory-motor circuit in the dorsal stream, which includes the vPMC, IPL, and S1 for internal motor programming and somatosensory feedback at the phonemic level. (5) a motor execution circuit in the dorsal stream and subcortical regions, which includes the M1, basal ganglia, cerebellum and vocal tract.

6.2.1 Higher-order cognitive control circuit

The higher-order cognitive control circuit involves cortical regions mainly in the PFC-PC network (the bottom row of circuits in Figure 6.1 B). It is supposed to be the source of top-down control over the ventral and dorsal streams in goal directioning, WM maintaining, predictive coding, motor planning, and behavioral monitoring, etc [213]. And the neural oscillatory mechanisms involved is supposed to via delta coupling with theta and alpha oscillations for short-/long-term memory encoding and selective attention inhibition [214, 215], which further couple with gamma and beta oscillations for low-level sensory and motor processing [216, 186]. A minimal requirement for a mechanism of top-down control is the convergence of diverse information, meaning it has access to diverse information about both the internal state of the system and the external state of the world [217]. And the PFC-PC circuit is anatomically and rhythmically well situated to meet this requirement [218]. For instance, the multisensory neurons in the PC receive feed-forward converging inputs from sensory-specific areas, allowing the merging of information from different senses [**RN3563**]. It connects with the PFC densely, which sends and receives projections from virtually all cortical sensory systems, motor systems, and many subcortical structures [219]. The slow delta waves carrying on these fibers could involve large neuronal populations over long time periods, which enables effective communication with long-distance assemblies in both directions [220]. Besides, the PFC is well-known for its capability of active WM maintenance [201], which is clearly important for language comprehension by keeping track of many levels of linguistic units and form syntactic structure, etc [201]. And the internal representation of goals in the WM is necessary for the anticipation of sensory input and the behavior they govern [213].

6.2.2 Ventral semantic comprehension and prediction circuit

The semantic comprehension and prediction circuit involves cortical regions in the ATL, pMTG/pITG and sensory input regions (the second bottom circuit in the SPAC model). This definition is basically consistent with classical localization of the ventral stream for comprehension, except that we consider both the visual and auditory modalities as the sensory input and emphasize the top-down syntactic influence. The neuronal function of this network mainly includes the bottom-up sensory input and recognition (via alpha-gamma CFC along the VC-pMTG/pITG path), semantic comprehension and WM encoding (via theta-gamma CFC along the pMTG/pITG-ATL path), as well as top-down regulation (via delta-alpha and delta-theta couplings along the ATL-pMTG/pITG-VC path). In this bidirectional process, the ATL holds an intermediate position that both converge the visual information in the VC and lexical information in the pMTG/pITG and forward the semantic associations to the PFC for high-level integration [221]. Meanwhile, it receives the regulatory commands and predictive coding from the PFC and descending them to the pMTG/pITG and VC to influence the low-level perceptual processing [203]. Such semantic comprehension and temporal expectation with syntactic constraints could be an automatic process[157], and have a profound influence on the structural EVS strategy and faster accessing times during sentence oral reading [222].

6.2.3 Dorsal motor-phonetic circuit

The dorsal motor-phonetic circuit involves cortical regions in the vPMC, Spt and sensory regions, including the VC for visual input and the STG/STS for auditory feedback (third row of circuits in the SPAC model). The circuits mainly serves for sensory-motor interaction for phonological processing in both forward and backward directions. On the one hand, the visual input from VC is transformed into phonological code in the Spt (via alpha-gamma coupling), which is further forwarded to the vPMC for speech motor programming at the syllabic level(via theta-gamma coupling). One the other hand, the vPMC functions under linguistic constraints from PFC (via delta-theta and delta-alpha couplings) and transforms them into phonological features in the Spt (via theta-gamma coupling), where they are compared with auditory feedback from the STG/STS (via alpha-gamma coupling). The error correction is then send back to the vPMC for future motor programming (via theta-gamma coupling) [127]. In this bottom-up - top-down/feedforward - feedback loop, the Spt region serves as a low-level sensory-motor interface that transforms between the visual/auditory inputs and the motor-related phonological codes [73]. Whereas the vPMC serves as a higher-level sensory-motor interface that collects the sensory features for speech programming and pass down the syntactic-constraint motor commands for prediction and error correction.

6.2.4 Dorsal motor-somatosensory circuit

The dorsal motor-somatosensory circuit involves cortical regions in the vPMC, IPL and SMC (the second row of circuits in the SPAC model). The function of this circuit forms a bidirectional triangle, in which the motor command is copied along the SMC (M1)-vPMC path (via delta-alpha coupling) and transformed into somatosensory features in the IPL sensory-motor interface predictively (via theta-gamma coupling). Meanwhile, the external somatosensory feedback reaches the SMC (S1) and extracted as somatosensory features in the IPL (via alpha-gamma and alpha-beta coupling). The differences between the predicted and feedback somatosensory differences would be error-coded and sent back to the vPMC for speech programming correction (via theta-gamma coupling). Again, in this circuit, the vPMC act as the higher-level of sensory-motor interface, while the IPL act as the low-level sensorimotor interface for motor feature programming. Its worth to mention that the motor programming part of the two dorsal motor circuits also involve the participation of the PFC, SMA, and the basal ganglia [41]. And the sensorimotor calibration parts of these two circuits also involve the participation of the PC and cerebellum [223].

6.2.5 Articulatory control circuit

The articulatory control circuit involves cortical regions in the SMC, and subcortical regions in the basal ganglia, thalamus and cerebellum (the top row of circuits in the SPAC model). An extention of this network could also include the PFC, PC, and vPMC [224, 225]. Beta oscillations have been found to be critical for the function of this circuit [226]. Traditionally, the basal ganglia were thought to be involved in the selection and inhibition of action commands, whereas the cerebellum was supposed to be involved in real-time fine tuning of movement [227]. More recent advances emphasize their nonmotor, cognitive functions [228]. From a computational viewpoint, it has been suggested that The cerebellum is specialized for supervised learning and is involved in externally driven (e.g. visually guided) activities [229]. The basal ganglia are specialized for reinforcement learning and is involved in internally generated (e.g. memory-guided) actions [230]. Moreover, the cerebral cortex is specialized for unsupervised learning based on neural plasticity and reciprocal connections within and between cortical areas [227]. Specific behaviors can be realized and benefit from the combination of multiple learning modules.

6.3 The interactive nature of the SPAC model

One prominent feature of the SPAC model is that it takes the active nature of speech processing into consideration, with forward and backward loops in both the ventral and dorsal streams. To avoid confusion, in the SPAC model, we define the information flow from the external (visual, auditory, somatosensory) input or feedback to the higher-cognitive center (the PFC, PC, ATL, and vPMC) as bottom-up (feedback), and the internal regulation and prediction over sensorimotor regions as top-down (feedforward). In the diagram of the SPAC model, the bottom-up and top-down directions are marked with black and red arrows, respectively. As we have described in the hierarchical organization of the SPAC model, all level of circuits (the articulatory control circuit, the dorsal motor-somatosensory/phonetic circuits and the ventral semantic comprehension and prediction circuit) involve bidirectional interactions that are regulated under the PFC-PC higher-cognitive control circuit. Particularly, in the two dorsal motor circuits, the vPMC plays a critical role as a high-level sensory-motor interface that communicates with the PFC via delta-alpha and delta-theta coupling for ascending the integrated sensory feedbacks and descending the high-level instructions (e.g., psycholinguistic constraints and motor control commands). Meanwhile, the vPMC interacts with the Spt and IPL for predicting the auditory and somatosensory features from the motor commands and receiving the feedback corrections via theta-gamma coupling in the opposite directions [132, 167]. In the ventral semantic circuit, the ATL takes a central role of intermediation, equivalent to the vPMC in the dorsal stream. On the one hand, the ATL associate lexical semantics retrieved from the pMTG/pITG via theta-gamma coupling, and forward the structured WM code to the PFC via delta-theta coupling. On the other hands, the ATL receives syntactic constraints and attentional instruction from the PFC via delta-theta and delta-alpha couplings, and pass down to the pMTG/pITG and VC for guiding the lexical retrieval and visual processes. Moreover, the ATL and vPMC also interact closely to exchange the linguistic constraints with the speech motor program. This added ventral semantic circuits fills up the black box of the conceptual system in the HSFC model [127] and linked the speech comprehension and production systems into an integrated speech chain.

6.4 Summary

In this chapter, we proposed a neurofunctional model of speech production and comprehension (SPAC) to summarize the results we found from Experiment 1 and Experiment 2. The SPAC model extended the current prevailing speech models, namely the dual-stream model and the HSFC model, from the lexical level to the sentence level. It is organized as five interconnected hierarchical neural circuits, including (1) a higher-level cognitive control circuit in the PFC and PC network for multisensory integration and internal regulation. (2) a semantic comprehension and prediction circuit in the ventral stream, which includes a sensory-semantic interface in the pMTG/pITG and a syntactic or compositional semantic center in the ATL; (3) a phonological-motor circuit in the dorsal stream, which includes the vPMC, Spt, and STG/STS for internal motor programming and auditory feedback at the syllabic level. (4) a somatosensory-motor circuit in the dorsal stream, which includes the vPMC, IPL, and S1 for internal motor programming and somatosensory feedback at the phonemic level. (5) a motor execution circuit in the dorsal stream and subcortical regions, which includes the M1, basal ganglia, cerebellum and vocal tract. Besides providing with spatiotemporal brain network network dynamics, we also explained with cross-frequency coupling mechanisms for the neural oscillatory operation. This CFC framework mediated the controversy of the passive-vs-active nature of the brain functions and linked the speech production and comprehension into a coherent speech chain.

Chapter7

Conclusion

7.1 Summary

Human speech communication, as an incredible manifestation of human intelligence, entails intricate temporal-spatial-spectral coordination of brain network dynamics. The combination of high-density electrophysiological data and behavioral data recording, together with advanced computational algorithms approved to be an effective way to trace dynamics brain activities and relate the behavioral patterns with cognitive operations. By using these innovative methods, we found that speech perception and production as a tightly linked speech chain, share overlapping networks and tend to function together in both listening and speaking tasks. They involve a ventral semantic comprehension and prediction circuit, a dorsal motor-phonetic circuit, a dorsal motor-somatosensory circuit, a sucortical (basal ganglia and cerebellum) circuit, and a higher-cognitive control circuit. These circuits are hierarchically organized and represent and process information at multiple linguistic levels and spatiotemporal scales. In this framework, information is not transferred unidirectionally or serially, rather it involves parallel and highly interactive channels carried on multi-frequency oscillations. Fast gamma oscillation carries the low-level fine features of the external stimuli bottom-up to the primary sensory regions. Alpha oscillations, originated from the thalamus inhibitory neurons, selectively suppress to allow the attended sensory information get through. Meanwhile, the alpha-gamma coupling serves in the thalamo-cortical feedback loops for extracting the sensory codes from long-term memory. Theta-gamma coupling, on the other hand, transforms the sensory codes into short-term or working memory based on the context. This function also relates to the temporal property of the theta cycle, which is close to the speech envelop and fitted for grouping (sub)phonemetic segments into interpretable lexical units. The above sensory perception and comprehension processes are performed along the ventral stream, where the semantic meaning associated in the ATL are then forwarded to the prefrontal lobe for the formation of syntactic structures and further instructions based on that. The bigger context associated with the syntactic structure is carried via slow delta oscillation and can be broad regulative via the prefrontal-parietal cognitive control network. The PFC-PC network modulate the excitation or inhibition of distinctive brain regions via delta-alpha coupling. The alpha oscillation could also couple with beta oscillation in the sensorimotor regions to coordinate the speech motor programming along the dorsal root. In this process, alpha desynchronization is associated with distributed sensory networks and contributes to establishing predictive codes during the planning phase of movement. Beta desynchronziation is predominant in the motor and somatosensory regions and sustained during the planning and execution process. After articulation, the auditory and somatosensory feedback is compared with the internal sensory predictions in the Spt and IPL respectively, and the error corrections would be send back to the vPMC for further speech motor programming. Underlying such complex coordination is a system of neural oscillations coupling with each other in a task-wise manner. With different temporal windows and spatial coverage, these frequencies are suitable for carrying multi-level information and grouping multi-scale networks distinctively. For information exchange and inter-regional functional transfer at different levels, the cross-frequency coupling (CFC) mechanisms are well-suited to bridge the cortical representational hierarchies. In short, the system of hierarchical oscillations and the mechanism of cross-frequency coupling are supposed to mediate across multi-spatiotemporal scales, build the bridge for effective communication between the external input and output with the internal mental lexicon, and form the feedforward and feedback loops between the ventral semantic and dorsal motor control circuits. Finally, based on the spatial architecture, temporal progression, and cross-frequency coupling patterns, we propose a neurofunctional model for the speech production and comprehension processes (SPAC). This comprehensive temporal-spatial-spectral framework promises to deepen our understanding of the fundamental human speech functions and be of referential values for the advancement of human-machine interactions.

7.2 Contributions

7.2.1 Brain network construction.

This study integrated the temporal, spatial, and spectral aspects of brain dynamics into a unified framework to explain the speech production and comprehension processes. These three aspects used to be investigated separately, which is not sufficient to provide a comprehensive view of the dynamic changing brain organizations. Our work bridged this gap and made it more intuitive to infer the neural mechanisms underlying the speech functions.

From the comprehensive brain network dynamics, we found that the motor system along the dorsal stream actively participated in the listening task (Experiment 1), and the semantic comprehension system along the ventral stream actively participated in the oral reading task (Experiment 2). These results indicate that speech production and comprehension are not separate systems that function on their own. Instead they form a tightly interactive speech chain that mutually develop and rely on each other in various speech tasks. Besides, we found that the hierarchical linguistic structure could be correspondingly processed by neural oscillations at different frequency bands, and the coordination of linguistic units at multiple scales are achieved via cross-frequency coupling mechanisms. These findings are not obviously available unless we inspect from a comprehensive and dynamic perspective. Thus it is crucial to integrate the spatial, temporal, and spectral aspects into a unified framework.

7.2.2 Neurofunctional model building.

This study constructed a neurofunctional model of speech production and comprehension (SPAC), which not only integrated and extended previous models for speech processing from lexical level to sentence level, but also incorporated the idea of feedforward/feedback (bottom-up/top-down) into the dynamic organizing system. More importantly, it gives a neural oscillatory explanation for the communication across the network structure, which is more fundamental in providing a deeper insight in the underlying neural mechanisms of speech functions. The CFC framework mediate the disputes in theories of sentence processing over who first (semantic or syntactic), parallel or serial, top-down or bottom-up, by assuming that both the interactions between visual input and semantic comprehension, as well as between semantic association and syntactic structure are bidirectional and parallel. These two-way communications are supported by the cross-frequency coupling between fast oscillation for forwarding fine-structure representations and slow oscillation for down-warding higher-level regulations. This is supposed to benefit our further understanding of the speech functions, and help us elucidate how our brain works ultimately.

7.2.3 Multi-modal data integration.

This study exemplified the feasibility of integrating multi-modal data acquisition system for simultaneous tracking of cognitive and behavioral patterns. The combination of high-density EEG recording with developed computational algorithms instantiated the successful application of these techniques into balancing the trade-off between temporal and spatial resolutions and exploring the highly complex speech functions. Therefore, it provides a technical reference for future studies in cognitive research.

7.3 Future Plans

For our future studies, we may apply the SPAC neurofunctional model to other speech production and comprehension tasks to test its validaty and applcability. Further, the insight from the SPAC model (e.g., the hierarchical feedforward and feedback organization, the top-down and bottom-up interactions and temporal-spatial-spectral dynamics) may have reference values in machine learning and applied to artificial neural network construction. Also it is important to create a two-way flow of insights between research on human and machine-based intelligence. We may also apply deep learning methods for analyzing increasingly large datasets and offers new models for exploring brain function.

7.3.1 Application to machine learning methods

7.3.1.1 hierarchical oscillatory framework

The SPAC model, different from other neurofunctional models that focuses on the regional functionality (functional separation) or inter-regional connectivity (functional integration), but provides a neural oscillatory framework explains the brain operation from a deeper underlying structure. The oscillatory framework may be referential for some of the bottlenecks faced in artificial neural networks. Early connectionism (from perceptron onwards) was inspired by the distributed nature of neural systems and the circuits in the neocortex and the cerebellum. These architectures led directly to the deep convolutional structures that revolutionized image processing. However, there are some unavoidable limitations in these structures. For example, the hierarchical structures in these artificial networks are embedded in the gradually deeper layers that could neither go limitless nor function in parallel. In contrast, the oscillatory framework could define oscillatory frequencies with arbitrary values and nesting these oscillations in a parallel functioning network. Moreover, in the deep learning networks, each node need to be assigned for each feature at every level. This will quickly lead to data explosion and limit further expansion. In contrast, the communication-through-coherence (CTC) in the neural oscillatory principle overcomes this bottleneck by assuming that neurons fire together wire together, thus could assemble neural network flexibly at arbitrary scales. Artificial neural networks could surely benefit from these neural oscillatory and network properties in their network construction.

7.3.1.2 Memory and attention

In the SPAC model, some of the important features that we emphasize include the top-down regulations based on memory and attention. These are achieved by alpha-gamma coupling along the thalamo-corical ascending path (alpha-inhibition for selective attention), alpha-gamma coupling in the thalamo-cortical feedback loops (long-term memory retrieval), and theta-gamma coupling in the hippocampo-cortical loops (short-term or working memory retrieval). Taking inspiration from the brain, the IBM Research team recently used machine learning techniques to develop computational models of attention and memory that are able to adapt to new environments while retaining their old memories [231]. This challenge can be broken down into short term adaptation, where there is little time to change a system and train it on what to pay attention to, and long term adaptation that is inspired by how the human brain forms memory and how neuroplasticity (e.g., adult neurogenesis) affects this process. In their design for the attention mechanism, the algorithm learns to quickly focus its attention on the right input based on a reward (i.e. feedback from its environment) obtained during the task. The higher the reward, the more attention it will place on a certain piece of input. This algorithm could further benefit from the alpha-gamma oscillatory mechanism by setting the external input with reward values (analogous to gamma amplitude) and presetting the internal attention deployment (analogous to alpha phase), and the over-threshold external input (high gamma amplitude and alpha trough) could be considered as prominent and processed with priority. This mechanism allows multiple sensory stimuli to be processed in parallel but with priority order in different phases. Another application for the long term learning is inspired by the memory forming process in the hippocampus. While synaptic plasticity, i.e. the changing strength of neuronal connections during learning, is the standard approach to neural net training, other types of plasticity, such as neurogenesis, can inspire novel learning methods, where the architecture of the network constantly adapts in response to the changing environment during lifelong learning. Beside the long-term memory, the working memory mechanism (theta-gamma coupling) as revealed in the SPAC model may provide more flexibility by coupling the current context and expectation (via internal theta entrainment) with retrieved long-term memory code (gamma oscillatory patterns) and form a specific code for the current purpose.

7.3.1.3 Supervised learning, unsupervised learning, and reinforcement learning

In the SPAC model, we mentioned that the cerebral cortex, the basal ganglia, and the the cerebellum are specialized for different types of learning. From a computational viewpoint, it has been suggested that the cerebral cortex is specialized for unsupervised learning based on Hebbian plasticity and reciprocal connections within and between cortical areas [227]. The basal ganglia is specialized for reinforcement learning (RL) based on the reward signal encoded in the dopaminergic fibers from the substantia nigra [230]. And the cerebellum is specialized for supervised learning based on the error signal encoded in the climbing fibers and the replication of arbitrary input–output mapping that was learned elsewhere in the brain.[229]. Currently, machine learning algorithms are using (1) supervised learning to train the modesl and predict outcomes (e.g., support vector machines (SVM), neural-network); (ii) unsupervised learning for data clustering, dimensionality reduction, graph network modeling, etc that are best suited for uncovering unknown statistical properties of the data set; and (3) reinforcement learning with the concept of better predictions based on dopamine-based reward system. There is no doubt that the principles from neuroscience and human behavior have led to groundbreaking ideas and practices in machine learning and could forward the progress by combining these specialized learning models into an organized artificial neural network.

7.3.2 Learning from machine learning methods

Machine Learning, in particular deep learning, has become an important tool in computational neuroscience to bridge the gap between models and data. Recent experimental techniques in neuroscience yield an increasing amount of rich data that can be fitted with complex models using novel machine learning techniques. This provides a great opportunity to strengthen the link between complex theoretical models and data. For instance, not only have deep networks set new standards in predicting responses of neural populations to arbitrary stimuli and the synthesis of novel stimuli for experimental manipulation, but novel probabilistic machine learning techniques also help to fit parameters of simulation-based (spiking) models to data. At the same time, deep learning serves as an inspiration for how biological networks could learn complex problems using biologically plausible learning rules. Our future studies could surely benefit form deep learning methods by relating high-dimensional neural data to high-dimensional behavior and using regression models to capture the behavioral or sensory features relate to neural activities.

Publications

Lists

Journal paper

* Zhao, B., Dang, J., Zhang, G. EEG source reconstruction evidence for the noun-verb neural dissociation along semantic dimensions. Neuroscience, 359, pp. 183–195, 2017.

International Conferences

* Zhao, B., Zhang, G., Dang, J. Temporal-Spatial-Spectral Investigation of Brain Network Dynamics in Human Speech Perception. Brain Informatics. BI 2020. Lecture Notes in Computer Science, vol 12241. LNAI, pp. 65–74, 2020.

* Zhao, B., Dang, J., Zhang, G., Unoki, M. Cortical oscillatory hierarchy for natural sentence processing. INTERSPEECH, 2020-October, pp. 125–129, 2020.

* Zhao, B., Huang, J., Zhang, G., Dang, J., Fu, Y., Wang, L. Brain network reconstruction of speech production based on electro-encephalography and eye movement. Acoustical Science and Technology, 41(1), pp. 349–350, 2020.

* Zhao, B., Zhang, G., Dang, J. Interactions Between Modal and Amodal Semantic Areas in Spoken Word Comprehension. ISSP, 10733 LNAI, pp. 198–207, 2018.

* Zhao, B., Huang, J., Zhang, G., Dang, J., Fu, Y., Wang, L. Revealing spatiotemporal brain dynamics of speech production based on EEG and eye movement. INTERSPEECH, 2018, 2018-September, pp. 1427–1431.

* Zhao, B., Dang, J., Zhang, G. EEG evidence for a three-phase recurrent process during spoken word processing. ISCSLP 2016, 2017, 7918465.

* Zhao, B., Dang, J., Zhang, G. Investigation of noun-verb dissociation based on EEG source reconstruction. APSIPA 2016, 2017, 7820817.

* Guo, T., Dang, J., Zhang, G., Zhao, B., Unoki, M. Frequency-specific Brain Network Dynamics during Perceiving Real Words and Pseudowords. ISCSLP 2021, 2021, 9362052.

* Huang, J., Zhao, B., Dang, J., Chen, M. Investigation of speech-planning mechanism based on eye movement and EEG. APSIPA ASC 2019, 2019, pp. 945–950, 9023038

Bibliography

[1] Steven L. Small and Gregory Hickok. “Chapter 1 - The Neurobiology of Language”. In: *Neurobiology of Language*. Ed. by Gregory Hickok and Steven L. Small. San Diego: Academic Press, 2016, pp. 3–9. ISBN: 978-0-12-407794-2. DOI: `https://doi.org/10.1016/B978-0-12-407794-2.00001-8`. URL: `https://www.sciencedirect.com/science/article/pii/B9780124077942000018`.

[2] P. Broca. “Remarques sur le siege de la faculte du langage aricule suivies d’une observation d7aphemie”. In: *Bull Soc Anat* 6 ().

[3] C Wernicke. “Der Aphasiche Symptomenkomplex”. In: (1874).

[4] G Pfurtscheller and A Aranibar. “Event-related cortical desynchronization detected by power measurements of scalp EEG”. In: *Electroencephalography and Clinical Neurophysiology* 42.6 (1977), pp. 817–826.

[5] V. Sakkalis. “Review of advanced techniques for the estimation of brain connectivity measured with EEG/MEG”. In: *Computers in Biology and Medicine* 41.12 (2011), pp. 1110–1117.

[6] C Price. “A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading.” In: *Neuroimage* 65 (2012), pp. 816–847. DOI: `10.1016/j.neuroimage.2012.04.062.Epub2012May12.`.

[7] C Wernicke. *The Symptom Complex of Aphasia*. 1969.

[8] Norman Geschwind. “The Organization of Language and the Brain”. In: *Science, New Series* 170.3961 (1970), pp. 940–944.

[9] Pascale Tremblay and Anthony Steven Dick. “Broca and Wernicke are dead, or moving past the classic model of language neurobiology”. In: *Brain Language* 162 (2016), pp. 60–71.

[10] D. G. Hillert. *Neurobiology of Language*. Encyclopedia of Evolutionary Psychological Science, 2016.

[11] Silvia Bunge and Itamar Kahn. “Cognition: An Overview of Neuroimaging Techniques”. In: vol. 2. 2009, pp. 1063–1067. ISBN: 9780080450469. DOI: `10.1016/B978-008045046-9.00298-9`.

[12] Van De Ven Vincent, Esposito Fabrizio, and Ingrid K Christoffels. “Neural network of speech monitoring overlaps with overt speech production and comprehension net-

works: a sequential spatial and temporal ICA study". In: *Neuroimage* 47.4 (2009), pp. 1982–1991.

[13] A. Prasad et al. "Current and future functional imaging techniques for post-traumatic stress disorder". In: *RSC Advances* (2019).

[14] F Pulvermuller. "Active perception: sensorimotor circuits as a cortical basis for language". In: *Nature Reviews Neuroscience* 11.5 (2010), pp. 351–360.

[15] Kestutis Kveraga, Avniel S. Ghuman, and Moshe Bar. "Top-down predictions in the cognitive brain". In: *Brain and Cognition* 65.2 (2007), pp. 145–168. ISSN: 0278-2626. DOI: `https://doi.org/10.1016/j.bandc.2007.06.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0278262607000954`.

[16] F. Roux and P. J. Uhlhaas. "Working memory and neural oscillations: alpha–gamma versus theta–gamma codes for distinct WM information?" In: *Trends in Cognitive ences* 18.1 (2014), pp. 16–25.

[17] Antonio Di Ieva. *The Fractal Geometry of the Brain.* 2016. ISBN: 978-1-4939-3995-4. DOI: `10.1007/978-1-4939-3995-4`.

[18] G. Buzsaki. *Rhythms of the Brain.* Rhythms of The Brain, 2009.

[19] Markku Penttonen and György Buzsáki. "Natural logarithmic relationship between brain oscillators". In: *Thalamus and Related Systems* (2003).

[20] György Buzsáki and Andreas Draguhn. "Neuronal Oscillations in Cortical Networks". In: *Science* 304.5679 (2004), pp. 1926–1929. DOI: `10.1126/science.1099745`. URL: `https://science.sciencemag.org/content/sci/304/5679/1926.full.pdf`.

[21] Gphld Silva. "Event-related EEG/MEG synchronization and desynchronization: basic principles". In: *Clinical Neurophysiology* (1999).

[22] . Varela F. et al. "The brainweb: phase synchronization and large-scale integration". In: *Nature Reviews Neuroscience* 2.4 (2001), pp. 229–239.

[23] Pascal Fries. "A mechanism for cognitive dynamics: neuronal communication through neuronal coherence". In: *Trends in Cognitive Sciences* 9.10 (2005), pp. 474–480. ISSN: 1364-6613. DOI: `https://doi.org/10.1016/j.tics.2005.08.011`. URL: `https://www.sciencedirect.com/science/article/pii/S1364661305002421`.

[24] Giraud Anne-Lise and Poeppel David. "Cortical oscillations and speech processing: emerging computational principles and operations". In: *Nature Neuroscience* 15.4 (2012), pp. 511–7.

[25] Erol Başar et al. "Gamma, alpha, delta, and theta oscillations govern cognitive processes". In: *International Journal of Psychophysiology* 39.2 (2001), pp. 241–248.

[26] P. L. Nunez and E. Harth. "Electric Fields of the Brain: The Neurophysics of EEG". In: *Physics Today* 35.6 (1982), p. 59.

[27] Ryan T. Canolty and Robert T. Knight. "The functional role of cross-frequency coupling". In: *Trends in Cognitive Sciences* 14.11 (2010), pp. 506–515. ISSN: 1364-6613. DOI: https://doi.org/10.1016/j.tics.2010.09.001. URL: https://www.sciencedirect.com/science/article/pii/S1364661310002068.

[28] Wolfgang Klimesch, Paul Sauseng, and Simon Hanslmayr. "EEG alpha oscillations: The inhibition–timing hypothesis". In: *Brain Research Reviews* 53.1 (2007), pp. 63–88. ISSN: 0165-0173. DOI: https://doi.org/10.1016/j.brainresrev.2006.06.003. URL: https://www.sciencedirect.com/science/article/pii/S016501730600083X.

[29] A. Engel. "Beta-band oscillations–signalling the status quo?" In: *Current Opinion in Neurobiology* (2010).

[30] Astrid von Stein and Johannes Sarnthein. "Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization". In: *International Journal of Psychophysiology* 38.3 (2000), pp. 301–313. ISSN: 0167-8760. DOI: https://doi.org/10.1016/S0167-8760(00)00172-0. URL: http://www.sciencedirect.com/science/article/pii/S0167876000001720.

[31] M Steriade, D. Mccormick, and T. Sejnowski. "Thalamocortical oscillations in the sleeping and aroused brain". In: *ence* 262.5134 (1993), pp. 679–85.

[32] G. G. Knyazev. "Motivation, emotion, and their inhibitory control mirrored in brain oscillations". In: *Neuroscience Biobehavioral Reviews* 31.3 (2007), pp. 377–395.

[33] Uwe Friese et al. "Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG". In: *NeuroImage* 66 (2013), pp. 642–647. ISSN: 1053-8119. DOI: https://doi.org/10.1016/j.neuroimage.2012.11.002. URL: http://www.sciencedirect.com/science/article/pii/S105381191201097X.

[34] G Buzsáki and E. I. Moser. "Memory, navigation and theta rhythm in the hippocampal-entorhinal system". In: *Nature Neuroscience* 16.2 (2013), pp. 130–138.

[35] Prasad R. Shirvalkar, Peter R. Rapp, and Matthew L. Shapiro. "Bidirectional changes to hippocampal theta–gamma comodulation predict memory for recent spatial episodes". In: 107.15 (2010), pp. 7054–7059. DOI: 10.1073/pnas.0911184107JProceedingsoftheNationalAca URL: https://www.pnas.org/content/pnas/107/15/7054.full.pdf.

[36] W. Klimesch. "Memory processes described as brain oscillations in the EEG-alpha and theta bands". In: *Psycoloquy* (1995).

[37] D Thornton et al. "Sensorimotor activity measured via oscillations of EEG mu rhythms in speech and non-speech discrimination tasks with and without segmentation demands". In: *Brain and Language* (2017).

[38] Riitta Hari, Sari Levänen, and Tommi Raij. "Timing of human cortical functions during cognition: role of MEG". In: *Trends in Cognitive Sciences* 4.12 (2000), p. 455.

[39] Elizabeth Heinrichs-Graham and Tony W. Wilson. "Is an absolute level of cortical beta suppression required for proper movement? Magnetoencephalographic evidence from healthy aging". In: *NeuroImage* 134 (2016), pp. 514–521. ISSN: 1053-8119. DOI: `https://doi.org/10.1016/j.neuroimage.2016.04.032`. URL: `https://www.sciencedirect.com/science/article/pii/S1053811916300726`.

[40] Douglas Cheyne, Leyla Bakhtazad, and William Gaetz. "Spatiotemporal mapping of cortical activity accompanying voluntary movements using an event-related beamforming approach". In: *Human Brain Mapping* 27.3 (2006), pp. 213–229. ISSN: 1065-9471. DOI: `https://doi.org/10.1002/hbm.20178`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.20178`.

[41] Bjørg Elisabeth Kilavik et al. "The ups and downs of beta oscillations in sensorimotor cortex". In: *Experimental Neurology* 245 (2013), pp. 15–26. ISSN: 0014-4886. DOI: `https://doi.org/10.1016/j.expneurol.2012.09.014`. URL: `https://www.sciencedirect.com/science/article/pii/S0014488612003767`.

[42] Pascal Fries. "Neuronal Gamma-Band Synchronization as a Fundamental Process in Cortical Computation". In: 32.1 (2009), pp. 209–224. DOI: `10.1146/annurev.neuro.051508.135603`. URL: `https://www.annualreviews.org/doi/abs/10.1146/annurev.neuro.051508.135603`.

[43] G. C. O'Neill et al. "Dynamics of large-scale electrophysiological networks: A technical review". In: *Neuroimage* (2017).

[44] Thomas Akam and Dimitri M. Kullmann. "Oscillations and Filtering Networks Support Flexible Routing of Information". In: *Neuron* 67.2 (2010), pp. 308–320. ISSN: 0896-6273. DOI: `https://doi.org/10.1016/j.neuron.2010.06.019`. URL: `https://www.sciencedirect.com/science/article/pii/S0896627310004770`.

[45] Andreas K. Engel, Pascal Fries, and Wolf Singer. "Dynamic predictions: Oscillations and synchrony in top–down processing". In: *Nature Reviews Neuroscience* 2.10 (2001), pp. 704–716. ISSN: 1471-0048. DOI: `10.1038/35094565`. URL: `https://doi.org/10.1038/35094565`.

[46] Scott Makeig et al. "Mining event-related brain dynamics". In: *Trends in Cognitive Sciences* 8.5 (2004), pp. 204–210. ISSN: 1364-6613. DOI: `https://doi.org/10.1016/j.tics.2004.03.008`. URL: `https://www.sciencedirect.com/science/article/pii/S1364661304000816`.

[47] A. S. Shah et al. "Neural dynamics and the fundamental mechanisms of event-related brain potentials". In: *CEREBRAL CORTEX -NEW YORK- OXFORD UNIVERSITY PRESS-* (2004).

[48] Charles E. Schroeder and Peter Lakatos. "Low-frequency neuronal oscillations as instruments of sensory selection". In: *Trends in Neurosciences* 32.1 (2009), pp. 9–18. ISSN: 0166-2236. DOI: `https://doi.org/10.1016/j.tins.2008.09.012`. URL: `https://www.sciencedirect.com/science/article/pii/S0166223608002506`.

[49] Peter Lakatos et al. "Neuronal Oscillations and Multisensory Interaction in Primary Auditory Cortex". In: *Neuron* 53.2 (2007), pp. 279–292. ISSN: 0896-6273. DOI: `https://doi.org/10.1016/j.neuron.2006.12.011`. URL: `https://www.sciencedirect.com/science/article/pii/S0896627306009962`.

[50] F. W. Ohl, H. Scheich, and W. J. Freeman. "Change in pattern of ongoing cortical activity with auditory category learning". In: *Nature* 412.6848 (2001), pp. 733–6. ISSN: 0028-0836 (Print) 0028-0836. DOI: `10.1038/35089076`.

[51] R. VanRullen and C. Koch. "Is perception discrete or continuous?" In: *Trends Cogn Sci* 7.5 (2003), pp. 207–213. ISSN: 1364-6613. DOI: `10.1016/s1364-6613(03)00095-0`.

[52] Kristijan Armeni et al. "Frequency-specific brain dynamics related to prediction during language comprehension". In: *NeuroImage* 198 (2019), pp. 283–295. ISSN: 1053-8119. DOI: `https://doi.org/10.1016/j.neuroimage.2019.04.083`. URL: `http://www.sciencedirect.com/science/article/pii/S1053811919304057`.

[53] Alexandre Hyafil. "Misidentifications of specific forms of cross-frequency coupling: three warnings". In: 9.370 (2015). ISSN: 1662-453X. DOI: `10.3389/fnins.2015.00370`. URL: `https://www.frontiersin.org/article/10.3389/fnins.2015.00370`.

[54] Alexandre Hyafil et al. "Neural Cross-Frequency Coupling: Connecting Architectures, Mechanisms, and Functions". In: *Trends in Neurosciences* 38.11 (2015), pp. 725–740. ISSN: 0166-2236. DOI: `https://doi.org/10.1016/j.tins.2015.09.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0166223615002088`.

[55] Roberto C. Sotero. "Topology, Cross-Frequency, and Same-Frequency Band Interactions Shape the Generation of Phase-Amplitude Coupling in a Neural Mass Model of a Cortical Column". In: *PLOS Computational Biology* 12.11 (2016), e1005180. DOI:

10.1371/journal.pcbi.1005180. URL: `https://doi.org/10.1371/journal.pcbi.1005180`.

[56] Ole Jensen and Laura L. Colgin. "Cross-frequency coupling between neuronal oscillations". In: *Trends in Cognitive Sciences* 11.7 (2007), pp. 267–269. ISSN: 1364-6613. DOI: `https://doi.org/10.1016/j.tics.2007.05.003`. URL: `http://www.sciencedirect.com/science/article/pii/S1364661307001271`.

[57] P. Lakatos et al. "An oscillatory hierarchy controlling neuronal excitability and stimulus processing in the auditory cortex". In: *J Neurophysiol* 94.3 (2005), pp. 1904–11. ISSN: 0022-3077 (Print) 0022-3077 (Linking). DOI: `10.1152/jn.00263.2005`. URL: `http://www.ncbi.nlm.nih.gov/pubmed/15901760`.

[58] John Lisman and György Buzsáki. "A Neural Coding Scheme Formed by the Combined Function of Gamma and Theta Oscillations". In: *Schizophrenia Bulletin* 34.5 (2008), pp. 974–980. ISSN: 0586-7614. DOI: `10.1093/schbul/sbn060`. URL: `https://doi.org/10.1093/schbul/sbn060`.

[59] MX Cohen. "Analyzing Neural Time Series Data". In: *MIT Press* (2014).

[60] G. Hickok and D. Poeppel. "The cortical organization of speech processing". In: *Nat. Reviews. Neurosci.* 8 (2007), pp. 393–402.

[61] J. R. Binder et al. "Human temporal lobe activation by speech and nonspeech sounds". In: *Cerebral Cortex* 10.5 (2000), pp. 512–528.

[62] Sophie K Scott et al. "Identification of a pathway for intelligible speech in the left temporal lobe". In: *Brain A Journal of Neurology* 123 Pt 12.12 (2000), p. 2400.

[63] C Price et al. "Segregating Semantic from Phonological Processes during Reading". In: *Journal of Cognitive Neuroscience* 9.6 (2014), pp. 727–733.

[64] F Pulvermuller et al. "Brain rhythms of language: nouns versus verbs". In: *European Journal of Neuroscience* 8.5 (1996), pp. 937–941.

[65] Timothy T. Rogers et al. "Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging". In: *Cognitive Affective and Behavioral Neuroscience* 6.3 (2006), p. 201.

[66] Angela D. Friederici. "Towards a neural basis of auditory sentence processing". In: *Trends in Cognitive Sciences* 6.2 (2002), pp. 78–84. ISSN: 1364-6613. DOI: `https://doi.org/10.1016/S1364-6613(00)01839-8`. URL: `https://www.sciencedirect.com/science/article/pii/S1364661300018398`.

[67] Gregory Hickok and David Poeppel. “Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language”. In: *Cognition* 92.1-2 (2004), pp. 67–99.

[68] G Hickok. “The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model”. In: *Journal of Communication Disorders* 45.6 (2012), p. 393.

[69] Alan D. Baddeley and Graham Hitch. “Working Memory”. In: *Psychology of Learning and Motivation* 255.5044 (1992), pp. 556–559.

[70] Lyn Frazier. *Theories of Sentence Processing.* Modularity in knowledge representation and natural-language understanding, 1987.

[71] J. L. Mcclelland and D. E. Rumelhart. “An interactive activation model of context effects in letter perception: I. An account of basic findings”. In: *Readings in Cognitive Science* 88.88 (1988), pp. 580–596.

[72] W. Marslen-Wilson and L. K. Tyler. “The temporal structure of spoken language understanding”. In: *Cognition* 8.1 (1980), pp. 1–71.

[73] G Hickok et al. “Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt”. In: *Journal of Cognitive Neuroscience* 15.5 (2003), pp. 673–682.

[74] John Houde and Srikantan Nagarajan. “Speech Production as State Feedback Control”. In: *Frontiers in Human Neuroscience* 5.82 (2011). ISSN: 1662-5161. DOI: `10.3389/fnhum.2011.00082`. URL: `https://www.frontiersin.org/article/10.3389/fnhum.2011.00082`.

[75] S. N. Baker. “Oscillatory interactions between sensorimotor cortex and the periphery”. In: *Current Opinion in Neurobiology* 17.6 (2007), pp. 649–655.

[76] Wjm Levelt, A. Roelofs, and A. S. Meyer. “A theory of lexical access in speech production”. In: *Behavioral and Brain Sciences* 22.1 (1999), pp. 1–38.

[77] F.H. Guenther. “Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production”. In: *Psychological Review* 102 (1995), pp. 594–621.

[78] Frank Guenther, Michelle Hampson, and Dave Johnson. “A Theoretical Investigation of Reference Frames for the Planning of Speech Movements”. In: *Psychological Review* 105 (1998), pp. 611–633.

[79] Eric R. Kandel, James Harris Schwartz, and Thomas M. Jessell. *Principles of neural science*. McGraw-Hill, Health Professions Division, 2000, p. 1414.

[80] Gregory Hickok. "Computational neuroanatomy of speech production". In: *Nature reviews. Neuroscience* 13.2 (2012), pp. 135–145. ISSN: 1471-0048 1471-003X. DOI: `10.1038/nrn3158`. URL: `https://pubmed.ncbi.nlm.nih.gov/22218206https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5367153/`.

[81] Vos M De et al. "Removal of muscle artifacts from EEG recordings of spoken language production". In: *Neuroinformatics* 8.2 (2010), pp. 135–150.

[82] D Jenson et al. "Temporal dynamics of sensorimotor integration in speech perception and production: independent component analysis of EEG data". In: *Frontiers in Psychology* 5.10 (2014), p. 656.

[83] Nai Ding et al. "Cortical Tracking of Hierarchical Linguistic Structures in Connected Speech". In: *Nature Neuroscience* 19.1 (2016), pp. 158–164.

[84] Dashiel Munding, Anne Sophie Dubarry, and F. Xavier Alario. "On the cortical dynamics of word production: a review of the MEG evidence". In: *Language Cognition and Neuroscience* 4 (2017), pp. 1–22.

[85] R. Oostenveld and P. Praamstra. "The five percent electrode system for high-resolution EEG and ERP measurements". In: *Clin Neurophysiol* 112.4 (2001), pp. 713–9. ISSN: 1388-2457 (Print) 1388-2457 (Linking). URL: `http://www.ncbi.nlm.nih.gov/pubmed/11275545`.

[86] B. Zhao, J. Dang, and G. Zhang. "EEG Source Reconstruction Evidence for the Noun-Verb Neural Dissociation along Semantic Dimensions". In: *Neuroscience* 359 (2017).

[87] C. A. Kothe and S. Makeig. "BCILAB: a platform for brain-computer interface development". In: *J Neural Eng* 10.5 (2013), p. 056014. ISSN: 1741-2552. DOI: `10.1088/1741-2560/10/5/056014`.

[88] C. Y. Chang et al. "Evaluation of Artifact Subspace Reconstruction for Automatic Artifact Components Removal in Multi-Channel EEG Recordings". In: *IEEE Transactions on Biomedical Engineering* 67.4 (2020), pp. 1114–1121. ISSN: 1558-2531. DOI: `10.1109/TBME.2019.2930186`.

[89] S Makeig. "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones". In: *Electroencephalography and Clinical Neurophysiology* 86.4 (1993), pp. 283–93.

[90] A Delorme and S Makeig. “EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis”. In: *Journal of Neuroscience Methods* 134.1 (2004), p. 9.

[91] Luca Pion-Tonachini, Ken Kreutz-Delgado, and Scott Makeig. “ICLabel: An automated electroencephalographic independent component classifier, dataset, and website”. In: *NeuroImage* 198 (2019), pp. 181–197. ISSN: 1095-9572 1053-8119. DOI: `10.1016/j.neuroimage.2019.05.026`. URL: `https://pubmed.ncbi.nlm.nih.gov/31103785https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6592775/`.

[92] Christoph M. Michel et al. “EEG source imaging”. In: *Clinical Neurophysiology* 115.10 (2004), pp. 2195–2222. ISSN: 1388-2457. DOI: `https://doi.org/10.1016/j.clinph.2004.06.001`. URL: `https://www.sciencedirect.com/science/article/pii/S1388245704002135`.

[93] C. Granger et al. “Investigationg causal relations by econometric models: cross spectral methods”. In: (1969).

[94] N. Wiener. “The theory of prediction”. In: *modern mathematics for the engineer* (1956).

[95] R. T. Canolty et al. “High Gamma Power Is Phase-Locked to Theta Oscillations in Human Neocortex”. In: *Science* 313.5793 (2006), pp. 1626–1628. DOI: `10.1126/science.1128115`. URL: `https://science.sciencemag.org/content/sci/313/5793/1626.full.pdf`.

[96] J. E. Peelle and M. H. Davis. “Neural Oscillations Carry Speech Rhythm through to Comprehension”. In: *Frontiers in Psychology* 3.320 (2012), p. 320.

[97] M. Joao et al. “Decoding Articulatory Features from fMRI Responses in Dorsal Speech Regions”. In: *The Journal of neuroscience : the official journal of the Society for Neuroscience* (2015).

[98] Friedemann Pulvermu¨ ller et al. “Motor cortex maps articulatory features of speech sounds”. In: *Proceedings of the National Academy of Sciences of the United States of America* 103.20 (2006), pp. 7865–7870.

[99] M. Riikka, D. Rebekah, and K. E. Watkins. “Auditory-Motor Processing of Speech Sounds”. In: *Cerebral Cortex* 5 (2013), pp. 1190–1197.

[100] A. M. Liberman et al. “Perception of the speech code”. In: *Psychological Review* 74.6 (1967), p. 431.

[101] A. Liberman and G. Mattingly. “The motor theory of speech perception revised”. In: *Cognition* 21 (1985), pp. 1–36.

[102] S. M. Wilson et al. "Listening to speech activates motor areas involved in speech production". In: *Nat. Neuro* 7.7 (2004), pp. 701–702.

[103] Alessandro D'Ausilio et al. "The Motor Somatotopy of Speech Perception". In: *Current Biology Cb* 19.5 (2009), pp. 381–5.

[104] R. Möttönen, J. Rogers, and K. E. Watkins. "Stimulating the lip motor cortex with transcranial magnetic stimulation". In: *Journal of Visualized Experiments Jove* e51665.88 (2014).

[105] R. Mottonen and K. E. Watkins. "Motor representations of articulators contribute to categorical perception of speech sounds". In: *Journal of Neuroscience* 29.31 (2009), pp. 9819–9825.

[106] I. G. Meister et al. "The Essential Role of Premotor Cortex in Speech Perception". In: *Current Biology* 17.19 (2007), pp. 1692–1696.

[107] M. R. Schomers et al. "Causal Influence of Articulatory Motor Cortex on Comprehending Single Spoken Words: TMS Evidence". In: *Cerebral Cortex* 10 (2015), pp. 3894–3902.

[108] Analia L. Arévalo, Juliana V. Baldo, and Nina F. Dronkers. "What do brain lesions tell us about theories of embodied semantics and the human mirror neuron system?" In: *Cortex* 48.2 (2012), pp. 242–254.

[109] L. Aziz-Zadeh and A. Damasio. "Embodied semantics for actions: Findings from functional brain imaging". In: *Journal of Physiology - Paris* 102.1-3 (2008), pp. 35–39.

[110] Gabriella Vigliocco et al. "Nouns and verbs in the brain: a review of behavioural, electrophysiological, neuropsychological and imaging studies". In: *Neuroscience and Biobehavioral Reviews* 35.3 (2011), pp. 407–426.

[111] R. C. Oldfield. "The assessment and analysis of handedness : the Edinburgh inventory". In: *Neuropsychologia* 9 (1971).

[112] Bin Zhao, Jianwu Dang, and Gaoyan Zhang. "A Neuro-Experimental Evidence for the Motor Theory of Speech Perception". In: *INTERSPEECH*, pp. 2441–2445.

[113] Guangting Mai, James W. Minett, and S. Y. Wang. "Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing". In: *Neuroimage* 133 (2016), pp. 516–528.

[114] Nai Ding and Simon Jonathan Z. “Cortical entrainment to continuous speech: functional roles and interpretations”. In: *Frontiers in Human Neuroscience* 8.2 (2014), p. 311.

[115] Adriano B L Tort et al. “Theta-gamma coupling increases during the learning of item-context associations”. In: *Proceedings of the National Academy of Sciences of the United States of America* 106.49 (2009), pp. 20942–20947.

[116] J. E. Lisman and O. Jensen. “The Theta-Gamma Neural Code”. In: *Neuron* 77.6 (2013), pp. 1002–1016.

[117] Mathias Vukelić et al. “Lateralized alpha-band cortical networks regulate volitional modulation of beta-band sensorimotor oscillations”. In: *Neuroimage* 87.4 (2014), pp. 147–153.

[118] H. Yuan et al. “Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: An EEG and fMRI study of motor imagery and movements”. In: *Neuroimage* 49.3 (2010), pp. 2596–2606.

[119] John Foxe and Adam Snyder. “The Role of Alpha-Band Brain Oscillations as a Sensory Suppression Mechanism during Selective Attention”. In: *Frontiers in Psychology* 2.154 (2011). ISSN: 1664-1078. DOI: `10.3389/fpsyg.2011.00154`. URL: `https://www.frontiersin.org/article/10.3389/fpsyg.2011.00154`.

[120] B. Zhao, G. Zhang, and J. Dang. “Temporal-Spatial-Spectral Investigation of Brain Network Dynamics in Human Speech Perception”. In: *International Conference on Brain Informatics*.

[121] A. Stuart et al. “Effect of delayed auditory feedback on normal speakers at two speech rates”. In: *Journal of the Acoustical Society of America* 111.5 (2002), pp. 2237–41.

[122] Houde and F J. “Sensorimotor adaptation in speech production”. In: *Science* 279.5354 (1998), pp. 1213–1216.

[123] M Fabbri-Destro and G Rizzolatti. “Mirror neurons and mirror systems in monkeys and humans”. In: *Physiology* 23.3 (2008), pp. 171–179.

[124] G Buccino et al. “Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study”. In: *European Journal of Neuroscience* 13.2 (2001), p. 400.

[125] A. J. Lotto, G. S. Hickok, and L. L. Holt. “Reflections on mirror neurons and speech perception”. In: *Trends in Cognitive Sciences* 13.3 (2009), pp. 110–114.

[126] Alessandro D"Ausilio et al. "The motor somatotopy of speech perception". In: *Current biology: CB* 19.5 (2009), pp. 381–385.

[127] G Hickok. "Computational neuroanatomy of speech production". In: *Nature Reviews Neuroscience* 13.2 (2012), pp. 135–45.

[128] Justin Riddle, Amber McFerren, and Flavio Frohlich. "Causal role of cross-frequency coupling in distinct components of cognitive control". In: *Progress in Neurobiology* 202 (2021), p. 102033. ISSN: 0301-0082. DOI: `https://doi.org/10.1016/j.pneurobio.2021.102033`. URL: `https://www.sciencedirect.com/science/article/pii/S0301008221000472`.

[129] Jonathan S. Brumberg et al. "Spatio-Temporal Progression of Cortical Activity Related to Continuous Overt and Covert Speech Production in a Reading Task". In: *Plos One* 11.11 (2016), e0166872.

[130] J. Sun et al. "Neural Encoding and Decoding With Distributed Sentence Representations". In: *IEEE Transactions on Neural Networks and Learning Systems* (2020), pp. 1–15. ISSN: 2162-2388. DOI: `10.1109/TNNLS.2020.3027595`.

[131] Astrid von Stein, Carl Chiang, and Peter König. "Top-down processing mediated by interareal synchronization". In: *Proceedings of the National Academy of Sciences* 97.26 (2000), pp. 14748–14753. DOI: `10.1073/pnas.97.26.14748`. URL: `https://www.pnas.org/content/pnas/97/26/14748.full.pdf`.

[132] Christian D. Márton et al. "Signature Patterns for Top-Down and Bottom-Up Information Processing via Cross-Frequency Coupling in Macaque Auditory Cortex". In: *eNeuro* 6.2 (2019), ENEURO.0467–18.2019. ISSN: 2373-2822. DOI: `10.1523/ENEURO.0467-18.2019`. URL: `https://pubmed.ncbi.nlm.nih.gov/31088914https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6520641/`.

[133] E. J. Beier et al. "Cortical Tracking of Speech: Toward Collaboration between the Fields of Signal and Sentence Processing". In: *Journal of Cognitive Neuroscience* 33.4 (2021), pp. 574–593. ISSN: 0898-929X. DOI: `10.1162/jocn_a_01676`.

[134] Marsel Mesulam. "Representation, inference, and transcendent encoding in neurocognitive networks of the human brain". In: *Annals of Neurology* 64.4 (2008), pp. 367–378. ISSN: 0364-5134. DOI: `https://doi.org/10.1002/ana.21534`. URL: `https://doi.org/10.1002/ana.21534`.

[135] L. Shuai and T. Gong. "Temporal relation between top-down and bottom-up processing in lexical tone perception". In: *Front Behav Neurosci.* 25.doi: 10.3389/fnbeh.2014.00097. eCollection 2014 (2014).

[136] Hyoe Tomita et al. "Top-down signal from prefrontal cortex in executive control of memory retrieval". In: *Nature* 401.6754 (1999), pp. 699–703. ISSN: 1476-4687. DOI: `10.1038/44372`. URL: `https://doi.org/10.1038/44372`.

[137] Lesya Y. Ganushchak and Yiya Chen. "Incrementality in Planning of Speech During Speaking and Reading Aloud: Evidence from Eye-Tracking". In: *Frontiers in Psychology* 7.33 (2016). ISSN: 1664-1078. DOI: `10.3389/fpsyg.2016.00033`. URL: `https://www.frontiersin.org/article/10.3389/fpsyg.2016.00033`.

[138] Agnieszka E. Konopka and Antje S. Meyer. "Priming sentence planning". In: *Cognitive Psychology* 73 (2014), pp. 1–40. ISSN: 0010-0285. DOI: `https://doi.org/10.1016/j.cogpsych.2014.04.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0010028514000279`.

[139] Agnieszka E. Konopka. "Planning ahead: How recent experience with structures and words changes the scope of linguistic planning". In: *Journal of Memory and Language* 66.1 (2012), pp. 143–162. ISSN: 0749-596X. DOI: `https://doi.org/10.1016/j.jml.2011.08.003`. URL: `https://www.sciencedirect.com/science/article/pii/S0749596X11000957`.

[140] H. Luo et al. "Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation". In: *PLoS Biology* 8.8 (2010), e1000445–.

[141] M. Gomez-Ramirez et al. "Oscillatory Sensory Selection Mechanisms during Intersensory Attention to Rhythmic Auditory and Visual Inputs: A Human Electrocorticographic Investigation". In: *Journal of Neuroscience* 31.50 (2011), pp. 18556–18567.

[142] Besle et al. "Tuning of the Human Neocortex to the Temporal Dynamics of Attended Events". In: *Journal of Neuroscience* 31.9 (2011), pp. 3176–3185.

[143] Lars Riecke et al. "Neural Entrainment to Speech Modulates Speech Intelligibility". In: *Current Biology* 28.2 (2017), pp. 161–169.

[144] J. Obleser and C. Kayser. "Neural Entrainment and Attentional Selection in the Listening Brain". In: *Trends in Cognitive Sciences* 23.11 (2019), pp. 913–926.

[145] N. Ding, M. Chatterjee, and J. Z. Simon. "Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure". In: *NeuroImage* 88 (2014), pp. 41–46.

[146] N. Ding and J. Z. Simon. "Cortical entrainment to continuous speech: functional roles and interpretations". In: *Frontiers in Human Neuroscience* 8 (2014), pp. 311–.

[147] K. Anne et al. “Neural Entrainment Determines the Words We Hear”. In: *Current Biology* 28 (2018), S0960982218309229–.

[148] P. Lakatos et al. “Entrainment of Neuronal Oscillations as a Mechanism of Attentional Selection”. In: *Science* 320.5872 (2008), pp. 110–113.

[149] K. B. Doelling et al. “Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing”. In: *NeuroImage* 85 (2014), pp. 761–768.

[150] J. E. Peelle, G. Joachim, and M. H. Davis. “Phase-Locked Responses to Speech in Human Auditory Cortex are Enhanced During Comprehension”. In: *Cerebral Cortex* 6 (2013), pp. 1378–1387.

[151] L Meyer. “The Neural Oscillations of Speech Processing and Language Comprehension: State of the Art and Emerging Mechanisms”. In: *European Journal of Neuroscience* 48.7 (2017).

[152] N. Ding et al. “Cortical tracking of hierarchical linguistic structures in connected speech”. In: *Nature Neuroence* (2016).

[153] Thilo Womelsdorf et al. “Modulation of Neuronal Interactions Through Neuronal Synchronization”. In: 316.5831 (2007), pp. 1609–1612. DOI: `10.1126/science.1139597JScience`. URL: `https://science.sciencemag.org/content/sci/316/5831/1609.full.pdf`.

[154] Paul Sauseng et al. “Cross-frequency phase synchronization: A brain mechanism of memory matching and attention”. In: *NeuroImage* 40.1 (2008), pp. 308–317. ISSN: 1053-8119. DOI: `https://doi.org/10.1016/j.neuroimage.2007.11.032`. URL: `http://www.sciencedirect.com/science/article/pii/S1053811907010592`.

[155] M Hassan et al. “Dynamic reorganization of functional brain networks during picture naming”. In: *Cortex; a journal devoted to the study of the nervous system and behavior* 73 (2015), p. 276.

[156] B Morillon et al. “Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study”. In: *Frontiers in Psychology* 3.248 (2012), p. 248.

[157] Corinna E. Bonhage et al. “Oscillatory EEG dynamics underlying automatic chunking during sentence processing”. In: *Neuroimage* 152 (2017), pp. 647–657.

[158] Satu Palva and J. Matias Palva. “New vistas for -frequency band oscillations”. In: *Trends in Neurosciences* 30.4 (2007), pp. 150–158. ISSN: 0166-2236. DOI: `https://doi.org/10.1016/j.tins.2007.02.001`. URL: `http://www.sciencedirect.com/science/article/pii/S0166223607000264`.

[159] Megan Cuellar et al. “Mu suppression as an index of sensorimotor contributions to speech processing: Evidence from continuous EEG signals”. In: *International Journal of Psychophysiology* 85.2 (2012), pp. 242–248.

[160] Han Yuan et al. “Negative covariation between task-related responses in alpha/beta-band activity and BOLD in human sensorimotor cortex: An EEG and fMRI study of motor imagery and movements”. In: *NeuroImage* 49.3 (2010), pp. 2596–2606. ISSN: 1053-8119. DOI: `https://doi.org/10.1016/j.neuroimage.2009.10.028`. URL: `http://www.sciencedirect.com/science/article/pii/S1053811909010994`.

[161] A. Keitel, J. Gross, and C. Kayser. “Perceptually relevant speech tracking in auditory and motor cortex reflects distinct linguistic features”. In: *PLoS Biology* 16.3 (2018), e2004473.

[162] Meyer and Lars. “The neural oscillations of speech processing and language comprehension: state of the art and emerging mechanisms”. In: *European Journal of Neuroence* (2018).

[163] B. Zoefel, A. Archer-Boyd, and M. H. Davis. “Phase Entrainment of Brain Oscillations Causally Modulates Neural Responses to Intelligible Speech”. In: *Current Biology* (2018), 401–408.e5.

[164] Lars Meyer, Yue Sun, and Andrea E. Martin. “Synchronous, but not entrained: exogenous and endogenous cortical rhythms of speech and language processing”. In: *Language, Cognition and Neuroscience* 35.9 (2020), pp. 1089–1099. ISSN: 2327-3798. DOI: `10.1080/23273798.2019.1693050`. URL: `https://doi.org/10.1080/23273798.2019.1693050`.

[165] Axmacher Nikolai et al. “Cross-frequency coupling supports multi-item working memory in the human hippocampus”. In: *Proceedings of the National Academy of Sciences of the United States of America* 107.7 (2010), pp. 3228–3233.

[166] Johne. Lisman and Ole Jensen. “The Theta-Gamma Neural Code”. In: *Neuron* 77.6 (2013), pp. 1002–1016.

[167] Sevada Hovsepyan, Itsaso Olasagasti, and Anne-Lise Giraud. “Combining predictive coding and neural oscillations enables online syllable recognition in natural speech”. In: *Nature Communications* 11.1 (2020), p. 3117. ISSN: 2041-1723. DOI: `10.1038/s41467-020-16956-5`. URL: `https://doi.org/10.1038/s41467-020-16956-5`.

[168] Ole Jensen and Ali Mazaheri. “Shaping Functional Architecture by Oscillatory Alpha Activity: Gating by Inhibition”. In: *Frontiers in Human Neuroscience* 4.186 (2010). ISSN: 1662-5161. DOI: `10.3389/fnhum.2010.00186`. URL: `https://www.frontiersin.org/article/10.3389/fnhum.2010.00186`.

[169] Antje Strauß, Malte Wöstmann, and Jonas Obleser. "Cortical alpha oscillations as a tool for auditory selective inhibition". In: *Frontiers in Human Neuroscience* 8.350 (2014). ISSN: 1662-5161. DOI: `10.3389/fnhum.2014.00350`. URL: `https://www.frontiersin.org/article/10.3389/fnhum.2014.00350`.

[170] R. C. Oldfield. "The assessment and analysis of handedness: the Edinburgh inventory". In: *Neuropsychologia* 9.1 (1971), pp. 97–113.

[171] Roger Levy and Christopher D. Manning. "Is it Harder to Parse Chinese, or the Chinese Treebank?" In: *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics* (2003), pp. 439–446.

[172] S. M. Doesburg et al. "Theta-Modulated Gamma-Band Synchronization Among Activated Regions During a Verb Generation Task". In: *Frontiers in Psychology* 3 (2012).

[173] N. E. Crone et al. "Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. I. Alpha and beta event-related desynchronization". In: *Brain* 121.12 (1998), pp. 2271–2299. ISSN: 0006-8950. DOI: `10.1093/brain/121.12.2271`. URL: `https://doi.org/10.1093/brain/121.12.2271`.

[174] Matthew J. Boring et al. "Left mid-ventral temporal cortex interacts with early visual cortex and the anterior temporal lobe to support word individuation". In: *bioRxiv* (2018), p. 411579. DOI: `10.1101/411579`. URL: `https://www.biorxiv.org/content/biorxiv/early/2018/09/25/411579.full.pdf`.

[175] P Hoffman, G. A. Evans, and M. A. Lambon Ralph. "The anterior temporal lobes are critically involved in acquiring new conceptual knowledge: evidence for impaired feature integration in semantic dementia". In: *Cortex; a journal devoted to the study of the nervous system and behavior* 50.100 (2014), p. 19.

[176] Lanfang Liu et al. "Auditory–Articulatory Neural Alignment between Listener and Speaker during Verbal Communication". In: *Cerebral Cortex* 30.3 (2019), pp. 942–951. ISSN: 1047-3211. DOI: `10.1093/cercor/bhz138JCerebralCortex`. URL: `https://doi.org/10.1093/cercor/bhz138`.

[177] Mortimer Mishkin, Leslie G. Ungerleider, and Kathleen A. Macko. "Object vision and spatial vision: two cortical pathways". In: *Trends in Neurosciences* 6 (1983), pp. 414–417. ISSN: 0166-2236. DOI: `https://doi.org/10.1016/0166-2236(83)90190-X`. URL: `https://www.sciencedirect.com/science/article/pii/016622368390190X`.

[178] J. F. Houde and S. S. Nagarajan. "Speech production as state feedback control". In: *Frontiers in Human Neuroscience* 5:82.doi: 10.3389/fnhum.2011.00082 (2011).

[179] M. Petrides. "Neuroanatomy of Language Regions of the Human Brain". In: *New York, NY: Academic Press.* (2014), 89–138.

[180] Rex E. Jung and Richard J. Haier. "The Parieto-Frontal Integration Theory (P-FIT) of intelligence: Converging neuroimaging evidence". In: *Behavioral and Brain Sciences* 30.2 (2007), pp. 135–154. ISSN: 0140-525X. DOI: `10.1017/S0140525X07001185`.

[181] Fedorenko Evelina, Duncan John, and Kanwisher Nancy. "Broad domain generality in focal regions of frontal and parietal cortex". In: *Proceedings of the National Academy of Sciences of the United States of America* 110.41 (2013), pp. 16616–16621.

[182] Jonathan Smallwood et al. "Cooperation between the default mode network and the frontal–parietal network in the production of an internal train of thought". In: *Brain Research* 1428 (2012), pp. 60–70. ISSN: 0006-8993. DOI: `https://doi.org/10.1016/j.brainres.2011.03.072`. URL: `http://www.sciencedirect.com/science/article/pii/S0006899311006676`.

[183] A. S. Shah et al. "Neural dynamics and the fundamental mechanisms of event-related brain potentials". In: *Cereb Cortex* 14.5 (2004), pp. 476–83. ISSN: 1047-3211 (Print) 1047-3211. DOI: `10.1093/cercor/bhh009`.

[184] L. Bressler Steven and K. Seth Anil. "Wiener–Granger Causality: A well established methodology". In: *NeuroImage* 58.2 (2011), pp. 323–329. ISSN: 1053-8119. DOI: `https://doi.org/10.1016/j.neuroimage.2010.02.059`. URL: `http://www.sciencedirect.com/science/article/pii/S1053811910002272`.

[185] Daniel M. Wolpert. "Computational approaches to motor control". In: *Trends in Cognitive Sciences* 1.6 (1997), pp. 209–216. ISSN: 1364-6613. DOI: `https://doi.org/10.1016/S1364-6613(97)01070-X`. URL: `https://www.sciencedirect.com/science/article/pii/S136466139701070X`.

[186] K Johari and R. Behroozmand. "Event-related Desynchronization of Alpha and Beta Band Neural Oscillations Predict Speech and Limb Motor Timing Deficits in Normal Aging". In: *Behavioural Brain Research* (2020).

[187] K. Bock and H. Loebell. "Framing sentences". In: *Cognition* 35.1 (1990), pp. 1–39. ISSN: 0010-0277 (Print) 0010-0277. DOI: `10.1016/0010-0277(90)90035-i`.

[188] Martin J. Pickering and Holly P. Branigan. "Syntactic priming in language production". In: *Trends in Cognitive Sciences* 3.4 (1999), pp. 136–141. ISSN: 1364-6613. DOI: `https://doi.org/10.1016/S1364-6613(99)01293-0`. URL: `https://www.sciencedirect.com/science/article/pii/S1364661399012930`.

[189] Martin J. Pickering and Victor S. Ferreira. "Structural priming: a critical review". In: *Psychological bulletin* 134.3 (2008), pp. 427–459. ISSN: 0033-2909. DOI: `10.1037/`

0033-2909.134.3.427. URL: `https://pubmed.ncbi.nlm.nih.gov/18444704https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2657366/`.

[190] Ricarda I. Schubotz. “Prediction of external events with our motor system: towards a new framework”. In: *Trends in Cognitive Sciences* 11.5 (2007), pp. 211–218. ISSN: 1364-6613. DOI: `https://doi.org/10.1016/j.tics.2007.02.006`. URL: `https://www.sciencedirect.com/science/article/pii/S1364661307000800`.

[191] Timothyj. Buschman et al. “Synchronous Oscillatory Neural Ensembles for Rules in the Prefrontal Cortex”. In: *Neuron* 76.4 (2012).

[192] Ryan T. Canolty, Karunesh Ganguly, and Jose M. Carmena. “Task-dependent changes in cross-level coupling between single neurons and oscillatory activity in multiscale networks”. In: *PLoS computational biology* 8.12 (2012), e1002809–e1002809. ISSN: 1553-7358 1553-734X. DOI: `10.1371/journal.pcbi.1002809`. URL: `https://pubmed.ncbi.nlm.nih.gov/23284276https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3527280/`.

[193] Mark O. Cunningham et al. “A role for fast rhythmic bursting neurons in cortical gamma oscillations ¡em¿in vitro¡/em¿”. In: *Proceedings of the National Academy of Sciences of the United States of America* 101.18 (2004), pp. 7152–7157. DOI: `10.1073/pnas.0402060101`. URL: `https://www.pnas.org/content/pnas/101/18/7152.full.pdf`.

[194] Peter J Uhlhaas et al. “Neural synchrony in cortical networks: history, concept and current status”. In: *Frontiers in Integrative Neuroscience* 3.17 (2009), p. 17.

[195] Pascal Fries. “Rhythms for Cognition: Communication through Coherence”. In: *Neuron* 88.1 (2015), pp. 220–235. ISSN: 0896-6273. DOI: `https://doi.org/10.1016/j.neuron.2015.09.034`. URL: `https://www.sciencedirect.com/science/article/pii/S0896627315008235`.

[196] G Pfurtscheller, Stancã K A Jr, and C Neuper. “Event-related synchronization (ERS) in the alpha band–an electrophysiological correlate of cortical idling: a review”. In: *International Journal of Psychophysiology* 24.1–2 (1996), pp. 39–46.

[197] Anna Mayer et al. “Expecting to See a Letter: Alpha Oscillations as Carriers of Top-Down Sensory Predictions”. In: *Cerebral Cortex* 26.7 (2015), pp. 3146–3160. ISSN: 1047-3211. DOI: `10.1093/cercor/bhv146`. URL: `https://doi.org/10.1093/cercor/bhv146`.

[198] J. M. Zumer et al. “Occipital Alpha Activity during Stimulus Processing Gates the Information Flow to Object-Selective Cortex”. In: *PLoS Biology* 12.10 (2014), e1001965–.

[199] W. Klimesch. “EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis”. In: *Brain Res Brain Res Rev* 29.2-3 (1999), pp. 169–195.

[200] Van Ede and Freek. “Mnemonic and attentional roles for states of attenuated alpha oscillations in perceptual working memory: a review”. In: *European Journal of Neuroscience* (2017).

[201] M. J. Frank, B. Loughry, and R. C. O’Reilly. “Interactions between frontal cortex and basal ganglia in working memory: A computational model”. In: *Cognitive Affective Behavioral Neuroscience* 1.2 (2001), pp. 137–160.

[202] Liina Pylkkänen. “The neural basis of combinatory syntax and semantics”. In: *Science* 366.6461 (2019), pp. 62–66. DOI: `10.1126/science.aax0050`. URL: `https://science.sciencemag.org/content/sci/366/6461/62.full.pdf`.

[203] A. Keitel et al. “Auditory cortical delta-entrainment interacts with oscillatory power in multiple fronto-parietal networks”. In: *Neuroimage* (2016), p. 32.

[204] Sjp Donoghue. “Oscillations in local field potentials of the primate motor cortex during voluntary movement”. In: *Proceedings of the National Academy of Sciences* 90.10 (1993), pp. 4470–4474.

[205] Fabian Klostermann et al. “Task-related differential dynamics of EEG alpha- and beta-band synchronization in cortico-basal motor structures”. In: *European Journal of Neuroence* 25.5 (2010), pp. 1604–1615.

[206] V. Chakarov et al. “Beta-range EEG-EMG coherence with isometric compensation for increasing modulated low-level forces”. In: *Journal of Neurophysiology* 102.2 (2009), pp. 1115–1120.

[207] M. Rosanova et al. “Natural frequencies of human corticothalamic circuits”. In: *Journal of Neuroscience the Official Journal of the Society for Neuroscience* 29.24 (2009), p. 7679.

[208] P. R. Roelfsema et al. “Visuomotor integration is associated with zero time-lag synchronization among cortical areas”. In: *Nature* 385.6612 (1997), pp. 157–161.

[209] T. H. Donner et al. “Buildup of Choice-Predictive Activity in Human Motor Cortex during Perceptual Decision Making”. In: *Current Biology Cb* 19.18 (2009), pp. 1581–1585.

[210] L. Meyer et al. “Linguistic Bias Modulates Interpretation of Speech via Neural Delta-Band Oscillations”. In: *Cerebral Cortex* 9 (2016), p. 9.

[211] S. Picazio et al. "Prefrontal Control over Motor Cortex Cycles at Beta Frequency during Movement Inhibition". In: *Current Biology* 24.24 (2014), pp. 2940–2945.

[212] G. Hickok, K. Okada, and J.T. Serences. "Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing". In: *Journal of Neurophysiology* 101.5 (2009), pp. 2725–2732.

[213] Thomas E Hazy, Michael J Frank, and Randall C O'Reilly. "Towards an executive without a homunculus: computational models of the prefrontal cortex/basal ganglia system". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 362.1485 (2007), pp. 1601–1613. DOI: `doi:10.1098/rstb.2007.2055`.

[214] Viej De et al. "Priority Switches in Visual Working Memory are Supported by Frontal Delta and Posterior Alpha Interactions". In: *Cerebral Cortex* (2018).

[215] P. Sauseng et al. "Fronto-parietal EEG coherence in theta and upper alpha reflect central executive functions of working memory". In: *International Journal of Psychophysiology* 57.2 (2005), pp. 97–103.

[216] T. V. Kerkoerle et al. "Alpha and gamma oscillations characterize feedback and feedforward processing in monkey visual cortex". In: *Proc Natl Acad Sci U S A* 111.40 (2014), pp. 14332–14341.

[217] A. Gazzaley and A. C. Nobre. "Top-down modulation: Bridging selective attention and working memory". In: *Trends in Cognitive Sciences* 16.2 (2011), pp. 129–135.

[218] Bradley Voytek et al. "Oscillatory dynamics coordinating human frontal networks in support of goal maintenance". In: *Nature Neuroscience* (2015).

[219] Earl K. Miller and Jonathan D. Cohen. "An Integrative Theory of Prefrontal Cortex Function". In: *Annual Review of Neuroscience* 24.1 (2001), pp. 167–202. ISSN: 0147-006X. DOI: `10.1146/annurev.neuro.24.1.167`. URL: `https://doi.org/10.1146/annurev.neuro.24.1.167`.

[220] B. Berger et al. "Dynamic regulation of interregional cortical communication by slow brain oscillations during working memory". In: *Nature Communications* (2019).

[221] Masha Westerlund and Liina Pylkkänen. "The role of the left anterior temporal lobe in semantic composition vs. semantic memory". In: *Neuropsychologia* 57 (2014), pp. 59–70. ISSN: 0028-3932. DOI: `https://doi.org/10.1016/j.neuropsychologia.2014.03.001`. URL: `http://www.sciencedirect.com/science/article/pii/S0028393214000761`.

[222] F Van Ede, M. Niklaus, and A. C. Nobre. “Temporal Expectations Guide Dynamic Prioritization in Visual Working Memory through Attenuated Oscillations”. In: *Journal of Neuroscience* 37.2 (2017), pp. 437–445.

[223] Heather L. et al. “Only Coherent Spiking in Posterior Parietal Cortex Coordinates Looking and Reaching”. In: *Neuron* (2012).

[224] Middleton et al. “Anatomical evidence for cerebellar and basal ganglia involvement in higher cognitive function”. In: *Science* (1994).

[225] Mpl Strick. “Basal ganglia and cerebellar loops: motor and cognitive circuits”. In: *Brain Research Reviews* (2000).

[226] Robert Schmidt et al. “Beta Oscillations in Working Memory, Executive Control of Movement and Thought, and Sensorimotor Function”. In: *Journal of Neuroscience* 39.42 (2019), pp. 8231–8238.

[227] Kenji Doya. “Complementary roles of basal ganglia and cerebellum in learning and motor control”. In: *Current Opinion in Neurobiology* 10.6 (2000), pp. 732–739. ISSN: 0959-4388. DOI: `https://doi.org/10.1016/S0959-4388(00)00153-7`. URL: `https://www.sciencedirect.com/science/article/pii/S0959438800001537`.

[228] R. C. O’Reilly and M. J. Frank. “Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia”. In: *Neural Computation* 18.2 (2006), pp. 283–328.

[229] M. Kawato and H. Gomi. “A computational model of four regions of the cerebellum based on feedback-error learning”. In: *Biological Cybernetics* 68.2 (1992), pp. 95–103.

[230] J. C. Houk, J. L. Adams, and A. G. Barto. “A Model of How the Basal Ganglia Generate and Use Neural Signals that Predict Reinforcement”. In: *Models of Information Processing in the Basal Ganglia* (1995).

[231] D. Bouneffouf. “Online learning with Corrupted context: Corrupted Contextual Bandits”. In: (2020).