THE NEURAL BASES OF AUDITORY LANGUAGE PROCESSING FOR CHINESE EFL STUDENTS: A COMBINED EVENT-RELATED POTENTIALS AND FUNCTIONAL MAGNETIC RESONANCE IMAGING STUDY BASED ON THE VERBOTONAL APPROACH

CAI XIRUI

A Thesis Submitted in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in English Language Studies
Suranaree University of Technology
Academic Year 2021

พื้นฐานของวงจรประสาทในการประมวลผลทางภาษาที่เกี่ยวกับการได้ยินของนักศึกษาจีนที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศ: ศึกษาโดยใช้ศักย์ไฟฟ้าสมองสัมพันธ์กับเหตุการณ์ร่วมกับการสร้างภาพโดยกิจด้วยเรโซแนนท์แม่เหล็กด้วยวิธีเวอร์เบอร์โทนอล

นายไช่ ซีรุย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปรัชญาดุษฎีบัณฑิต
สาขาวิชาภาษาอังกฤษศึกษา
มหาวิทยาลัยเทคโนโลยีสุรนารี
ปีการศึกษา 2564

THE NEURAL BASES OF AUDITORY LANGUAGE PROCESSING FOR CHINESE EFL STUDENTS: A COMBINED EVENT-RELATED POTENTIALS AND FUNCTIONAL MAGNETIC RESONANCE IMAGING STUDY BASED ON THE VERBOTONAL APPROACH

Suranaree University of Technology has approved this thesis submitted in partial fulfillment of the requirements for the Degree of Doctor of Philosophy.

Thesis Examining Committee

____

(Assoc. Prof. Dr. Pham Vu Phi Ho)

Chairperson

____

(Prof. Dr. Andrew Lian)

Member (Thesis Advisor)

____

(Dr. Nattaya Puakpong)

Member (Thesis Co-Advisor)

____

(Asst. Prof. Dr. Jeffrey Dawala Wilang)

Member

____

(Dr. Zhenhui Li)

Member

____

(Dr. Suksan Suppasetseree)

Member

____

(Assoc. Prof. Dr. Chatchai Jothityangkoon)

Vice Rector for Academic Affairs and Quality Assurance

____

(Assoc. Prof. Dr. Thara Angskun)

Dean of Institute of Social Technology

ไช่ ซีรุย : พื้นฐานของวงจรประสาทในการประมวลผลทางภาษาที่เกี่ยวกับการได้ยินของนักศึกษาจีนที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศ: ศึกษาโดยใช้ศักย์ไฟฟ้าสมองสัมพันธ์กับเหตุการณ์ร่วมกับการสร้างภาพโดยกิจด้วยเรโซแนนท์แม่เหล็กด้วยวิธีเวอร์เบอร์โทนอล (THE NEURAL BASES OF AUDITORY LANGUAGE PROCESSING FOR CHINESE EFL STUDENTS: A COMBINED EVENT-RELATED POTENTIALS AND FUNCTIONAL MAGNETIC RESONANCE IMAGING STUDY BASED ON THE VERBOTONAL APPROACH) อาจารย์ที่ปรึกษา : Professor Dr. Andrew Lian และอาจารย์ที่ปรึกษาร่วม : อาจารย์ ดร.ณัฏฐญา เผือกผ่อง, 238 หน้า.

คำสำคัญ: การประมวลผลทางภาษา/ เวอร์โบโทโนลิซึ่ม/ การฟังในรูปแบบไดโคติกและไดโอติก/ ศักย์ไฟฟ้าสมองสัมพันธ์กับเหตุการณ์/ การสร้างภาพโดยกิจด้วยเรโซแนนท์แม่เหล็ก/ ความถนัดของซีกสมอง/ ข้อมูลทางการได้ยินที่เหมาะสมที่สุด

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาว่าคุณสมบัติทางกายภาพของสัญญาณทางภาษาที่ถูกส่งไปยังผู้เรียนในรูปแบบไดโคติกหรือไดโอติกมีอิทธิพลอย่างไรต่อกิจกรรมของสมองที่นำไปสู่การรับรู้ภาษา ลักษณะของสัญญาณข้อมูลรับเข้าของภาษามาจากหลักการของทฤษฎีเวอร์โบโทโนลิซึ่มและบางส่วนจากการค้นพบอื่น ๆ ทางประสาทวิทยา ด้วยการทดลองที่ผนวกระหว่างศักย์ไฟฟ้าสมองสัมพันธ์กับเหตุการณ์และการสร้างภาพโดยกิจด้วยเรโซแนนท์แม่เหล็กช่วยทำให้รูปแบบเชิงเวลาและเชิงพื้นที่ของกิจกรรมของระบบประสาทของนักศึกษาจีนที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศในขณะที่ฟังภาษาจีน (ภาษาที่หนึ่ง) และภาษาอังกฤษ (ภาษาที่สอง) ชัดเจนขึ้น นอกจากนี้ยังได้มีการศึกษาความคิดเห็นของนักศึกษาเกี่ยวกับสัญญาณเสียงโดยใช้การสัมภาษณ์แบบกึ่งโครงสร้าง การวิจัยแบบผสมผสานที่รวมเอาการวิจัยเชิงคุณภาพและเชิงปริมาณมาใช้ร่วมกันได้ถูกนำมาใช้ในการระบุชี้สัญญาณข้อมูลทางการได้ยินที่เหมาะสมที่สุดสำหรับผู้เรียนภาษาภายใต้หลักการเวอร์โบโทโนลิซึ่ม ความถนัดของซีกสมอง และทฤษฎีภาระการทำงานทางปัญญา

นักศึกษาที่ถนัดมือขวาที่มีระดับภาษาอังกฤษอยู่ในระดับปานกลางจำนวน 30 คนในมหาวิทยาลัยแพทย์ที่ตั้งอยู่ในภาคตะวันตกเฉียงใต้ของประเทศจีนเข้าร่วมในงานวิจัยนี้ จากหลักการเวอร์โบโทโนลิซึ่ม ตัวกระตุ้นทางเสียงที่อยู่ในรูปแบบประโยคภาษาจีนและภาษาอังกฤษที่ผ่านการกรองให้ความถี่ย่านเสียงต่ำ 320 เฮิร์ตสผ่านได้โดยที่ข้อมูลทางฉันทลักษณ์ถูกเก็บไว้เพื่อทดสอบสมมติฐานว่าสัญญาณข้อมูลทางการได้ยินที่เหมาะสมที่สุดที่เป็นไปได้นั้นเป็นไปตามความถนัดของซีกสมองในกระบวนการทางฉันทลักษณ์และทางภาษาศาสตร์ ตัวกระตุ้นที่ได้ผ่านการกรองและไม่ได้ผ่าน

การกรองได้ถูกนำมาจัดการภายใต้เงื่อนไขการฟังแบบไดโคติกและไดโอติก ทั้งนี้ได้มีการกำหนดค่าตัวกระตุ้นเป็นสี่แบบในแต่ละภาษา: ตัวกระตุ้นที่ผ่านการกรองสำหรับหูทั้งสองข้าง (FL-FR); ตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างซ้ายและไม่ผ่านการกรองสำหรับหูข้างขวา (FL-R); ตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างขวาและไม่ผ่านการกรองสำหรับหูข้างซ้าย (L-FR); ตัวกระตุ้นที่ไม่ผ่านการกรองสำหรับหูทั้งสองข้าง (NL-NR) การบันทึกศักย์ไฟฟ้าสมองสัมพันธ์กับเหตุการณ์และการสแกนการสร้างภาพโดยกิจด้วยเรโซแนนท์แม่เหล็กได้ถูกดำเนินการแยกกันในขณะที่ผู้เข้าร่วมวิจัยฟังสัญญาณเสียง หลังการทดลองเสร็จ ได้มีการสัมภาษณ์กึ่งโครงสร้างกับผู้เข้าร่วมวิจัย

จากศักย์ไฟฟ้าสมองสัมพันธ์กับเหตุการณ์ที่เกี่ยวข้องกับภาษาและบริเวณสมองที่ถูกกระตุ้นสัญญาณภาษาที่หนึ่งที่เป็นตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างซ้ายและไม่ผ่านการกรองสำหรับหูข้างขวาลดภาระงานทางจิตใจสำหรับการประมวลผลด้านอรรถศาสตร์และการประมวลผลเฉพาะด้านโครงสร้าง และไม่เกี่ยวข้องกับบริเวณสมองที่เพิ่มเติมสำหรับการประมวลผลด้านอรรถศาสตร์และวากยสัมพันธ์ ในแง่ของภาษาที่สอง ตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างซ้ายและไม่ผ่านการกรองสำหรับหูข้างขวาลดภาระงานด้านจิตใจสำหรับการจัดการด้านอรรถศาสตร์และวากยสัมพันธ์โดยไม่ต้องใช้พื้นที่สมองเพิ่มเติมสำหรับการประมวลผล นอกจากนี้ผู้เข้าร่วมวิจัยส่วนใหญ่แสดงความชื่นชอบต่อสัญญาณภาษาที่หนึ่งและภาษาที่สองที่เป็นตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างซ้ายและไม่ผ่านการกรองสำหรับหูข้างขวาเนื่องจากมีความชัดเจนมากกว่าและไม่ก่อให้เกิดความรู้สึกไม่สบาย อีกทั้งอาจช่วยให้เข้าใจประโยคได้ดีขึ้น ดังนั้นสัญญาณที่เป็นตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างซ้ายและไม่ผ่านการกรองสำหรับหูข้างขวาจึงนับได้ว่าเป็นสัญญาณข้อมูลทางการได้ยินที่เหมาะสมที่สุดสำหรับนักศึกษาชาวจีนที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศ ในขณะเดียวกัน สัญญาณภาษาที่สองที่เป็นตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างขวาและไม่ผ่านการกรองสำหรับหูข้างซ้ายที่ทำให้ภาระการประมวลผลของภาระงานด้านจิตใจมีมากขึ้นสำหรับการจัดการด้านอรรถศาสตร์และวากยสัมพันธ์จะเป็นสัญญาณการได้ยินที่ไม่เหมาะสมสำหรับนักศึกษาชาวจีน แต่ก็ยังไม่ชัดเจนว่าภาษาที่หนึ่งที่เป็นตัวกระตุ้นที่ผ่านการกรองสำหรับหูข้างขวาและไม่ผ่านการกรองสำหรับหูข้างซ้ายนั้นเป็นสิ่งที่ไม่เหมาะสม เมื่อเปรียบเทียบกับการประมวลผลต่อภาษาที่หนึ่งที่เป็นตัวกระตุ้นที่ไม่ผ่านการกรองสำหรับหูทั้งสองข้าง ผลกระทบของภาษาที่สองพบว่าการประมวลผลของภาษาที่สองที่เป็นตัวกระตุ้นที่ไม่ผ่านการกรองสำหรับหูทั้งสองข้างต้องใช้พื้นที่ของสมองมากขึ้นในการควบคุมหรือสลับภาษา อย่างไรก็ตามสัญญาณภาษาที่สองที่มีเสียงกรองในสภาพการฟังแบบไดโคติกหรือไดโอติดสามารถลดผลกระทบของภาษาที่สองได้ สำหรับความคิดเห็นต่อวิธีการฟังแบบนี้นักศึกษาแสดงความสนใจและความเต็มใจที่จะใช้สัญญาณเหล่านี้ในการเรียนรู้ภาษาอังกฤษ

โดยรวมแล้วงานวิจัยนี้ศึกษาถึงคุณค่าของการรับรู้และใช้ประโยชน์จากแนวคิดทางประสาทวิทยาศาสตร์สำหรับการประมวลผลภาษาและคุณสมบัติทางกายภาพของตัวป้อนทางภาษาเพื่อวัตถุประสงค์ในการปรับปรุงการรับรู้ภาษาและการเรียนรู้ภาษาซึ่งผลการวิจัยมีผลกระทบอย่างยิ่งต่อการศึกษาภาษาทั้งทางทฤษฎีและการปฏิบัติ

สาขาวิชาภาษาต่างประเทศ ลายมือชื่อนักศึกษา Xiaoqi Cai

ปีการศึกษา 2564 ลายมือชื่ออาจารย์ที่ปรึกษา ________

ลายมือชื่ออาจารย์ที่ปรึกษาร่วม ________

CAI XIRUI: THE NEURAL BASES OF AUDITORY LANGUAGE PROCESSING FOR CHINESE EFL STUDENTS: A COMBINED EVENT-RELATED POTENTIALS AND FUNCTIONAL MAGNETIC RESONANCE IMAGING STUDY BASED ON THE VERBOTONAL APPROACH. THESIS ADVISOR: PROF. ANDREW LIAN, Ph.D. AND THESIS CO-ADVISOR: NATTAYA PUAKPONG, Ph.D., 238 PP.

Keyword: Language Processing/ Verbotonalism/ Dichotic and Diotic Listening/ ERP/ FMRI/ Hemispheric Specialization/ Optimal Auditory Input

The current study proposed to explore how the physical quality of the language signals sent to learners either dichotically or diotically influenced brain activity leading to language perception. The actual nature of the language input signals was derived in part from principles of verbotonal theory and partly from other findings from neuroscience. By implementing a combined Event-Related Potential (ERP) and functional Magnetic Resonance Imaging (fMRI) experiment, this study unraveled Chinese university EFL students' temporal and spatial patterns of neural activity while listening to Chinese (L1) and English (L2) signals. Further, students' opinions about the signals were investigated through semi-structured interviews. A mixed-method design integrating quantitative and qualitative methods was employed in order to identify an optimal auditory input signal for language learners based on the principles of verbotonalism, hemispheric specialization, and cognitive load theory.

Thirty right-handed students in a medical university in southwestern China with an intermediate level of English proficiency took part in the current study. According to the verbotonal principles, auditory stimuli (Chinese and English sentences) were 320 Hz low-pass filtered, and prosodic information was retained. To test the hypothesis that the possible optimal auditory input signal was in line with hemispheric specialization for linguistic and prosodic processing, filtered and unfiltered stimuli were organized in dichotic and diotic listening conditions. Four configurations of stimuli were therefore obtained in each language: both-ear-filtered stimuli (FL-FR); filtered stimuli in the left ear and unfiltered in the right ear (FL-R); filtered stimuli in the right ear and

unfiltered in the left ear (L-FR); both-ear-unfiltered stimuli (NL-NR). ERP recording and fMRI scanning were performed separately while the participants were listening to the signals. After the experiments, semi-structured interviews were conducted.

From the language-related ERPs and the activated brain regions, the L1 FL-R signal lowered the mental workload for semantic processing and the later structure-specific processing and did not involve additional brain regions for semantic and syntactic processing. In terms of L2, FL-R reduced the mental load for semantic and syntactic manipulations without recruiting additional brain areas for processing. In addition, most respondents expressed preferences for the L1 and L2 FL-R signals as they were clearer, did not evoke the feeling of discomfort, and might help understand the sentences. Thus, the FL-R signals could be identified as optimal auditory input signals for Chinese EFL students/listeners. Meanwhile, L2 L-FR, imposing more mental processing load for semantic and syntactic manipulations, appears to be the non-optimal auditory signal for Chinese students. But it was unclear whether L1 L-FR was non-optimal. Compared to L1 NL-NR processing, the L2 effect was found that the processing of L2 NL-NR involved more brain regions for language control/switching. However, the L2 signals with low-pass filtered sounds in either dichotic or diotic listening condition could reduce the L2 effect. For the opinions of the signals, students expressed an interest and willingness to use the signals for learning English.

Overall, the present study addresses the value of recognizing and exploiting the neurobiological bases for language processing and the physical features of language input for the purposes of improving language perception and language learning. The findings have far-reaching theoretical and practical implications for language education.

School of Foreign Languages　　　　Student's Signature ____________

Academic Year 2021　　　　Advisor's Signature ____________

Co-advisor's Signature ____________

ACKNOWLEDGEMENTS

I am extremely lucky to have many individuals who have been continuously encouraging me along the way. First and foremost, it is a genuine pleasure to express my deepest and sincerest gratitude to my thesis advisors Professor Dr. Andrew Lian and Dr. Nattaya Puakpong. Professor Andrew Lian's scholarly advice, illuminating insights, and meticulous scrutiny have helped me to a great extent to accomplish this study. He opened the door of the Verbotonal world for me four years ago, and it was at that time that he inspired me to begin this exciting and adventurous Ph.D. journey. He made me believe that I could do better and that life could be so much better as we kept innovating and changing our minds. In addition, I truly appreciate having Dr. Nattaya Puakpong as my co-advisor. Her patience, enthusiasm, and professional expertise helped me in all the time of research and writing this thesis. The completion of this work could not be possible without her guidance and support.

It is my privilege to thank my committee members: Assoc. Prof. Dr. Pham Vu Phi Ho (who served as chair), Dr. Suksan Suppasetseree, Asst. Prof. Dr. Jeffrey Dawala Wilang, and Dr. Zhenhui Li. Their constructive feedback and suggestions helped me strengthen my thesis.

I would like to extend my thanks to all teachers, staff, and friends at the School of Foreign Languages, Suranaree University of Technology for all the invaluable help and inspiration. I am highly indebted to the teachers and students, Prof. Dr. Yunfa Fu and Mr. Yuyang Dong in particular, at the Faculty of Information Engineering and Automation, Kunming University of Science and Technology for providing me with the EEG machine and helping me immensely in the ERP experiment. Also, I owe a deep sense of gratitude to the teachers and students, especially Assoc. Prof. Dr. Yin Mo and Mr. Yaoping Shi, at the Department of Medical Imaging, The First Affiliated Hospital of Kunming Medical University for technical support and assistance during the fMRI experiment. My special thanks go to all the participants for their active participation.

Finally, to my caring, loving, and supportive parents, Hong Cai and Changyun Yao: my heartfelt thanks. Thank you for everything!

Cai Xirui

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure **Page**

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
AUC	Area under the curve
BA	Brodmann area
BOLD	Blood oxygenation level dependent
dB	Decibel
ECFS	Extreme capsule fiber system
EEG	Electroencephalography
EFL	English as a foreign language
ERP	Event-related potential
ESL	English as a second language
F_0	Fundamental frequency
FFR	Frequency-following response
FL-FR	Filtered in left and right channels
FL-R	Filtered in the left channel and unfiltered in the right channel
fMRI	Functional magnetic resonance imaging
FOP	Frontal operculum
FWHM	Full width at half maximum
Hz	Hertz
IFC	Inferior frontal cortex
IFG	Inferior frontal gyrus
IFOF	Inferior-fronto-occipital fascile
IPC	Inferior parietal cortex
IPL	Inferior parietal lobule
ITS	Inferior temporal sulcus
k	Cluster size (number of voxels)
L1	The first language (Mandarin Chinese in this study)
L2	The second language (English in this study)
L-FR	Unfiltered in the left channel and filtered in the right channel

LIST OF ABBREVIATIONS (Continued)

LH/L	Left hemisphere
LI	Lateralization index
MFG	Middle frontal gyrus
MMN	Mismatch negativity
MNI	Montreal Neurological Institute
ms	Millisecond
MTC	Middle temporal cortex
MTG	Middle temporal gyrus
NL-NR	Unfiltered in left and right channels
PMC	Premotor cortex
PoCG	Postcentral gyrus
PPI	Psycho-physiological interaction
PrCG	Precentral gyrus
RH/R	Right hemisphere
ROI	Region of interest
s	Second
SD	Standard deviation
SPL	Superior parietal lobule
Spt	Sylvian parietal-temporal
STG	Superior temporal gyrus
STS	Superior temporal sulcus
T	T value
UF	Uncinate fascile
VTA	The Verbotonal approach

CHAPTER 1
INTRODUCTION

The present study aims to identify an optimal auditory language input for Chinese EFL university students based on the Verbotonal approach (VTA). A combined ERP and fMRI experiment and semi-structured interviews are used to investigate students' brain responses while listening to different auditory language signals and opinions on these signals. This introductory chapter presents the background of the study first. Then, the statement of the problem, the significance, objectives, and research questions are delineated. At last, definitions of key terms and a summary of the chapter are outlined.

1.1 Background of the study

Educators and scholars see the language learning process as more likely to be mental processes. It means that teachers concentrate more on the "mind" that requires students to use existing knowledge and skills to cope with unanswered questions or tough situations, which are higher-order/complex cognitive skills such as reasoning, decision-making, problem-solving, and thinking (Galotti, 2017; Levine, 2009). In a typical foreign language classroom, teachers ask students to brainstorm a topic or a report, memorize words and structures, communicate in oral and written forms, and self-manage their learning time and materials during and after class. Teachers put too much emphasis on higher-level mental processes, while they almost forget about the student's "body." The problem is that the mind (mental processes) is overemphasized, but the body is usually neglected by foreign language teachers. Considering the "mind-body" connection, our body is not only connected to the mind but also has impacts on the mind (McNerney, 2011). The body (i.e., our brains, bodies, and bodily experiences) constrains, regulates, and structures the mind by allowing sensorimotor systems to receive sensory input and produce behavioral output (Foglia & Wilson, 2013). In other words, the body gives rise to the development of cognitive skills, known as "embodied cognition" (Foglia & Wilson, 2013; Leitan & Chaffey, 2014; McNerney, 2011). Thus, learning not only involves the student's mind, the teacher, and the environment but, foremost, the student's biological basis (i.e., the body), especially the brain. Only when the brain is exposed to the appropriate input and gets effectively involved in information processing could learning outcomes be improved. As a primary connector of the "mind-body" problem, auditory input is one initial and significant source of sensory input since fetal life, which is the basis of the proprioceptive memory and auditory-memory development (Asp, 2006). But how the brain processes the incoming signal is under-valued or under-researched by foreign language teachers. Thus, neurobiological-based listening/auditory perception is the focus of the current study.

This is due to the fact that all human activity, including learning a language, is inevitably biological in nature or, at the very least, closely connected with biological activity, especially cerebral activity. For centuries, language, the biological basis of language, in particular, has been a heated discussed topic in scientific fields, especially in fields of medicine and linguistics (Small & Hickok, 2016). Since the striking finding of brain localization of language production (the left posterior inferior frontal gyrus, Broca's area) (Broca, 1861a, 1861b) from the patients of Pierre Paul Broca in the middle of the 19th century, the "seat of language" sparked the focus of attention in human brain's unique capacity for managing speech understanding and production (Small & Hickok, 2016). After Broca's discovery, lesion-based analysis of brain localization for language blossomed. Investigators at that time held simple views of a link between brain pathology and behavior, represented by one of the contributors Wernicke (1874) who identified that the left superior temporal gyrus is responsible for language comprehension. Due to advances in psychology, information processing, and linguistic theories in the middle to late 20th century, researchers were able to access more detailed descriptions of language performance than before (Caramazza & Berndt, 1978; Chomsky, 1965). In addition, the technological development of structural brain imaging, particularly computed tomography (CT) and magnetic resonance imaging (MRI), made the research of brain pathology in vivo possible (e.g., Cappa & Vignolo, 1983; Metter et al., 1984). Although lesion analysis usually connects one specific focus of a brain lesion with one single linguistic or psychological phenomenon, this approach has provided a heuristic understanding of the neurobiology of language and identified functions of brain areas successfully (Small & Hickok, 2016).

Applications and wide acceptance of brain-cognition relation and high-resolution functional brain imaging (Petersen et al., 1988, 1989) in the late 20th century have dramatically redirected the way to brain-behavior investigations in language studies. Since then, brain activity in response to language stimuli can be measured by neuroscience techniques. Task-dependent electroencephalography (EEG), also known as event-related potentials (ERP), reflects electrical activity that is time-locked to the presentation of a sensory stimulus or a cognitive process, moreover, provides the high temporal-resolution measurement of the activity in neural networks (Kuhl & Rivera-Gaxiola, 2008; Small & Hickok, 2016). By contrast, functional magnetic resonance imaging (fMRI), as a noninvasive approach, allows in vivo investigation and provides high spatial-resolution localizations of neural activation induced by the given stimulus, which detects the blood-oxygenation changes in an MRI scanner (Small & Hickok, 2016). The increasing number of psychological and linguistic studies on measuring brain responses leads to a new field of cognitive neuroscience, in which researchers attempt to understand the relationship between mind and brain by examining the biological bases and computations for language (Small & Hickok, 2016).

From Broca's first attempt in the brain basis of language to analysis between behavior and brain pathology, then to cognitive neuroscience, the development of neuroscience, linguistic theories, and the advent of neuroscience techniques always push forward the explorations of brain-language relations for many centuries, which makes neurobiology of language possible to examine *"the biological implementation and linking relations for representations and processes necessary and sufficient for production and understanding of speech and language in context"* (Small & Hickok, 2016, p. 5). The research findings in this area not only reveal the brain basis of language but necessarily facilitate language learning and the physiological approach to therapy for speech/language disorder.

How to get the student's brain actively and properly involved in the language learning process is the question of how to raise student's awareness and change their perceptions of the foreign symbol system (Lian, 2004; Lian & Pineda, 2014). To answer this question, the Verbotonal approach (VTA) provides an optimal solution, which primarily considers the concept of neuroplasticity as a sound principle in language learning and therapy for the hearing impaired, and seeks an optimal auditory signal to the individual brain (Guberina & Asp, 1981, 2013). Since VTA concerns how to raise language learners' awareness and restructure the ways that they perceive the auditory signals of the language being learned (Lian & Sussex, 2018), it has been viewed as an effective approach for improving listening and speaking skills by emphasizing the essential role of hearing as an act of meaning-making and the fundamental value of prosody as a part of this process, i.e., stress, rhythm, and intonation (Asp et al., 2012; Guberina, 1972; He et al., 2015; Kim & Asp, 2002; Lian, 1980; Yang, 2016). This is because low-frequency stimulation, according to VTA, with the highlight of rhythm and intonation patterns is effective for structuring/re-structuring the perception of speech for hearing-impaired subjects and language learners (Asp et al., 2012; Guberina, 1972; He et al., 2015; Kim & Asp, 2002; Lian, 1980; Yang, 2016).

But the neural mechanisms for processing the auditory signals on the principle of VTA are not quite clear. Hence, this study focuses on the neural processing and students' opinions of L1 and L2 signals in different physical forms based on VTA. The hypothesis that there exists one possible optimal auditory language input signal for Chinese EFL students is tested by a combined ERP and fMRI experiment, as well as students' opinions of these language signals. As a result, this form of optimal language input can be adopted in the pedagogical instruction for EFL students in a Chinese context.

In order to identify an optimal auditory language signal for Chinese university EFL students, low-pass filtered stimuli and the methods of dichotic and diotic listening are used in the current study. As VTA initiates, low-pass filtered signals with rhythm and intonation patterns facilitate (foreign) language learning, which provides the foundations for speech perception and production and improving listening skills (Asp, 2006). In addition, the methods of dichotic and diotic listening can be used to reveal ear advantages for language perception and language skills (Asbjornsen & Helland, 2006). Thus, the current study uses filtered and unfiltered language signals under dichotic and diotic listening conditions to explore one type of optimal language input for Chinese EFL students, which is derived in part from principles of verbotonal theory and partly from other findings from neuroscience. To measure participants' brain responses while dichotically and diotically listening to different language signals in L1 and L2, neuroscience techniques, ERP and fMRI, are used to investigate the temporal-spatial neural signatures of auditory language processing. ERP is used to measure reaction time and mental workload during the processing of auditory stimuli, and fMRI is implemented for the spatial localization of processing auditory language signals by tracking blood-oxygenation changes during neural activation (Antonenko et al., 2014; Kuhl & Rivera-Gaxiola, 2008). Besides, participants' opinions of different signals are sought through semi-structured interviews. By creating signals in Verbotonal-based dichotic and diotic listening conditions, the present study aims to seek one optimal auditory language signal that students could make the best sense of and to optimize students' cognitive load during auditory language processing. Further, to better understand the processing mechanisms of auditory language signals for Chinese, this research takes both L1 (Chinese) and L2 (English) into consideration. All the auditory signals, including L1 and L2, are low-pass filtered and unfiltered stimuli based on VTA and arranged in dichotic and diotic listening conditions, so as to uncover the processing specializations employed by L1 and L2 and unveil Chinese students' neural networks for language processing.

All in all, this study highlights Chinese EFL students' neurobiological bases, especially the neural networks for auditory language processing (both in L1 and L2), which are usually ignored by foreign language teachers. By revealing students' temporal-spatial neural signatures of auditory language processing and their opinions to these signals, the present study seeks to create an optimal, physical, auditory stimulation in Verbotonal-based dichotic and diotic listening conditions. Ultimately, an optimal auditory language signal would enable EFL students to best make sense of it so as to restructure their perceptual mechanisms and facilitate language-learning.

1.2 Statement of the problem

1.2.1 The lack of awareness-raising

To improve students' listening, speaking, reading, and writing skills, English teachers in the research site – the Department of Foreign Languages of Kunming Medical University, China, have tried various teaching methods. According to a survey conducted among the total number of thirty-six English teachers in the Department by the researcher, the communicative approach (91.67%, 33 out of 36), the whole language approach (80.56%, 29 out of 36), the audio-lingual method (72.22%, 26 out of 36), and the grammar-translation method (61.11%, 22 out of 36) are predominant in teaching English to students in the university. Teachers claimed that they would use different teaching methods based on the textbook, students' English proficiency level, and exam content. Since the textbook they were using was designed to incorporate communicative activities with audio/video materials, most teachers followed the textbook and adopted the communicative approach (91.67%) and the audio-lingual method (72.22%). Teachers also reported that the whole language approach and the grammar-translation method were used because of the exam-driven curriculum and students' intermediate or lower level of English proficiency that required them to pay more attention to grammar/structure, reading, writing, and translation skills. Thus, mixed teaching methods were adopted by the English teachers in the research site.

In terms of the communicative approach, teachers set up real-life situations that students would encounter in real life, in which students are motivated by communicating in meaningful ways about meaningful topics (Brumfit & Johnson, 1979). But critics state that the communicative approach prioritizes the "function" over the "structure" of the language, and students would be the owner of "communicative competence" without being capable of making full, or adequate, use of the language (Ridge, 1992; Swan, 1985). The whole language approach emphasizes comprehension of the text meaning and critical thinking strategies (Freeman & Freeman, 1992). Since the whole language approach regards language as a meaning-making system, the phonetic, syntactic, semantic, and pragmatic aspects function in relational ways, in which students are required to make meaning in reading and express meaning in writing (Freeman & Freeman, 1992). As for the audio-lingual method, it aims at language proficiency in listening and speaking skills, which is tightly related to behaviorism, and, therefore, takes drilling, repetition, and memorization as central elements of instruction (Larsen-Freeman, 2000). Proponents of the audio-lingual method claim that language learning is habit-driven, particularly pronunciation, thus, continual repetition and accurate imitation of pronunciation and grammatical structures are effective and successful in improving aspects of language; while critics assert that it is a mechanical method that overemphasizes repetition and accuracy and pays no attention to students' communicative competence in the target language (Larsen-Freeman, 2000). Regarding the grammar-translation method, students are taught the grammatical structures by memorizing, doing grammar drills, applying grammar rules, and translating sentences to and from the target language, which is also criticized for little attention to the language aspects of pronunciation and communication (Damiani, 2003; Larsen-Freeman, 2000).

English teachers have tried hard to employ these methods and approaches with multiple learning materials to motivate students and improve their English proficiency. But students' performance is not always satisfactory as 76.32% (29 out of 38) of English teachers in the research site are unsatisfied with students' outcomes after adopting these teaching methods/approaches. Students' unsatisfactory performance may result from the in-class and out-of-class activities/tasks with the teaching methods/approaches mentioned above, as students need to memorize/rote, apply rules, communicate in oral and written forms, translate to or from the target language, and even read and think critically in a foreign language. It is obvious that all these activities/tasks are higher-order cognitive processes (Galotti, 2017; Levine, 2009), in which the language information *"that has been previously received, processed, and stored by basic cognitive processes gets used, combined, reformatted, or manipulated"* (Galotti, 2017, p. 339). These activities/tasks seem demanding for the students with the intermediate or lower level of English. Instead of these higher-order cognitive processes, one possible solution to the problem is to raise students' awareness and change their perceptions of a foreign language before doing these tasks. In order to raise their awareness and change their perceptions of a language, student's biological basis, i.e., the body, specifically the brain, is inevitably fully involved. But teachers almost forget that students are more biological beings than learners. It is, therefore, more important for teachers to prioritize students' perceptions of language signals derived from the neurobiological bases than class activities and tasks that call for higher-level cognitive skills. As learning is a meaning-making process, pedagogical instruction is to enable students to make better sense of language signals in order to structure/re-structure their perception of the language being learned, as a result, to facilitate language learning.

1.2.2 The lack of the mention of optimal language input

As the key to learning is to raise student's awareness, here comes the question that what kind of language input or, to be specific, what the physical characteristics of the auditory signal would be optimal for EFL students. An optimal language input signal should be best suited for language processing in the student's brain and able to maximally reorganize neural connections to make neuroplasticity (or learning) happen. Regarding language input, the most prominent current hypothesis is the input hypothesis (Krashen, 1982, 1985, 2003). Krashen (1982) places primary importance on the "comprehensible input" that can lead to acquisition. Language learners should be exposed to the "comprehensible input" containing *i+1*. In this hypothesis, *i* stands for previously acquired linguistic competence, and the *+1* refers to "the next increment" of new knowledge or language structure that is a bit beyond the acquirer's current level, meanwhile, it is still within the student's capacity to acquire (Krashen, 1982, 1985, 2003). However, as McLaughlin (1987) criticizes that the concept of "comprehensible input" is not precisely defined, which gives rise to the untestability of the hypothesis. It is ambiguous that what "*i+1*" signifies and what "comprehensible input" means (McLaughlin, 1987). Furthermore, no mention is made by proponents of "input," and the literature in general, of the physical features of an optimal auditory input signal that may assist with language learning. Therefore, to make the mention of the "optimal language input" clearer and to identify the reason why the input signal is optimal for language processing for Chinese university EFL students, this study explores the neural signatures for processing the low-pass filtered stimuli based on VTA and unfiltered signals under dichotic and diotic listening conditions.

1.2.3 The lack of neurobiological evidence

VTA considers the concept of neuroplasticity a key principle and seeks the optimal stimulation to the individual brain (Guberina & Asp, 1981, 2013), which has been applied to the fields of therapy to the hearing impaired as well as foreign language learning. Prior studies on EFL learning based on VTA indicate that low-pass filtered signals restructure the ways that students perceive auditory signals and improve students' English proficiency (He, 2014; He et al., 2015; Lian & Sussex, 2018; Luu, 2021; Wen et al., 2020; Yang, 2016; Yang et al., 2017). After learning English based on VTA, Chinese EFL students achieve improvements in pronunciation (He, 2014; He et al., 2015), speaking (word-reading, sentence-reading, singing, and oral interview) (Yang, 2016; Yang et al., 2017), listening and working memory capacity (Luu, 2021), and other subskills of vocabulary, grammar, fluency, and comprehensibility (He, 2014; He et al., 2015; Yang, 2016; Yang et al., 2017). Another study conducted by Wen et al. (2020) shows that the low-frequency, the prosodic signal is more effective for correcting both perception and production of English pronunciation. All these empirical studies show that low-pass filtered signals are significantly effective in improving Chinese EFL students' English proficiency in general. Since these studies are based on the framework of VTA, which targets rewiring the brain's neural connectivity, the exploration of neural mechanisms for processing the low-pass filtering signal should be indispensable. But there has been a lack of knowledge about it up till now. This study aims to provide neurobiological evidence for processing low-pass filtered prosodic signals and unfiltered signals, which is realized by a combined ERP and fMRI experiment to uncover the temporal-spatial neural signatures for processing filtered and unfiltered L1 and L2 signals while dichotically and diotically listening.

1.3 Significance of the study

This study is based on previous successful verbotonal-based studies using low-pass filtered language signals to facilitate perception and raise awareness of language phenomena resulting in enhanced language-learning (He, 2014; He et al., 2015; Lian & Sussex, 2018; Luu, 2021; Wen et al., 2020; Yang, 2016; Yang et al., 2017). It seeks to refine these findings by performing instrumental studies of brain activity and load by securing biological data in response to the changing nature of the physical signal and the nature of the language. This will help determine the best physical signal (if any) that should be provided as input to students in order to maximize perception and, more generally, enhance the learning of a second language. Its findings

1) will serve to provide insights into how language is processed. This is basic scientific information about cerebral processing and will provide clues about how the brain deals with language bearing audio signals.

2) will serve to develop new ways of providing language learners (and possibly L1 speakers) with the best possible physical audio input to facilitate perception, comprehension, and learning. Much more research will be needed in order to fine-tune our findings to maximize the audio language processing of L2 learners. As a side issue, the research could also have an impact on L1 speakers who may be experiencing language difficulties or who may be suffering from language or perceptual deficits as a result, for instance, of cerebral accidents such as strokes.

In summary, the significance of this study is that it will provide basic scientific information on the biological processing of language-bearing audio signals for both L2 learner-listeners (e.g., Garcia-Sierra et al., 2011; Rüschemeyer et al., 2005; Shi et al., 2006) and L1 listeners (e.g., Chien et al., 2021; Friederici, 2012; Parker et al., 2005; Plante et al., 2002). Once that processing has been understood, the results may be applied to optimizing the physical audio input for language learners.

An additional benefit may be that the results could also be applied to L1 speakers with pathological conditions relating to language. This last aspect will, however, not form the primary focus of the current research project.

Finally, this study represents only a small step in the iterative process of progressive refinement of research into the nature of the physical signal as it affects L2 learning. Once general processing patterns have been established, future research (beyond this thesis) will delve into the details of how else to manipulate the physical audio signal so as to enhance the L2 learning process.

1.4 Objectives of the study

This study is intended to explore an optimal auditory language input signal for Chinese university EFL students. To be specific, the following objectives are targeted:

1) To identify students' brain activity while dichotically and diotically listening to filtered and unfiltered L1 and L2 signals in the combined ERP and fMRI experiment;
2) To investigate students' opinions of the auditory language input signals in Verbotonal-based dichotic and diotic listening conditions;
3) To seek an optimal auditory language input signal according to students' brain responses while listening and views after listening to the signals in Verbotonal-based dichotic and diotic listening conditions.

1.5 Research questions

Based on the objectives of the current study, research questions are formulated as follows:

1) What are students' brain responses while dichotically and diotically listening to filtered and unfiltered L1 and L2 signals in the combined ERP and fMRI experiment?
2) What are students' opinions of the auditory language input signals in Verbotonal-based dichotic and diotic listening conditions?
3) Does there exist an optimal auditory language input signal with filtered and unfiltered stimuli under any dichotic or diotic listening condition(s) for university students who speak Chinese as L1 and English as L2? If yes, in what ways?

1.6 Definitions of key terms

1) EFL students: EFL students in the present study refer to the students in Kunming Medical University, Kunming City, Yunnan Province, China, who have been learning English as a foreign language for over ten years and have passed the College English Test (CET) Band 4 (CET is the national standardized English

proficiency test for undergraduate and postgraduate students in China; Band 4 refers to the medium level of the test).

2) Dichotic listening: Dichotic listening is a non-invasive procedure, which can be used, inter alia, to measure/test cerebral hemispheric specialization of auditory processing. In a dichotic approach, subjects listen to one kind of signal through one ear and, simultaneously, a different signal through the other ear. Through this approach, a general right ear advantage for verbal stimuli and a left ear advantage for non-linguistic stimuli (such as melody, or prosody of language) have been reported in healthy individuals (Gandour et al., 2004; Meyer et al., 2002; Sammler et al., 2015; Tervaniemi & Hugdahl, 2003; Vigneau et al., 2006).

3) Diotic listening: The diotic listening condition means that the same signals are presented binaurally to both ears simultaneously.

4) Low-pass filtered stimuli: This study adopts the low-pass filtering of auditory stimuli, a technique where an audio recording is modified by an audio filter that only allows frequencies under 320 Hz (or some other appropriate frequency) to be preserved. This filtered stimulus preserves the fundamental frequency (F_0) of the sentences being studied, together with their stress, rhythm, loudness, and intonation features, while the higher frequencies that help to define words are removed (Lian & Sussex, 2018). The filtered stimulation that consists of a stream of sound is reminiscent of a hummed sentence, where the prosody becomes salient and available in a way that is unusual, stimulating, and awareness-raising (He, 2014; He et al., 2015; Lian & Sussex, 2018; Luu, 2021; Wen et al., 2020; Yang, 2016; Yang et al., 2017).

5) Unfiltered stimuli: The unfiltered stimuli in the current study are audio language signals that maintain all frequencies of the sounds without being filtered at any cut-off frequency, which are sounds in our daily speech/communication. Compared with 320 Hz low-pass filtered stimuli that make prosody salient, the unfiltered stimuli preserve frequencies under and above 320 Hz that help listeners define words (Lian & Sussex, 2018).

6) The neural bases of auditory language processing: A neural scaffolding that is constructed by speech-sensitive regions from the temporal cortex (the auditory cortex) to frontal and parietal cortices, which is assumed to be engaged in spoken language processing (Peelle, 2019). These anatomical and functional connections lay the basis for a dual-stream model for speech processing (Hickok, 2012; Hickok & Poeppel, 2007; Rauschecker & Scott, 2009), to be specific, a dorsal stream and a ventral stream. The dorsal stream is associated with speech perception-production, which begins from the auditory cortex, through the arcuate fasciculus and parietal lobe, to the dorsal premotor cortex; the ventral stream is correlated with speech object recognition, which connects the auditory cortex, through the anterior temporal lobe and the uncinate fasciculus, to ventral inferior frontal gyrus (Hickok, 2012; Hickok & Poeppel, 2007). The dual-stream model is used as the framework in the mainstream cortical studies of speech processing (Peelle, 2019).
7) Temporal and spatial neural signatures of auditory processing: The neural signatures of auditory processing represent one’s collective experience with sound throughout life (Kraus & Anderson, 2015; Kraus & Nicol, 2014), which can inform the fundamental mechanisms of sound-themed cognitive tasks (Kraus & Nicol, 2014). To illustrate the neural signatures for processing filtered and unfiltered speech signals, this study measures the temporal and spatial signatures of the brain’s responses to these signals by using ERP and fMRI techniques. ERP provides precise time resolution in milliseconds and the amplitude of brain waves, making it well suited for measuring reaction time and load/capacity of speech processing to reveal the neural networks in response to the auditory language signal (Friederici, 2005; Kuhl, 2004; Kuhl & Rivera-Gaxiola, 2008). To illustrate the brain localization of the stimulus processing, fMRI offers high spatial-resolution maps of the neural activity across the entire brain by detecting the changes of blood oxygenation elicited by neural activation/firing (Gernsbacher & Kaschak, 2003).

1.7 Summary

In this chapter, an introduction to the research background is presented. Following the background, statement of the problem, the significance of the study, research objectives, and research questions are demonstrated. In addition, some relevant key terms are introduced or defined to avoid misunderstanding. In the following chapter, relevant conceptual pieces of work and studies will be reviewed to construct a theoretical framework for the current research.

CHAPTER 2
LITERATURE REVIEW

This chapter offers a review of literature and theories that are relevant to the present study. It consists of five sections, in which language and language learning, neuroscience techniques used with brain responses to linguistic signals, neurobiological theories of speech perception, the Verbotonal approach, and cognitive load theory are discussed respectively. Lastly, a theoretical framework for the current study is constructed based on the above-mentioned review and the research context.

2.1 Language and language learning

The notions of language and language learning are presented in this section, which frames the concepts of language perception and development/learning from neurobiological and cognitive perspectives.

2.1.1 What is language?

Language is considered a mental phenomenon (Jackendoff, 2002). According to Noam Chomsky's *Aspects of the Theory of Syntax* (1965), language is a structure or model of something in a language speaker's mind that the speaker can speak or hear, to be specific, a mental representation of structures including phonological, syntactic, semantic/conceptual, and spatial structures (Chomsky, 1965). The term "representation" means representing something to someone, meanwhile, something could represent something else in someone's mind (Jackendoff, 2002). It is believed that language is a semiotic system in the human mind, and it is the symbols that symbolize the entities in the human mind. The entities in the mind, however, do not symbolize anything. For instance, the entity of the color white in the mind does not symbolize the word *white* but the mental entity makes the word what the color it is. In addition, a symbol becomes a symbol only when it is shared by a perceiver or community of perceivers (Jackendoff, 2002).

The term "mind" is conventionally known as the seat of consciousness, and the problem of "mind-body" touches upon the relationship between consciousness and the physical world. Since Freud, the "unconscious mind" has been well-known and more frequently used as what we are not aware of. This view of the "unconscious mind" is then often thought to be phenomena described as "mental" (Jackendoff, 2002). Under the "mind," it is about "body," specifically, the brain. When processing linguistic expressions, there is no room in the mind to elaborate the phonological, syntactic, semantic/conceptual, and spatial structures of each expression, which is far beyond available introspection. *"It leaves room only for neurons firing and thereby activating or inhibiting other neurons through synaptic connections"* (Jackendoff, 2002, p. 21). In the field of modern cognitive neuroscience from the 1990s, the term "mind" has been depicted as the functional organization and activity of the brain, only a small part of which appears consciously but a majority of which does not (Jackendoff, 2002). When talking about parsing a linguistic expression from the mind/brain point of view, *"we are speaking in functional terms; this functional organization is embodied in a collection of neurons engaging in electrical and chemical interaction"* (Jackendoff, 2002, p. 22).

2.1.2 What is language learning?

Chomsky (1965) states that language learning or grammar construction is innate, which implies "present at birth." But it is more widely accepted that language develops with the maturation or development of a human organism (Harris, 2006; Jackendoff, 2002). The term "innate" indicates that something, determined by genes, automatically develops after birth, such as tooth growth. The other term "acquisition" that is in contrast with "innate" highlights the influence of the environment (Harris, 2006; Jackendoff, 2002). In fact, most organisms develop on the basis of a dynamic interaction of innate and environmental influences (Harris, 2006; Jackendoff, 2002). Jackendoff (2002) draws an analogy between language learning and muscle building that the muscular system, like the locations of muscles, is innate, yet the strength and the size of muscles rely on exercise and nutrition. This is analogous to language learning, the ability to acquire a language, i.e., the neurobiological basis of language perception and production, is determined by genes and may well develop in the brain of a two- or three-year-old kid. But the child's performance of speaking and understanding a language is dependent on the interaction between the neurobiological basis of language and the language input in the environment (Harris, 2006; Jackendoff, 2002).

From a neurobiological perspective at the structural level, learning, including language learning, occurs when neuroplasticity happens. The term "neuroplasticity" means *"the capacity of nervous tissue to change in structure and function in response to factors which can be described as 'environmental'. These factors act on the level of the neuron or its substructures such as dendrites and dendritic spines but also - on a more macroscopic level - at the level of aggregates of neurones (e.g., neuronal networks)"* (de Jong et al., 2009, p. 51). Learning a language, as one of the factors, can restructure and reorganize the neuron and neuronal assemblies, and thus, can lead to the structural and functional changes of the brain (de Jong et al., 2009; Kolb & Whishaw, 2015). Neuroplasticity, as an intrinsic property of neural tissue, is the process of brain development and maturation (de Jong et al., 2009), as well as the basis for a human to learn languages. More and more studies show language exposure and language learning induce neuroplasticity in bilingual brains, second language learners, and even deaf-blind subjects measured by neuroimaging techniques (e.g., Li et al., 2014; Obretenova et al., 2010; Tu et al., 2015).

From the viewpoint of interactive specialization, there is increasing regional specialization and pathways in the human brain that develop systematically during language learning (Booth, 2010). According to behavioral and neuroimaging literature, increasing specialization shows that activation becomes more and more focal over language development, implying greater activation in task-relevant areas and less activation with the irrelevant task (Berl et al., 2006; Bunge & Wright, 2007; Casey et al., 2005; Kadosh & Johnson, 2007). To be specific, neuroimaging findings reveal that the left superior temporal area is increasingly activated during phonological processing and becomes more specialized over phonological development (Bitan et al., 2009; Bitan et al., 2007; Booth, 2010), which indicates the link between learning new auditory information (i.e., sound-based representations) and superior temporal regions. It is observed that developmental increases of activation in the middle temporal gyrus during semantic processing in both visual and auditory modalities, and greater involvement of the left inferior frontal gyrus during syntactic processing (Booth, 2010; Brauer & Friederici, 2007). Scholars claim that these increases of activation are not only associated with phonological, semantic, and syntactic development but with students' proficiency, since lower proficiency is correlated with greater activation of brain regions (Booth, 2010). It indicates that students with lower proficiency have underdeveloped phonological, semantic, and syntactic manipulations, which leads to a greater degree of controlled processing for retrieval and selection (Booth, 2010).

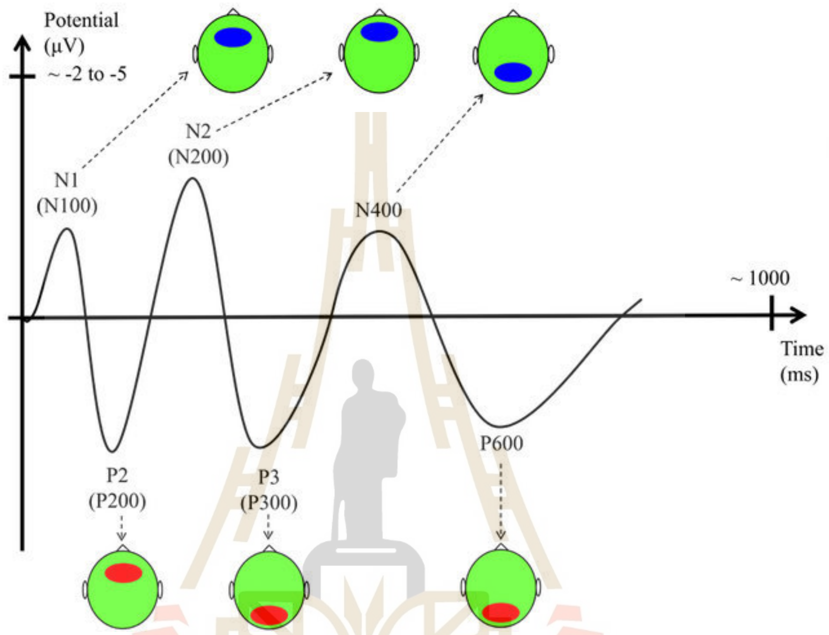
Language, determined by genetic factors and constrained by the language input in the environment, develops over a student's lifetime (Harris, 2006). From a neurobiological standpoint, learning a language leads to neuroplasticity and increasing specialization with the development of the human organism and language exposure (Booth, 2010; de Jong et al., 2009; Kolb & Whishaw, 2015). The process of learning a language is thought to be a mental process from a cognitive perspective that a student makes sense of the semiotic system by virtue of the dynamic interaction between the neurobiological basis and the language input in the context (Harris, 2006; Jackendoff, 2002; Lian, 2004). Therefore, the present study takes this interaction as the focus, aiming at uncovering the neural mechanism of auditory language processing and seeking an optimal auditory language input signal that Chinese university EFL students are sensitive to and they can make the best sense of. As a result, this optimal auditory language input signal could restructure students' perceptual mechanisms to facilitate language-learning.

2.2 Neuroscience techniques used with brain responses to linguistic signals

This section presents a general introduction to two neuroscience techniques, EEG/ERP and fMRI, used in the current study. Related research and findings on the neural bases of language processing measured by EEG/ERP and fMRI will be embedded in the next section 2.3.

2.2.1 EEG/ERP

EEG (electroencephalography) is a noninvasive method to monitor and record electrical activity of the brain. By detecting EEG activity, Event-related potential (ERP) is used to measure electrophysiological responses to a cognitive-related event in sensory, cognitive, or motor processes (Luck, 2014). Systematic deviations from brain waves/fluctuations exhibited by ERPs precede, accompany, or follow the experimental events (Antonenko et al., 2014). These elicited potentials reflect electrical activities that are time-locked and phase-locked presentations of sensory stimuli (e.g., visual or auditory stimuli), or a cognitive process (e.g., recognition of a semantic/syntactic violation in a sentence) (Morgan-Short & Tanner, 2014; Swaab et al., 2012). The different potentials are represented by a letter showing polarity (P: positive; N: negative) and the following number refers to the typical latency in milliseconds, e.g., P300 and N100 mean a positive wave peaking at 300 ms and a negative wave peaks at 100 ms respectively (Antonenko et al., 2014). Compared with other functional imaging techniques, the observed potentials with their latency, amplitude, duration, and distribution can be used as indicators to understand the cognitive processes elicited by the experimental task (Antonenko et al., 2014). In other words, ERP can provide indices not only of lateralization (demonstrated by the distribution of potentials) in processing but also of dynamic changes in mental operations (i.e., the processing load/capacity shown by amplitude) and speed (reaction time illustrated by latency) of information processing at different points in the information processing stream (Kutas et al., 2012; Luck, 2014). Therefore, ERP has been widely used to study speech and language processing not only in healthy adults and children but in language disorders (Friederici, 2005; Goswami, 2004; Kuhl, 2004; Kuhl & Rivera-Gaxiola, 2008). By placing sensors on a subject's scalp, *"the activity of neural networks firing in a coordinated and synchronous fashion in open field configurations can be measured, and voltage changes occurring as a function of cortical neural activity can be detected"* (Kuhl & Rivera-Gaxiola, 2008, p. 513). Since ERP provides precise time resolution of brain activity during stimulus processing in milliseconds, it is quite suited for studies on the high-speed and temporally ordered language structure (Kuhl & Rivera-Gaxiola, 2008; Kutas et al., 2012). The main ERP components are illustrated with their functional interpretations, latencies, and scalp topographies regarding language processing in Figure 2.1. However, ERP is limited in spatial resolution so it cannot reveal the source of brain localization accurately during processing (Kuhl & Rivera-Gaxiola, 2008; Kutas et al., 2012).



Notes: N1 refers to preattentive perceptual processing in the primary cortex; P2 stands for preattentive perceptual processing in the primary cortex; N2 depicts stimulus detection; P3 indicates stimulus categorization and memory updating; N400 represents semantic/conceptual processing; P600 describes syntactic processing.

Figure 2.1 Daltrozzo and Conway's (2014) illustration of functional interpretation, latencies, and scalp topographies for main language-related ERP components

2.2.2 fMRI

As the mainstay of neuroimaging at present, functional magnetic resonance imaging (fMRI) is noninvasive and provides high spatial-resolution images of neural activity over the brain (Logothetis, 2012), which is widely adopted to map human's neural network for language processing (e.g., Gernsbacher & Kaschak, 2003). Different from ERP, fMRI is less sensitive to detect neural activity directly since it measures the changes in blood oxygenation in response to neural activation/firing, called the blood oxygenation level dependent (BOLD) signals (Kuhl & Rivera-Gaxiola, 2008; Logothetis, 2012). fMRI works based on the fact the active neurons need more oxygen, at least more than is required for the metabolic needs, thus, the brain sends extra oxygen to the active nerve cells (i.e., the activated regions) because of a specific task/activity (Poldrack, 2018). As a result, the number of oxyhemoglobin (oxygen-carrying hemoglobin) would be increased in the activated brain region (Indefrey, 2012; Poldrack, 2018). Different concentrations of hydrogen in oxygenated and deoxygenated blood of different cerebral regions can be manipulated by the MRI magnet, and then the computerized images can be read and translated (Indefrey, 2012). According to BOLD signals, the cerebral regions that are most active during cognitive or perceptual tasks could be determined (Indefrey, 2012). Neural events occur in milliseconds but it may take seconds to make blood-oxygenation changes, therefore, fMRI is limited in temporal resolution (Kuhl & Rivera-Gaxiola, 2008).

As ERP and fMRI techniques have their strengths and limits, this study adopts both techniques to collect data. ERP is used to measure reaction time and processing load. In addition, fMRI is used for brain localization for the processing of language stimuli. Therefore, by implementing ERP and fMRI techniques, this study explores temporal-spatial neural signatures for processing the filtered and unfiltered signals under dichotic and diotic listening conditions and aims to look for an optimal auditory language input signal that is best suited for processing in Chinese EFL students' neural networks.

2.3 Neurobiological theories of speech perception

This section introduces the neurobiological theories of language/speech perception, which includes the active and perceptual mechanisms for language processing, the cortical and subcortical bases of speech perception, and the neural networks of syntactic, semantic, and linguistic prosody processing. Related neuroscience studies of language perception and processing are reviewed as well.

2.3.1 Mechanisms of speech processing

2.3.1.1 Active cognitive processing of speech

Whether speech sounds are processed by specific neural mechanisms or by general cognitive/perceptual processes has been a debate since the 1960s (Lotto & Holt, 2016). The motor theory of speech perception (Liberman et al., 1964; Liberman et al., 1967) supports that the acoustic stimulus and speech perception are mediated by articulatory movements and their sensory feedback, but the critique of the theory by Lane (1965) holds that identification and discrimination functions for speech stimuli are not different from those for nonspeech signals under comparable conditions. Although researchers defend their "all-or-non" positions on specialized language processes and explain the phenomena that require language-specific or general processing mechanisms (Diehl et al., 2004; Lotto & Holt, 2016), the debate of "speech-is-special" (Fowler, 2008; Lotto et al., 2009; Massaro & Chen, 2008) drives the focus of the field towards new directions. Since then, more and more scholars have been investigating distinctions, relative roles, and interactions of the motor, perceptual, cognitive, and linguistic systems during speech/language processing, which provides more reasonable and comprehensive models of speech perception and production (Lotto & Holt, 2016).

At the basic level, the general auditory system is adequate for speech perception, but the ways that human use speech to communicate are constrained and constructed by the operational histories (Lian, 2004; Lian & Pineda, 2014) of the auditory system and language experience, which is due to the information-carrying signal must be encoded by the auditory system and discriminated based on the listener's previously acquired language knowledge (Lotto & Holt, 2016). Based on this, one view considers speech perception as a process of pattern matching that acoustic signals are transformed to representations so as to match the stored mental representations of linguistic structures (Heald et al., 2016; Heald & Nusbaum, 2014; Lotto & Holt, 2016). This statistical pattern-matching process assumes that *"relatively stable linguistic categories are characterized by neural representations related to auditory properties of speech that can be compared to speech input"* (Heald & Nusbaum, 2014, p. 1). Hence, it is a passive process that inputs are linked directly to outputs without hypothesis testing or information-contingent manipulations as automatized cognitive processes (Heald & Nusbaum, 2014; Shiffrin & Schneider, 1977), which suggests *"rigidity of processing with few demands on cognitive processing"* (Heald & Nusbaum, 2014, p. 1). However, language use and understanding require flexibility and generativity for the contingent response (Heald & Nusbaum, 2014) rather than pattern matching or simple detection/discrimination of speech sounds. Thus, the current study holds the view that speech perception, basically as a cognitive process (Neisser, 1967), is an active process that is attentionally guided and is similar to hypothesis testing (Heald et al., 2016; Heald & Nusbaum, 2014; Nusbaum & Henly, 1992; Nusbaum & Schwab, 1986). The active process of speech relies on the active cognitive system that has a control structure allowing *"'information contingent processing' or the ability to change the sequence or nature of processing in the context of new information or uncertainty"* (Heald & Nusbaum, 2014, p. 1). The active system, theoretically, constructs hypotheses to be tested when new information comes or is derived, as a result, cognitive and perceptual flexibility is built to respond to signals even in novel situations (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986). Active cognitive processing on speech regards attention and plasticity as important factors in considering listeners' perception of speech (Heald & Nusbaum, 2014).

Attention in the active process of speech does not imply that speech analysis is consciously guided but the contingent changes of acoustic information in the early stage of auditory encoding can influence the listener's perception of speech and experience (Heald & Nusbaum, 2014). Since listeners' experience structures the basic neural patterns obtained from the acoustic signals and constrains early encoding, auditory encoding in the brain, reflected by the auditory neural response, is formed from *"the top-down under active and adaptive processing of higher-level of knowledge and attention"* (Heald & Nusbaum, 2014, p. 4). The mechanism for language comprehension, to a broad extent, ought to be plastic, indicating that speech recognition fundamentally means learning (Heald et al., 2016; Heald & Nusbaum, 2014). But the plasticity of speech processing remains a controversial issue of how long-term memory structures operating speech processing are reshaped/adjusted to permit the plasticity, meanwhile, protecting and retaining the previously stored/learned information when newly acquired information is inconsistent/irrelevant with the information in the long-term memory system (Born & Wilhelm, 2012; Carpenter & Grossberg, 1988). It is a debate that will continue. From the view of speech perception as an active cognitive process, learning a language or plasticity is not merely higher-level auditory processing positioned above word recognition but the mechanisms that can shift attention to relevant acoustic signals for speech perception and then tune speech perception to manipulations of the specific vocal features derived from variant speakers, contexts, and distortion of speech sounds (Heald et al., 2016; Heald & Nusbaum, 2014).

The fundamental problem challenging theories of speech perception is the lack of invariance between the acoustic presentations of speech and the linguistic interpretation of these presentations (Heald et al., 2016; Heald & Nusbaum, 2014). From the view of the simple recognition system, specifically passive processing of speech, recognition of the stimulus/input is a simple process of matching the pattern of the incoming stimulus to the most similar previously-stored pattern in a mathematical way of comparing (Heald et al., 2016). The recognition will succeed if the stored pattern is closest to the input pattern, however, pattern-matching does not always work successfully for speech recognition because of the lack of invariance problem in speech (Heald et al., 2016; Heald & Nusbaum, 2014). The problem of the many-to-many mapping between acoustic patterns and perceptual interpretations has existed for a long time (e.g., Liberman et al., 1967), but the distinctions of the other two issues of many-to-one and one-to-many mappings, which are of significance to understand the mechanisms for language/speech processing, are unclear (Heald et al., 2016; Heald & Nusbaum, 2014). The many-to-many mapping problem occurs when one specific pattern can be interpreted or categorized in various ways. Nusbaum and Magnuson (1997) consider the many-to-one mapping as a simple deterministic mechanism that creates the one-to-one mapping allowing passive manipulations of feature detection between inputs and outputs, thus, the many-to-one mapping can be achieved and represented by a set of one-to-one mappings for different stimuli. But the problem of the one-to-many mapping can be solved only by non-deterministic mechanisms (Nusbaum & Magnuson, 1997). There may be ambiguity about the interpretation of the input (e.g., the formant pattern of the vowel can be interpreted as either *bit* or *bet*), but one could eliminate some alternative interpretations according to additional information such as the speaker or speech context. This supports what has been mentioned above that there is no passive or automatic processing on speech, but it is an active cognitive process that tests hypotheses about interpretations and identifies constraints for interpretations, which is essential to achieve flexibility and generativity for language comprehension (Heald et al., 2016; Heald & Nusbaum, 2014; Nusbaum & Magnuson, 1997; Nusbaum & Schwab, 1986).

Therefore, the current study insists that speech perception is an active process of testing a battery of candidate hypotheses based on the pattern structure (mainly acoustic pattern) of an utterance. During the process of speech recognition, listeners make predictions of possible categories of utterances according to signal information, their experience/histories, attention, and speech context. The perceptual system then operates tests to distinguish the differences of the alternative classifications from information of the signal, which results in increased cognitive load due to the increased number of alternatives to be examined and diagnostic tests. Consequently, the constancy of perception on speech sounds is the outcome of hypothesis testing to differentiate the alternative interpretations of linguistic signals (Lotto & Holt, 2016). It is the active cognitive process that mediates the processes of hypothesis construction and testing.

2.3.1.2 Perceptual adaptation to speech

Traditional theories of speech perception concentrate more on the stability of categories and recognition of speech patterns, such as phonemes, to avoid mentioning the balance between plasticity and stability, in which speech is assumed to be developed in the early life, modified by exposure, and manipulated by the passive detection system (Fodor, 1983; McClelland & Elman, 1986). However, since there always exists variability of speech context and speakers, mechanisms for speech perception should be dependent on adaptive processing to flexibly respond to a wide range of speech patterns. Hence, learning or understanding a language should be supported by perceptual mechanisms that are adaptive to cope with contextual variability, speaker variability, or other forms of speech variation (Heald et al., 2016).

As learning speech sounds is relevant to recognizing variability of phonetic properties from different speakers in various contexts such as variations in accent and speaking rate, discriminating specific speech sounds basically relies on shifts in perceptual attention (Heald et al., 2016). Based on this point of view, the previous study conducted by Schwab et al. (1985) examined the effects of training on students' perception of synthetic speech. Compared with the other two groups trained by natural speech and without training, participants trained by synthetic speech performed better in the word recognition task and long-term retention of knowledge. Moreover, in the six-month follow-up investigation, the experimental group also performed better due to the experience acquired from previous exposure to synthetic speech. This study is informative that students' perception of synthetic speech results in adaption to variations in speaker and context during speech processing. Furthermore, the study of Greenspan et al. (1988) supported the effects of stimulus structure and variability on students' perception of speech, in which subjects were trained by either isolated words or sentences of synthetic speech with either novel stimuli or fixed and repeated list of stimuli. Results showed that subjects trained with stimuli of synthetic speech could better generalize beyond the training examples to novel stimuli and contexts than the group trained with repeated stimuli.

In addition, Francis et al. (2000) claimed that learning new phonetic categories of a foreign language could be regarded as learning to direct perceptual attention on the acoustic-phonetic properties of speech that were phonologically related to any given context. Two groups of participants in this research were trained with category-level stimuli as feedback – cooperating-cue stimuli (i.e., syllables in which the cued perception of the burst release of a consonant was consistent with the origin of the formant transitions) and conflicting-cue stimuli (i.e., the burst release was different from the formant transitions) – to learn to shift their attention from one acoustic cue to another. Results indicated that learning with such stimuli redirected students' attention to acoustic cues, moreover, this refocus of attention could generalize to more untrained and novel phonetic contexts. The findings also supported that students could learn a variety of speech sounds outside their first language, which required the ability to restructure existing knowledge and was further confirmed by

Francis and Nusbaum (2002). According to attention-to-dimension models, Francis and Nusbaum (2002) examined students' use of acoustic-phonetic dimensions before and after training. Though the models assumed that the structure of students' perceptual space was stable and learning was the process of re-weighting the existed dimensions to emphasize/de-emphasize changes in selective attention, their results indicated that training restructured students' perceptual space and students were able to redirect their perceptual attention to other acoustic-phonetic cues that were not previously used to form distinctions in their mother tongue.

From the perspective of cognitive load on working memory, Francis and Nusbaum (2009) investigated the effects of intelligibility on working memory capacity required by the perception of synthetic speech. A primary task of speeded word recognition and a secondary task with varied demands of working memory were paired in three experiments. The variable of speech intelligibility was determined by acoustic cues with poorer and better speech quality. The results showed that such training improved students' speech intelligibility and recognition speed significantly but the increased load on working memory slowed recognition. In the following experiment, participants received no training, and intelligibility was operated by changing synthesizers. The findings indicated that poorer quality of acoustic cues (without training) improved recognition accuracy and the increased load on memory decreased the accuracy rate, but more intelligible speech (i.e., better speech quality) did not spare more available working capacity for efficient use. This study is informative that perceptual learning increases students' intelligibility of speech, which is caused by the shift of perceptual attention to more phonetically-related cues, which in turn lowers cognitive load on working memory for speech recognition. Regarding cognitive load mentioned here, the following section 2.5 will introduce cognitive load theory in detail, including the benefit of reducing cognitive load and how previous studies optimize cognitive load while learning English as a foreign language.

All in all, due to adaptability and flexibility of the active cognitive process in speech perception, students are able to recognize and understand the signal despite the variability of speech through previously acquired acoustic-phonetic representations and redirect attention to information that has been cued by perception, so as to propose a solution to *"the one-to-many lack of invariance problem"* (Heald et al., 2016, p. 199) and uncertainty of speech interpretation would be reduced. It is believed that this non-deterministic computational mechanism for speech processing allows students to shift perceptual attention to the cues that can be discriminated among a set of alternative interpretations of one acoustic representation under the circumstances of contextual variability (e.g., speaker variability of pitch and speaking rate). Therefore, the current study holds that speech perception is achieved by an active cognitive process (Heald et al., 2016; Heald & Nusbaum, 2014; Nusbaum & Morin, 1992; Nusbaum & Schwab, 1986). In particular, students' perception of language signals is fundamental in recognizing and learning a (foreign) language, thus, developing or changing students' perception of language/speech should be prioritized in the learning process. Taken together, this study seeks to create an optimal auditory stimulation to enable EFL students to best make sense of that stimulation, as a result, to restructure their perceptual mechanisms to facilitate language-learning (Guberina & Asp, 1981; Lian & Sussex, 2018).

2.3.2 The neural bases of speech perception

2.3.2.1 The cortical bases of speech perception

Among neurobiological theories of speech perception, some models of neural networks are regarded as passive processing because they describe speech processing as a complete bottom-up recognition process, in which specific brain regions decode auditory language signals from acoustic coding to phonetic/phonological coding then to word recognition in a linear way. In addition, these passive processing models that ignore contextual variability in speech seldom address the lack of invariance problem (Heald et al., 2016; Hickok & Poeppel, 2016). Other neurobiological theories propose that speech perception involving neural processing networks with working memory, attention control, sensory perception, etc. is an active cognitive process. These models allow top-down projections from brain regions to modify/change the perception of auditory language signals, in which the neural networks of language/speech processing are perceptual (Heald et al., 2016; Hickok & Poeppel, 2016). Some mainstream neurobiological models of speech processing are reviewed as follows, and this part discusses the cortical bases of speech perception.

Hickok and Poeppel (2007) put forward a neurobiological model that consists of a dual-stream neural network of speech processing, namely a ventral stream identifying and recognizing speech and a dorsal stream perceiving and producing speech (the dual-stream model is shown in Figure 2.2). This model (Hickok & Poeppel, 2000, 2004, 2007) suggests that speech perception at the early stages involves bilateral auditory cortices (areas in green) on the dorsal superior temporal gyrus (STG) for spectrotemporal analysis of the auditory language signal and the superior temporal sulcus (STS) for phonological analysis (in yellow). Then, two streams go in two directions: one stream leads to a temporal ventral pathway for language/speech comprehension in areas (in pink) of the posterior middle temporal gyrus (pMTG) and the posterior inferior temporal sulcus (pITS) for lexical access (laterally but with weak left-dominant bias), and for combinatorial processing in the anterior middle temporal gyrus (aMTG) and the anterior inferior temporal sulcus (aITS) (may be dominated by the left hemisphere). The other dorsal stream is strongly left-hemisphere dominant for sensorimotor integration and speech production (in blue), which involves sensory and motor systems in the regions of the Sylvian parietal-temporal (Spt) area and the frontal areas (the posterior inferior frontal gyrus, pIFG; the premotor cortex, PMC; and the anterior insula). During the processes of speech perception and production, the conceptual network (the grey box) distributes widely over the cortex.

Hickok and Poeppel (2007) innovatively separate speech perception-production from speech recognition by a dual-process model based on cortical functions, in which speech perception is treated as the sublexical task that requires discriminating or categorizing auditory input (Heald et al., 2016). Although they claim that this model is an active process requiring working memory and executive control, it is, in fact, not adequate to solve the lack of invariance problem in speech to better understand spoken language in various contexts. According to Hickok and Poeppel, the dorsal stream puts emphasis on word learning and metalinguistic performance instead of recognizing speech, while recognition and understanding speech/language signals are only manipulated in the ventral stream that transforms acoustic representations into mental lexical representations (Heald et al., 2016). Therefore, this dual-stream model of speech perception is a passive processing model that does not address how the dorsal stream copes with contextual variability, i.e., how the dorsal stream compares acoustic alternatives, maps onto working memory, and shifts perceptual attention between acoustic cues. Additionally, the model does not explain the interaction between the ventral and dorsal streams and how the conceptual network affects the two streams, which leaves the contextual-based attention shift unanswered (Heald et al., 2016).

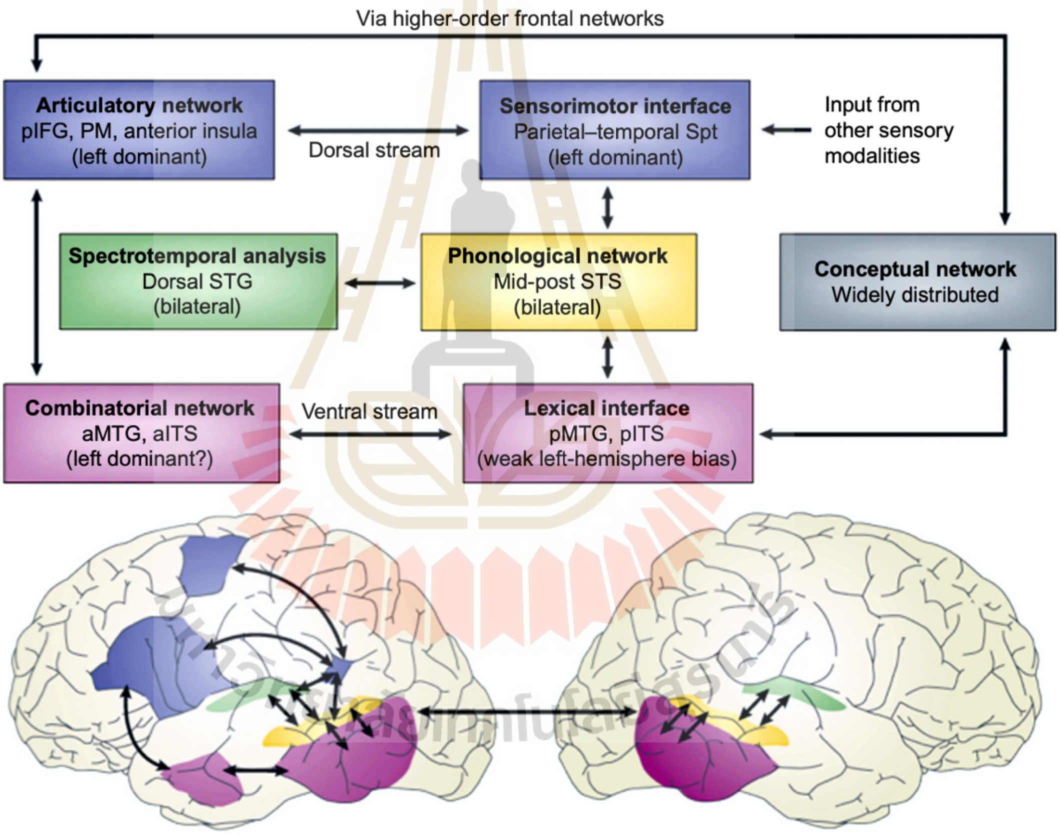


Figure 2.2 Hickok and Poeppel's (2007) dual-stream model of speech perception

Compared with Hickok and Poeppel's model, other models discuss the interaction between cortical areas and neural pathways in the perceptual processing of speech. Rauschecker and Scott (2009) propose a new model for active processing of speech, which is a forward mapping and an inverse mapping. In the pathway of forward mapping, the speech signal is decoded based on categories by the inferior frontal cortex, and sent to motor-articulatory movements in the premotor cortex, then transmitted to the inferior parietal lobe and the posterior superior temporal cortex as a reference copy. The inverse mapping is in a reverse direction of the same pathway, in which attentional/intentional shifts in the inferior parietal lobe impact context-dependent action plans in the prefrontal cortex and the premotor cortex, and these predictions are sent to the inferior frontal cortex and compared with the sensory input. The forward and inverse pathways in this model are dynamic and simultaneously active, in which the lack of invariance problem of speech perception may be resolved by the inferior frontal cortex or by the interaction between the inferior frontal cortex and the anterior superior temporal cortex (Rauschecker & Scott, 2009).

Friederici (2012) presents a model of speech/language processing with four pathways. The processing of language signals starts from the primary auditory cortex to the anterior STG through the ventral stream to the frontal cortex. The ventral back-projection from pars triangularis to the anterior STG and MTG accounts for the top-down semantic processing, and the dorsal back-projection from pars opercularis to posterior STG and STS is related to top-down processing of grammatical relations. The dorsal stream from the primary auditory cortex through posterior STG/STS to the premotor cortex explains auditory-motor mapping. In addition, the anterior and posterior regions of the temporal cortex are linked by the inferior and middle longitudinal fasciculi, which supports the flow travels from/to mid-MTG. This model adds complexity to the dual-stream model, which emphasizes syntactic and semantic processing at the sentence level but the lack of invariance problem remains.

Davis and Johnsrude (2007) put forward a model of top-down interactive mechanisms in auditory networks that account for active speech processing. Lexical representations of speech signals are interpreted in the inferior frontal cortex through the ventral stream and compared with echoic representations of new-incoming speech in the temporal cortex, which permits top-down projections from the inferior frontal cortex to modify/change the perception of the signal via the auditory pathway. In addition, motoric/somatotopic representations of speech are projected to the inferior frontal cortex in the dorsal stream and to the temporal areas for perceptual retune of the signal. Two parallel representations – echoic and motoric/somatotopic representations of auditory speech signals support top-down interactive mechanisms for speech/language perception, which is considered as an active process.

Although Davis and Johnsrude (2007) have developed a model of active processing in speech perception, this model, together with other above-mentioned models, only focuses on cortical connections. However, the auditory language signal is processed in the cochlea from the lower-level regions of the nervous system before the signal arrives in the primary auditory cortex (Huffman & Henson, 1990). To better explain the neural mechanisms for speech perception, the following part will introduce the subcortical bases of speech perception.

2.3.2.2 The subcortical bases of speech perception

After speech sounds enter ears, auditory processing begins from the cochlea, and the acoustic signal flows along the path from the cochlea nucleus, the superior olive, the inferior colliculus, and the medial geniculate to the primary auditory cortex (Heald et al., 2016). In fact, auditory processing has been well-established before the signal reaches the primary auditory cortex since the connections between these subcortical structures of the peripheral nervous system contain more descending projections than ascending projections (Heald et al., 2016; Huffman & Henson, 1990). More descending projections mean that the auditory cortex modulates acoustic processing more frequently at lower levels of the auditory system, adjusting/readjusting the processing from the auditory cortex to the medial geniculate and then to the cochlea nucleus (Mellott et al., 2014). Therefore, the subcortical structures play a significant role in speech perception and provide perceptual and adaptive models for speech/language processing.

The brainstem shows the obvious top-down influence of cortical processing on subcortical processes. EEG studies showed that the brainstem frequency-following response (FFR) (Krishnan, 2007), a phase-locked response to the fundamental frequency (F_0) of a stimulus, was found to indicate early encoding of the acoustic signal from the auditory system and top-down perceptual pre-processing mediated by descending projections from the cortex (Galbraith et al., 2003). Specifically, this experiment showed that the FFR was influenced by selective attention that attended tones elicited larger amplitudes of the FFR while smaller amplitudes were elicited by the ignored tones in detection tasks of dichotic tones. In addition, attentional allocation in auditory and visual modalities also impacted the FFR. As auditory and visual signals were presented to participants simultaneously, participants were asked to only focus on either the visual or the auditory signal. Results showed that the signal-to-noise ratio of the FFR decreased as attention was allocated in the visual stimulus, and increased when attention was shifted to the auditory signal (Galbraith et al., 2003). Though only tones were adopted in the study rather than speech, results identified that selective attention could modify the processing of auditory signals in the brainstem and then influence the processing in the cortex.

As to the FFR to speech stimuli, Krishnan (2002) used synthetic English back vowels to investigate top-down processes of linguistic categorization at the cortical level. Spectrum analysis of the FFR showed that obvious peaks emerged at the first formant and smaller peaks appeared at the harmonics between other formants. The prominent peak at the first formant indicated that information categorization began in the auditory brainstem, in which selective suppression enhanced the important cues of the signal and these cues were retuned to be strengthened and more salient. Besides, the interpretation of the signal's lexical presentation at the cortical level might also affect the way that the auditory brainstem processed the signal and transmitted it to higher-level pathways.

Apart from the FFR to vowel formants, lexical tones of Mandarin Chinese were also adopted in the FFR study. Krishnan et al. (2005) investigated Chinese speakers' and English speakers' FFRs to four Chinese tones. FFR findings indicated that Chinese native speakers had more robust pitch representations and smoother pitch tracking. To be specific, spectral data of FFR exhibited that Chinese speakers had stronger representations of the second formant (F_2) compared with English speakers across four lexical tones. This study supported that language experience possibly induced neural plasticity at the level of the brainstem, in which the speech input might be enhanced or primed by linguistic-relevant features before processed at the cortical level. Not only did language experience such as lexical tones inducing the differences in FFRs, but also musical training that induced earlier, larger FFRs, and better phase-locked responses to F_0 in both music signals and subjects' native language signals (Musacchia et al., 2007; Wong et al., 2007).

In the periphery, the cochlea is connected and functions with the vestibular organ for the perception of feeling and hearing speech (Asp, 2006). As the central part of the vestibular system, the cochlea plays a significant role in integrating and organizing multiple sensory input for perception from the brainstem to the cortex (Asp, 2006). After the recognition pattern of speech is established and stabilized, auditory processing turns into the dominant modality for speech perception and production, which is perceptually based on the tonotopic organization (i.e., the processing pattern of different frequencies) from the cochlea to the cortical level (Asp, 2006). However, as mentioned above, there are more descending projections from the cortex to the periphery than ascending projections in auditory speech processing. Auditory speech processing exhibits the effects of high-level cognitive processing on the subcortical system from the cochlea. The enhancing effect of auditory selective attention on spectral peaks of evoked otoacoustic emissions has been identified (Giard et al., 1994; Maison et al., 2001), and enhanced suppression of click-evoked otoacoustic emissions is produced by discrimination training (Maison et al., 2001).

Except for the function of cortical organizations in speech perception and production in different ways, neurological evidence also suggests that perceptual and adaptive processing of speech/language stimuli results from subcortical structures such as the brainstem and the cochlea. Descending projections from the primary auditory cortex to the periphery can be regarded as the context-driven top-down process. In turn, plasticity of subcortical structures induced by the individual's language experience (or even musical experience) influences language processing at the cortical level, which can be considered as the stimulus-driven bottom-up process. Both processes of context-driven top-down and stimulus-driven bottom-up together influence the processing of tones, vowels, speech, and other acoustic signals, reflecting an active cognitive process of speech/language (Heald et al., 2016). Tracing back to the problem of the lack of invariance in speech, the top-down processing at the subcortical level may offer the solution, which allows the auditory system to perceive and adapt to the variability of speaker and context. To make perceptual and adaptive processing of speech possible, subcortical structures play a crucial role in enhancing the prominent spectral cues of the incoming auditory signal as pre-processing before the processing of speech recognition at the cortical level (Heald et al., 2016). The reason the current study adopts the Verbotonal approach is that it emphasizes perceptual learning of language as a holistic and multi-sensory learning system and highlights language learning as an active cognitive process with both bottom-up and top-down processes. The detailed review of VTA will be in the following section 2.4.

2.3.3 The neural networks of syntactic and semantic processing

In terms of sentence-level processing beyond single word processing, the existing neuroimaging techniques such as fMRI, diffusion-weighted MRI (dMRI), and ERP allow us to map out the neural networks for syntactic and semantic processes by identifying neuron connections between brain regions. Although the localization of syntactic and semantic processing in sentence comprehension cannot be identified in one specific brain region, the definable and functional neural networks for syntactic and semantic processes can be formulated based on the neuroanatomical and neuroimaging evidence (Friederici, 2016). The neuroanatomical pathway model of language processing proposed by Friederici and Gierhan (2013) is displayed in Figure 2.3, in which two dorsal and two ventral pathways are identified in the left hemisphere.

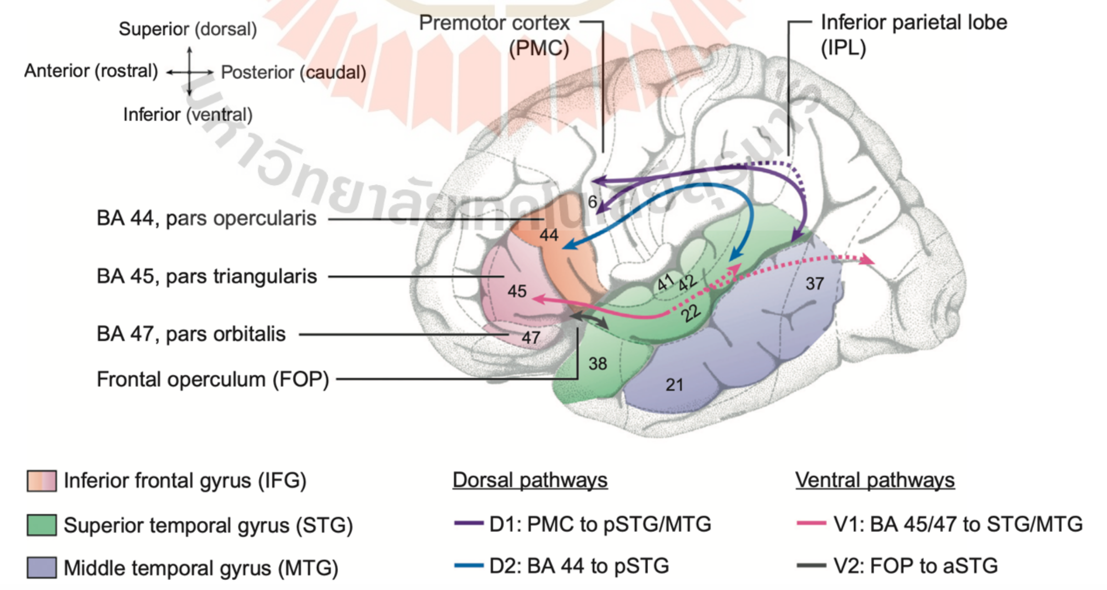


Figure 2.3 Friederici and Gierhan's (2013) neuroanatomical pathway model for language processing

In this figure, numbers are defined by the cytoarchitecture and presented according to Brodmann areas (BAs). The pathways are the connections between cerebral regions linked by white matter fiber bundles from fMRI and dMRI evidence (Friederici & Gierhan, 2013). The dorsal pathway D1 links the premotor cortex (PMC) to the posterior superior temporal gyrus (pSTG)/middle temporal gyrus (MTG), and the dorsal pathway D2 links BA 44 to the pSTG. The ventral pathway V1 links BA 45/47 to the superior temporal gyrus (STG)/MTG through the extreme capsule fiber system (ECFS)/longitudinal inferior fronto-occipital fasciculus (IFOF), and the ventral pathway V2 links the frontal operculum (FOP) to the anterior superior temporal gyrus (aSTG) via the uncinate fasciculus (UF). Among these four pathways, D1 is identified as a sensory-to-motor pathway established in the early life and related to language acquisition (Hickok & Poeppel, 2004, 2007), especially engaged in speech repetition according to fMRI-dMRI findings (Friederici & Gierhan, 2013; Saur et al., 2008), hence, D1 is not directly relevant to sentence comprehension. The other three pathways, namely D2, V1, and V2, are involved in sentence comprehension. D2 is responsible for processing the complicated hierarchical structures based on the existing D1 that children's output can be modified/changed by the language signals they have perceived during early acquisition (Friederici, 2016). This part reviews how these three pathways D2, V1, and V2 work for language processing, including the processes of auditory perception, recognition of word, structure building, and comprehension within milliseconds, which is outlined as an active process with top-down and bottom-up processes, especially in syntactic and semantic processing (Friederici, 2016).

2.3.3.1 The neural networks of syntactic processing

The neural networks for syntactic processing are identified to be a ventral pathway V2 and a dorsal pathway D2. Syntactic processing begins with building structures, manipulated by previously acquired grammatical knowledge of adjacent dependencies in the language being processed, e.g., determiner and prepositional phrases. Adults or proficient language users can operate the structure processing more automatically by getting the FOP and the aSTG involved (Friederici et al., 2003). But for second language learners or less proficient users, this process may be less automatized and BA 44 is involved as well (Brauer & Friederici, 2007; Rüschemeyer et al., 2005). In the ventral pathway V2, the aSTG develops templates of local structures such as the prepositional phrase that new-incoming information can be mapped on after receiving information from the primary auditory cortex. As the information of the new-incoming phrase meets the local structures in the aSTG, the structures in the aSTG become available (Bornkessel & Schlesewsky, 2006). Then, information is sent to the FOP from the aSTG through the uncinate fascicle. V2 is considered as a pathway of local syntactic computation, which is fundamental in syntactic processing (Friederici, 2016).

As to syntactically complicated sentences, studies show that BA 44 is activated when processing hierarchical dependencies such as sentences with different word-order (Friederici et al., 2006; Grewe et al., 2005; Meyer et al., 2012; Röder et al., 2002), sentences with different embedded clauses (Makuuchi et al., 2009), sentences with "pseudowords + grammatical particles" versus morphosyntactic information alone (Ohta et al., 2013), and the processing of these sentences interacted with working memory. From these studies, the syntactic hierarchy processing leads to significant activation of BA 44, demonstrating a process of global computations in syntactic processing. Another region involved in syntactic complexity is the pSTG/STS (e.g., Kinno et al., 2008; Newman et al., 2010; Santi & Grodzinsky, 2010). The pSTG/STS is activated when the verb is semantically inconsistent with its argument (Friederici et al., 2003; Obleser & Kotz, 2010), in addition, the pSTG/STS is relevant to verb class and argument order (Bornkessel et al., 2005). Research shows that the pSTG/STS is related to syntactic processing and semantic verb-argument processing (Grodzinsky & Friederici, 2006). Then the dorsal pathway D2 connecting BA 44 to the pSTG/STS is associated with the processing of complex sentences, within which BA 44 is responsible for hierarchical structures and the pSTG/STS integrates semantic and syntactic processing of complicated sentences (Friederici, 2016).

MRI studies conducted with patients confirm the ventral and dorsal pathways in syntactic processing. Griffiths et al. (2013) stated that patients with lesions in the areas of the ventral and dorsal pathways in the left hemisphere had deficits in syntactic processing compared with the healthy controls. This study identified a causal role of the left inferior frontal cortex (IFC) and the posterior middle temporal cortex (pMTC) in syntactic processing. Another study investigating patients with progressive aphasia by Wilson et al. (2011) showed that degeneration of the fiber tract in the dorsal pathway that links Broca's area (BA 44/45) with the temporal cortex resulted in a deficiency of syntactically-complex sentence processing, which supported the understanding of the dorsal pathway D2 in processing complex sentences. As to the ventral pathway V2 concerning syntactic processing, a study showed that stroke patients with lesions only in the left anterior temporal regions had deficits in syntactic comprehension (Dronkers et al., 2004). Also in stroke patients, the correlation between white matter integrity and patients' behavioral performance was identified that both dorsal and ventral pathways were involved in syntactic processing (Rolheiser et al., 2011).

The ERP component – P600, a positive-going waveform, is a slow late component relating to syntactic manipulation in language comprehension and production (Swaab et al., 2012). It demonstrates an onset around 500 ms, peaks at about 600 ms after the stimulus presentation, and lasts for several hundred milliseconds (Friederici, 2002; Gouvea et al., 2010). Both auditory and visual stimulus modalities can elicit the P600 component with widespread distribution over the scalp and mostly over the centro-parietal areas (Hagoort, 2008). Studies show that a P600 effect is elicited by syntactic violations at the sentence level in adults and children, in which grammatical errors, syntactic ambiguity, and other syntactic anomalies are included (Daltrozzo & Conway, 2014; Silva-Pereyra et al., 2007; Swaab et al., 2012).

An ERP study conducted by Christiansen et al. (2012) used a sequential learning task with complex structured sequences and natural language stimuli to investigate the time course and distribution of the P600 component. Results show that structural incongruities (syntactic violations) elicited P600 effects in both artificial grammars and natural language with similar topographical distributions, indicating that identical or similar underlying neural correlates are involved in difficult or ungrammatical syntactic processing of non-linguistic and natural language stimuli (Christiansen et al., 2012).

As to students' syntactic processing of the second language, Tokowicz and MacWhinney (2005) used a grammar judgment task to investigate native English speakers who learn Spanish as a second language. Three syntactic conditions were used as stimuli: tense-marking (similar in L1 and L2), determiner number agreement (different in L1 and L2), and determiner gender agreement (only in L2). P600 effects were elicited after the onset of the critical word (violation or matched control) in the sentence, which supported that grammatical violations elicited P600 effects in both L1 and L2 speakers. In addition, results also revealed that students were sensitive to violations of structures in L2 that were similar to L1, but less sensitive to violations of the structures that were different from L1. Further, a strong grammatical effect illustrated by the P600 effect was detected in the unique structure of L2, indicating that students were sensitive to these violations implicitly. This study offered evidence that students were sensitive and capable of processing L2 syntax at the early stage of language learning, in the meantime, syntactic processing was also influenced by the similarity between L1 and L2.

2.3.3.2 The neural networks of semantic processing

One of the two ventral pathways V1 linking BA 45/47 to the STG/MTG through the ECFS/IFOF is identified to be engaged in semantic processing in fMRI and dMRI studies (Saur et al., 2008; Wilson et al., 2011). In this pathway, BA 45 and 47 are considered to be involved in semantic processing (Patterson et al., 2007), which are activated as subjects are asked to do the tasks of semantic relatedness and congruence judgment (e.g., Dapretto & Bookheimer, 1999; Kuperberg et al., 2000; Newman et al., 2010). Additionally, the temporal, especially the anterior temporal lobe, plays an essential role in lexical-semantic processing even at the word level. Studies of patients with semantic dementia, a neurodegenerative syndrome in the anterior temporal lobes bilaterally, show semantic degradation/deficits in single words across different modalities and conceptual knowledge (e.g., Lambon Ralph & Patterson, 2008; Patterson et al., 2007).

Although it is clear that brain regions of the anterior temporal lobe, the IFC, and the posterior temporoparietal areas are activated in semantic processing of sentences, the model and the interaction between these areas during semantic processing of sentences are still not clear, and fMRI studies on semantic processing of sentences are scant. Newman et al. (2010) investigated semantic relatedness of the nouns in the sentence influencing sentential processing. Results showed stronger semantic-related activation in the IFG, especially the area of BA 45/47. In another study by Obleser and Kotz (2010), semantic expectancy (i.e., cloze probability) in different sentences was examined. fMRI results showed that activation of the STS extended anteriorly and posteriorly in low-cloze sentences, and converged to the mid STG/STS area in more predictable sentences, i.e., high-cloze context. In addition, activation of the left IFG increased as intelligibility increased, and activation of the left inferior parietal cortex (IPC) was related to successful speech comprehension due to either increased stimulus quality or facilitation of semantic processing. This study indicates that semantic expectancy can regulate speech processing by allocating fewer neural resources to highly predictable sentences (Obleser & Kotz, 2010). Although studies have confirmed the importance of the ventral pathways in semantic processing, it is still unclear how the anterior temporal, the inferior frontal, and the temporal-parietal areas interact with each other in language comprehension (Friederici, 2016).

ERP studies have detected the component N400, a negative-going wave reaching the peak around 400 ms after the stimulus onset, as an indicator of language processing (Kutas & Federmeier, 2011), specifically correlated with semantic manipulation (Morgan-Short & Tanner, 2014). The N400 is an evoked potential of the healthy brain's response to language signals and other meaningful (or potentially meaningful) signals in both visual and auditory modalities. The topographic distribution of the N400 is typically over centro-parietal regions (Kutas & Federmeier, 2011; Swaab et al., 2012). Prior research showed that the N400 component could be identified in different experimental paradigms ranging from simple sequential designs such as oddballs to complex sequential stimuli such as artificial and natural grammars, which might be helpful to better understand cognitive mechanisms for language processing (Daltrozzo & Conway, 2014). Kutas and Federmeier (2000) revealed that the nature and time course of semantic memory were related to the N400 component during language comprehension. Results showed that the N400 varied systematically while processing semantic information, and semantic memory was inherently relevant to sentence processing. Additionally, the left hemisphere was assumed to organize semantic memory for pre-processing the meaning of forthcoming words whether this pre-processing would fail or not. The N400 amplitude changed systematically with the processing of meaningful and potentially meaningful stimuli, demonstrating the ongoing semantic processing (Kutas & Federmeier, 2000; Kutas & Federmeier, 2011).

A longitudinal ERP study on novice L2 learners may better explain language learners' lexical and semantic processing and development. McLaughlin et al. (2004) recruited native English speakers who had been enrolled in an introductory French course as participants. After 14 hours, 63 hours, and 138 hours of instruction, ERP experiments were conducted while the participants are performing French lexical decision tasks, in which stimuli were comprised of three conditions: semantically-related word pair, semantically-unrelated word pair, and word-pseudoword pair. In the first ERP session after 14-hour's learning, there was no significant difference between semantically-related and semantically-unrelated word pairs, but a significant N400 amplitude appeared in the word-pseudoword pair. After 138 hours' instruction, the largest N400 effect was shown to pseudowords, the intermediate effect was caused by semantically-unrelated word pairs, and the smallest effect to semantically-related word pairs, which were similar to native speakers' N400 effects produced by these conditions of word pairs. This longitudinal research indicated that the N400 component was highly sensitive to lexical and semantic processing, moreover, students' brain could distinguish word and pseudoword soon after the instruction but more instruction time did lead L2 learners to incorporate lexical information into semantic on-line processing system (McLaughlin et al., 2004).

Besides, a positive-going deflection peaking around 400 ms (P400) was detected to be related to semantic processing. Liu et al. (2009) identified that the P400 was sensitive to the semantic process of mismatched multisensory information by using the stimuli with visual and auditory modalities. The results suggested that the P400 was an index of semantic processing and reflected the cognitive integration process of the multisensory signals.

The results of the P400 component were confirmed by other studies. Researchers found that the N400 semantic effect could be separated into two portions – the negative-going wave N400 at the Pz site (N400Pz) and the positive-going deflection P400 at the Cz site (P400Cz) by using the visual sentences as stimuli. The N400Pz was related to semantic processing in the area of the bilateral anterior medial temporal lobe, but the P400Cz reflected a sequential integration and expectancy system in the medial parietal area (Dien, 2009; Dien et al., 2010). Dien et al. (2010) claimed that the N400Pz reflected lexical access and the P400Cz was related to the manipulation of sequential representations that could be taken as a kind of context. Since the stimuli were sentences, the sentences were a sequential set of stimuli and created such context in which the words of the sentence were integrated. The P400Cz could be regarded as an index of integrating the sequential representations. In addition, the P400Cz reflected an expectancy process for sequential probabilities (Dien et al., 2010). Both the N400Pz and the P400Cz components were sensitive to semantic processing (Dien et al., 2010).

Another ERP component – the N600 (a negative-going waveform peaking around 600 ms) was also identified to be related to semantic processing. The N400 mentioned above was the earlier process that was dependent on the previously-established meaningful semantic representations (not necessarily lexical), but the N600 reflected a later process of the explicit interpretation of stimulus semantics (Cummings et al., 2006). Both of the fronto-central negative waves, the N400 and the N600, were sensitive and larger for semantic tasks, which were generated in the auditory cortex and frontal areas (Shahin et al., 2006). In addition, the N600 component contributed to the required response (Cummings et al., 2006) or decision-making (Shahin et al., 2006) during the experiment, such as responding/deciding whether the stimulus was a standard or a deviant. The findings showed that the N600 was indicative of a stimulus-general process that maintained task demands and monitored responses (Halgren et al., 1994). The late fronto-central negative wave suggested the general/non-specific cognitive processes relating to working memory and attention (Cummings et al., 2006; Itoh et al., 2005; King & Kutas, 1995; Koelsch et al., 2003).

2.3.4 The neural bases of linguistic prosody processing

Apart from syntactic and semantic processing, another important factor influencing speech perception and production is linguistic prosody or suprasegmental features of speech. Linguistic prosody (suprasegmental features), including pitch, loudness, tempo, and voice quality, can convey lexical meanings (the change of stress leads to different lexical meanings, e.g., *green card* and *green card*), emotional information (e.g., expressing happy or angry feeling by changing vocal pitch), and discourse/context information (e.g., intonation changes impact syntactic and semantic interpretations of the sentences). Therefore, prosody is an indispensable factor in both linguistic and non-linguistic processing (Paulmann, 2016). But the function of linguistic prosody in dynamic language comprehension has always been undervalued in the studies of neural mechanisms for language processing.

Clinical evidence from studies of patients with brain lesions shows there is a lack of consensus regarding whether the right hemisphere (RH) or the left hemisphere (LH) processes linguistic prosody. A meta-analysis on activation likelihood estimation (ALE) conducted by Witteman et al. (2011) reported that lesions in the LH or RH led to similar detrimental impacts on the perception of linguistic prosody. In addition, studies revealed that both cortical and subcortical structures were engaged in processing linguistic prosody (Brådvik et al., 1991; Cancelliere & Kertesz, 1990; Rymarczyk & Grabowska, 2007; Starkstein et al., 1994). Meanwhile, it lacks clear evidence to support the hypothesis that prosody is mainly mediated by subcortical structures (Paulmann, 2016). Therefore, it seems that there is insufficient evidence for the hypotheses of the hemispheric pattern and cortical-subcortical pattern for processing linguistic prosody from clinical trials alone.

A neuroimaging study of positron emission tomography (PET) conducted by Gandour et al. (2000) compared pitch processing in Chinese, Thai, and English native speakers, in which Chinese and Thai were tonal languages that pitch changes led to distinctive lexical meanings, and English was a non-tonal language. Five participants from each language group were asked to perform three discrimination tasks: pitch patterns of Thai lexical tones in the tonal condition, pitch patterns of nonspeech signals in the pitch condition, and syllable-initial consonants in the consonant condition. Results showed that only the Thai group had significant activation in the left FOP comparing the tonal condition with the pitch condition, indicating that suprasegmental and segmental units were processed in Broca’s area for the Thai group. But in non-speech pitch discrimination, Thai participants did not show the same activation in LH, indicating that the lateralization pattern of prosodic processing was influenced by linguistic relevance/knowledge. This study supported that prosodic processing was manipulated by a top-down process from the higher cortical level based on the linguistic features/functions of the specific language. Later, Gandour et al. (2004) used fMRI to confirm the effect of hemispheric laterality in the processing of tone pitch. Chinese and English speakers received fMRI scanning while discriminating four Chinese tones presented in one- and three-syllable pairs. Findings showed that Chinese subjects had left-lateralized activation when processing the signals, especially in the inferior parietal, the posterior superior temporal, the anterior temporal, and frontopolar areas, but not the English speakers. Moreover, activation of the right STS and the right middle frontal gyrus (MFG) increased in both English and Chinese participants. This study identified that the pattern of lateralization during language processing resulted from language experience and the internal prosodic representation was established based on the input of auditory signals. The perception of linguistic prosody was fundamentally manipulated by the RH; but as processing language signals was more complicated than the auditory analysis of tones or pitch, the LH got more involved, which showed the task-dependent characteristic. These two studies suggested that language experience of tones could impact activation patterns of the cerebral regions and further influenced language processing from acoustic representations to internal representations of suprasegmentals in the language signals. Thus, speakers without tonal language experience did not have higher-order prosodic representations. Therefore, Gandour et al. (2004) claimed that the left-lateralized activation was related to higher-order processing for prosody, and the right-lateralized activation was associated with lower-order processing for acoustic information.

In addition, another study by Wong et al. (2012) recruited two groups of subjects with congenital amusia who had difficulty processing music (pitch and rhythm). For comparison, one group was native speakers of a tonal language (Cantonese) and the other was non-tonal languages (Canadian French and English). Results showed that tonal language speakers had enhanced pitch perception ability in an identification test compared to speakers of non-tonal languages. This study further supported that tonal language experience tuned the cortex for different processing mechanisms of prosodic information and language signals.

In order to yield the neural correlates of lexical tone perception in the tonal language – Chinese, Kwok et al. (2016) did an fMRI study with strongly right-handed native Mandarin Chinese speakers. Participants were auditorily presented with two syllables of a bi-syllabic Chinese word and were asked to decide whether the two syllables had the same tone in a tone perception task, during which participants received fMRI scanning. The findings indicated that the left IFG, the right MTG, and the bilateral STG were activated while processing pitch in linguistic signals. The authors elucidated that the left STG was responsible for basic acoustic analysis of auditory stimuli, but the right MTG and the left inferior frontal area were engaged in tonal and semantic processing of Chinese language stimuli.

As to the prosodic processing of sentences, neuroimaging research finds that the RH is dominant in the processing of sentential prosody. By using fMRI, Meyer et al. (2002) investigated the cerebral regions involved in slow prosodic processing of sentences. Three types of stimuli were adopted: normal speech, syntactic speech (grammatically correct pseudo sentences), and prosodic speech (filtered sentences without segmental or lexical information). All stimuli induced significant blood oxygenation signals in the supratemporal areas of both hemispheres. Stronger activation in the Heschl gyrus of the LH appeared during the processing of normal speech and syntactic speech, indicating that the left superior temporal area was activated while lexical-semantic and syntactic processing at the sentence level. But the right superior temporal area, particularly the planum temporale, was more activated by speech prosody. The fronto-opercular cortices were activated by syntactic speech and became more activated by prosodic speech, but this area did not obviously engage in the processing of normal speech. This study confirmed that the RH was more related to the processing of speech prosody than the LH.

Similar to this research, Plante et al. (2002) explored the processing mechanism of sentential prosody by using fMRI. Low-pass filtered and unfiltered sentences were used as stimuli. Twelve adult participants were asked to listen to the stimuli passively (just listening) and listen to the signals actively (remember and recognize information) while scanning. Comparing brain activation patterns induced by prosodic stimuli (filtered sentences) with normal sentences (unfiltered signals) in the passive listening condition, filtered signals induced significant activation in the bilateral STG. As to activation patterns in the active listening task, both filtered and unfiltered stimuli induced bilateral activation in the frontal lobes but filtered signals induced stronger right-lateralized activation than unfiltered signals. This study revealed that the processing of sentential prosody led to stronger temporal activation in the passive listening condition, while memory demands were related to different lateralization patterns of frontal activation induced by filtered and unfiltered stimuli, indicating that the interaction between memory demands and sentential/prosodic processing induced different cerebral activation patterns.

In another fMRI study conducted by Tracy et al. (2011), two paradigms were adopted: spoken sentences (variations in the pitch change of the internal phrases) and tone-sequence phrases with pitch changes. Results demonstrated that pitch processing in tone-sequence signals resulted in significant activation in the right frontal and temporal cortices. But left-lateralized activation was significantly greater in pitch change from lexical stimuli, including the cingulate gyrus, MTG, and STG. This study indicates that pitch processing is task-dependent according to the complexity of the stimulus.

As for ERP studies on speech prosodic processing, Li et al. (2011) used Chinese question-answer dialogues as stimuli and investigated congruent and incongruent question-answer dialogue pairs with violations in prosodic prominence, boundary, and both. This study reported a late negative-going waveform from 270 to 510 ms over the fronto-central areas in response to manipulations of prosodic prominence. And a later negative ERP effect occurring from 270 to 660 ms showed brain responses to violations of the prosodic boundary. These findings indicated there was an ERP component that was prosody-related. Findings from Böcker et al. (1999) also supported this prosody-related component. The N325 was detected in extracting metrical stress of bisyllabic words, which was elicited in both discrimination (active listening) and listening (just listening, passive listening) tasks. The time course of this component 325 ms was similar to the time window detected by Li et al. (2011), which indicated a prosody-relevant ERP component.

In addition, ERP evidence also showed that prosodic processing was task-dependent. The P800, a left temporoparietal positive component, was identified by Astésano et al. (2004). The prosodic mismatch, elicited by the sentences beginning with statements and ending with a question, and vice versa, resulted in the P800 effect when participants were asked to distinguish whether the sentences were prosodically congruent in two attention conditions. This study indicated that the P800 was elicited by prosodic mismatch, moreover, prosodic processing interacted with semantic processing since larger prosodic mismatch occurred in the semantically incongruent condition than in the semantically congruent condition.

ERP studies on speech prosody supported that the prosodic processing of sentences interacted with language processing regarding semantic (e.g., Paulmann et al., 2012) and syntactic processing (e.g., Steinhauer et al., 1999), which were together involved in on-line language comprehension. Since prosody-related ERP effects showed different time courses and distributions in different conditions, these variations reflected temporal dynamics that the processing mechanisms for linguistic prosody were multifaceted, task-dependent, and language-experience-based processes (Paulmann, 2016). As a result, the precise processing of speech prosody might recruit more cerebral regions/structures bilaterally.

Although previous studies have identified the neural bases of auditory language processing for both L1 users and foreign language learners measured by neuroscience techniques such as ERP and fMRI, these findings only focus on neural processing mechanisms of syntactic, semantic representations, tones, or prosody, in addition, comparisons are made between conditions of one language or speakers of two languages. There is scant research emphasizing the seeking of an optimal language signal through the brain activity in response to the alternative signals, especially by changing the physical characteristics of the auditory stimulus. This research attempts to fill the gap to explore an optimal auditory language input signal for Chinese EFL students by revealing their neural processing mechanisms and opinions of different forms of signals via a combined ERP and fMRI experiment and semi-structured interviews.

In sum, speech/language processing as an active cognitive process requires an adaptable and flexible processing mechanism to cope with the variability of speech (e.g., talker and contextual variability). The establishment of this mechanism is based on perception building of a language. The reason is that in order to recognize and understand language signals, one sieves information from previously acquired acoustic-phonetic representations and redirects attention to information cued by perception. Therefore, building the perception of a language is of great significance in learning the language, as this is the first and foremost step for later processing of language stimuli. But how to build students' perception of language and what kind of language signal is optimal to build their perception? These questions will be answered and demonstrated in the next section 2.4 introducing the Verbotonal approach because this approach regards the perception of language as an essential role in learning a language and provides an optimal language input signal for building language perception and facilitating language learning. This section offers a theoretical framework based on current evidence obtained from neuroimaging techniques, which indicates that speech/language processing including syntactic, semantic, and prosodic processing is multifaceted and influenced by the factors of the stimulus type (linguistic/non-linguistic signal), listening condition (active/passive listening), memory demands (high/low), and language experience (tonal/non-tonal language background). In turn, this dynamic neural mechanism of speech processing leads to different hemispheric patterns and cortical-subcortical processing patterns that influence the processing of the above factors. Therefore, speech perception is an active process that dynamically and constantly structures/re-structures students' neural networks (i.e., neuroplasticity) for language processing through both context-driven top-down (cortical-to-subcortical) and stimulus-driven bottom-up (subcortical-to-cortical) processes.

2.4 The Verbotonal approach

In this section, an introduction to the Verbotonal approach (VTA) is proposed first. Then the implementations of VTA in foreign language teaching and learning are presented.

2.4.1 General introduction to the Verbotonal approach

The Verbotonal approach (VTA) was initiated by Dr. Petar Guberina in the 1950s, who was interested in speech perception and developed this method of rehabilitation for people with communication problems (Guberina & Asp, 1981; Renard & Van Vlasselaer, 1976). Fundamentally, this approach takes language/speech development as a process of "meaning-making" (Guberina & Asp, 1981). The "meaning" of speech is presented not only by linguistic expressions but by auditory and visual information including the rhythm, intonation, loudness, tempo, and gestures of the speaker (Guberina & Asp, 1981). The speaker is a producer and a perceiver of speech, more importantly, the speaker's production of speech reflects how he/she perceives speech (Guberina & Asp, 1981, 2013). But because of speakers' diverse operational histories (Lian, 2004; Lian & Pineda, 2014), the perceptions of a sound vary from individual to individual. As to second language learning, students' perceptions of the sounds in the target language are influenced by the phonological systems of their native languages, described as "phonological sieve," which is a mechanism to keep the sounds that are recognizable and to reject those that are not (Lian & Lian, 1997). Learning a foreign language is analogous to rehabilitating and developing the native language for hearing-impaired subjects. Foreign language learners have been trained and attuned to perceive and produce sounds of their native languages well, but they may be insensitive, or "deaf," to the sounds of the foreign language. Therefore, if the speaker's perception of speech changes, his/her speech production will change, in other words, the correction of the speaker's speech is based on the correction of speech perception (Guberina & Asp, 1981, 2013).

How to raise students' awareness of the new language and change their perceptions of the foreign sounds are essentials for foreign language learning and teaching. VTA provides a solution. Guberina emphasizes that rhythm and intonation, perceived through low frequencies, convey meanings and play key roles in speech perception and production (Guberina & Asp, 1981, 2013; Kim & Asp, 2002). The reason is that the cochlea and the vestibular organ are sensitive to the changes of pitch, rhythm, and intonation via the low-frequency pattern of speech. The development of the auditory organs begins from the feeling of the speech rhythm with low-frequency sounds in the uterus, and the proprioceptive memory develops along with feeling the speech rhythms after birth, which provides the basis of the auditory-memory development later on (Asp, 2006). From clinical practice, Guberina concluded that the brain would function best if it received the low-frequency auditory signals that the ears were most sensitive to, and these optimal stimuli would enhance the brain function with time and training so that the subjects could prepare to deal with more difficult tasks or less favorable sounds (Guberina & Asp, 1981, 2013). The optimal stimuli mentioned in Guberina's study are the low-frequency signals that are low-pass filtered. The low-pass filtered stimulus preserves the fundamental frequency (F_0) of the sentences, together with prosodic features, i.e., stress, rhythm, loudness, and intonation, while the higher frequencies that help to define words are removed (Lian & Sussex, 2018). The low-pass filtered stimuli stimulate subjects' multi-sensory experience, i.e., vestibular, tactile, and proprioceptive senses, but lighten the processing load for lexical and meaning processing and release more attentional resources for other processes (de Jong et al., 2009; Lian & Sussex, 2018). In this way, VTA can restructure or reorganize the way the hearing-impaired perceive language signals effectively and raise their awareness of language; therefore, it is believed that it works for foreign language learners as well.

As mentioned above, since the brain reorganizes neural connections constantly, i.e., neuroplasticity, VTA considers it as a key to language learning/development and therapy to the hearing impaired, aiming to seek an optimal stimulation to maximize neuroplasticity in the individual brain (Guberina & Asp, 1981, 2013). For rehabilitation of impaired hearing, similar to learning a foreign language, VTA views speech perception and production as a multi-sensory and whole-body experience via rewiring and restructuring the subject's neural networks (Asp et al., 2012; Guberina & Asp, 1981). Because neuroplasticity occurs as the vestibular system develops consistently with body movement and vocalization, which enhances speech perception through proprioceptive memory and the vestibular end-organ (Asp, 2006). To achieve the ultimate goal of neuroplasticity, VTA capitalizes on the perception of low-frequency stimuli to restructure the brain's neural connectivity, to which human ears are most sensitive (Asp et al., 2012; Guberina & Asp, 1981; Kim & Asp, 2002). Thus, the low-frequency stimulation that consists of the linguistic prosody is stimulating and awareness-raising; further, it facilitates the rewiring of neural connectivity in the brains of subjects with hearing loss and foreign language learners (Guberina & Asp, 2013; Lian & Sussex, 2018).

2.4.2 The Verbotonal approach to foreign language teaching and learning

Since the advent of VTA, it has been mainly applied in the rehabilitation of subjects with hearing impairment, treatment of language/speech disorders, and foreign language teaching. VTA used in foreign language teaching is an adaptation of the original Verbotonal principles, called the Structuro-Global-Audio-Visual (SGAV) methodology (Renard & Van Vlasselaer, 1976).

VTA is considered an effective approach to phonetic correction from previous studies. As early as 1980, Andrew Lian introduced VTA to teach French pronunciation in his book *Intonation Patterns of French (Teacher's Book).* Lian (1980) created optimal situations to reinforce students' perceptions and develop their articulatory abilities by adopting sensitization and reinforcement sessions during teaching. In the sensitization session, it included several steps such as *"relaxation phase," "audition of filtered sentences," "the importance of movement and gesture," "humming along," "interpretation of the intonation patterns," "mouthing the words," "repetition on a background of filtered patterns," "humming the patterns," "adding the words to them,"* etc. And the reinforcement session comprised *"self-testing"* and *"sensitization and intensive practice"* (Lian, 1980, p. 3). These procedures were designed under the principles of VTA by highlighting the prosodic features of French and releasing the processing load for vowels and consonants, which prioritized the students' multi-sensory and whole-body experience to shape their perceptions of French.

Studies showed that VTA was effective in teaching pronunciation and phonetic correction for students learning French as a foreign language. Ludovic (2010) had a strong point of view that verbotonalism was an effective and comprehensive method in correcting French rhythm and intonation, phonemes, and vowels in Japanese students. In addition, Alazard et al. (2011) compared VTA with the traditional articulatory approach in phonetic correction by establishing a multiphonia corpus with audio-video classroom recordings. Results showed that repetition of prosodic patterns with gestures facilitated phoneme perception and re-production, which helped students access the lexicon and the segment of speech flow in French effectively. Moreover, this study indicated that VTA had an advantage over the traditional articulatory approach in teaching pronunciation of other foreign languages, not only French.

As the current study is focusing on VTA-based EFL in the Chinese context and discussing VTA-based L1 (Chinese) and L2 (English) processing mechanisms, the following part reviews literature on teaching and learning of Chinese and English based on VTA. In terms of teaching Chinese as a foreign language under the principles of VTA, Hu and Uno (2005) taught Chinese tones to Japanese beginners. Thirty-five university students in a basic Chinese class were involved in the research. After seven-week Verbotonal training of the tones of monosyllabic and disyllabic words, students distinguished four Chinese tones successfully and their performances in the tasks of dictating and pronouncing the tones were improved significantly.

Zhang (2006) applied VTA that she called “the Somatically Enhanced Approach (SEA)” to teach Australian students Mandarin Chinese prosody. In her research, twenty-two students were grouped into experimental and control groups. Students in the experimental group received the multi-sensory communicative approach, SEA based on VTA, while the control group was taught by the communicative approach, a non-multi-sensory method. The teaching procedures contained *“the sensitization session”* and *“repetition exercises”* (p. 150). At the beginning stage of learning, students in both experimental and control groups encountered difficulties with Chinese tones in similar order and made similar errors due to interference of their first language. But after the implementation, results indicated that SEA, a VTA-based approach, was effective in teaching Mandarin Chinese tones by lowering interference of students’ native language.

As for teaching English as a foreign language by adopting VTA, several studies have been conducted in the Chinese context. He (2014) integrated a CALL-based (Computer-Assisted Language Learning) autonomous learning environment with VTA, named "CALL-VT" system, to examine the effectiveness of VTA in teaching pronunciation to Chinese university EFL students. Ninety-five first-year English majors of two intact classes were involved in her study and selected as one experimental group and one control group randomly. The pedagogical intervention comprised *"classroom and out-of-classroom activities"* (p. 3). Classroom activities aimed to make students more sensitive to English sounds by listening to low-frequency stimuli based on VTA, and out-of-classroom activities were designed to develop students' autonomous learning of pronunciation. After the intervention, the experimental group outperformed the control group in the tests of phonemes, word-reading, passage-reading, and oral interview. In addition, the experimental group did better in pronunciation, comprehensibility, and fluency according to the feedback of English native raters.

Yang (2016) developed a VTA-based intervention to investigate the effectiveness of VTA in improving Chinese primary EFL students' speaking skills and enhancing their phonological working memory. A quasi-experiment was conducted with two intact classes of eighty students in Grade 3, which were randomly taken as the experimental group taught by VTA, and the control group received the traditional teaching method for speaking skills. The pedagogical procedures in her study were *"in-class sensitization and out-of-class reinforcement sessions"* (p. 88), which lasted eighteen weeks. Findings showed that the experimental group outperformed the control group in all aspects, which included overall speaking proficiency and the individual tests of word-reading, sentence-reading, singing, and oral interview. Moreover, the subskills of vocabulary, grammar, pronunciation, fluency, and comprehensibility in the experimental group were improved significantly. Simultaneously, the capacity of phonological working memory in the experimental group was also improved substantially, since participants had considerable progress in repeating lengthy non-words and non-words with low wordlikeness.

As the goal of VTA is to seek an optimal stimulus for individuals' language development, Wen et al. (2020) refined the original model of VTA and had a trial to test alternative optimal frequency bands for Chinese university EFL students' perception and production of English, especially correcting pronunciation. By comparing the frequency bands of three vowel contrasts, such as /ɪ/ and /iː/, he discovered that optimal frequencies varied from student to student due to their processing of language signals. And a discontinuous frequency band was more effective in correcting pronunciation than a continuous frequency band. Additionally, a narrow frequency band (e.g., 1/3 of an octave) was more effective for reshaping speech perception and correcting production than a full octave. Then the multiple optimal frequencies were used to diagnose each student's optimal frequency for each of the vowels being studied. Each participant in the experimental group used his/her own optimal frequency and did listen-repeat exercises with no teaching of any kind, and the control group did the same listen-repeat exercises without any kind of teaching. The only difference was that the control group was not exposed to filtered stimulation. Results indicated that the experimental group outperformed the control group in both perception and production, including increased intelligibility. His research was innovative to identify each student's optimal frequency for perception, production, and correction of English pronunciation; furthermore, the low-frequency, the prosodic signal was examined to be an optimal stimulation for individual's pronunciation correction.

To understand how students perceive and produce speech sounds and how they can improve (foreign) language learning, VTA provides an answer that the low-frequency stimulation can effectively raise students' awareness and change their perceptions of language signals since human ears are most sensitive to low frequencies (Guberina & Asp, 2013; Kim & Asp, 2002). The pedagogical intervention of low-pass filtered stimuli with body movement fosters a holistic and multi-sensory model for foreign language learners. Literature shows that learners of French/Chinese/English as a foreign language improve their pronunciation significantly, specifically, French phoneme (Alazard et al., 2011), Chinese tones (Hu & Uno, 2005; Zhang, 2006, 2012), English speaking skills in general (He, 2014; He et al., 2015; Yang, 2016; Yang et al., 2017), listening (Luu, 2021), pronunciation correction (Wen et al., 2020), as well as phonological working memory (Yang, 2016; Yang et al., 2017). Previous studies provide empirical evidence that VTA is an optimal model for language perception and production, however, research on the neural processing mechanisms for the low-pass filtered stimulation in Chinese EFL students is still scarce. Hence, the present study aims to unveil students' brain activity in response to filtered and unfiltered stimuli by using a combined ERP and fMRI experiment. In previous studies, low-pass filtered stimuli are presented binaurally to the students, which does not seem to be an optimal stimulation for each student's brain. In fact, human ears have a general right ear advantage for verbal signals and a left ear advantage for melody and speech intonation in healthy individuals (Gandour et al., 2004; Meyer et al., 2002; Sammler et al., 2015; Tervaniemi & Hugdahl, 2003; Vigneau et al., 2006). Therefore, the current research aims to explore an optimal physical language stimulation by adopting low-pass filtered and unfiltered signals under dichotic and diotic listening conditions, which is consistent with ear advantages and assumed to be best suited for neural processing. Previous studies used low-pass filtered stimulation in L1 for the rehabilitation of subjects with hearing impairment or treatment of speech disorder. For healthy (foreign) language learners, prior research concentrated more on L2 teaching and learning by using filtered stimuli, i.e., using filtered signals to teach/learn Chinese/English/French as a second/foreign language, while researchers neglected students' perception and production of the filtered stimulation in students' L1. Therefore, this study attempts to take both L1 (Chinese) and L2 (English) into consideration to uncover the processing mechanisms of filtered and unfiltered language signals in L1 and L2. All in all, to fill in these gaps, the current study explores an optimal, physical language stimulation based on VTA and ear advantages, which is assumed to be the low-pass filtered and unfiltered language signals under a certain dichotic or diotic listening condition. To reveal an optimal language input signal and the neural mechanisms underlying L1 and L2 processing in Chinese EFL university students, a combined ERP and fMRI experiment and semi-structured interviews are used to investigate students' brain responses and opinions to these stimuli.

2.5 Cognitive load theory

For this section, cognitive load theory is introduced, including a general introduction to the theory, measurements of cognitive load, and a review of studies on optimizing EFL learners' cognitive load in listening.

2.5.1 General introduction to cognitive load theory

Cognitive load theory is concerned with how the components of working memory and long-term memory are organized as cognitive architecture and comprise human cognition (Sweller, 2010; Sweller et al., 2011; Sweller et al., 1998). Learning requires students to process instructional information in their working memory, as a result, the instructional information imposes load on their working memory, which can be categorized into intrinsic, extraneous, and germane cognitive load (Sweller et al., 2011). Intrinsic cognitive load refers to the working memory load that is imposed by the fundamental structure of the instructional information, moreover, the information is what the learning goal requires a student to acquire. Extraneous cognitive load represents the load imposed only by the instructional procedures, which means the way how the information is presented or what activities students are involved in. Different from intrinsic and extraneous cognitive load, germane cognitive load is another category, which does not imply the imposed load by learning materials but refers to working memory resources allocated to or related to learning (Sweller et al., 2011). Intrinsic and extraneous cognitive load are assumed to add to the total cognitive load, which *"determine the total cognitive load imposed by material that needs to be learned. That total cognitive load determines the required working memory resources needed to process the information with some resources dealing with intrinsic cognitive load (germane resources) and other resources dealing with extraneous cognitive load (extraneous resources)"* (Sweller et al., 2011, p. 58). Cognitive load theory describes and predicts how working memory resources are allocated and how working and long-term memory interact with each other (Antonenko et al., 2014).

In the cognitive process, working memory is limited in capacity and time span to hold and process new-incoming information, but the processing capacity and duration will be unlimited if coping with familiar information that is stored in long-term memory previously (Sweller et al., 2011). According to Miller (1956), working memory keeps 7 ± 2 information elements. But Cowan (2001) further claims that the limit is four and the number decreases if the information has to be remembered (e.g., memorizing vocabulary) and processed (e.g., problem-solving) at the same time. Once the information has been stored in long-term memory, it can be processed as one element in working memory, as a result, the cognitive load is reduced. Thus, the cognitive load imposed by a certain task will be lowered if the student has previous knowledge or expertise of the task to free up more capacity and time for other processes, especially for deeper elaboration (de Jong et al., 2009). In addition, repeated practice, or increasing expertise, of a task results in automated cognitive schemata, which releases working memory resources and makes controlled processing unnecessary (de Jong et al., 2009; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Therefore, previous knowledge and increasing expertise lower cognitive load to process the information more efficiently (de Jong et al., 2009).

The goal of the theory is to guide and develop the instructional design based on the effects of cognitive load during the learning process, consequently, to optimize the level of cognitive load in different learning contexts (Antonenko et al., 2014; Sweller et al., 2011). In other words, the first concern of the instructional design is to optimize students' cognitive load in the learning process since cognitive overload may have a negative impact on task performance and outcomes (de Jong et al., 2009; Sweller et al., 2011), further, to facilitate knowledge acquisition and transformation from working memory to long-term memory. Evidence has shown that the more problem-solving tasks the learning process involves, the poorer learning outcomes the students get. As more working memory resources are occupied and required to hold and process more information in working memory, working memory load is increased, i.e., an increased extraneous cognitive load (Ayres & Sweller, 1990; Sweller, 1988). Additionally, studies have revealed that error rates become higher in the tasks with a higher requirement of cognitive load during learning, so the error rate is considered to be an indicator of working memory demands and intensity of decision-making (Sweller et al., 2011; Sweller & Cooper, 1985). Increased cognitive load influences learning outcomes negatively, in contrast, instructional procedures that can reduce cognitive load are assumed to lead to better outcomes (Sweller et al., 2011). Therefore, researchers working in the context of cognitive load theory devote themselves to devising learning strategies and materials to optimize or reduce students' cognitive load in the learning process. The following section 2.5.3 will explain more, especially in the field of EFL teaching and learning.

It is worth mentioning that the term "cognitive load" used in this study is different from the term "working memory load." The tasks adopted in the studies of working memory load do not interact with students' previous knowledge and long-term memory (Antonenko et al., 2014). For example, in the n-back task, frequently used as a measure of working memory load in psychological research, participants repeat a given stimulus (it can be a letter or a symbol) after a sequence of stimuli, which are not related to participants' prior knowledge or long-term memory. As to cognitive load discussed in the present study, it concentrates on the tasks with more complicated and less structured information that requires students' prior knowledge and interactions between working memory and long-term memory (Antonenko et al., 2014), for instance, listening to a conversation and then comprehending the language information.

2.5.2 Measures of cognitive load

To achieve the goal of analyzing the effects of cognitive load on learning and designing instructional procedures and materials, researchers have endeavored to measure cognitive load in different ways. But since cognitive load is multidimensional in nature, it is difficult to be defined and measured (Seeber, 2011). According to Eggemeier (1988), Paas et al. (1994), and Sweller et al. (2011), four major methods/techniques for measuring cognitive load are summarized as follows.

2.5.2.1 Indirect measures

Cognitive load was indirectly measured at the early stage of research in this field. The relation of problem solving and learning was first investigated, on which the assumption of cognitive load was based (Sweller et al., 2011). Techniques such as computational models, performance during acquisition, and error profiles are adopted.

The first attempt in cognitive load theory was to test the hypothesis that high problem-solving search, as a learning strategy, would result in a higher load of working memory. After a series of studies, Sweller and his colleagues claimed that considerable problem-solving tasks imposed a high extraneous load and students achieved poorer outcomes, but cognitive load was reduced when dealing with low problem-solving search (Ayres & Sweller, 1990; Sweller, 1988). They further explained that the higher level of search occupied more working memory resources, which required a more complicated model to hold and process more information in working memory (Ayres & Sweller, 1990; Sweller, 1988). There is no doubt that the computational model is an important trial as a starting point of measuring cognitive load. For these studies, empirical and analytical data were analyzed to estimate cognitive load, in which data was obtained from experts' opinions and rating, and analyzed with mathematical models (Paas et al., 2003).

Performance during the learning process provides more evidence through observations of the effects of cognitive load. Scholars reported that when a student was asked to learn the new information by employing a learning strategy, this process imposed an increased cognitive load that would exert influence on not only test scores but also future performance in the whole learning process (Chandler & Sweller, 1991, 1992). Errors have also been used as an indicator to measure cognitive load during learning because increased cognitive load may have a negative influence on both learning time and task accuracy (Sweller et al., 2011). Evidence has proved the assumption that students made errors when working memory reached at a relatively high point while doing a problem-solving task (Ayres & Sweller, 1990). It indicates that the high error rate at a particular point represents the location where working memory capacity reaches its limit, which can be employed as an indicator of working memory demands (Sweller et al., 2011).

Indirect measures estimate students' cognitive load by assessing their performance and learning outcomes during the learning process, for instance, task completion time and error rates. When indirect measures are taken as a way of evaluating cognitive load, learning materials are the same but the instructional design differs. The focus of the indirect measurement is extraneous load since the intrinsic load of learning materials remains constant and the extraneous load of the instructional design is assessed (Schmutz et al., 2009). However, the current study uses different auditory language signals as stimuli and intends to compare participants' brain responses to different listening materials, rather than the effects of the instructional design on cognitive load. Furthermore, indirect measures set criteria for measuring cognitive load such as task completion time and error rates. These criteria seem objective to evaluate students' cognitive load while doing learning tasks, but these measures concentrate more on performance/learning outcomes and ignore students' diverse mechanisms of information processing and knowledge learning. Therefore, the present study will not use indirect measures to evaluate participants' cognitive load while listening to different auditory language signals. Instead, this study focuses on the neural processing of different kinds of listening materials and investigates participants' opinions on the signals they hear during the experiment.

2.5.2.2 Subjective measures

As the theory develops, cognitive load is not solely aimed to predict the effectiveness of pedagogical instruction by using error rates and learning time to measure, but more direct measures are needed to investigate instructional effects on cognitive load. Paas (1992) put forward subjective measures of mental effort and difficulty to measure cognitive load, which was a crucial breakthrough.

Paas (1992) asserted that the amount of mental effort, or "intensity of effort" (p. 429), that students exerted in the learning process could be introspected and reported by the students, which could be used to measure cognitive load. According to Paas et al. (2003), mental effort being measured here is *"the aspect of cognitive load that refers to the cognitive capacity that is actually allocated to accommodate the demands imposed by the task: thus, it can be considered to reflect the actual cognitive load"* (p. 64). This subjective measure, a 9-point Likert scale from 1 (very, very low) to 9 (very, very high mental effort), was developed for students to self-rate their mental effort at different phases of learning and testing processes. The correlation between self-rating of mental effort and performance was identified after the situations of increased and lowered cognitive load in instructional procedures were compared. As hypothesized, the results showed that students taught by the instructional design with lower cognitive load had lower self-rating of mental effort and achieved better learning outcomes than the students exposed in a hypothesized high cognitive load condition did. After this pioneering study, some replication research supported that the measure of mental effort by students' self-rating was effective and sensitive in predicting their cognitive load (Paas & van Merriënboer, 1994; Paas et al., 1994).

Apart from measuring mental effort, the measure of difficulty self-rating is an indicator of cognitive load as well, which requires students to rate how difficult or easy they feel during the tasks. Some experiments revealed that self-rating of difficulty could vary in a task due to various element interactions, but this subjective measure could detect the variations in the task (Ayres, 2006; Marcus et al., 1996). Similar to the self-rating of mental effort mentioned above, subjective rating of difficulty is also considered as a sensitive measure, which can be used to detect and distinguish the load that is imposed by pedagogical procedures (Sweller et al., 2011). As for L2 learning studies, self-rating of L2 task difficulty has been adopted by a number of researchers (e.g., Gilabert & Barón, 2013; Gilabert et al., 2011; Levkina & Gilabert, 2012, 2014; Malicka & Levkina, 2012; Robinson, 2007; Sasayama, 2013). Most researchers used a multidimensional scale raised by Robinson (2001), which covered the relations of task complexity, difficulty, and production but the item of perceived mental effort was excluded. Results indicated that subjective self-rating of difficulty in L2 tasks potentially measured and predicted students' perceived difficulty (as teachers' anticipation) while doing the tasks. Sasayama (2013) adopted both self-rating measures of mental effort and difficulty and found that the two measures had convergent outcomes that the task designed to be more complicated was rated as more difficult and more mental effort required than other tasks designed with lower complexity. In addition, Révész et al. (2016) did a validation study on self-rating measures of mental effort and difficulty, dual-task, and expert judgments regarding the mental manipulation of task complexity in the EFL context. 96 students (48 English native speakers and 48 ESL students) and 61 ESL teachers were recruited. Students were asked to perform simple and complex versions in each of the three oral tasks, which included picture narrative task, map task, and decision-making task. In each condition, students rated their perceived mental effort and task difficulty on the 9-point Likert scale. In addition, the ESL teachers offered their expert judgments and anticipation of mental effort and difficulty. The results showed that self-rating of mental effort and difficulty was a potentially valid research tool to estimate task-generated cognitive load, which further supported that the more complex tasks were, the greater cognitive effort was measured as predicted.

However, subjective measures of mental effort and difficulty in measuring cognitive load are still controversial, which seems not to be completely reliable and valid (de Jong, 2010). Several variations are challenging the reliability and validity of the subjective measures. The first variation is that the questions asking for "effort" and "difficulty" in the questionnaires of different studies are unclear, whether asking for the learning material, the learning process, or the understanding of the learning material (de Jong, 2010). Kester et al. (2006) asked two separate questions of mental effort and the understanding of materials by adopting a 9-point scale and identified an effect of pedagogical interventions on perceived mental effort during practice, but no effect was exerted on mental effort to understand the subject. The problem of asking different questions in different studies is that the outcomes/results may vary greatly due to the questions being asked (van Gog & Paas, 2008). The second variation is that the wording on the scales such as "very, very" and "very" is hard for students to define and distinguish. And the extremes on the scales range from "very, very low mental effort" to "very, very high mental effort" and from "very, very easy" to "very, very difficult," which requires students to clearly understand the limits of their cognitive load and their ability to do the learning task. It is doubtful whether students can accurately understand the extremes on the scales and distinguish them unless they have known the material being presented, the pedagogical procedures, and their processing capacity. The third variation is the timing of distributing the questionnaire. The majority of studies send the questionnaire once after the pedagogical intervention, but some studies distribute it several times during and after the instruction (de Jong, 2010). It is assumed that the measure will be more precise if cognitive load is measured more often (de Jong, 2010). However, whether a student can estimate an average load or not is not certain; whether a researcher should calculate the average value or not is also unclear. In addition, whether the measure should be instantaneous (measure several times after each task/procedure) or an average estimated by students after the instruction is still vague (de Jong, 2010). The fourth variation that influencing the reliability and validity of the measure is the use of the one-item questionnaire. The one-item questionnaire is problematic because values on the scales lead to different interpretations across studies (de Jong, 2010). For instance, the first experiment in Pollock et al.'s (2002) study suggested that the score of about 3 on a scale of 7 was the ''best'' cognitive load condition, but Kablan and Erden (2008) stated that of 2.5 on a scale of 7 was the ''poorest'' condition in the separated text group. The scores achieved from the one-item questionnaire seem that *"there is no consistency in what can be called a high (let alone a too high) cognitive load score or a low score"* (de Jong, 2010, p. 116).

The above-mentioned factors challenge the reliability and validity of measuring cognitive load. Thus, the current research will not use students' self-ratings of mental effort and difficulty as measures of cognitive load.

2.5.2.3 Task-dependent measures

Traditionally, working memory load is assessed by dual-task methodology, which uses a secondary task combined with a primary task (Britton & Tesser, 1982; Kerr, 1973). The secondary task is normally dissimilar to the primary task, in which students are involved in an extra cognitive activity that is secondary to the primary task and occupies less working memory resources during the learning or problem-solving processes (Sweller et al., 2011). For instance, the primary task is to solve a mathematical problem, meanwhile, the secondary task requires students to respond to a sound during the process. The underlying rationale is that a high cognitive load in the primary task may impede the performance of the secondary task, on the contrary, a relatively low cognitive load in the primary task may lead to better performance in the secondary task (Sweller et al., 2011). Since the dual-task method stems from psychological research, the secondary task usually comprises activities of detecting a visual or auditory stimulus, and performance of the secondary task is evaluated by reaction time and accuracy, indicating that slower reaction or inaccurate performance result from more working memory resources, or cognitive capacity, are consumed by the primary task (de Jong, 2010; Sweller et al., 2011).

In terms of using the dual-task method in foreign language learning, this measurement is rarely adopted. DeKeyser (1997) first used both single-task and dual-task methods to evaluate students' knowledge development of explicitly taught morphosyntactic rules of a second language. For the dual-task experiment, a number between 100 and 1,000 was shown to students on the screen before a picture of the target linguistic rule was presented. Another part of the dual-task was the beeps at irregular intervals that were also presented before they were asked to perform production and comprehension tasks of the target rule. Students were asked to memorize the number, count the beeps, and calculate the difference between the number shown on the screen and the number of beeps in the secondary task. DeKeyser (1997) hypothesized that interference from the secondary task would be reduced as the automatization of morphosyntactic rules was gradually developed. But results showed the difference between the two conditions was not as expected and he speculated that it might be due to the secondary task was not demanding enough.

Declerck and Kormos (2012) adopted the dual-task measure to investigate encoding mechanisms and lexical selection during the L2 speaking task. The primary task is speaking while the secondary task was finger-tapping. Participants were required to press one button numbered from 1 to 10 as randomly as possible in every second during an L2 speaking task. Findings demonstrated that the number that the participants tapped became less random while they were speaking. The authors concluded that the performance of the secondary finger-tapping task was influenced by the primary speech task as hypothesized.

In the study of Révész et al. (2016) as mentioned earlier, the dual-task method, together with subjective measures of task difficulty, mental effort, and expert judgments, was used to assess cognitive load generated by task complexity for ESL students. Half of the participants (24 English natives and 24 ESL students) were selected in the dual-task condition and the other half in the single-task condition. For the dual-task design, the primary tasks contained simple and complex versions in each of the narrative, map, and decision-making tasks. And the secondary task was to detect color changing on the screen, in which the students were asked to react as fast as they could when the color was changed to green rather than red. Accuracy and reaction time were analyzed, and results revealed that participants performed much accurately and faster in the secondary task when the primary task was simple as expected. Further, the findings showed that accuracy and reaction time could be predictors of secondary-task performance in dual-task conditions.

The reasons why the dual-task method is rarely used in (foreign) language teaching and learning may be provided by Sweller et al. (2011) that *"secondary tasks require much more planning and, depending on the nature of the secondary task, may require equipment. They can interfere with normal classroom practice to a much greater extent"* (p. 79). Furthermore, as mentioned in the study of DeKeyser (1997), the variable of complexity/difficulty of the secondary task is hard to control, in which results show no difference between tasks due to the secondary task may be not sufficiently demanding. Because of participants' diverse perceptions of difficulty and working memory load, it is controversial how difficult or complex the task could be presented as a secondary task. More importantly, the dual-task method is more frequently used in psychological research, especially focusing on the allocation of attention and working memory. However, this is not the focus of the present study. Thus, this study will not use the dual-task method as a measure of cognitive load in the research design.

2.5.2.4 Physiological measures

As discussed above, the total/average cognitive load is measured by subjective and task-dependent measures after or during tasks are performed. In order to provide a more accurate measure of instantaneous cognitive load, the physiological index as a method of measuring cognitive load was proposed. One of the physiological measures was heart rate (Paas & van Merriënboer, 1994). Paas and van Merriënboer (1994) compared heart rate analysis with a subjective measure. They stated that the measure of heart rate was not valid and sensitive but the subjective measure had more potential in measuring cognitive load. Besides, Nickel and Nachreiner (2003) also claimed that the measure of heart rate could not be an effective indicator for mental workload. Pupillary responses were then adopted as a measure of cognitive load, which were identified as a sensitive tool (Paas et al., 2003; van Gog et al., 2009). Researchers used a series of tasks requiring different levels of memory load, and findings suggested that pupil dilation increased as memory load increased. It was concluded that pupil size could be correlated with memory load. However, there was a limited number of studies using this measure (e.g., Schultheis & Jameson, 2004). The main disadvantage of pupillary reaction is the sensitivity of this measure (de Jong, 2010; Sweller et al., 2011). Research showed that pupillary response had an age limitation because the changes of pupillary reactions gradually reduced as participants grew old, specifically, it did not show a correlation between the load of cognitive tasks and pupillary reactions in elderly participants (van Gerven et al., 2004).

Due to the limitations of heart rate and pupillary responses in measuring cognitive load, more and more researchers suggested the use of neuroscience techniques EEG/ERP (Antonenko et al., 2010) and fMRI (Paas et al., 2008; Whelan, 2007) assess cognitive load in the learning process. Since EEG/ERP can measure brain activity in natural settings noninvasively and has a high temporal resolution, which is sensitive enough to provide physiological evidence as an index of cognitive load changes in milliseconds (Antonenko et al., 2010). EEG/ERP can also help demonstrate the effects of pedagogical instructions by detecting subtle fluctuations of brain waves that indicate instantaneous cognitive load (Antonenko et al., 2010). Antonenko and Niederhauser (2010) used the indirect measure, subjective measure, and EEG to investigate learning with hypertexts, in which mental effort was self-rated and EEG was used to collect brain waves. Performance scores showed learning with the hypertext outperformed learning without hypertext, and self-rated mental effort indicated no between-group differences. But EEG was very sensitive to measure the differences in the brain waves of alpha, beta, and theta, indicating a significantly low load in hypertext processing. Compared with other measures, EEG had an advantage in reflecting different types of cognitive load, e.g., instantaneous, peak, average, and overall cognitive load (Antonenko & Niederhauser, 2010). Schultheis and Jameson (2004) did an ERP study using novelty P300 (Friedman et al., 2001) to measure cognitive load. Findings indicated that potentials elicited by the event were sensitive to detect the differences between tasks, and the amplitude of P300 could be an indicator to represent cognitive load that was needed for the difficult text processing. In another study, Stevens et al. (2007) investigated students' cognitive load when dealing with science problems in a multimedia environment by using EEG. Results revealed that cognitive load increased when students were facing more difficult problems as hypothesized. But findings also indicated that students' cognitive load did not decline as their skills increased, which meant that greater mental effort might be involved in strategic refinement (Stevens et al., 2007). These studies of EEG/ERP used to measure cognitive load during the learning process indicate that EEG/ERP is sufficiently sensitive and reliable to measure cognitive load of language processing, therefore, it can be adopted in the present study.

As for another promising neuroimaging technique fMRI, it is usually used to map the brain localization during language processing since it features spatial resolution. To measure cognitive load, Whelan (2007) argued that neuroimaging techniques might lead to an improvement in cognitive load measurement, but anyone technique of current neuroimaging measurements used to assess cognitive load could not accurately demonstrate the differences in various types of cognitive load under different learning conditions, since its limits in precision and methodology (Whelan, 2007). Therefore, more and more scholars suggest a combination of neuroscience techniques with high temporal and spatial resolution such as EEG-fMRI (e.g., Ullsperger & Debener, 2010) to measure cognitive load.

In order to better measure and interpret participants' cognitive load, this study will not adopt the indirect measure, subjective measure (i.e., self-ratings of mental effort and task difficulty), and task-dependent measure, due to the limitations of these measures that influence the reliability and validity of measurement results as mentioned above. The current study adopts the physiological measure – a combined ERP and fMRI experiment, together with semi-structured interviews. Among all the measurements of cognitive load, the physiological measures of ERP and fMRI are sufficiently sensitive and reliable in detecting instantaneous cognitive load (Antonenko & Niederhauser, 2010; Stevens et al., 2007; Ullsperger & Debener, 2010), and the interview as a method for qualitative research allows participants to express their opinions toward different auditory language signals they have heard during the experiment. As a result, the ERP and fMRI data would demonstrate participants' brain activity in response to different signals, including amplitudes, areas under the curve (mental workload), latencies (reaction time) of potentials, and the brain regions involved in processing these stimuli. Further, semi-structured interviews are conducted to investigate participants' opinions and perceptions of the stimuli. Taken together, an optimal auditory language signal for Chinese EFL students to process would be identified.

2.5.3 Studies on optimizing EFL students' cognitive load in listening

To optimize cognitive load while EFL students are listening, researchers have had different trials in designing listening modalities, especially creating multimedia listening environments as assistance to facilitate students' listening comprehension and language learning (Mohsen, 2016). Diao et al. (2007) examined the effectiveness of simultaneous written-texts (on-screen texts) on listening comprehension for EFL students. After receiving the pedagogical intervention in three modalities: auditory listening only, listening + full script, and listening + simultaneous subtitles, three groups of participants were asked to perform tasks of translation and multiple-choice questions. By comparing the results, researchers concluded that the full script and subtitles helped students understand the passages. But the full script and subtitles hindered students' construction and automatization in listening comprehension because they might impose additional cognitive load to interfere with listening.

Chen (2011) used SynctoLearn, an automatic video-transcript synchronization system, to evaluate the effectiveness of the synchronized video-subtitle modality on improving EFL students' listening skills. Two university intact classes at similar proficiency levels were divided into two groups. The experimental group watched the synchronized video with transcript and the control group watched the same video without synchronized subtitles. After the instrument, participants were asked to make a summary according to the video as a test. Results showed that the experimental group performed better in summarizing key points of the video. Moreover, researchers claimed that synchronized subtitles helped reduce cognitive load and anxiety for EFL students while they were watching authentic videos.

Another study conducted by Aldera and Mohsen (2013) used annotations, captions, and animation modes to investigate whether annotations for captioned animation was effective for facilitating listening skills. Fifty Arabic-speaking EFL adult students were recruited in the research and divided into three groups. Three groups of participants watched an animated story under three conditions respectively: animation + captions + keyword annotation, animation + captions, and animation only. After the instrument, participants needed to complete a comprehension test and recall the story. However, findings indicated that captioning hindered students' listening comprehension and recall. Researchers further explained that these multiple modalities might create more cognitive load for students.

Although researchers have tried to optimize EFL students' cognitive load while listening, learning in multimedia environments does not always lead to positive effects on reducing cognitive load. It is still controversial what listening modality is effective for lowering cognitive load for EFL students. Regardless of the merit of multimedia learning environments as adopted in previous studies, the use of different listening modalities, such as on-screen texts, synchronized subtitles, captioned animation, etc., may not benefit all EFL students, and, more importantly, may impose additional cognitive load that hinders listening comprehension in a foreign language (Mohsen, 2016). From the literature, prior studies mainly focus on changing listening modalities, but little research concentrates on the auditory signal per se, i.e., changing the physical characteristics of the auditory language signal to lower EFL students' cognitive load. Therefore, the present study intends to fill this gap. To optimize students' cognitive load in listening, VTA is supposed to be one solution. VTA aims to restructure the ways that students perceive auditory signals via low frequencies that their ears are most sensitive to (Guberina & Asp, 1981, 2013; Kim & Asp, 2002). The stimuli are low-pass filtered that only preserves prosodic features and the higher frequencies defining words are removed, so students do not have to manipulate lexical and meaning processing, which lightens the processing load and releases more attentional resources for other (deeper) processes (de Jong et al., 2009; Lian & Sussex, 2018). Thus, it is assumed to be an optimal stimulus in optimizing EFL students' cognitive load while listening. Further, the current study adopts low-pass filtered and unfiltered stimuli under dichotic and diotic listening conditions, which ought to be an optimal auditory language signal that is consistent with ear advantages. Therefore, this study uses the physiological measure (a combined ERP and fMRI experiment) to assess Chinese university EFL students' cognitive load while dichotically and diotically listening to filtered and unfiltered stimuli, and conducts semi-structured interviews to investigate students' opinions toward these signals. The present study aims to seek an optimal auditory language stimulus that can lower mental workload, as a result, to enable EFL students to best make sense of that stimulation and to restructure their perceptual mechanisms to facilitate language-learning (Guberina & Asp, 1981; Lian & Sussex, 2018).

2.6 Theoretical framework of the current study

The major concern of the current study is to explore an optimal auditory language signal for Chinese university EFL students, which is sought through neuroimaging techniques and semi-structured interviews. This study is fundamentally based on the neurobiological basis of language processing. Learning a language, or learning in general, is a neurobiologically-based process of neuroplasticity that structures/re-structures student's neural networks for speech/language processing constantly and dynamically (de Jong et al., 2009). Language learning, as an act of meaning-making (Lian, 2004; Lian & Pineda, 2014), is achieved by changing/transferring the perception of new-incoming representations of language input into meaning. Furthermore, because of students' various operational histories (Lian, 2004) and the lack of invariance problem in speech (Heald et al., 2016; Heald & Nusbaum, 2014), students' perceptions of language are idiosyncratic. Therefore, speech/language processing requires an active mechanism that enables students to adaptively, flexibly, and perceptually make sense of language input.

To facilitate (foreign) language learning, the optimal solution is to focus on building/changing the perception of language signals from the neurobiological perspective. The Verbotonal approach is proposed on the basis of neuroplasticity, which emphasizes restructuring the neural networks for language perception and production via a holistic and multi-sensory system of learning (Asp et al., 2012; Guberina & Asp, 1981). To maximally reorganize neural connections, the low-frequency stimulus used as language input plays a crucial role (Guberina & Asp, 1981, 2013; Kim & Asp, 2002). Since the development of subcortical organs, the proprioceptive memory, and the auditory-memory development begin from the perception of the pitch, rhythm, and intonation carried by the low-frequency pattern of speech, subcortical organs are sensitive to low-frequency sounds (Asp, 2006). That is why a student's speech production will change if his/her perception of speech changes (Guberina & Asp, 1981, 2013). The low-frequency stimulus implemented by low-pass filtering (< 320 Hz) is the starting point of building/changing students' perception of language, which further determines the stimulus-driven bottom-up process in speech/language processing.

In addition, the primary concern of pedagogical design is to optimize students' cognitive load because cognitive overload negatively influences task completion and performance outcomes during the whole learning process (de Jong et al., 2009; Sweller et al., 2011). Therefore, optimizing cognitive load is aimed to assist students to maintain an optimal level in mental processing load in various learning contexts, as a result, to facilitate knowledge acquisition by transforming information from working memory to long-term memory effectively (Antonenko et al., 2014; Sweller et al., 2011). Since it is still controversial to optimize students' cognitive load while listening to foreign language signals by changing modalities of the multimedia instruction, the present study intends to change the physical characteristics of the auditory language signal to lower EFL students' cognitive load. This attempt is based on the assumption that low-pass filtered stimuli (i.e., low-frequency signals), containing only prosodic information, lightens students' cognitive load for lexical and meaning processing and releases more attentional resources for other deeper and higher-order processes (de Jong et al., 2009; Lian & Sussex, 2018).

According to hemispheric specialization, a general right ear advantage for verbal processing and a left ear advantage for non-linguistic processing (such as melody, or prosody of language) have been reported (Gandour et al., 2004; Meyer et al., 2002; Sammler et al., 2015; Tervaniemi & Hugdahl, 2003; Vigneau et al., 2006). Thus, the current study employs low-pass filtered (i.e., prosodic) and unfiltered stimuli in dichotic and diotic listening conditions to investigate temporal-spatial neural signatures for linguistic and prosodic processing (Alho et al., 2012; Hugdahl, 2003; Vanhoucke et al., 2013). To measure Chinese EFL students' cognitive load while listening and explore an optimal auditory language stimulus, the physiological measure (a combined ERP and fMRI experiment) (Ullsperger & Debener, 2010) is utilized to measure participants' cognitive load during the processing of different signals. After the combined ERP and fMRI experiment, semi-structured interviews are conducted to investigate participants' opinions toward these auditory language stimuli. Consequently, students' brain responses and views of filtered and unfiltered stimuli under different dichotic and diotic listening conditions are yielded to identify an optimal auditory language signal for Chinese EFL students.

To sum up, the present study holds that neurobiological bases and active cognitive processing of speech constitute the processing mechanism of speech. To raise EFL students' awareness and change their perceptions of foreign language signals during the learning process, low-pass filtered and unfiltered stimuli are used as auditory language input based on the principle of low-frequency stimuli in VTA; further, to explore temporal-spatial neural signatures for processing filtered and unfiltered stimuli, four configurations under dichotic and diotic listening conditions (i.e., filtering in one/both ear(s) and unfiltering in one/both ear(s)) are adopted. In other words, signals in Verbotonal-based dichotic and diotic listening conditions are created in the current study. To identify an optimal auditory language input signal, a combined ERP and fMRI experiment and semi-structured interviews are implemented to reveal participants' brain activity and opinions to these signals. As a result, the stimulus with low-pass filtered and unfiltered sounds under a certain dichotic or diotic listening condition can be considered as an optimal auditory language input signal according to mental workload, reaction time, hemispheric lateralization, and brain localization while processing the signals, together with participants' views toward the signals. The overarching theoretical framework of the present study is demonstrated in Figure 2.4.

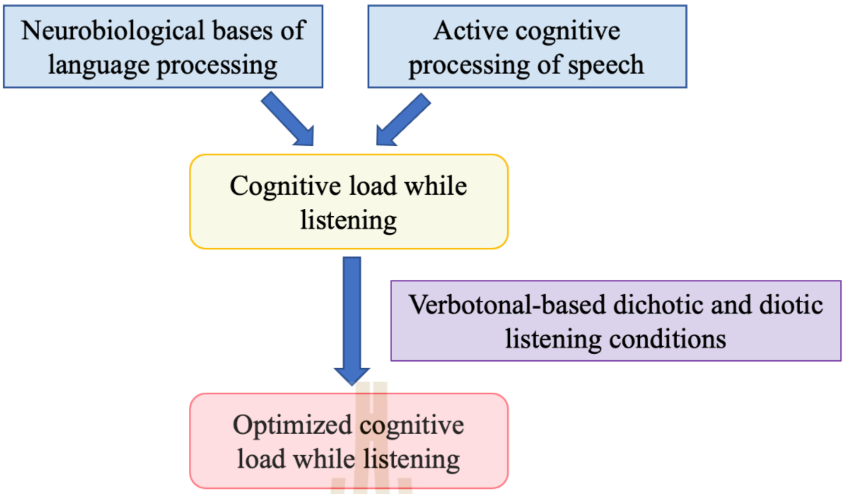


Figure 2.4 Theoretical framework of the current study

2.7 Summary

This chapter presents a review of concepts and theories that are relevant to the current study. They are mainly about language and language learning, neuroscience techniques used with brain responses to linguistic signals, neurobiological theories of speech perception, the Verbotonal approach, and cognitive load theory. In the end, a theoretical framework based on the review is constructed to underpin the current study. In the next chapter, the research methodology of the study will be presented.

CHAPTER 3
RESEARCH METHODOLOGY

This chapter presents the research methodology of the study. It introduces the research design and participants first. In the following section, the experimental paradigm is discussed, including auditory language stimuli and the stimulus presentation paradigm adopted in the combined ERP and fMRI experiment. After that, data collection procedures and data analysis methods of the experiment and semi-structured interviews are presented. The last section touches on ethical considerations and the pilot study.

3.1 Research design

The current study adopted a mixed methods research design, which integrated quantitative and qualitative research data. As Creswell and Creswell (2018) noted, a quantitative method provided closed-ended answers from questionnaires, pedagogical instruments, etc., but qualitative data offered open-ended and non-predetermined responses. In addition, mixed methods triangulated data sources across quantitative and qualitative methods, and neutralized bias and weaknesses of each method to seek convergence between the two methods (Creswell & Creswell, 2018; Jick, 1979). Thus, mixed methods integrating quantitative and qualitative methods were more effective for the current study in answering the research questions raised in Chapter 1. In order to answer the first research question, quantitative data obtained from the combined ERP and fMRI experiment would provide participants' brain responses to different auditory language signals, and qualitative data describing their opinions and views toward the signals would answer the second research question. Then, the answer to the third research question was the interpretation of the findings from comparing and relating the results of quantitative and qualitative data. As a result, an optimal auditory language signal for EFL students would be uncovered.

Further, integration of quantitative and qualitative methods could check the validity of each database to make sure that the results of the study were valid (accurate) (Creswell & Creswell, 2018). Database of the combined ERP and fMRI experiment or semi-structured interviews would help explain the other database to better answer the research questions, thus, to ensure the validity and trustworthiness of the findings in the current study.

3.2 Research context and participants

Kunming Medical University (KMU), located in Kunming, China, was established in 1933. There are 18 schools in KMU, offering 33 Bachelor's degree programs, 17 Master's degree programs, and 4 Doctoral programs. Over 14,800 undergraduate and 4,200 postgraduate students are studying medicine, science, management, and social science at KMU currently. Students of KMU are non-English major students with an intermediate or lower level of English proficiency. Two English courses are provided for the first- and second-year undergraduate students, namely an integrated course (reading, writing, and translation) and a listening-speaking course (listening and speaking). Students are required to take two 80-minute classes in the first two years and take the College English Test (CET) Band 4 (the national standardized English proficiency test for undergraduate and postgraduate students in China, and Band 4 refers to the intermediate difficulty level). As to teachers in the Department of Foreign Languages, KMU, there are thirty-six English teachers now.

The population referred to a group of individuals sharing one characteristic that could be distinct from other groups. The research sample was selected and studied from the population by researchers (Creswell & Guetterman, 2019). In the current study, the population was Chinese university EFL students who speak Mandarin Chinese (a tonal language) as their first language and English as a foreign language with an intermediate level of proficiency. But practically speaking, it was impossible for a researcher to investigate the entire population. Thus, a target population or sampling frame was chosen. After the target population was determined, researchers selected a sample from the target population to study (Creswell & Guetterman, 2019). Since the aim of sampling was to attain representativeness, a sample ought to be assembled in a way to be representative of the population, and the results from the sample could be generalized to the population (Babbie, 2016; Creswell & Guetterman, 2019; Fowler, 2009; Welman et al., 2005). Thus, thirty students of KMU, who met the inclusion criteria described in 3.2.1, were selected as a sample, and the findings of the sample could be generalized to the population of Chinese university EFL learners with an intermediate level of English proficiency. Due to the importance of selecting a sample, the target population and the sampling design were demonstrated respectively.

3.2.1 The target population

In order to select a homogeneous group of individuals as the target population, a questionnaire including students' personal information and handedness (shown in Appendix A) was distributed first, which was designed based on the inclusion criteria listed as follows.

Students in KMU

The target population or sampling frame listed individuals in the population that researchers could access (Creswell & Guetterman, 2019). Regarding the availability of the target population (Creswell & Creswell, 2018), students aged 22 to 28 in Kunming Medical University, China, majoring in medicine and health such as clinical medicine, nursing, rehabilitation, etc., were selected. In order to ensure the target population had a similar level of English proficiency, students were selected as they had passed the College English Test (CET) Band 4. Those who had passed CET Band 4 and self-evaluated English proficiency as an intermediate level in the questionnaire met this inclusion criterion.

Normal hearing

This study concerned auditory signals, so normal-hearing participants were required. The frequency range that the healthy ear could process was about 20 – 20,000 Hz, and the frequencies for speech were mainly between 500 and 4000 Hz (Hain, 2006). As for loudness, the audible range for the healthy ear was from 0 to 180 dB, but people who were exposed to sounds louder than 90 dB all the time would have chronic hearing damage, and sounds louder than 130 dB could lead to acute hearing loss (Institute for Quality and Efficiency in Health Care, 2008). For the current study, it was impractical for the researcher to screen for hearing loss in each student. Therefore, self-evaluation of the hearing was included in the questionnaire of personal information (Questions 3, 4, and 5 listed in Appendix A), which contained three questions: “Do you have difficulty hearing what teachers say in the classroom?”, “Do you have difficulty hearing a whispered voice, a finger rub, or a clock tick?”, and “Have you ever worn a hearing aid?” Students who denied the statements in the three questions would be regarded as normal-hearing subjects whose hearing was within normal hearing thresholds of frequency and loudness.

Normal or corrected-to-normal vision

As language and vision were primary systems for human perception and cognition, linguistic and visual information were integrated to recognize and interpret the language signal through a person’s holistic multi-sensory system (Ferreira & Tanenhaus, 2007). According to Cortiella and Horowitz (2014), learning disabilities were most commonly led by processing difficulty in visual and auditory perception. A child with a visual or auditory perceptual disorder would have a disadvantage in certain stages of the learning process. Therefore, visual and auditory perceptions were of equal importance in the process of language learning. For the current study, two main questions with two sub-questions (Questions 6 and 7 shown in the questionnaire in Appendix A) were designed to evaluate students’ vision, which are “Without wearing glasses, can you see the regular print clearly in books/newspapers? If no, can you read books/newspapers with glasses?”, and “Without wearing glasses, can you see street signs clearly while walking on the street? If no, can you see them clearly with glasses?” The questions assessed whether the student had normal vision who could see both close and faraway objects clearly, or he/she was hyperopic/myopic but the vision was corrected-to-normal after wearing glasses. Students who had normal or corrected-to-normal vision met the inclusion criterion.

No history of neurological, psychiatric, or vascular disease

Neurological and psychiatric disorders were driven by brain dysfunction or structural brain diseases, which induced damage in movement, sensation, and speech/language, in addition, led to impairments in brain areas that controlled cognition, perception, mood, emotion, etc. (Lyketsos et al., 2007). Evidence also showed that vascular lesions influenced cognitive processes, including language processing (Cummings, 1993). A cohort study conducted by Copland (2003) suggested that cortical vascular lesions and vascular/degenerative basal ganglia impairments led to lexical ambiguity in the priming task, which indicated that vascular and basal ganglia lesions interfered with the selective attentional engagement of the semantic network. Therefore, neurological, psychiatric, and vascular diseases impacted language processing or resulted in speech/language disorder. Students who had a history of neurological, psychiatric, or vascular disease would be excluded from the target population of the current study after they completed the questionnaire.

Not taking any psychotropic or anti-hypertensive medications

Psychiatric disorders (e.g., schizophrenia, bipolar disorder, and depression) influenced higher brain functions (Schulz & Steimer, 2000), which were closely related to language disorders and learning disabilities (Sundheim & Voeller, 2004). Though the direct link between the symptoms of psychiatric disorders, the action mechanisms of the psychotropic medications, and higher brain functions was not quite clear, most of these medications were not target-selective, which meant they diffused throughout the brain and had an overall impact on the brain (Schulz & Steimer, 2000). From the perspective of pharmacological effects, psychotropic medications, similar to anti-hypertensive medications, would exert influence on various brain systems and brain functions (Schulz & Steimer, 2000). As the present study focused on students' neural processing of languages, students without a history of psychiatric disorders and without taking any psychotropic or anti-hypertensive medications satisfied this inclusion criterion.

Language experience

As mentioned in Chapter 2, language experience had an effect on the neural processing of language at both cortical and subcortical levels. Language experience was another important factor of the current study, which considered language processing of both L1 and L2. In order to recruit a homogeneous group of participants regarding language experience, students' L1 and L2 should be at a similar level of proficiency. For L1, Mandarin Chinese was the target language being studied so Mandarin should be participants' mother tongue, excluding Cantonese, Hakka, Min

(Hokkien), Xiang (Hunan), Gan (Jiangxi), Wu (Zhejiang) dialects, and other ethnic languages. As to L2, English was the foreign language that they had been learning for over ten years. Students, as mentioned above, who had passed CET Band 4 and self-evaluated English proficiency as intermediate level in the questionnaire were considered as being at a similar proficiency level and would be recruited in the target population of the current study.

Have not received professional musical training

Since musical experience changed the neural representation of pitch processing that influenced both musical and linguistic processing in the brain stem (Wong et al., 2007), students who had received professional musical training were excluded from this study.

Handedness

There had been a debate on the association between handedness and hemispheric lateralization for language since the finding of Broca's area in the patient with right-handedness and left-hemisphere language regions in the 19^{th} century. However, more and more studies employing neuroimaging techniques and psychometric tests suggested that language lateralization was independent of either left- or right-handedness, indicating that the mechanisms for hemispheric language lateralization were similar in left- and right-handed subjects (Knecht et al., 2000; Mazoyer et al., 2014; Szaflarski et al., 2012). Although there were exceptions that a very small number of left-handed subjects with the right hemisphere dominating language (Mazoyer et al., 2014), scholars believed that hemispheric lateralization for language was not determined by handedness but a natural phenomenon (Knecht et al., 2000). To ensure the validity of the current study and avoid the exception of right-hemisphere language dominance in left-handed students, the current study selected strongly right-handed students as participants. A handedness questionnaire adapted from the Edinburgh Handedness Inventory (Oldfield, 1971) (included in Appendix A) was distributed before the study. Students who were strongly right-handed with a laterality index ≥ 90 were included in the target population.

The above-mentioned inclusion criteria excluded the factor of intelligence quotient (IQ). The intelligence quotient was correlated with cognitive control abilities

(Checa & Fernández-Berrocal, 2015), but language processing was a higher cognitive process related to several aspects of knowledge-comprehension, short-term memory, long-term retrieval, auditory processing, visual-spatial processing, reading-writing ability, etc. (Mackintosh, 2011). A single score of the IQ test could not adequately explain students' multidimensional aspects of intelligence, and students with similar IQ test scores could be quite different in their performance and talents (Kaufman & Lichtenberger, 2005; Sattler, 2001). In addition, the students in the target population of this study had been enrolled in the same university studying medicine and health at a similar language proficiency level of L1 and L2. Thus, they could be considered as the homogeneity of the IQ level. The present study would not take the IQ test score as an inclusion criterion for the target population.

3.2.2 The sampling design

After selecting the target population, random sampling was conducted to choose the participants for the current study. According to Creswell and Creswell (2018), a random sample allowed each individual in the population to have an equal possibility to be chosen, which was representative of the population and ensured the results of the study could be generalized to the population. Therefore, thirty university students from Kunming Medical University, China were randomly selected as participants, or as a sample for the present study. The sample size of thirty was decided for the reason of statistical analysis, which was the minimum sample size for the analysis requirement (Ross & Willson, 2018). As a result, the findings could be generalized to the population of Chinese university EFL students whose first language was Mandarin Chinese and English language proficiency was at the intermediate level.

3.2.3 Participants

Students who completed the questionnaire and met the inclusion criteria mentioned above were selected as participants in the current study. Thirty students at Kunming Medical University, China, with the average age being 24.77 (SD ± 1.48), are strongly right-handed with the laterality index = 90.00 (SD ± 4.21) (Oldfield, 1971), who denied a history of neurological, psychiatric, or vascular disease, denied receiving professional music training before, and had normal hearing and vision. Their first language was Mandarin Chinese, and they had been learning English as a foreign language. As to English language proficiency, they had passed the College English Test Band 4 and self-evaluated English proficiency as pre-intermediate or intermediate level in the questionnaire. The participants gave written informed consent before participating in the study, as approved by the Institutional Review Board of Suranaree University of Technology (COA No. 108/2564). The demographic data of the participants are listed in Table 3.1.

Table 3.1 Demographic data of the participants

Gender	Male	14 (46.67%)
	Female	16 (53.33%)
Age	24.77 (SD ± 1.48)	
Education	Master students	30 (100%)
Handedness	Strongly right-handed	
	Laterality index = 90.00 (SD ± 4.21)	

3.3 Experimental paradigm

3.3.1 Auditory language stimuli

The listening materials, presented in the combined ERP and fMRI experiment, were complete Chinese (L1) and English (L2) sentences. This was because the research focus of the current study was to investigate prosodic processing at the sentence level rather than the lexical tones or intonation at the word level. Therefore, sentences with different sentence types and intonation patterns were the auditory language stimuli in the experiment. According to sentence types, there were four major types of both Chinese and English sentences, namely declarative, interrogative, imperative, and exclamative sentences (Greenbaum & Nelson, 2002, 2009; Huang & Liao, 2017; Xing, 2016). For the uses in communication, declarative sentences contained positive and negative statements in Chinese and English. As to interrogative sentences, the main types in Chinese were yes/no questions, wh- questions, alternative questions, and rhetorical questions (Huang & Liao, 2017; Xing, 2016), and four main types of English interrogatives were yes/no questions, wh- questions, alternative questions, and tag questions (Greenbaum & Nelson, 2002, 2009). Thus, auditory language stimuli in the current study were classified into eight sentence types in each language, which were declarative (positive), declarative (negative), yes/no questions, wh- questions, alternative questions, rhetorical questions (in Chinese)/tag questions (in English), imperative, and exclamative sentences. The classification based on the sentence types ensured that the intonation patterns of eight sentence types were covered.

Since all the participants had passed CET Band 4 and self-evaluated their English proficiency as intermediate level, the participants' actual English listening proficiency was basically consistent with the difficulty level of the Cambridge English qualifying exam – Cambridge English Preliminary (PET) with B1/intermediate difficulty level. Thus, the vocabulary used in the English auditory stimuli was selected from *Objective PET Teacher's Book (4th Edition)* (Hashemi & Thomas, 2013). All sentences were simple sentences consisting of one independent clause without ambiguity. As a result, L2 listening materials with the B1/intermediate difficulty level would not impose too much additional mental workload for the participants to process. As to L1 auditory stimuli, the vocabulary used in the Chinese sentences was selected from the Chinese Proficiency Test (HSK) Level 4 in *Chinese Proficiency Test Syllabus Level 4* (Confucius Institute Headquarters, 2009). HSK was a standardized test of Chinese language proficiency for speakers of other languages and Level 4 meant the intermediate difficulty level. The reason for selecting vocabulary from HSK Level 4 was to make sure the difficulty levels of L1 and L2 auditory stimuli were similar. Since the two tests were standardized language proficiency exams with the intermediate difficulty level, listening materials with vocabulary from HSK Level 4 and PET were used as the auditory signals in the current study.

All the Chinese and English listening materials were pronounced naturally by two male native speakers (one Chinese and one English) and recorded at a 44.1 kHz sampling rate in a 32-bit stereo *.wav file by using Adobe Audition CC (Version 11.1.0)[1], which was a comprehensive software tool for audio recording and editing. Chinese and English sentences were pronounced at a rate of 200 words per minute approximately.

---

[1] Adobe Audition CC (Version 11.1.0) is a comprehensive toolset that includes multitrack, waveform, and spectral display for creating, mixing, editing, and restoring audio content. Retrieved from adobe.com/products/audition

The current study created signals as listening materials in the experiment, in which the auditory stimuli included filtered and/or unfiltered sounds in both ears, categorized into four configurations of stimuli: (a) filtered stimuli in both channels (FL-FR); (b) filtered stimuli in the left channel and unfiltered stimuli in the right channel (FL-R); (c) unfiltered stimuli in the left channel and filtered in the right channel (L-FR); (d) unfiltered stimuli in both channels (NL-NR). All auditory stimuli were edited by using Adobe Audition CC (Version 11.1.0). Filtering was performed with a cut-off frequency set at 320 Hz as per standard practice in verbotonalism. This meant that frequencies of filtered stimuli above 320 Hz were removed from the audio recording and only frequencies of 320 Hz or below were maintained. As to the unfiltered stimulation, all the frequencies were preserved. An example of the stimulus with FL-R configuration is shown in Figure 3.1. All unfiltered stimuli were normalized to 70% in amplitude, and all filtered signals were normalized to 85% in amplitude to guarantee equal intensity since filtered sounds were limited in bandwidth compared to the unfiltered ones (Meyer et al., 2002).

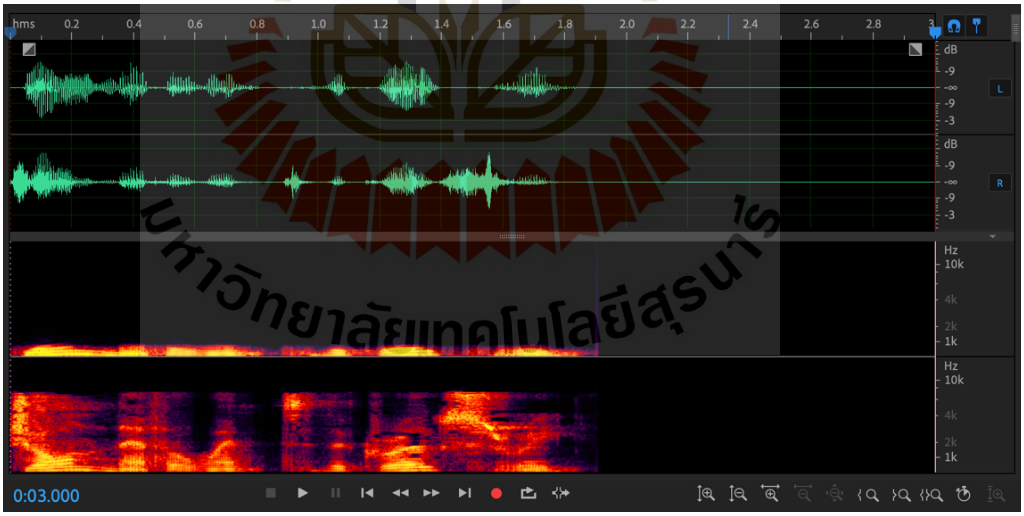


Figure 3.1 Spectrogram of the FL-R stimulus "John never went to the grocery."

According to four configurations of auditory stimuli: (a) FL-FR, (b) FL-R, (c) L-FR, and (d) NL-NR, there were four sets of stimuli in L1 and four sets in L2, namely L1 FL-FR, L1 FL-R, L1 L-FR, L1 NL-NR, L2 FL-FR, L2 FL-R, L2 L-FR, and L2 NL-NR. In each set of stimuli, there were eight sentences corresponding to eight sentence types in Chinese and English as mentioned above. Further, these eight sentences in one stimulus set (e.g., L1 FL-FR) differed from the sentences in another set, but the sentence types, together with the intonation patterns, were consistent among sets of stimuli in each language. The reason for doing so was to better spot the differences of prosodic processing mechanisms for different configurations of auditory language signals, meanwhile, to try to avoid interference of participants' prediction of the incoming signals during the experiment since they would listen to different language signals in different experimental sessions. However, as the research focus, the sentences with the same sentence types and similar intonation patterns remained consistent in L1 and L2 stimulus sets. Four sets of L1 auditory stimuli with eight sentences in each set and four sets of English sentences with eight sentences in each were presented respectively in the eight runs of ERP recording and fMRI scanning. Thus, signals in Verbotonal-based dichotic and diotic listening conditions were created to identify an optimal language signal for processing among the four configurations of stimuli in L1 and L2.

To identify an optimal auditory language signal, the between-group comparisons among the four configurations of stimuli for L1 and L2 were made in the current study. The same group of participants took part in the combined ERP and fMRI experiment. During the experiments, L1 and L2 sentences (i.e., the auditory stimuli) in each stimulus set (e.g., L1 FL-FR) were presented randomly to make sure that the participants did not know which sentence would be presented next. The stimulus presentation paradigm in the experiments will be explained in detail in the next section. L1 and L2 sentences, used as auditory stimuli in the present study, are shown in Table 3.2.

Table 3.2 L1 and L2 sentences used as auditory stimuli

L1 sentences			
FL-FR	FL-R	L-FR	NL-NR
她今天下午要上历史课。 She is going to take a history class this afternoon.	爸爸每天开车上下班。 Dad drives to work every day.	他要给爷爷买一个蛋糕。 He wants to buy a cake for grandpa.	我下个月要去北京出差。 I'm going on a business trip to Beijing next month.
我不想去办信用卡了。 I don't want to apply for a credit card.	他们不愿意跟团旅行。 They are not willing to travel with tourist groups.	妈妈不满意小明的成绩。 Mom is not satisfied with Xiao Ming's exam results.	我同学都不喜欢唱歌。 My classmates don't like singing.
他们都是大学生吗？ Are they college students?	你们去过上海吗？ Have you been to Shanghai?	他吃过北京烤鸭吗？ Has he eaten Peking duck?	你下课要回宿舍吗？ Are you going back to the dormitory after class?
你是怎么去图书馆的呢？ How did you get to the library?	妈妈什么时候下班呢？ When does mom leave work?	你手里拿的是什么呢？ What are you holding?	小王家住在哪儿呢？ Where does Xiao Wang live?
难道你没有听天气预报吗？ Don't you listen to the weather forecast?	你不是知道这星期不上课吗？ Don't you know that there will be no classes this week?	这样放弃不觉得可惜吗？ Isn't it a pity to give up like this?	难道你没有给手机充电吗？ Didn't you charge the phone?
你喜欢文学还是喜欢历史？ Do you like literature or history?	你跑得快还是他跑得快？ Can you or he run faster?	这本书好还是那本书好？ Is this book better or that book?	晚饭出去吃还是在家吃？ Eating out or eating at home?
要好好听老师的话。 Listen carefully to what the teacher says.	请给我们多提意见。 Please give us more opinions.	这个问题你来回答。 Please answer this question.	你们可得抓紧时间。 You have to hurry up.
多么聪明的姑娘呀！ What a clever girl she is!	多么晴朗的天呀！ What a fine day!	他跑得真快啊！ How fast he can run!	谁能想得到啊！ Who would have thought of it!

Table 3.2 L1 and L2 sentences used as auditory stimuli (Cont.)

L2 sentences			
FL-FR	FL-R	L-FR	NL-NR
They went to the café last week.	I went to the bookstore yesterday.	She went to the market this morning.	He went to the station just now.
Paul did not call me yesterday.	John never went to the grocery.	Dan does not like online courses.	Lynn could not find any paper.
Is she working very hard?	Were they traveling together?	Did you go to the concert?	Have they visited here before?
Where are the keys to the back door?	What have they decided to do next?	How did the accident happen?	What time are your friends coming?
Karen plays the piano, doesn't she?	Mary will be here soon, won't she?	Jim should pass the exam, shouldn't he?	There was a lot of traffic, wasn't there?
Would you like some tea or coffee?	Do you like to play football or basketball?	Does she work a full-time or a part-time job?	Will they buy a house or rent somewhere?
Open the window a little more, please.	Just give me a minute, please.	Everybody sit down, please.	Turn on the computer, please.
What a beautiful day it is!	What good luck they had!	How wonderful it is to see you!	How beautifully she sang!

In terms of the stimuli, the mean number of syllables in L1 sentences was 9 (SD ± 1.39) and the mean length of L1 audio signals was 1675.63 (SD ± 324.33) ms. For L2 stimuli, the mean number of syllables in the sentences was 8.19 (SD ± 1.20) and 1904.69 (SD ± 351.44) ms for the mean length.

3.3.2 Stimulus presentation paradigm

3.3.2.1 Stimulus presentation paradigm for the ERP experiment

Stimulus presentation for the ERP experiment was controlled by E-Prime 3.0 (https://pstnet.com/products/e-prime), which was a comprehensive software for computerizing the experimental design, data collection, and analysis in psychological, cognitive, and behavioral experiments. There were eight runs of ERP recording corresponding to the eight sets of auditory language stimuli with four auditory configurations in L1 and L2 (i.e., L1 FL-FR, L1 FL-R, L1 L-FR, L1 NL-NR, L2 FL-FR, L2 FL-R, L2 L-FR, and L2 NL-NR), which meant one set of stimuli in one language was examined in each run. Eight sets of L1 and L2 stimuli were presented to the participants in eight separate runs of ERP recording. In one run of the ERP experiment, there were eight Chinese/English sentences of eight sentence types (as discussed in the previous section). Each sentence was repeated six times, therefore, there were forty-eight auditory stimuli in total in one run of the ERP experiment. These forty-eight auditory language signals were presented randomly and continuously to the participants during each run as shown in Figure 3.2. One auditory stimulus was one sentence with one certain auditory configuration in L1/L2. All auditory stimuli were designed to be 3 seconds in length, during which the sentence signals lasted 1600 ms on average, followed by silence for around 1400 ms. Since the current study concentrated on sentential processing, the ERP components elicited within 1000 ms after the last word of each sentence were analyzed for the research purpose. One run of recording for one set of language signals took 144 seconds, thus, the whole ERP experiment took around 40 minutes.

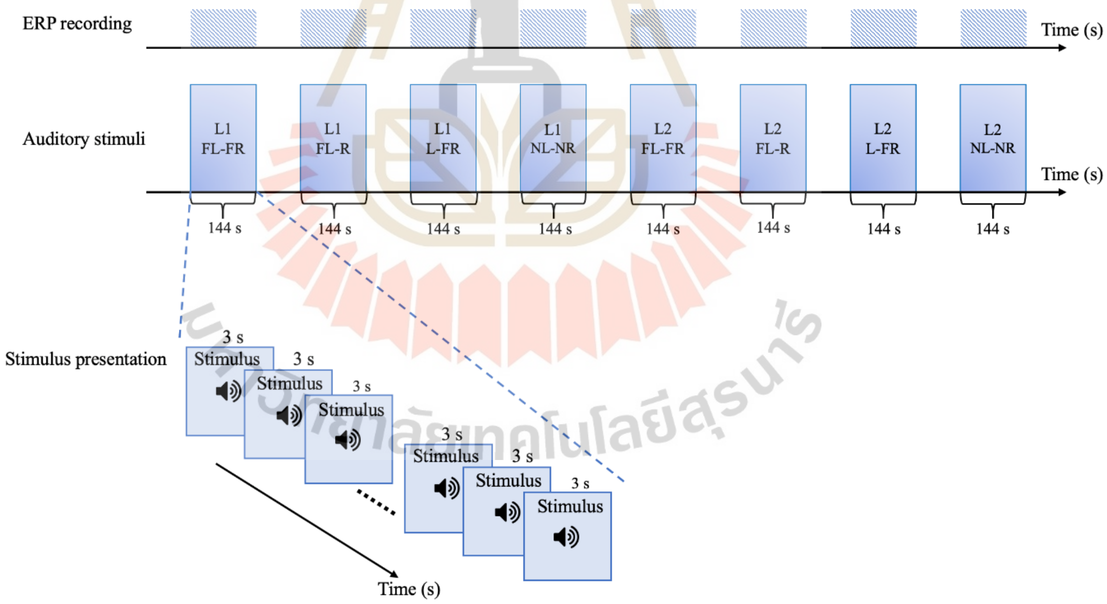


Figure 3.2 ERP experimental design

3.3.2.2 Stimulus presentation paradigm for the fMRI experiment

For the stimulus presentation in the fMRI experiment, a block design was adopted as the paradigm. The block design was one of the major paradigms for fMRI experiments, in which *"a condition is presented continuously for an extended* *time interval (block) to maintain cognitive engagement, and different task conditions are usually alternating in time"* (Tie et al., 2009, p. T108). A number of advantages and benefits that the block design conveyed to the current study were as follows. The block design showed efficiency to get an adequate signal-to-noise ratio via collapsing across trials (Bandettini et al., 1993; Dale & Buckner, 1997). As a result, it demonstrated the strength of robustness that task blocks induced changes in the blood-oxygen-level-dependent (BOLD) signal compared with the baseline in rest blocks (Brockway, 2000; Buxton et al., 1998; Rombouts et al., 1997), meanwhile, it strengthened the statistical analysis due to a large number of trials (Friston et al., 1999). Another advantage that was important for the present study was that the block design was innately suited for the detection of the brain regions activated by particular tasks/stimuli compared to other paradigms (Donaldson, 2004; Petersen & Dubis, 2012). Due to the reasons above, the block design was adopted as the stimulus presentation paradigm to investigate the activated cerebral regions that were induced by different configurations of auditory language signals in the current study.

According to the eight sets of auditory language signals (i.e., L1 FL-FR, L1 FL-R, L1 L-FR, L1 NL-NR, L2 FL-FR, L2 FL-R, L2 L-FR, and L2 NL-NR), there were eight runs of fMRI scanning. Each run of fMRI scanning examined one set of auditory language signals. The block design in the fMRI experiment contained rest and stimulus blocks in each run, in which one rest/stimulus block lasted 24 seconds. Six rest blocks and six stimulus blocks were presented continuously and alternately, which took 288 seconds in each run of fMRI scanning for each set of auditory language signals. The fMRI experiment took 40 minutes approximately. The block design is shown in Figure 3.3.

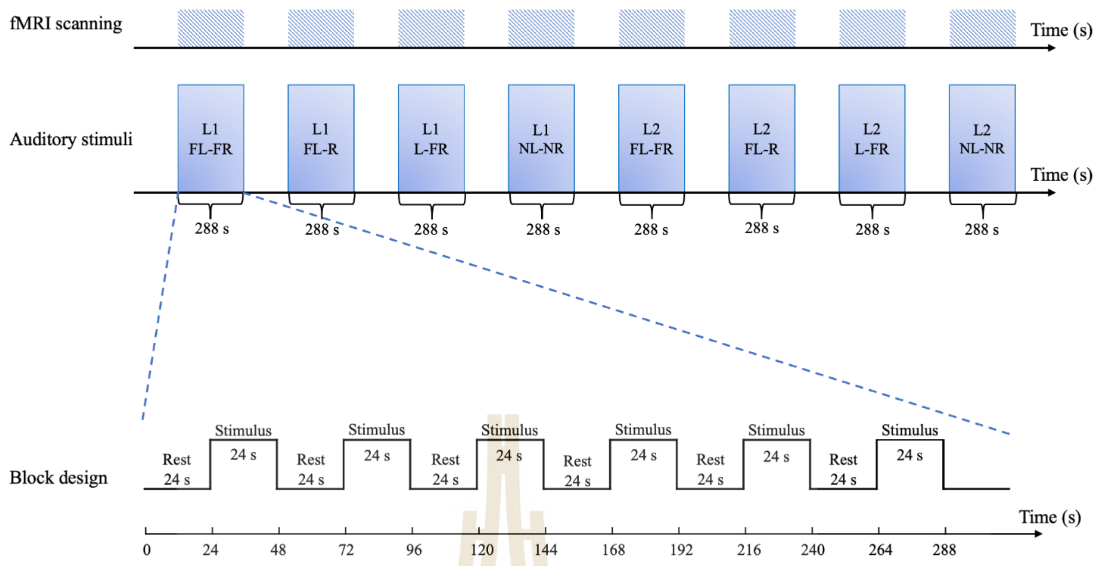


Figure 3.3 Block design for the fMRI experiment

A twelve-block design, i.e., six rest blocks and six stimulus blocks, was adopted as the paradigm (as illustrated in Figure 3.4A). In each stimulus block, there were eight sentences corresponding to eight sentence types. That meant one stimulus block contained eight auditory trials, and each auditory trial was 3 seconds, thus, a stimulus block lasted 24 seconds. Each of the eight sentences was presented six times in six stimulus blocks respectively, in addition, the presentation order of the eight stimuli in each stimulus block was totally random. Between stimulus blocks, there was 24-second silence designed to be rest blocks as the baseline. Therefore, six stimulus blocks contained forty-eight sentences, i.e., forty-eight auditory trials, and 24-second rest blocks were intervals between the stimulus blocks. The stimulus-rest block design for the fMRI experiment is shown in Figure 3.4B.

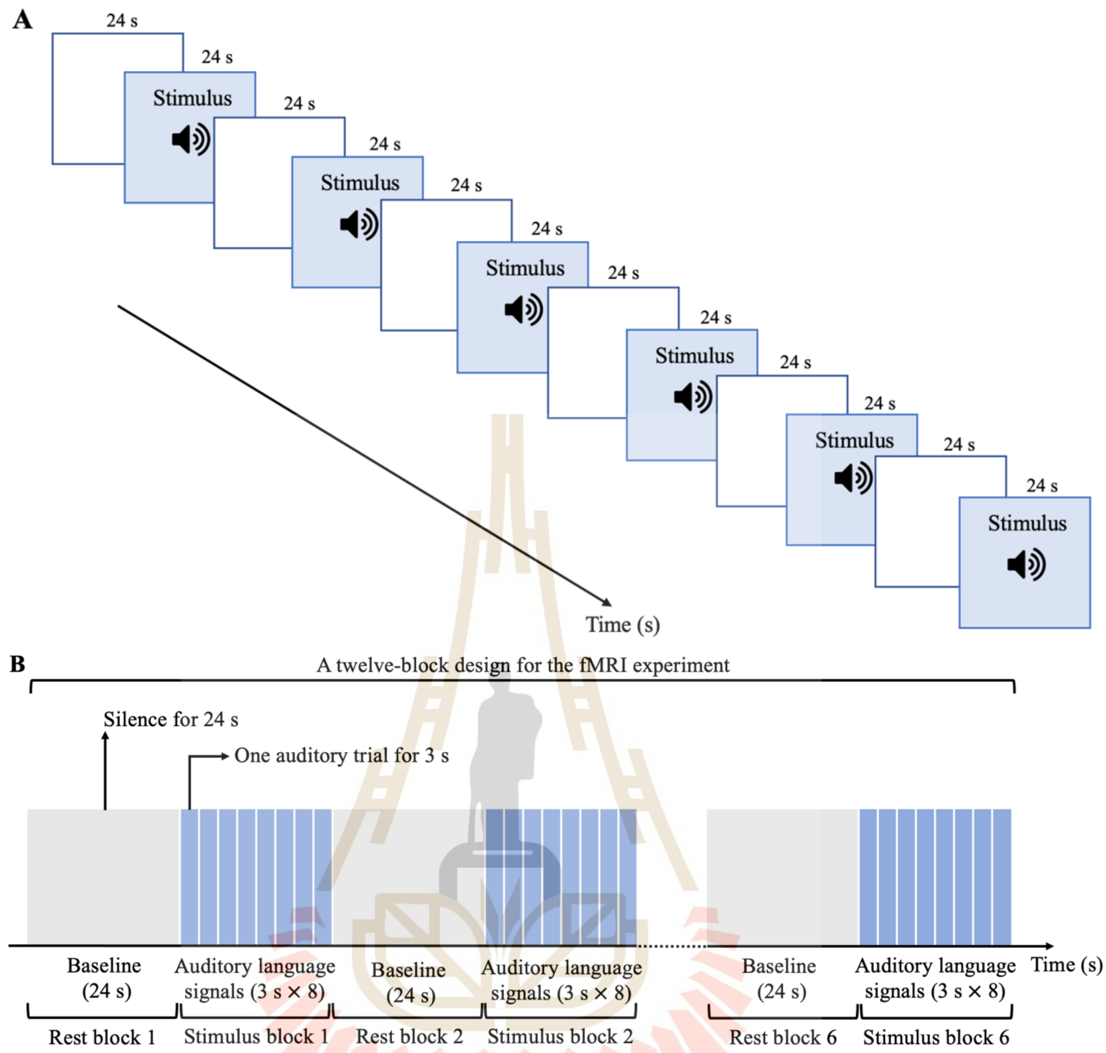


Figure 3.4 A twelve-block design for the fMRI experiment

3.4 Experimental procedures

ERP recording and fMRI scanning were performed sequentially in two separate experiments, at least five days apart. Since the listening materials in ERP and fMRI experiments were the same sentences, two experiments were separated by at least five days. Thus, fMRI results would not be influenced by the previously-conducted ERP experiment as the participants might forget what they had heard in the ERP experiment. ERP and fMRI experiments included eight runs of recording/scanning corresponding to eight sets of auditory stimuli with four filtering/unfiltering configurations in L1 and L2 as mentioned in 3.3.2. It took around 40 minutes for ERP recording/fMRI scanning.

Quantitative analysis of ERP and fMRI data would reveal the temporal-spatial neural signatures for processing the auditory language signals, and uncover cognitive load while listening to different signals. After the combined ERP and fMRI experiment, semi-structured interviews were conducted. Qualitative interviews allowed participants to express their feelings after listening to the signals and views on using the signals for learning a language. With quantitative and qualitative data taken into account, an optimal auditory language signal for Chinese university EFL students would be unveiled. Figure 3.5 illustrates the experimental procedures of the current study.

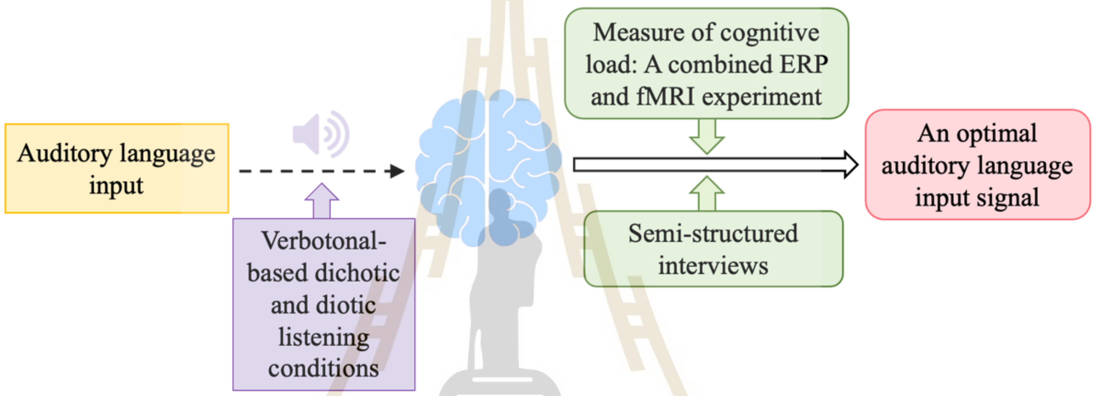


Figure 3.5 Experimental procedures

3.5 Data collection

3.5.1 ERP recording

To collect ERP data, participants were seated comfortably in a quiet room, wearing an electrode cap (Greentek Gelfree-S3 cap, China) and headphones to enable the participants to hear the signals. During the experiment, participants were asked to listen carefully to the stimuli and close their eyes to avoid blinks. Meanwhile, data were recorded from scalp electrodes based on the standard electrode placement – International 10-20 System (Klem et al., 1999) at sixteen sites: F1, F2, FC5, FCz, FC6, T7, C5, C6, T8, CP3, CPz, CP4, P7, P8, P9, and P10 (as shown in Figure 3.6). These sites recorded brain activity from frontal (F), temporal (T), parietal (P), and central area (C). The reason for choosing these sites was that the sixteen sites had been proved to be the optimal 16 channels and positions of the electrodes to measure neural activity of speech (Montoya-Martínez et al., 2019). All electrode impedances were kept below 5 kΩ before data collection. The EEG data were analyzed by using EEGLAB (Version 15.0.0b)[2]. 30-second recording of resting-state EEG data before eight runs of the ERP experiment were used as a control (baseline) signal. The data collection for each auditory trial lasted for 3000 ms. As the research focus of the current study was the processing of sentential stimuli, the ERP components elicited within 1000 ms after the end time of the last word in the sentences were analyzed by investigating their latencies and amplitudes. The off-line data were filtered with a band-pass filter of 0.1-70.0 Hz and artifacts above 100 μV were rejected.

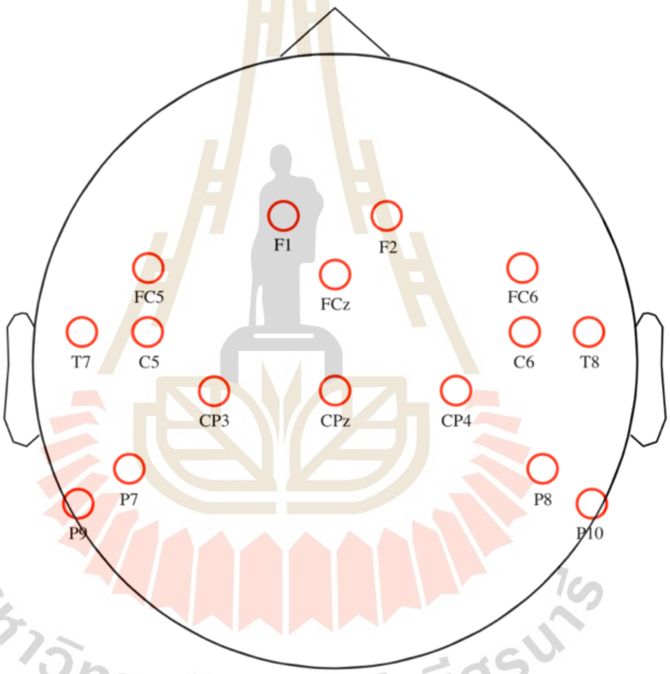


Figure 3.6 Location of electrode sites

3.5.2 fMRI scanning

All images were acquired by using a General Electric MR750w 3.0T MRI scanner (GE, USA). Participants lay in the scanner with their head position secured with foam padding, wearing MRI-compatible pneumatic in-ear headphones. An optical response box was placed in each participant's dominant right hand and a compression

[2] EEGLAB is an interactive Matlab toolbox for processing continuous and event-related EEG, MEG, and other electrophysiological data. Retrieved from https://sccn.ucsd.edu/eeglab/index.php

alarm ball in the left hand. The head coil was positioned over each participant's head to avoid the motion during scanning. During scanning, there was no task for the participants. The participants were asked to be relaxed and listen carefully to each signal.

In order to get a high resolution, T1 weighted 3D images were obtained via a 3D magnetization-prepared GRE sequence with the following parameters: TR = 8.5 ms, TE = 3.2 ms, flip angle = 12°, matrix size = 256 × 256, field of view = 256 mm × 256 mm, slice thickness = 1 mm, gap = 0 mm, number of slices = 148, resulting voxel size = 1 mm × 1 mm × 1 mm. Gradient echo (GRE) localizer images were acquired to determine the placement of the functional slices. For the functional images, a susceptibility-weighted single-shot echo planar imaging (EPI) method with blood oxygenation level-dependency (BOLD) was used with the following scan parameters: TR = 2000 ms, TE = 30 ms, flip angle = 90°, matrix size = 64 × 64, field of view = 224 mm × 224 mm, slice thickness = 3.0 mm, gap = 1 mm, number of slices = 36. These parameters led to a 3.5 mm × 3.5 mm × 4.0 mm voxel size. By using an interleaved bottom-to-top sequence, 72 whole-brain volumes were acquired for each run. The acquisition of the anatomical scan and eight runs of fMRI scanning took approximately 40 minutes.

3.5.3 Semi-structured interviews

The interview, as the gold standard for qualitative research (Silverman, 2018), is one of the basic and most widely used methods to investigate individuals' opinions, feelings, and beliefs about their experiences and situations (Ary et al., 2018; Heigham & Croker, 2009). Instead of testing hypotheses, researchers employed interviews to understand participants' experiences and how individuals make meaning of them (Ary et al., 2018). Three types of interviews are the structured interview, the open interview, and the semi-structured interview (Heigham & Croker, 2009). The questions in the structured interview are precisely designed, from which a researcher tends to get specific information without much variation. On one hand, the controlled form of the structured interview is accurate to gain the answers to research questions and easy to compare between respondents; on the other hand, it lacks in-depth and rich information (Heigham & Croker, 2009). In the open interview, on the contrary, questions are unplanned, unstructured, and not pre-determined, which provides rich information. But it is difficult for a researcher to compare, and it is possible that interviewees' responses do not answer what they are being interviewed or cannot fulfill the purpose of the interview. Between the structured and the open interview, the semi-structured interview allows the interviewer to formulate questions ahead of time and modify them during the interview. In addition, questions are designed to be open-ended in order to reveal interviewees' views about their experiences for the study purpose. The interviewer in the semi-structured interview knows the topics that should be covered, meanwhile, probes in-depth information due to the great flexibility of the semi-structured interview (Ary et al., 2018; Heigham & Croker, 2009). The current study not only employed ERP and fMRI as physiological measures to reveal the neural representation of the signal processing but adopted the semi-structured interview to investigate participants' opinions on different auditory language signals after listening, which could not be measured in the combined ERP and fMRI experiment.

Based on the objectives of the current study, the topics that should be covered were determined, then guiding questions in semi-structured interviews were formulated as outlined in Appendix C. The main questions included the feelings when listening to the signals, the signal(s) that would help/hinder understanding of the sentence meanings and structures, the signal(s) that students preferred or disliked, and views on using the signals for language learning. In order to ensure that the questions were relevant to the research objectives and clear enough to understand, the Item Objective Congruence (IOC) index was adopted to validate the interview questions of the current study. Five experts, who had experience in EFL research and had been teaching English for over five years, did the IOC analysis (see Appendix D). After the IOC analysis, revisions were made according to experts' comments and suggestions to improve the quality of the semi-structured interview.

All the interviews were conducted after the combined ERP and fMRI experiment because they had listened to all kinds of stimuli and were familiar with four kinds of signals in L1/L2. Interviews were carried out in participants' native language – Chinese as they could better understand the questions and express their views more freely without language barriers. Each participant was asked about the feelings of listening to the signals after the combined ERP and fMRI experiment, and a focus group of twelve participants who had multiple points of view on the signals and were enthusiastic about what they had listened to was selected as interviewees in the semi-structured interview. All the interviews were recorded for further transcription and analysis with the interviewees' permission.

3.6 Data analysis

3.6.1 ERP data analysis

The measures of mean peak amplitudes, latencies, and the area under the curve (AUC) of the peak of interest were analyzed after obtaining the ERP data.

For the amplitude analysis, the peak amplitudes of both negative- and positive-going deflections around 400 ms and 600 ms were taken into consideration. The N400 was sensitive to language manipulations especially for semantic processing, which became smaller when the information was more predictable or easier to process (Kutas & Federmeier, 2011). The P400 was identified as the semantic integration of sequential representations and an expectancy process for sequential probabilities (Dien, 2009; Dien et al., 2010; Liu et al., 2009). The syntactically more difficult or less-preferred sentences evoked larger P600 amplitudes compared to sentences that were easier to parse or in some other way preferred (Swaab et al., 2012). In addition, the N600 reflected the explicit interpretation of stimulus semantics and a stimulus-general process that involved non-specific cognitive processes such as attention and working memory (Cummings et al., 2006; Shahin et al., 2006). The peak amplitudes were reliable local negative and positive maximums in the detected ERP deflections (Kappenman & Luck, 2011), in which the waveforms around 400 ms and 600 ms reflected semantic and syntactic manipulations in the brain (Cummings et al., 2006; Dien et al., 2010; Kutas & Federmeier, 2011; Liu et al., 2009; Shahin et al., 2006; Swaab et al., 2012).

As to the latency analysis, language ERPs occurred relatively late and were long-latency components, possibly due to multiple brain areas contributing to their generation (Swaab et al., 2012). The peak latencies of the negative and positive waveforms around 400 ms and 600 ms were analyzed for reaction time after the end time of the last word in the sentences. The reaction time reflected stimulus evaluation processes, and it also reflected the relative timing of response processes (Swaab et al., 2012).

The measure of AUC reflected not only the maximal value of neuronal resources but also the amount of neural processing of the signal over a span of time (Mahmoudian et al., 2013; McDowell et al., 2003). The AUC was calculated according to the formula (1).

$$S = \int_a^b F(t)dt \quad (1)$$

Among them, *S* is the desired area, *a* stands for the start point of the curve, *b* refers to the endpoint of the curve, and *F(t)* represents the difference wave.

In addition, studies showed that the N400, P400, N600, and P600 effects were maximal at midline centroparietal sites, indicating a typical centro-parietal or fronto-central scalp distribution (Brouwer et al., 2012; Brouwer & Hoeks, 2013; Dien et al., 2010; Shahin et al., 2006; Swaab et al., 2012; van Herten et al., 2005). Thus, the measures of amplitudes, latencies, and AUC of the ERP components within 1000 ms after the stimulus presentation at the midline sites (Cz and Pz) were taken for main analyses in this study.

ERP data analysis included the individual analysis and the group analysis. The individual analysis was adopted to investigate mean peak amplitudes, latencies, and AUC of the ERP components around 400 ms and 600 ms induced by eight types of stimuli in L1 and L2. Each stimulus was calculated and analyzed separately, and the grand average waveforms for each stimulus were plotted. To better compare with each other, the group analysis was used for comparisons between stimuli. Statistical analyses using ANOVA (analysis of variance) tests between eight stimuli at two electrode sites (Cz and Pz) were carried out with SPSS (Version 22.0) to check if the mean amplitudes of the ERP components elicited by all auditory stimuli were significantly different from each other. Further, mismatch negativity (MMN) (defined as the difference wave) illustrating the differences between potentials induced by target stimuli and non-target stimuli were used for the group analysis. The MMN difference wave was a subtraction of the ERP elicited by the non-target stimulus from the ERP elicited by the target stimulus (i.e., target stimuli minus non-target stimuli), which was a versatile method to detect the discrepancy and discriminate the alterations of the ERPs between two stimuli (Beauchemin & De Beaumont, 2005; Cowan et al., 1993; Kappenman & Luck, 2011). As a result, difference waves between stimuli would provide more information on brain responses of language manipulations rather than the absolute voltage values (Morgan-Short & Tanner, 2014). As no assumptions about the target and non-target stimuli were made, all possible difference waves between stimuli were plotted and analyzed. MMN difference waves were analyzed in terms of peak amplitudes and AUC, which were two appropriate and objective ways to analyze the difference waves with high variability (Beauchemin & De Beaumont, 2005). The amplitude and AUC were quantitative measures of the activated neurons involved in the processing of auditory signals (Mahmoudian et al., 2013). The AUC of MMN indexed the neural processing difference elicited by the target stimulus when compared to that of the non-target stimulus (Beauchemin & De Beaumont, 2005). The AUC was calculated by using the formula (1) as mentioned above.

3.6.2 fMRI data analysis

Changes in blood oxygenation level-dependent (BOLD) contrast associated with listening to different types of stimuli were assessed on a pixel-by-pixel basis, using the general linear model (Friston et al., 1995; Friston et al., 1994) and the theory of Gaussian fields (Worsley & Friston, 1995) as implemented in the Statistical Parametric Mapping (SPM12) software[3] (Penny et al., 2011). This method took advantage of multivariate regression analysis and corrected for temporal and spatial autocorrelations in the fMRI data.

The fMRI data were analyzed by using SPM12 running in the Matlab Image Processing Toolbox (R2019a, Natick, MA, USA). All image data were preprocessed by the following procedures shown in Figure 3.7.

---

[3] Statistical Parametric Mapping (SPM12) software is designed to analyze brain imaging data sequences, and to construct and assess spatially extended statistical processes for functional imaging data. Retrieved from https://www.fil.ion.ucl.ac.uk/spm/software/spm12/

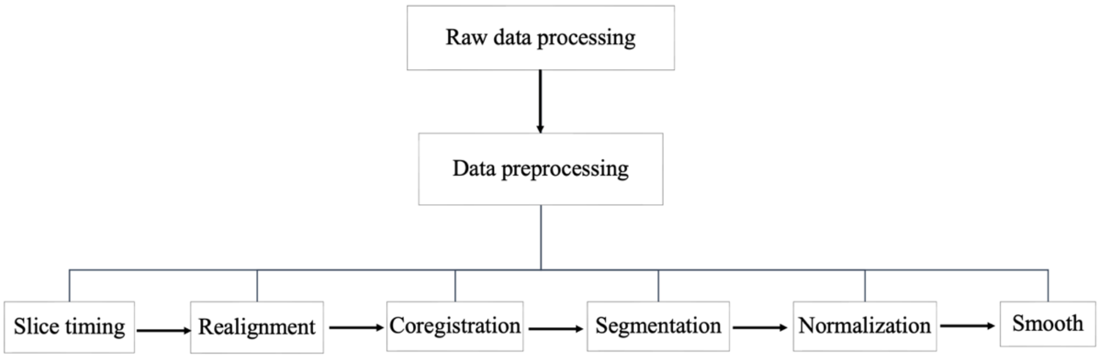


Figure 3.7 Flowchart of BOLD-fMRI data processing

Raw data, the functional images in the DICOM (Digital Imaging and Communications in Medicine) format, were first converted to SPM-compatible image rolls.

(1) Slice timing correction: the method of interleaved slice acquisition was used to create volumes.

(2) Realignment: the rejection threshold for the excessive motion was a three-dimensional motion exceeding 3 mm or a three-dimensional rotation exceeding 3 degrees.

(3) Coregistration: the fMRI images were coregistered to the structural images (T1-weighted anatomical images) after realignment.

(4) Segmentation: segmentation of white matter, gray matter, and cerebrospinal fluid was performed for accurate coregistration.

(5) Normalization: image data were normalized into the Montreal Neurological Institute (MNI) stereotactic space for anatomical localization of the activated brain areas (Talairach & Tournoux, 1988). The voxel size was 3.5 mm × 3.5 mm × 4.0 mm.

(6) Smooth: each functional image was spatially smoothed by using a Gaussian kernel filter of 7 mm × 7 mm × 8 mm full width at half maximum (FWHM) to improve the signal-to-noise ratio.

fMRI data analysis also contained individual and group analyses. The individual analysis was performed by separate one-sample *t*-tests for all stimuli. The purposes of the individual analysis were to identify anatomical localization of significant activation induced by the stimuli and to determine the regions of interest (ROIs) for further analysis. Activations below a threshold of $p < .001$, uncorrected, cluster size ≥ 10, T ≥ 3 were reported. The group analysis was used for comparisons between stimuli, which was conducted by contrasting eight configurations of stimuli by using two-sample *t*-tests. Since there were no assumptions about the directions of effects in the contrasts, all possible contrasts were performed for both directions. All reported results were at uncorrected $p < .006$, cluster size ≥ 20, T ≥ 3.

In addition, the lateralization index (LI) was performed to quantify the degree of language dominance and lateralization in the ROIs observed in the individual analysis (Lee et al., 2010). The formula (2) used for calculating LI was as follows:

$$LI = \frac{(V_L - V_R)}{(V_L + V_R)} \quad (2)$$

Where V_L indicates the number of voxels activated for the left hemisphere and V_R means the number of voxels activated in the right hemisphere (Springer et al., 1999). LIs were classified as left hemisphere language dominant with LI > 0.20, symmetric / no clear hemispheric preference with -0.20 ≤ LI ≤ 0.20, and right hemisphere language dominant with LI < -0.20 (Springer et al., 1999).

The xjView toolbox (Version 9.7; https://www.alivelearn.net/xjview) was adopted to visualize cerebral activations induced by the auditory language stimuli in the current study.

3.6.3 Analysis of semi-structured interviews

Qualitative data collected from semi-structured interviews were analyzed by content analysis, which was a systematic and rule-governed process of summarizing and reporting written data (Cohen et al., 2017). There are some advantages of content analysis. First, content analysis is unobtrusive as what is being observed is not influenced by the observer (Ary et al., 2018; Krippendorf, 2004). And the rules for content analysis are explicit and public (Mayring, 2004). By adopting content analysis, language/linguistic features and meaning in the context can be analyzed systematically and verifiably through codes and categories (Mayring, 2004). In addition, the data in the text can be kept for a long time, which makes re-analysis and replication possible

(Cohen et al., 2017). Since the content analysis has several strengths and can be performed with any written text, content analysis of the interview transcription in the current study is feasible.

Before coding and analyzing the qualitative data sets, the researcher got familiar with the data in the process of transcribing, reading, and re-reading the transcripts. Salient and recurring ideas, especially the feelings and views of listening to different signals, were first identified. Coding was be carried out to make data locatable. The initial step of coding involved the identification of participants' feelings and views. Then codes were assigned to the feelings and views. After coding, an attempt was made to discover categories between the codes. Across the categories, the relationships or patterns were explored to identify themes. Finally, meaning and insights extracted from qualitative data were interpreted and related to the research questions, theories or concepts in the literature review (Creswell & Guetterman, 2019; Heigham & Croker, 2009).

Another trained researcher double-coded a sample of the interview data (approximately 30%) to make sure that the inter-rater reliability for all the data was achieved over 90%. Disagreements were discussed and final decisions were made by two researchers together.

3.7 Ethical considerations

As researchers, ethical issues must be given priority consideration in planning and conducting research and we have obligations to subjects and the profession (Ary et al., 2018). Especially when human subjects are involved in the research, their rights, privacy, dignity, and sensitivities must be respected (Ary et al., 2018). Attention to ethical issues should be paid in every step of conducting research (Creswell & Creswell, 2018). Before carrying out the current study, the researcher sought approval from the university by an institutional review board (IRB), asked for permission from the research site and participants, and negotiated authorship for future publications. At the beginning of doing research, the problem and the purpose of the study were specified so that participants' benefit from the study could be identified. After contacting participants, the researcher told them information about the study, including the general purposes, their rights, procedures, discomfort and risks that might occur, confidentiality, etc. Participants voluntarily took part in the study and signed the consent form (shown in Appendix B). In the process of collecting data, the researcher respected participants and site staff, disrupted them as little as possible, ensured that all participants did the same tasks in the experimental procedure, and avoided collecting sensitive and harmful information. When analyzing data, the researcher reported both positive and negative findings from multiple perspectives, and maintained anonymity and confidentiality of participants, meanwhile, avoided "going native." In terms of reporting, sharing, and storing data, the researcher reported data, findings, and conclusions honestly without information that would do harm to participants, avoided biased and vague language, would keep raw data for five years, would share data with others, and would provide proof regarding ethics and no conflict of interest (Creswell & Creswell, 2018). Thus, the current study was conducted in accordance with the standards and principles of the Declaration of Helsinki (World Medical Association, 2013).

3.8 Pilot study

A pilot study evaluates the appropriateness and practicability of the proposed research procedures especially the data-collection instruments within a small group of participants, in which the researcher can find out any problems that can be solved and make any changes (if needed) to ensure the actual research can be conducted properly (Ary et al., 2018; Fraenkel et al., 2012). It is worth the time and effort to do the pilot study since the actual study will be performed more smoothly if some unanticipated problems can be solved during piloting (Ary et al., 2018).

Two students took part in the pilot study. The combined ERP and fMRI experiment went well and the researcher did not encounter unexpected problems during the process of data analysis. Thus, no changes were made for the experimental design in the combined ERP and fMRI experiment. Results of the pilot study were reported as follows.

3.8.1 Results of the pilot study

3.8.1.1 ERP results of the pilot study

Table 3.3 listed mean peak amplitudes and latencies of the ERP components elicited by the L1 and L2 stimuli at the Cz and Pz electrode sites in the pilot study. Grand average ERP waveforms in response to the auditory stimuli at the Cz and Pz electrode sites were plotted in Figure 3.8.

Table 3.3 Mean peak amplitudes and latencies of the ERPs for the L1 and L2 stimuli at the Cz and Pz electrode sites in the pilot study

			300-500 ms		500-900 ms	
	Stimulus	Electrode site	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Latency (ms)
L1	FL-FR	Cz	-0.35	406	0.56	684
		Pz	-0.61	408	1.02	668
	FL-R	Cz	-1.96	662	3.17	976
		Pz	-2.69	682	4.99	878
	L-FR	Cz	-2.38	504	4.77	878
		Pz	-3.20	558	5.14	974
	NL-NR	Cz	-1.47	486	2.30	716
		Pz	-1.47	480	2.28	750
L2	FL-FR	Cz	-0.17	456	-0.03	754
		Pz	-0.17	424	-0.03	776
	FL-R	Cz	-0.14	530	-0.01	694
		Pz	-0.12	740	0.04	688
	L-FR	Cz	-0.23	742	1.84	998
		Pz	-0.15	604	0.08	982
	NL-NR	Cz	-0.03	610	0.05	774
		Pz	-0.03	602	0.05	786

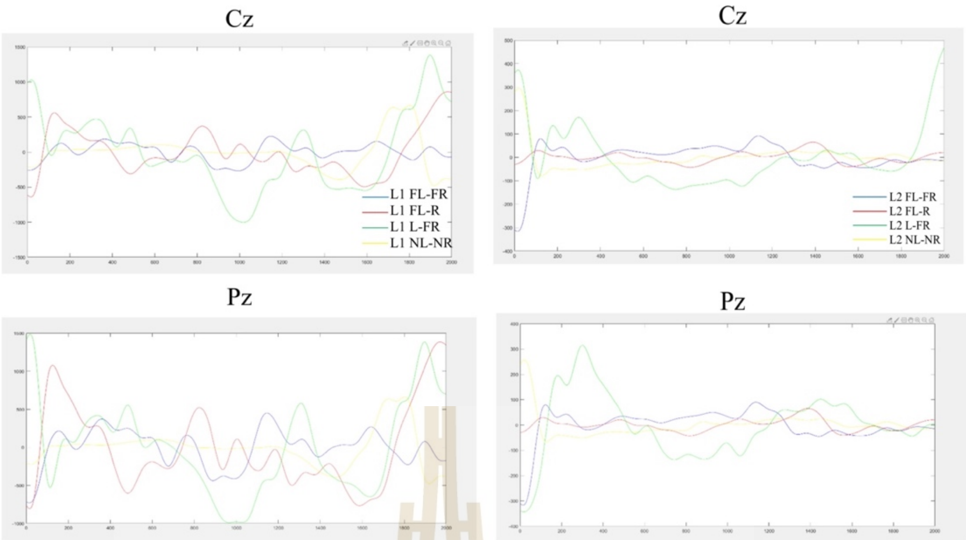


Figure 3.8 Grand average ERP waveforms in response to the auditory stimuli at the Cz and Pz electrode sites in the pilot study

3.8.1.2 fMRI results of the pilot study

Activations below a threshold of *p* < .001, uncorrected, cluster size ≥ 10, T ≥ 3 were reported in Table 3.4.

Table 3.4 Brain activations for the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions in the pilot study

Stimulus	Brain region	BA	k	T	MNI coordinates		
					x	*y*	*z*
L1							
FL-FR	L STG	42	117	9.16	-67	-27	10
	L IPL	40	42	8.84	-63	-44	22
	L MFG	46	10	4.11	-50	42	26
	R STG	32	138	10.5	62	-35	6
	R MTG	22	27	9.34	69	-33	1
	R MFG	46	81	8	52	49	22
FL-R	L STG	42	108	7.91	-63	-22	10
	L IPL	40	25	7.34	-58	-44	22
	R STG	22	19	5.49	62	-12	4
L-FR	L STG	22	139	7.56	-61	4	-6
	L IPL	40	32	7.3	-65	-43	23

Table 3.4 Brain activations for the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions in the pilot study (Cont.)

	R STG	42	141	7.67	60	-32	6
	R PoCG	40	48	3.87	65	-24	19
	R MTG	22	20	7.11	65	-32	2
NL-NR	L STG	42	80	6.87	66	-31	6
	R STG	42	17	4.79	59	-35	6
L2							
FL-FR	L STG	42	71	8.88	-65	-23	10
	R STG	22	104	9.85	62	-35	6
	R PoCG	40	34	6.66	66	-22	19
	R MTG	22	23	8.84	66	-30	4
	R IPL	40	18	6.68	48	-53	54
	R MFG	46	63	7.00	48	53	22
	L MFG	10	25	5.59	-40	53	26
FL-R	R STG	22	109	8.08	66	-9	9
	R PoCG	40	62	6.45	62	-28	18
	R IPL	40	28	6.02	48	-54	54
	L STG	22	104	9.5	62	-6	10
	L IPL	40	56	9.41	-66	-40	22
	R Cerebellum		10	4.89	45	-70	-30
L-FR	L STG	42	106	6.36	-61	-21	10
	L IPL	40	55	10.69	-54	-42	22
	R STG	42	130	9.38	69	-32	10
	R PoCG	40	64	5.6	57	-24	17
NL-NR	L STG	22	109	6.59	-65	-20	5
	L IPL	40	48	8.93	-61	-46	22
	R STG	42	87	7.1	67	-30	6

Notes: Clusters are thresholded at $p < .001$ (uncorrected), $k \geq 10$, $T \geq 3$. Coordinates are given for the stereotactic space of Talairach and Tournoux (1988). BA, Brodmann area; *k*, cluster size (number of voxels); T, t value; L, left hemisphere; R, right hemisphere; STG, superior temporal gyrus; MTG, middle temporal gyrus; MFG, middle frontal gyrus; IPL, inferior parietal lobule; PoCG, postcentral gyrus.

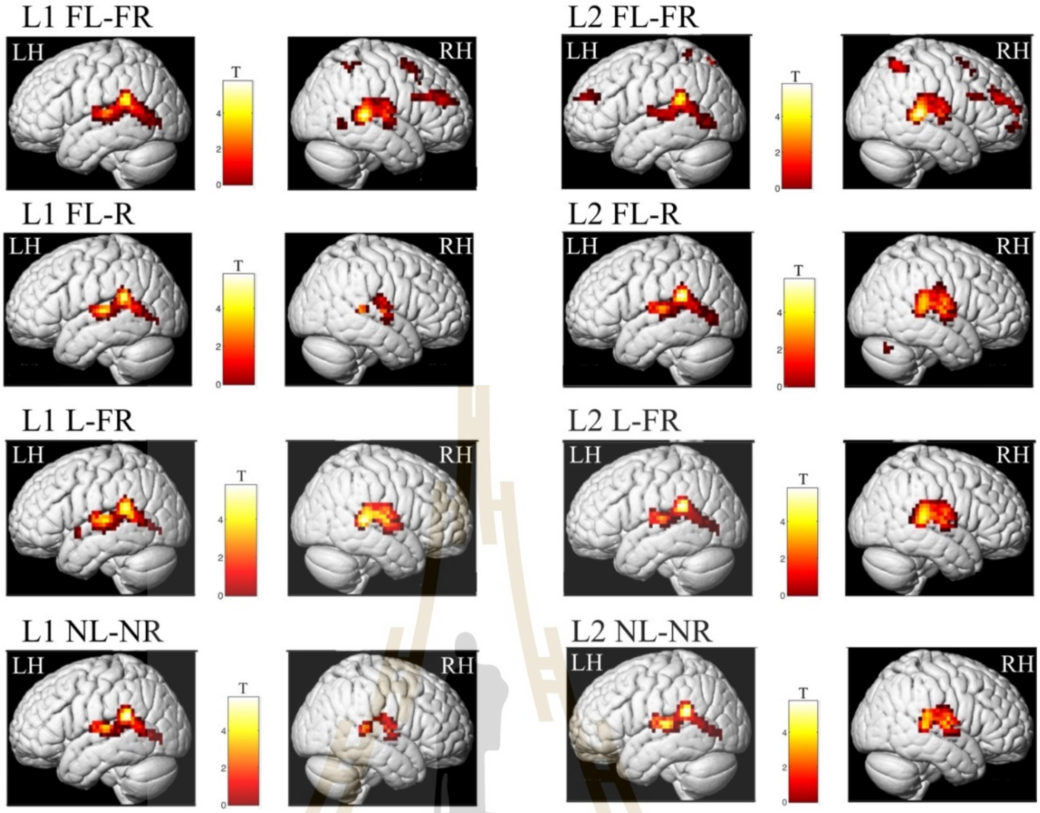


Figure 3.9 Brain activation maps for the L1 and L2 stimuli in the pilot study

3.8.1.3 Results of semi-structured interviews

Two students were asked about their feelings and views after listening to the signals in the pilot study. The signals were perceived differently in the left or right ear. To be specific, students claimed that FL-FR was unclear, but NL-NR was the clearest of all. FL-R and L-FR sounded like monophonic sounds that the sounds came from the right ear for FL-R and from the left ear for L-FR. The feeling of discomfort was not reported.

As for understanding the sentences, students reported that they had difficulties in understanding FL-FR. However, they could understand the meanings and structures of the FL-R, L-FR, and NL-NR signals. Regarding L1 and L2, it was much easier to understand L1 than L2 signals.

Two students showed a curiosity for the listening materials and expressed their interest in using the dichotic signals when learning English. But the students admitted that they almost forgot the signals and the order of the signal presentation when asked about their feelings in the interviews. Thus, the next section will discuss how to improve the experimental processes in the actual study.

3.8.2 How to improve the study

The participants in the pilot study reported that they felt tired and distracted while listening to eight sets of signals and staying still for around forty minutes. Considering the problem presented by the participants, there would be a short break after the first four runs of ERP recording and fMRI scanning, i.e., after four sets of L1 signals were presented. Instead of listening to eight sets of L1 and L2 signals at a time, participants would relax for a while after the presentation of the first four sets of L1 signals and they could talk to the researcher through the intercom in the scanner at break time if needed. The reason for doing so was to ensure the participants to be more attentive during the experiment.

As for the semi-structured interview, they could not remember what they had heard in the experiment. To be specific, they could not remember the order of the signal presentation and the number of signal sets, which caused trouble for them to answer the questions of "which kind of signal…" in the interviews. To make participants recall what they had heard in the experiment and express their views more accurately in semi-structured interviews, eight types of auditory signals would be presented to the interviewee as retrospectives during the interview. The reasons for presenting the signals again in the interview were to recall interviewees' memory and to confirm the configuration/type of the signal that they were referring to, as a result, to conduct more effective interviews.

3.9 Summary

This chapter introduces the research design, participants, the experimental paradigm, data collection procedures, data analysis methods, and ethical considerations regarding the current study. In the last section, the pilot study is presented, including the improvements to be made.

CHAPTER 4
RESULTS

This chapter presents the findings of the current study. The first section provides the quantitative results of the ERP and fMRI experiments. After that, the qualitative results from the semi-structured interviews are given in the second section. At last, a summary of the chapter is made.

4.1 Results of the combined ERP and fMRI experiment

4.1.1 ERP results

4.1.1.1 Results of the individual analysis

Table 4.1 demonstrates the mean peak amplitudes, latencies, and AUC of the ERPs in the 300-500 ms and the 500-900 ms time windows for all auditory stimuli over the centro-parietal sites (recorded from the Cz and Pz electrode sites). Figure 4.1 shows grand average ERP waveforms in response to different signals over the centro-parietal sites.

Table 4.1 Mean peak amplitudes, latencies, and AUCs of the ERPs for the L1 and L2 stimuli at the Cz and Pz electrode sites

Stimulus		Electrode site	300-500 ms Amplitude (µV (± SD))	300-500 ms Latency (ms)	300-500 ms AUC	500-900 ms Amplitude (µV (± SD))	500-900 ms Latency (ms)	500-900 ms AUC
L1	FL-FR	Cz	1.08 (0.71)	373	1.30	0.35 (0.42)	589	1.04
		Pz	-0.81 (0.78)	415	1.59	2.26 (0.53)	620	4.66
	FL-R	Cz	0.37 (0.66)	393	0.05	2.11 (0.64)	586	9.51
		Pz	1.85 (0.74)	461	0.86	0.72 (0.69)	619	0.54
	L-FR	Cz	-0.79 (0.66)	414	2.40	-0.49 (0.67)	611	0.94
		Pz	-2.41 (0.72)	391	1.60	1.07 (0.58)	621	0.62
	NL-NR	Cz	0.50 (0.64)	416	0.18	-0.21 (0.46)	622	0.28
		Pz	-2.38 (0.73)	401	2.90	1.48 (0.68)	627	1.16

Table 4.1 Mean peak amplitudes, latencies, and AUCs of the ERPs for the L1 and L2 stimuli at the Cz and Pz electrode sites (Cont.)

Stimulus		Electrode site	300-500 ms Amplitude (μV (± SD))	300-500 ms Latency (ms)	300-500 ms AUC	500-900 ms Amplitude (μV (± SD))	500-900 ms Latency (ms)	500-900 ms AUC
L2	FL-FR	Cz	-0.19 (0.72)	421	0.30	0.12 (0.56)	625	2.13
		Pz	-0.67 (0.73)	363	1.83	0.83 (0.67)	573	0.29
	FL-R	Cz	-0.06 (0.77)	462	0.52	-0.49 (0.47)	570	0.24
		Pz	-2.74 (0.74)	377	1.22	1.29 (0.63)	620	0.39
	L-FR	Cz	0.26 (0.80)	376	0.94	0.19 (0.71)	572	0.77
		Pz	-2.05 (0.73)	365	2.16	0.53 (0.64)	642	0.54
	NL-NR	Cz	-0.08 (0.84)	451	0.99	0.15 (0.79)	589	0.99
		Pz	-1.37 (0.73)	423	1.64	1.09 (0.84)	588	0.33

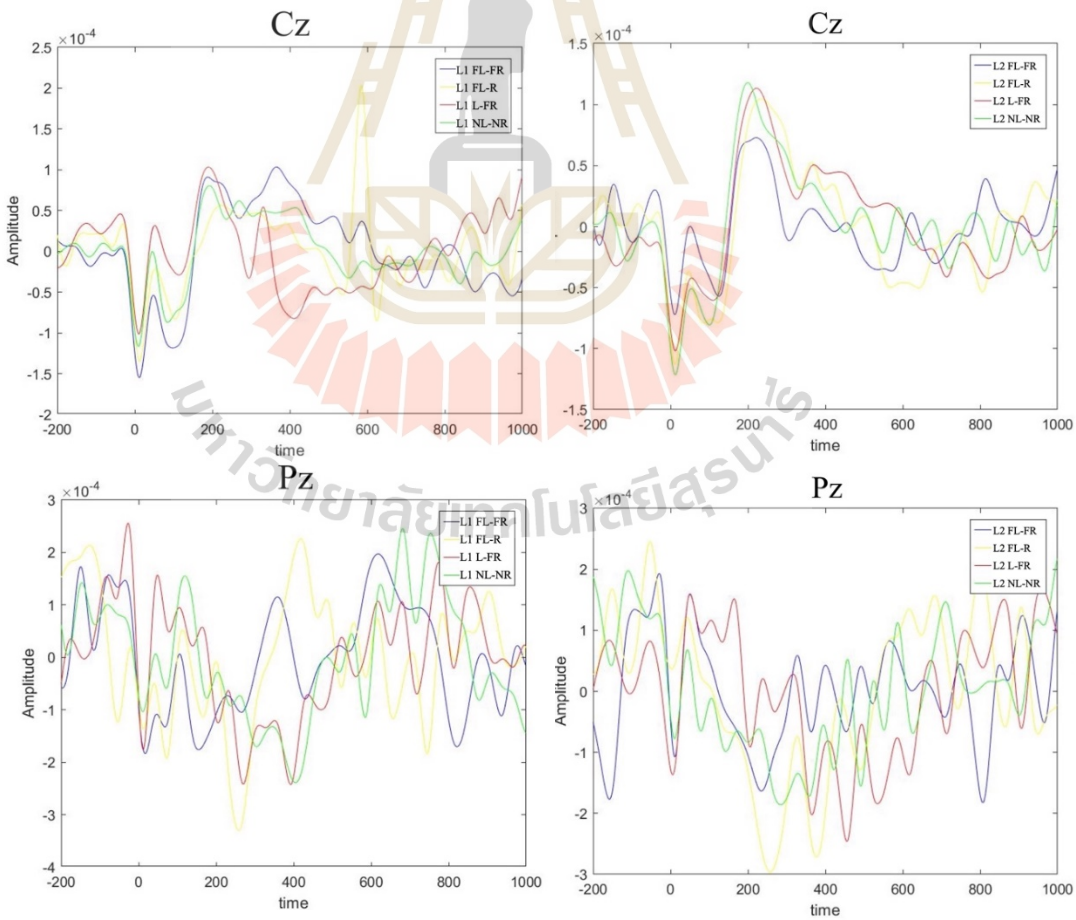


Figure 4.1 Grand average ERP waveforms in response to the auditory stimuli at the Cz and Pz electrode sites

(1) L1

In the 300-500 ms time window, the positive-going waves peaking around 400 ms were elicited by FL-FR, FL-R, and NL-NR signals at the Cz site, indicating a P400 component at the Cz site (the P400Cz is related to the semantic integration of sequential representations and expectancy processes). Only the L-FR signal elicited a negative-going deflection at the Cz site, indicating an N400 at the Cz site (the N400Cz is related to the semantic manipulation). Regarding the processing load and latencies, the mean peak amplitudes showed that FL-FR induced the highest amplitude 1.08 (SD ± 0.71) μV with the shortest reaction time of 373 ms, and the AUC 1.30 was larger than that of FL-R and NL-NR. FL-R induced the lowest peak amplitude 0.37 (SD ± 0.66) μV and the smallest AUC 0.05 with a shorter reaction time compared with L-FR and NL-NR. As to the L-FR stimulus, an N400Cz was elicited with a higher peak amplitude (-0.79 ± 0.66 μV) relative to FL-R and NL-NR, and the AUC of 2.40 was the largest at the Cz site. The results indicated that FL-R elicited the semantic integration and expectancy processes (the P400Cz) with a lowered mental workload and a shortened reaction time, but L-FR elicited the process of semantic manipulation (the N400Cz) with the highest processing load and a more latent reaction time.

During 300-500 ms at the Pz site, the N400Pz (the semantic manipulation especially lexical access) was elicited by FL-FR, L-FR, and NL-NR, but FL-R elicited the P400 for semantic integration. L-FR induced the largest peak amplitude -2.41 (SD ± 0.72) μV and the shortest reaction time of 391 ms, in addition, the AUC 1.60 was larger than that of FL-FR and FL-R. The FL-R signal induced the P400 with the lowest peak amplitude -0.58 (SD ± 0.74) μV and the smallest AUC 0.86 with a more latent reaction time of 461 ms at the Pz site. NL-NR induced a relatively higher peak amplitude -2.38 (SD ± 0.73) μV and the largest AUC of 2.90. The results of the N400Pz indicated that NL-NR resulted in more effortful processing of the lexical meanings with a shorter reaction time. But FL-R elicited the lowest processing load for semantic integration though the reaction time was more latent.

In the 500-900 ms time window, a P600 was elicited by FL-FR and FL-R at the Cz site (the P600Cz is related to syntactic processing) and an N600 was elicited by L-FR and NL-NR at the same site (the N600Cz is sensitive to stimulus semantics and is related to the stimulus-general process). In terms of the mental workload and reaction time, FL-R induced the P600Cz with the highest peak amplitude of 2.11 (SD ± 0.64) μV, the largest AUC 9.51, and the shortest reaction time 586 ms. L-FR elicited the N600Cz with a relatively higher peak amplitude compared with FL-FR and NL-NR, and a relatively smaller AUC compared with FL-R and FL-FR. The NL-NR stimulus elicited the N600Cz with the lowest peak amplitude -0.21 (SD ± 0.46) μV, the smallest AUC 0.28, and the longest reaction time 622 ms. The results at the Cz site during 500-900 ms indicated that FL-R elicited the P600Cz, resulting in more effortful syntactic processing. The NL-NR signal induced the lightest load in stimulus semantics and stimulus-general processes (the N600Cz) at the central site.

In 500-900 ms at the Pz site, four stimuli elicited the P600Pz. FL-FR induced the highest peak amplitude of 2.26 (SD ± 0.53) μV and the largest AUC 4.66 with a relatively shorter reaction time. FL-R induced the lowest peak amplitude 0.72 (SD ± 0.69) μV, the smallest AUC 0.54, and the least latent reaction time of 619 ms at the Pz site. Compared with NL-NR, L-FR elicited a lower peak amplitude of 1.07 (SD ± 0.58) μV, a smaller AUC 0.62, and a shorter reaction time of 621 ms. The results of the P600Pz indicated that FL-R induced the lowest processing load of syntactic manipulation with the shortest reaction time, but FL-FR resulted in more effortful syntactic processing.

(2) L2

In the time window of 300-500 ms at the Cz site, FL-FR, FL-R, and NL-NR elicited the N400Cz, but the L-FR signal elicited the P400Cz. As to the processing load and reaction time, the N400Cz induced by the FL-R stimulus showed the lowest peak amplitude -0.06 (SD ± 0.77) μV in a more latent reaction time than that of other stimuli, and a smaller AUC 0.52 compared with L-FR and NL-NR. L-FR induced the P400Cz with the highest peak amplitude of 0.26 (SD ± 0.80) μV and a larger AUC 0.94 than that of FL-FR and FL-R, but the reaction time was the shortest. NL-NR elicited the N400Cz with a relatively lower peak amplitude -0.08 (SD ± 0.84) μV, but the AUC 0.99 was the largest at the Cz site. But FL-FR resulted in the lightest semantic processing load with AUC 0.30. Compared with NL-NR, FL-R lowered the mental load of semantic processing. In addition, the L-FR stimulus induced the semantic integration and expectancy processes with a relatively high mental workload and the shortest reaction time.

In 300-500 ms at the Pz site, four stimuli elicited the N400Pz. FL-R induced the highest peak amplitude -2.74 (SD ± 0.74) μV, but the AUC was 1.22 that was the smallest with a longer reaction time relative to FL-FR and L-FR. L-FR induced a quite high peak amplitude -2.05 (SD ± 0.73) μV and the largest AUC of 2.16 with a less latent reaction time compared with FL-R and NL-NR. The peak amplitude induced by FL-FR was the lowest -0.67 (SD ± 0.73) μV and the reaction time was the shortest in 363 ms. The results at the Pz site indicated that FL-R induced the semantic manipulation with the lowest processing load, but L-FR resulted in the heaviest mental load in semantic processing.

In the 500-900 ms time window, FL-FR, L-FR, and NL-NR elicited the P600 at the Cz site (P600Cz), but FL-R elicited the N600Cz. FL-FR induced the lowest peak amplitude of 0.12 (SD ± 0.56) μV and the longest reaction time of 625 ms, resulting in the largest AUC of 2.13 at the Cz site. Compared with FL-FR and NL-NR, the peak amplitude of the P600Cz elicited by L-FR was higher, the reaction time was shorter, and the AUC was smaller. The N600Cz elicited by FL-R showed a quite high peak amplitude -0.49 (SD ± 0.47) μV with the shortest reaction time 570 ms, making the AUC 0.24 the smallest. The results of the Cz site indicated that FL-R induced the lowest processing load in the semantic and stimulus-general processes with the least latent reaction time, but FL-FR induced more effortful syntactic processing.

During 500-900 ms at the Pz site, four stimuli elicited the P600Pz. Though L-FR induced the lowest peak amplitude 0.53 (SD ± 0.64) μV, the reaction time 642 ms was the longest, and the AUC 0.54 was the largest. FL-R induced the highest peak amplitude of 1.29 (SD ± 0.63) μV. But compared with L-FR, FL-R induced a shorter reaction time of 620 ms and a smaller AUC of 0.39. The smallest AUC 0.29 was induced by FL-FR with a lower peak amplitude of 0.83 (SD ± 0.67) μV than that of FL-R and NL-NR, and FL-FR induced the least latent reaction time of 573 ms. The results indicated that L-FR induced the heaviest mental workload for syntactic processing with the most latent reaction time, and FL-FR resulted in the lightest load and the shortest reaction time for syntactic processing.

(3) ANOVAs on the mean peak amplitudes

Results of ANOVAs on the mean peak amplitudes of the elicited ERPs at the Cz and Pz electrode sites in the time windows of 300-500 ms and 500-900 ms are listed in Table 4.2.

Table 4.2 ANOVAs on the mean peak amplitudes of the ERPs at the Cz and Pz electrode sites

		300-500 ms		500-900 ms	
		Cz	Pz	Cz	Pz
L1	FL-FR vs. FL-R	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-FR vs. L-FR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-FR vs. NL-NR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-R vs. L-FR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-R vs. NL-NR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	L-FR vs. NL-NR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
L2	FL-FR vs. FL-R	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-FR vs. L-FR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-FR vs. NL-NR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-R vs. L-FR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	FL-R vs. NL-NR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	L-FR vs. NL-NR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
L1 vs. L2	L1 FL-FR vs. L2 FL-FR	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	L1 FL-R vs. L2 FL-R	$p < .001$	$p < .001$	$p < .001$	$p < .001$
	L1 L-FR vs. L2 L-FR	$p < .001$	$p < .001$	$p < .001$	$p < .001$

The ANOVA tests showed that the mean peak amplitudes of the ERP components elicited by the filtered and unfiltered L1 and L2 auditory stimuli were significantly different from each other at the central and parietal electrode sites in the 300-500 ms and the 500-900 ms time windows ($p < .001$). Specifically, the mean peak amplitudes of the ERPs elicited by the four configurations of the auditory signals (FL-FR, FL-R, L-FR, and NL-NR) in both L1 and L2 were compared with each other, and the contrasts indicated that the differences between the mean peak amplitudes of the ERPs were significant ($p < .001$). In addition, differences of the mean ERP peak amplitudes elicited by L1 and L2 stimuli of the same configurations were significant (*p* < .001) at the central and parietal electrode sites. The results showed that the brain responded to L1 and L2 filtered and unfiltered stimuli in dichotic and diotic conditions variously, indicating that the change of the physical features of the auditory language signals led to different activities in the brain. Further, brain activity in response to L1 and L2 signals differed significantly, indicating that language proficiency led to differences in language processing load.

4.1.1.2 Results of the group analysis

For the group analysis, the mismatch negativity (MMN) difference waves were used to analyze the processing differences between the target and the non-target stimuli in terms of peak amplitudes and AUC. Thus, all stimuli were regarded as the target and the non-target stimuli respectively in the analyses of MMN difference waves. The MMN differences within-L1/L2 signals and between-L1 and L2 stimuli were calculated and analyzed. Results of the mean peak amplitudes and AUC of MMN in the time windows of 300-500 ms and 500-900 ms over the central and parietal electrode sites are listed in Table 4.3. The MMN difference waves of the L1 stimuli are plotted in Figure 4.2, L2 in Figure 4.3, and between-L1 and L2 in Figure 4.4.

(1) Comparisons of the L1 stimuli

In the time window of 300-500 ms, results of the MMN difference waves between L1 stimuli at the Cz electrode site revealed that the largest P400 effect appeared when FL-FR – L-FR with the largest peak amplitude (1.70 μV) and AUC (0.54), indicating that FL-FR elicited more positive P400 amplitudes as compared to that of L-FR. Meanwhile, the largest N400 effect at the Cz site was elicited as L-FR –

Table 4.3 Mean peak amplitudes and AUCs of the MMN differences between the L1 and L2 auditory stimuli at the Cz and Pz electrode sites

MMN difference between stimuli	Cz				Pz			
	300-500 ms		500-900 ms		300-500 ms		500-900 ms	
(1) L1	Amplitude (μV)	AUC	Amplitude (μV)	AUC	Amplitude (μV)	AUC	Amplitude (μV)	AUC
FL-FR – FL-R	0.51	0.10	-1.39	0.36	-0.45	0.02	0.91	0.55
FL-R – FL-FR	-0.68	0.19	1.14	0.88	0.44	0.02	-0.91	0.55
FL-FR – L-FR	1.70	0.54	1.74	0.22	-0.28	0.03	0.89	0.46
L-FR – FL-FR	-1.66	0.45	-1.37	0.06	0.09	0.08	-0.89	0.46
FL-FR – NL-NR	0.26	0.13	1.11	0.08	-0.78	0.32	1.08	0.52
NL-NR – FL-FR	-0.46	0.18	-0.94	0.08	0.01	0.19	-1.08	0.52
FL-R – L-FR	1.15	0.21	2.87	1.05	-0.12	0.05	0.23	0.18
L-FR – FL-R	-1.11	0.32	-2.87	1.05	-0.36	0.06	-0.23	0.18
FL-R – NL-NR	-0.42	0.12	2.25	0.96	0.45	0.17	0.29	0.19
NL-NR – FL-R	0.16	0.02	-2.25	0.96	-0.45	0.17	-0.29	0.19
L-FR – NL-NR	-1.38	0.36	-0.33	0.17	0.24	0.13	0.20	0.08
NL-NR – L-FR	1.43	0.31	0.65	0.15	-0.24	0.13	0.15	0.03

Table 4.3 Mean peak amplitudes and AUCs of the MMN differences between the L1 and L2 auditory stimuli at the Cz and Pz electrode sites (Cont.)

MMN difference between stimuli	Cz				Pz			
	300-500 ms		500-900 ms		300-500 ms		500-900 ms	
(2) L2	Amplitude (μV)	AUC	Amplitude (μV)	AUC	Amplitude (μV)	AUC	Amplitude (μV)	AUC
FL-FR – FL-R	-0.36	0.03	0.84	0.30	-0.12	0.07	-0.40	0.03
FL-R – FL-FR	-0.26	0.11	-0.84	0.30	-0.35	0.13	0.37	0.08
FL-FR – L-FR	-0.50	0.03	-0.65	0.22	-0.01	0.02	-0.39	0.01
L-FR – FL-FR	0.43	0.03	0.65	0.22	-0.13	0.08	0.56	0.03
FL-FR – NL-NR	-0.73	0.23	0.46	0.29	-0.16	0.05	-0.41	0.12
NL-NR – FL-FR	0.43	0.03	0.65	0.22	-0.13	0.08	0.56	0.03
FL-R – L-FR	-0.50	0.05	-0.98	0.23	-0.37	0.06	-0.16	0.10
L-FR – FL-R	0.13	0.01	0.98	0.23	0.11	0.10	0.41	0.01
FL-R – NL-NR	0.09	0.04	-0.74	0.20	-0.21	0.07	0.06	0.02
NL-NR – FL-R	-0.50	0.11	0.98	0.23	0.11	0.10	0.41	0.01
L-FR – NL-NR	0.59	0.18	0.32	0.02	0.15	0.08	0.46	0.06
NL-NR – L-FR	-0.29	0.18	0.52	0.20	-0.15	0.08	0.46	0.06

Table 4.3 Mean peak amplitudes and AUCs of the MMN differences between the L1 and L2 auditory stimuli at the Cz and Pz electrode sites (Cont.)

MMN difference between stimuli	Cz				Pz			
	300-500 ms		500-900 ms		300-500 ms		500-900 ms	
(3) L1 and L2	Amplitude (μV)	AUC	Amplitude (μV)	AUC	Amplitude (μV)	AUC	Amplitude (μV)	AUC
L1 FL-FR – L2 FL-FR	0.79	0.16	1.21	0.43	-0.62	0.20	0.76	0.45
L2 FL-FR – L1 FL-FR	-0.83	0.07	-0.42	0.27	0.06	0.10	-0.76	0.45
L1 FL-R – L2 FL-R	-0.23	0.06	2.71	0.04	-0.19	0.08	-0.06	0.12
L2 FL-R – L1 FL-R	0.03	0.07	-2.71	0.04	-0.51	0.03	0.17	0.14
L1 L-FR – L2 L-FR	-1.30	0.11	-0.58	0.14	-0.25	0.08	-0.38	0.16
L2 L-FR – L1 L-FR	-0.51	0.52	1.15	0.06	-0.04	0.35	0.61	0.40
L1 NL-NR – L2 NL-NR	0.12	0.17	0.07	0.14	0.21	0.24	-0.05	0.05
L2 NL-NR – L1 NL-NR	-0.45	0.21	0.29	0.08	0.39	0.19	0.37	0.17

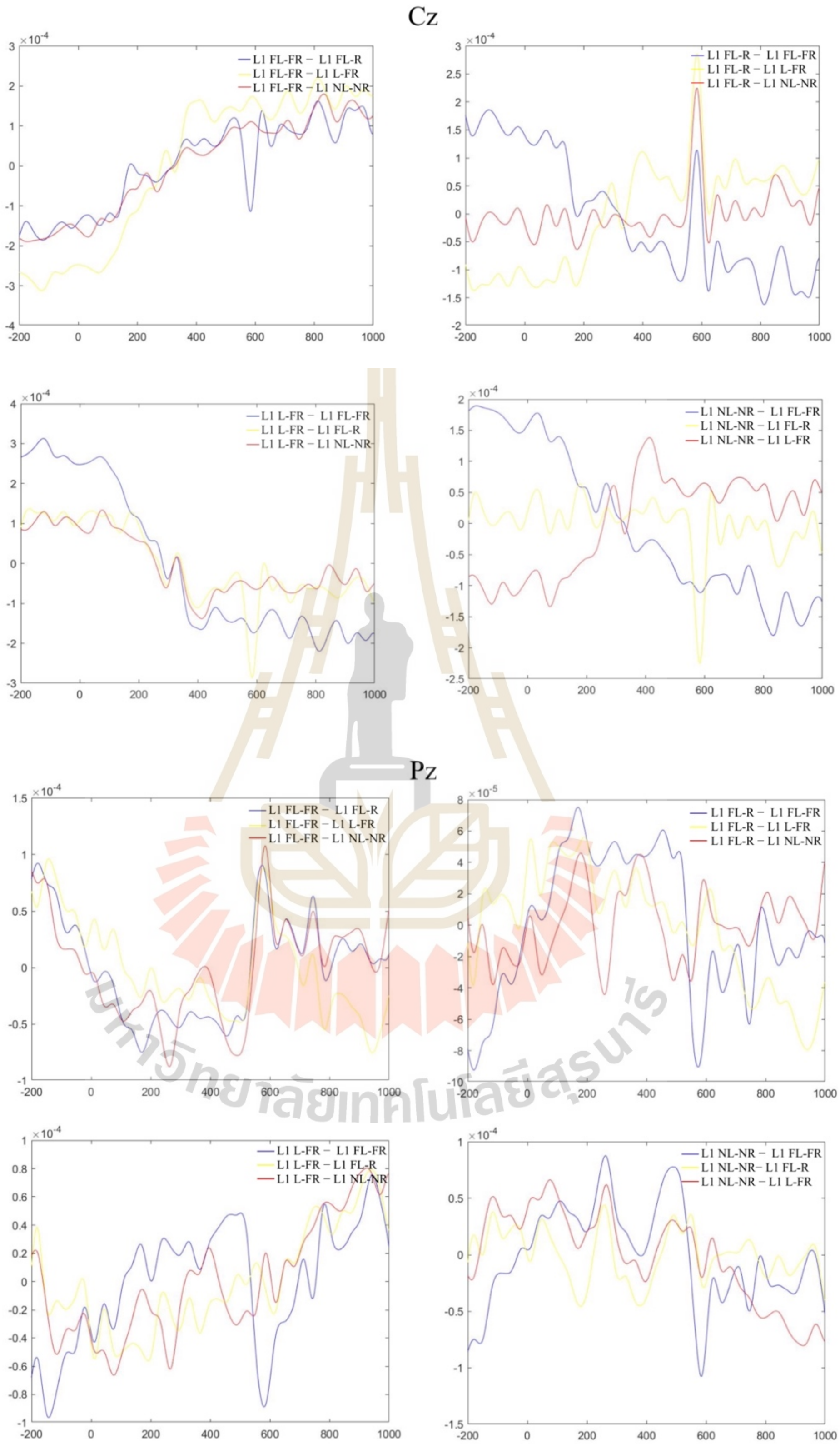


Figure 4.2 MMN difference waves between L1 stimuli at the Cz and Pz electrode sites

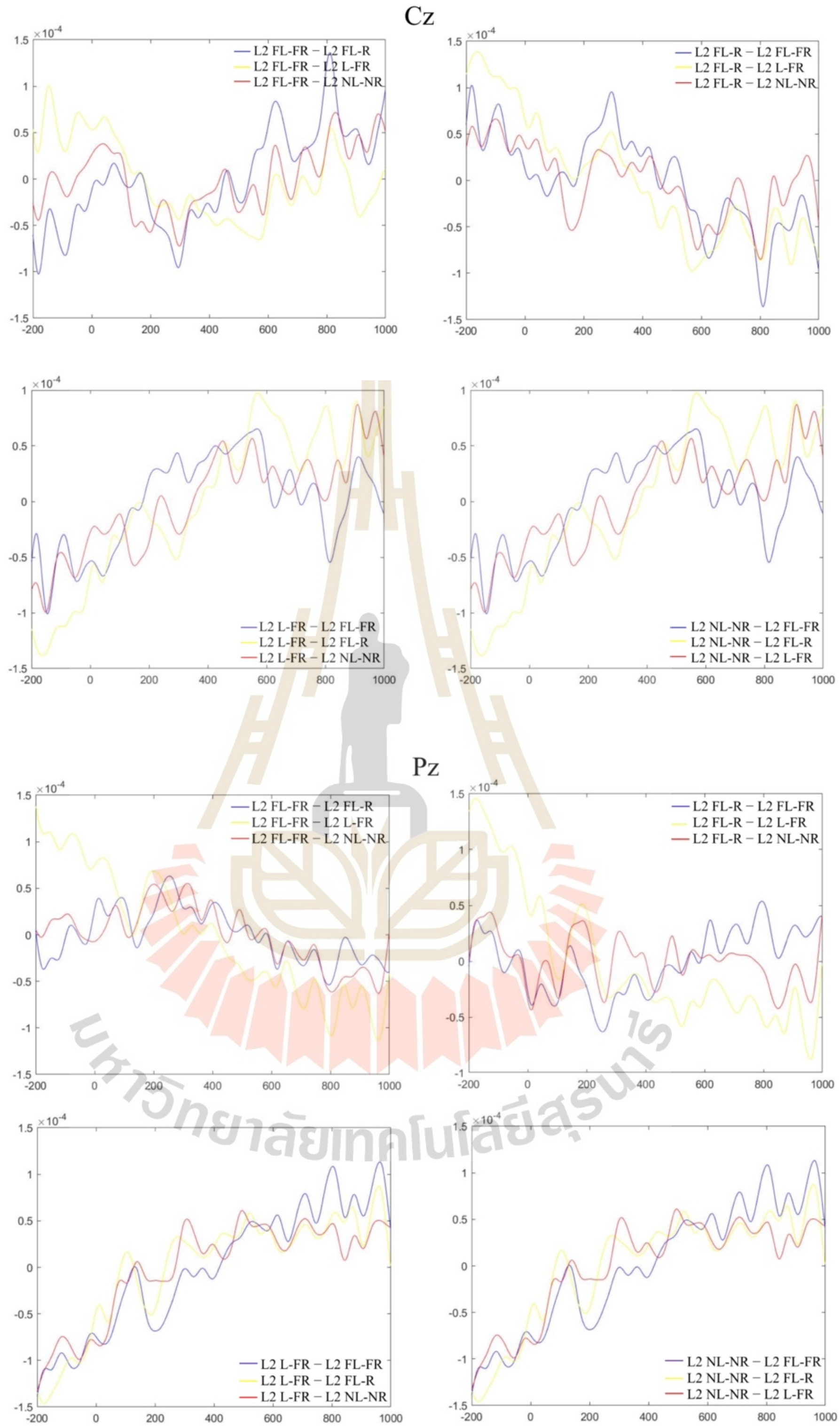


Figure 4.3 MMN difference waves between L2 stimuli at the Cz and Pz electrode sites

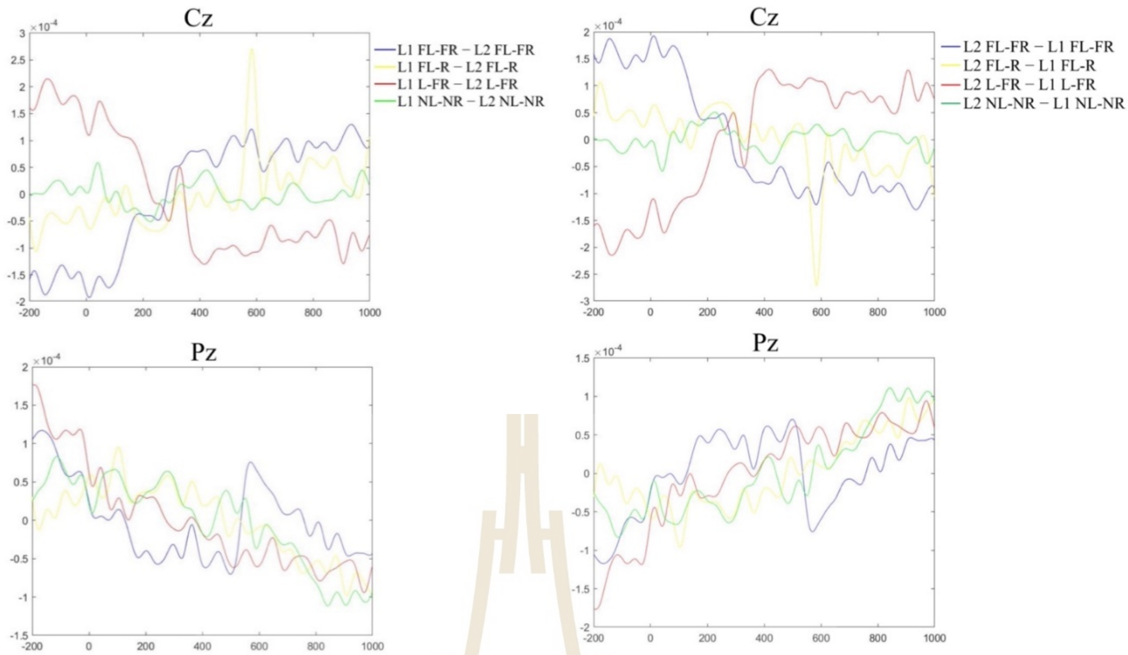


Figure 4.4 MMN difference waves between L1 and L2 stimuli at the Cz and Pz electrode site

FL-FR with the largest peak amplitude of -1.66 μV and AUC of 0.45, showing L-FR induced more negative N400 amplitudes as compared to FL-FR. The results of the largest P400 and N400 effects were consistent with the results in the previous individual analysis that FL-FR elicited the P400Cz with the highest peak amplitude but only L-FR elicited the N400Cz with the largest AUC, thus making the discrepancies between the two stimuli greater. It indicated that FL-FR induced higher amplitudes and more involvement of the neuronal resources for semantic integration of sequential representations and expectancy processes as compared to L-FR. On the contrary, L-FR induced higher amplitudes for lexical access when compared to FL-FR. The smallest N400 and P400 effects were found between FL-R and NL-NR. Specifically, the MMN difference wave of FL-R – NL-NR showed the smallest peak amplitude (-0.42 μV) and AUC (0.12) of the N400, suggesting the discrepancy between the two stimuli for semantic processing was small. Further, NL-NR – FL-R demonstrated the smallest P400 effect with 0.16 μV in the peak amplitude and 0.02 in AUC, indicating the difference of the neural processing of semantic integration and the expectancy processes between the two stimuli was small.

In the time window of 300-500 ms at the Pz site, the largest N400 effect occurred in the MMN difference of FL-FR – NL-NR with the highest peak amplitude of -0.78 μV and the largest AUC of 0.32, indicating that there was a large discrepancy between FL-FR and NL-NR for lexical access. In another word, FL-FR elicited greater amplitudes for the processing of lexical meanings as compared to NL-NR. As to the smallest effects, FL-FR – FL-R showed the smallest AUC 0.02 of the N400 with the peak amplitude -0.45 μV, and FL-R – FL-FR elicited the smallest AUC 0.02 of the P400 with 0.44 μV in the peak amplitude. It indicated that the differences between FL-FR and FL-R in the amplitude of lexical processing and semantic integration were small.

Regarding the MMN differences in the 500-900 ms time window at the Cz site, FL-R – L-FR and L-FR – FL-R induced the largest P600 and N600 effects respectively. To be specific, the MMN difference between FL-R and L-FR showed the largest P600 with the highest peak amplitude of 2.87 μV and the largest AUC of 1.05, and L-FR – FL-R induced the largest N600 with the peak amplitude -2.87 μV and the AUC 1.05. It indicated that FL-R resulted in higher amplitudes and more involvement of the neuronal resources for syntactic manipulation as compared to L-FR, which was consistent with the results of the individual analysis that FL-R induced the highest P600 peak amplitude and the largest AUC for syntactic processing. It also revealed that L-FR elicited higher amplitudes for stimulus semantics and stimulus-general processes when compared to FL-R. On the contrary, the smallest N600 effect was found in the MMN difference of L-FR – FL-FR with -1.37 μV in the peak amplitude and 0.06 in AUC. It indicated that there was a small discrepancy between L-FR and FL-FR for semantic manipulation and stimulus-general processes.

At the Pz site during 500-900 ms, the MMN differences of FL-FR – FL-R and FL-R – FL-FR induced the largest AUC (0.55) of the P600 and N600 respectively. Compared to FL-R, FL-FR induced more neuronal resources to get involved in syntactic processing. But FL-R led to more involvement of neuronal resources for stimulus semantics and stimulus-general processes as compared to FL-FR. It is noteworthy that FL-FR – NL-NR and NL-NR – FL-FR induced the highest peak amplitude (1.08 and -1.08 μV) and the second-largest AUC (0.52 and 0.52) of the P600 and N600 respectively. FL-

FR resulted in more neural involvement for syntactic manipulation as compared to NL-NR. NL-NR elicited higher amplitude for stimulus semantics and stimulus-general processes when compared to FL-FR. The smallest P600 effect appeared when NL-NR – L-FR with the lowest peak amplitude 0.15 μV and the smallest AUC 0.03, which indicated the difference of syntactic neural processing between NL-NR and L-FR was small at the parietal site.

(2) Comparisons of the L2 stimuli

In the 300-500 ms time window at the central site, the largest N400 effect was observed in FL-FR – NL-NR with the highest peak amplitude -0.73 μV and the largest AUC 0.23. FL-FR elicited higher amplitudes and more neural resources involved for semantic manipulation at the Cz site as compared to NL-NR. In addition, NL-NR – L-FR induced the second-largest N400 effect with -0.29 μV in the peak amplitude and 0.18 in the AUC, suggesting that NL-NR elicited higher amplitudes and led to more involvement of the neural resources for semantic processing when compared to L-FR. For the largest P400 effect, L-FR – NL-NR induced the highest peak amplitude of 0.59 μV and the AUC of 0.18, indicating that L-FR led to higher amplitudes for the processes of semantic integration and expectancy updating as compared to NL-NR. But the smallest P400 effect was observed in L-FR – FL-R with 0.13 μV in the peak amplitude and 0.01 in the AUC. It indicated that the differences in semantic integration and expectancy processes between L-FR and FL-R were small.

In the time window of 300-500 ms at the Pz site, the MMN difference of FL-R – FL-FR induced the largest N400 effect that the peak amplitude was -0.35 μV and the AUC was 0.13. It revealed that FL-R resulted in higher amplitudes and more neuronal resources involved for lexical access as compared to FL-FR. For the largest P400 effect, L-FR – FL-R and NL-NR – FL-R induced the highest peak amplitudes of 0.11 μV and the AUCs of 0.10. It indicated that L-FR led to more involvement of the neuronal resources for semantic integration as compared to FL-R, and NL-NR induced more involvement of the neural resources for semantic integration when compared to FL-R. As to the smallest N400 effect, FL-FR – L-FR elicited -0.01 μV in the peak amplitude and 0.02 in the AUC, suggesting that the neural processing discrepancy between the two stimuli in semantic manipulation was small.

In terms of the time window of 500-900 ms at the Cz site, the MMN differences of FL-FR – FL-R and FL-R – FL-FR showed the largest P600 and N600 effects respectively. FL-FR – FL-R induced the largest P600 effect with the peak amplitude of 0.84 μV and the AUC of 0.30, indicating that there was a great difference in neural processing of syntactic manipulation between the FL-FR and FL-R stimuli. It was observed in the individual analysis that FL-FR elicited the largest AUC of the P600 and only FL-R elicited the N600 with the smallest AUC. Thus, the MMN difference between FL-FR and FL-R revealed a quite large P600 effect, suggesting that FL-FR led to higher amplitudes and more involvement of the neural resources for syntactic processing as compared to FL-R. Conversely, FL-R – FL-FR resulted in the largest N600 effect with the peak amplitude of -0.84 μV and the AUC of 0.30. Compared to FL-FR, FL-R led to more neuronal resources involved and higher amplitudes for the processing of stimulus semantics and the stimulus-general processes at the Cz site. In addition, FL-FR – NL-NR induced the second-largest P600 effect with 0.46 μV in the peak amplitude and 0.29 in the AUC, suggesting that FL-FR required higher amplitudes and more neural resources for syntactic manipulation as compared to NL-NR. Regarding the smallest P600 effect, L-FR – NL-NR elicited 0.32 μV in the peak amplitude and 0.02 in the AUC. It revealed a small difference in syntactic processing between the L-FR and NL-NR stimuli.

During 500-900 ms at the parietal site, the MMN difference of FL-R – FL-FR elicited the largest P600 effect with the peak amplitude of 0.37 μV and the AUC of 0.08, indicating that FL-R induced higher amplitudes and more neuronal resources were required for syntactic processing as compared to FL-FR. The MMN difference of FL-FR – NL-NR induced the largest N600 effect that the peak amplitude was -0.41 μV and the AUC was 0.12, suggesting that FL-FR elicited higher amplitudes and resulted in more involvement of the neural resources for stimulus semantics and stimulus-general processes when compared to NL-NR. As for the smallest P600 effects, L-FR – FL-R and NL-NR – FL-R elicited the peak amplitudes of 0.41 μV and the AUCs of 0.01. It indicated that the difference of syntactic processing of the L-FR and FL-R stimuli was small, and there was a small discrepancy between NL-NR and FL-R for syntactic manipulation as well. The smallest N600 effect was observed in the MMN difference of FL-FR – L-FR with -0.39 μV in the peak amplitude and 0.01 in the AUC, which suggested the difference of neural processing for stimulus semantics and stimulus-general processes between the two stimuli was quite small.

(3) Comparisons of the L1 and L2 stimuli

L1 and L2 stimuli of the same configuration were compared by using the MMN difference waves. In the time window of 300-500 ms at the Cz site, the largest N400 effect was found in L2 L-FR – L1 L-FR with -0.51 μV in the peak amplitude and the largest AUC of 0.52, indicating that higher amplitudes and more neuronal resources were required for L2 L-FR semantic manipulation than that of L1 L-FR. L1 NL-NR – L2 NL-NR induced the largest P400 effect at the Cz site with 0.12 μV in the peak amplitude and the AUC of 0.17, indicating that L1 NL-NR resulted in higher amplitudes and more involvement of the neural resources for semantic integration and expectancy processes as compared to L2 NL-NR. But the smallest N400 effect was induced by the MMN difference of L1 FL-R – L2 FL-R with -0.23 μV in the peak amplitude and the AUC of 0.06, suggesting that there was a small difference between L1 FL-R and L2 FL-R in semantic processing at the central site.

At the Pz site during 300-500 ms, L2 L-FR – L1 L-FR elicited the N400 effect with -0.04 μV in the peak amplitude and the largest AUC of 0.35, suggesting that L2 L-FR induced more involvement of the neural resources and amplitudes for lexical access than that of L1 L-FR. The largest P400 effect was observed in L1 NL-NR – L2 NL-NR with 0.21 μV in the peak amplitude and 0.24 in the AUC. It revealed that L1 NL-NR led to more involvement of the neuronal resources and higher amplitudes for semantic integration of sequential representations and expectancy processes as compared to L2 NL-NR. The smallest N400 effect was in the MMN difference of L2 FL-R – L1 FL-R, which elicited -0.51 μV in the peak amplitude and 0.03 in the AUC. It indicated that the difference of lexical access between the stimuli of L2 FL-R and L1 FL-R was small.

In the time window of 500-900 ms at the Cz site, L1 FL-FR – L2 FL-FR and L2 FL-FR – L1 FL-FR elicited the largest P600 and N600 effects respectively. L1 FL-FR – L2 FL-FR induced the P600 peak amplitude of 1.21 μV and the AUC of 0.43, showing a great difference between L1 FL-FR and L2 FL-FR in syntactic processing. The largest P600 effect indicated that L1 FL-FR resulted in more involvement of the neural resources and higher amplitudes for syntactic manipulation than that in L2. On the contrary, L2 FL-FR – L1 FL-FR elicited the largest N600 effect with -0.42 μV in the peak amplitude and the AUC of 0.27. It demonstrated that L2 FL-FR led to more neuronal resources and higher amplitudes to get involved in stimulus-semantic and stimulus-general processes as compared to L1 FL-FR. In addition, the smallest P600 and N600 effects were in L1 FL-R – L2 FL-R and L2 FL-R – L1 FL-R respectively with 2.71 and -2.71 μV in the peak amplitudes and the AUCs of 0.04. It showed the differences of syntactic, stimulus semantics, and stimulus-general manipulations between the L1 and L2 FL-R stimuli.

In the 500-900 ms time window at the Pz site, L1 FL-FR – L2 FL-FR elicited the largest P600 effect with 0.76 μV in the peak amplitude and the AUC of 0.45, suggesting that a larger difference between the two stimuli in the processing of syntax. L1 FL-FR induced higher amplitudes and more neural resources involved for syntactic processing than that in L2. But L2 FL-FR – L1 FL-FR elicited the largest N600 effect with -0.76 μV in the peak amplitude and the AUC of 0.45. It revealed that L2 FL-FR resulted in more involvement of the neural resources and higher amplitudes for stimulus semantics and stimulus-general processes as compared to the L1 FL-FR stimulus. As for the smallest effects, L2 FL-R – L1 FL-R elicited the smallest P600 effect with 0.17 μV in the peak amplitude and the AUC of 0.14, demonstrating that there was a small discrepancy between the two stimuli in syntactic processing. Besides, L1 NL-NR – L2 NL-NR induced the smallest N600 effect with -0.05 μV in the peak amplitude and the AUC of 0.05, which indicated that the difference of stimulus semantics and stimulus-general processes between L1 and L2 NL-NR stimuli was small.

4.1.2 fMRI results

4.1.2.1 Results of the individual analysis

One-sample *t*-tests for the L1 and L2 low-pass filtered and unfiltered stimuli in dichotic and diotic conditions were performed. Activations below a threshold of $p < .001$, uncorrected, cluster size ≥ 10, T ≥ 3 are reported in Table 4.4. The brain activation maps for the L1 and L2 stimuli are plotted in Figures 4.5 and 4.6.

Table 4.4 Brain activations for the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions

Stimulus	Brain region	BA	k	T	MNI coordinates		
					x	*y*	*z*
L1							
FL-FR	L STG	42	144	5.19	-64	-32	10
	L MTG	21	75	5.15	-64	-25	-6
	L IFG	-	47	5.21	-50	28	10
	L MFG	-	15	4.04	-47	0	54
	R MFG	-	254	5.43	45	7	46
	R STG	-	172	5.51	62	-25	6
	R MTG	21	129	5.82	66	-14	-10
	R IFG	9	122	5.15	45	11	30
FL-R	L STG	21	144	5.16	-64	-25	-2
	L MTG	22	76	5.32	-61	-32	2
	L IFG	44	41	4.14	-57	18	18
	L MFG	-	19	3.67	-47	7	50
	R STG	42	115	5.06	69	-18	6
	R MTG	21	54	5.29	66	-14	-6
	R IFG	45	17	4.01	52	25	22
	R MFG	-	12	3.53	45	21	26
L-FR	L STG	-	96	4.62	-61	-28	6
	L MTG	-	33	4.11	-64	-21	-6
	L IFG	45	12	4.18	-58	21	22
	R STG	-	125	5.00	66	-14	6
	R MTG	-	52	4.72	62	-28	-2
	R MFG	-	31	3.86	52	7	42
	R IFG	9	16	3.82	52	18	26
NL-NR	L STG	22	258	4.89	-57	-4	-2
	L MTG	21	70		-55	-1	-9
	R STG	-	268	4.88	52	7	-6
	R MTG	22	85	5.50	55	-11	-6

Table 4.4 Brain activations for the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions (Cont.)

Stimulus	Brain region	BA	k	T	MNI coordinates		
					x	*y*	*z*
L2							
FL-FR	L STG	22	108	3.91	-58	-10	6
	L MTG	21	17	4.16	-64	-14	-10
	R STG	-	182	4.60	54	4	-10
	R MTG	21	57		64	-15	-12
	R MFG	46	15	3.64	52	21	30
FL-R	L STG	-	242	4.49	-50	-25	6
	L MTG	21	45		-54	-2	-12
	L IFG	45	10	3.79	-46	23	8
	R STG	-	144	4.70	69	-14	10
	R MTG	21	59	4.59	55	-14	-2
	R MFG	8	51	3.99	52	11	42
	R IFG	-	37	3.96	55	11	30
L-FR	L STG	22	251	5.78	-47	-7	-6
	L MTG	21	31		-63	-18	-10
	R STG	38	254	4.82	52	8	-14
	R MTG	21	89	4.74	65	-20	-14
	R IFG	-	17	4.09	55	18	18
NL-NR	L STG	42	250	4.59	-66	-13	6
	L MTG	21	26		-64	-16	-3
	L Precentral gyrus	6	19	5.15	-50	-7	6
	R STG	38	261	5.11	52	-4	-10
	R MTG	21	58		62	-14	-12
	R Postcentral gyrus	3	57	3.89	34	-35	66
		5		3.84	34	-46	62
		2		3.82	45	-28	50

Notes: Clusters are thresholded at $p < .001$ (uncorrected), $k \geq 10$, T ≥ 3. Coordinates are given for the stereotactic space of Talairach and Tournoux (1988). BA, Brodmann area; *k*, cluster size (number of voxels); T, t value; L, left hemisphere; R, right hemisphere; STG, superior temporal gyrus; MTG, middle temporal gyrus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; PrCG, precentral gyrus; PoCG, postcentral gyrus.

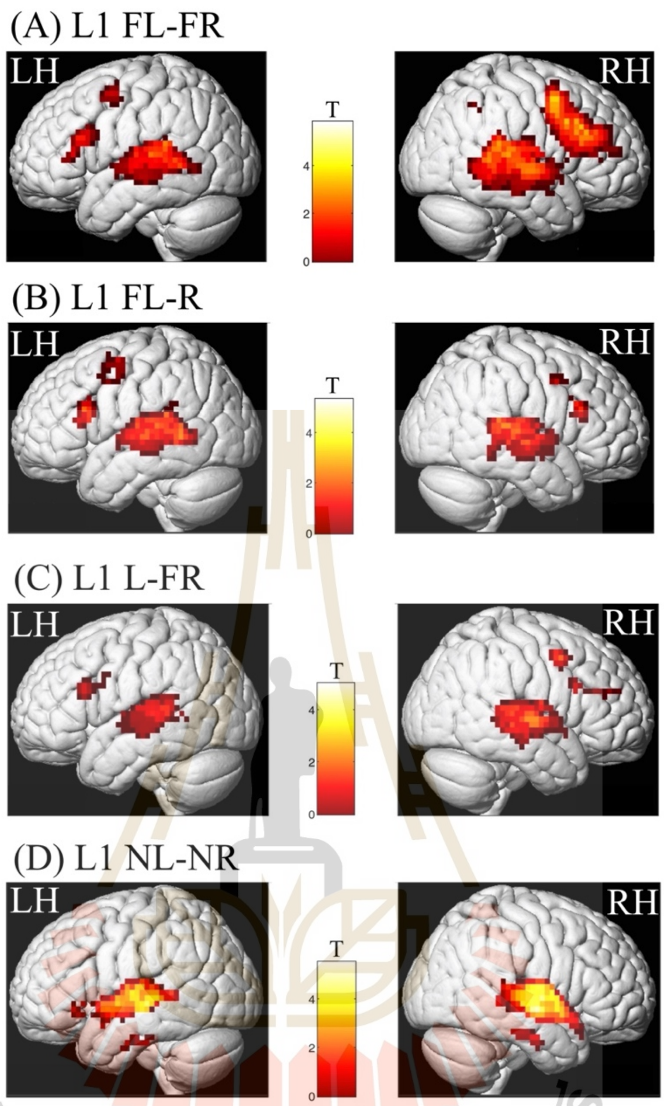


Figure 4.5 Brain activation maps for the L1 stimuli

Notes: Clusters are thresholded at $p < .001$ (uncorrected), $k \geq 10$, $T \geq 3$. LH, left hemisphere; RH, right hemisphere.

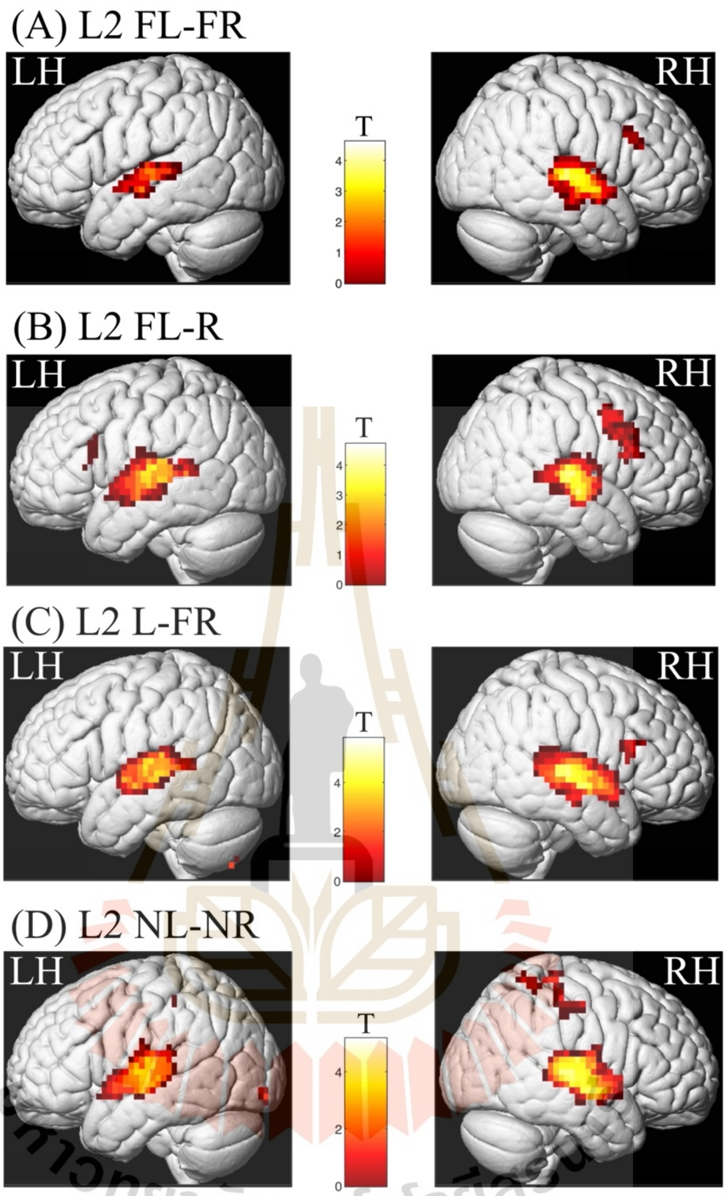


Figure 4.6 Brain activation maps for the L2 stimuli

Notes: Clusters are thresholded at $p < .001$ (uncorrected), $k \geq 10$, T ≥ 3. LH, left hemisphere; RH, right hemisphere.

(1) L1

The one-sample *t*-test for the FL-FR stimulus revealed a strong involvement of bilateral frontotemporal regions, including the bilateral superior temporal gyri (STG), middle temporal gyri (MTG), inferior frontal gyri (IFG), and middle frontal gyri (MFG) ($p < .001$, uncorrected).

For FL-R, increased activity was observed in the bilateral STG, MTG, IFG, and MFG ($p < .001$, uncorrected).

L-FR induced activations in the bilateral STG, MTG, and IFG, in addition, the right MFG was also activated ($p < .001$, uncorrected).

The NL-NR stimulus showed stronger activity in the bilateral STG and MTG ($p < .001$, uncorrected).

(2) L2

Regarding the L2 stimuli, FL-FR induced activity in the brain regions of the bilateral STG and MTG, and the right MFG ($p < .001$, uncorrected).

FL-R revealed increased activity in the bilateral STG, MTG, IFG, and the right MFG ($p < .001$, uncorrected).

As to L-FR, a stronger involvement of the bilateral STG, MTG, and the right IFG was found ($p < .001$, uncorrected).

In terms of NL-NR, activations were identified in the bilateral STG, MTG, the left precentral gyrus (PrCG), and the right postcentral gyrus (PoCG) ($p < .001$, uncorrected).

According to the findings of the one-sample *t*-tests for the L1 and L2 low-pass filtered and unfiltered stimuli in dichotic and diotic conditions, regions of interest (ROIs) were determined, which were the bilateral STG, MTG, IFG, MFG, PrCG, and PoCG. These twelve ROIs were used for the further analysis of hemispheric lateralization for processing the stimuli in 4.1.2.3.

4.1.2.2 Results of the group analysis

To better compare the activation patterns of the L1 and L2 low-pass filtered and unfiltered stimuli in dichotic and diotic conditions, two-sample *t*-tests were performed in the group analysis to compare the stimuli within-L1/L2 and between-L1 and L2. There were no assumptions about the directions of effects in the contrasts, so all possible contrasts were performed for both directions. All reported results were at uncorrected $p < .006$, cluster size ≥ 20, T ≥ 3, shown in Table 4.5 and Figures 4.7 to 4.9.

Table 4.5 Comparisons of the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions

Contrast	Brain region	BA	k	T	MNI coordinates		
					x	y	z
(1) L1 vs. L1							
FL-FR vs. FL-R	-						
FL-R vs. FL-FR	L PrCG	6	46	3.52	-22	-7	66
	L PoCG	-	22	3.10	-42	-33	61
FL-FR vs. L-FR	-						
L-FR vs. FL-FR	L PrCG	6	72	3.39	-23	-9	66
	L IPL	40	52	3.00	-50	-32	46
	L SPL	7	20	3.25	-19	-67	45
FL-FR vs. NL-NR	R MFG	-	97	3.23	52	28	22
	R PrCG	6	28	3.22	41	11	62
NL-NR vs. FL-FR	-						
FL-R vs. L-FR	-						
L-FR vs. FL-R	-						
FL-R vs. NL-NR	-						
NL-NR vs. FL-R	-						
L-FR vs. NL-NR	-						
NL-NR vs. L-FR	-						
(2) L2 vs. L2							
FL-FR vs. FL-R	-						
FL-R vs. FL-FR	L MFG	-	21	3.23	-43	18	30
FL-FR vs. L-FR	-						
L-FR vs. FL-FR	-						
FL-FR vs. NL-NR	R IPL	40	46	3.11	52	-56	46

Table 4.5 Comparisons of the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions (Cont.)

Contrast	Brain region	BA	k	T	MNI coordinates		
					x	y	z
	R MFG	10	27	3.60	31	56	22
	L IPL	-	20	3.15	-47	-63	46
NL-NR vs. FL-FR	R PoCG	2	1016	4.26	45	-25	50
	L PrCG	6		3.96	-8	-21	66
	L PoCG	43		3.56	-48	-14	18
	R PrCG	6	32	3.68	66	-4	38
	R MFG	9		3.12	62	11	30
FL-R vs. L-FR	-						
L-FR vs. FL-R	-						
FL-R vs. NL-NR	-						
NL-NR vs. FL-R	-						
L-FR vs. NL-NR	-						
NL-NR vs. L-FR	R PrCG	6	250	4.00	31	-18	66
	R PoCG	3		3.71	31	-35	66
	L PrCG	6	22	3.55	-29	-18	66
(3) L1 vs. L2 & L2 vs. L1							
L1 FL-FR vs. L2 FL-FR	-						
L2 FL-FR vs. L1 FL-FR	-						
L1 FL-R vs. L2 FL-R	-						
L2 FL-R vs. L1 FL-R	-						
L1 L-FR vs. L2 L-FR	-						
L2 L-FR vs. L1 L-FR	-						
L1 NL-NR vs. L2 NL-NR	-						
L2 NL-NR vs. L1 NL-NR	L PrCG	-	245	3.29	-29	-21	70
	R PrCG	-		3.01	27	-21	70
	L PoCG	3	55	3.52	-64	-18	34

Notes: Clusters are thresholded at $p < .006$ (uncorrected), $k \geq 20$, T ≥ 3. Coordinates are given for the stereotactic space of Talairach and Tournoux (1988). BA, Brodmann area; *k*, cluster size (number of voxels); T, t value; L, left hemisphere; R, right hemisphere; PrCG, precentral gyrus; PoCG, postcentral gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; MFG, middle frontal gyrus.

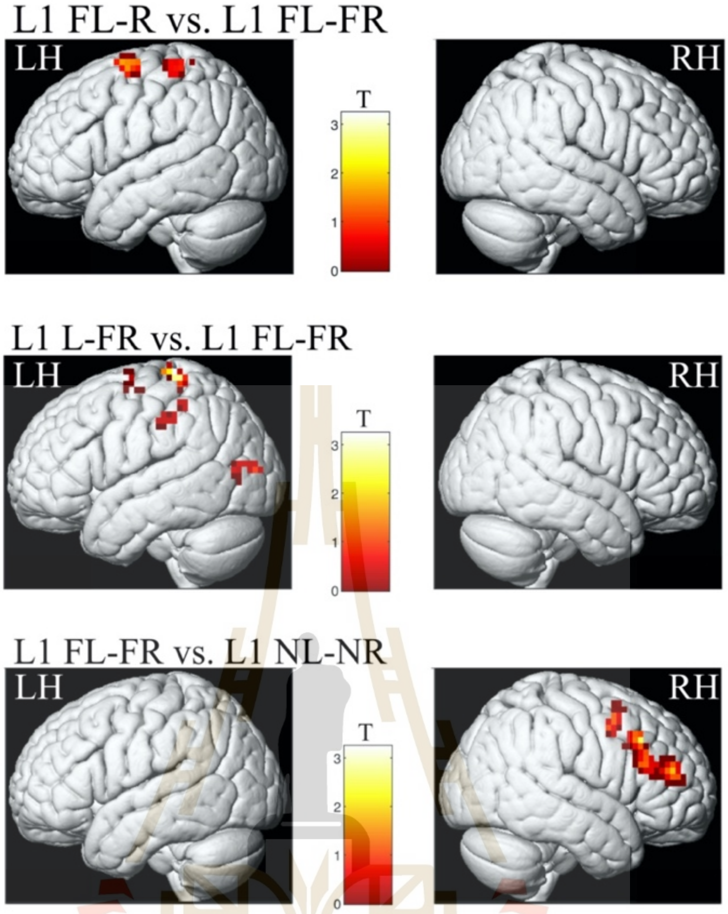


Figure 4.7 Comparisons of brain activation patterns within the L1 stimuli

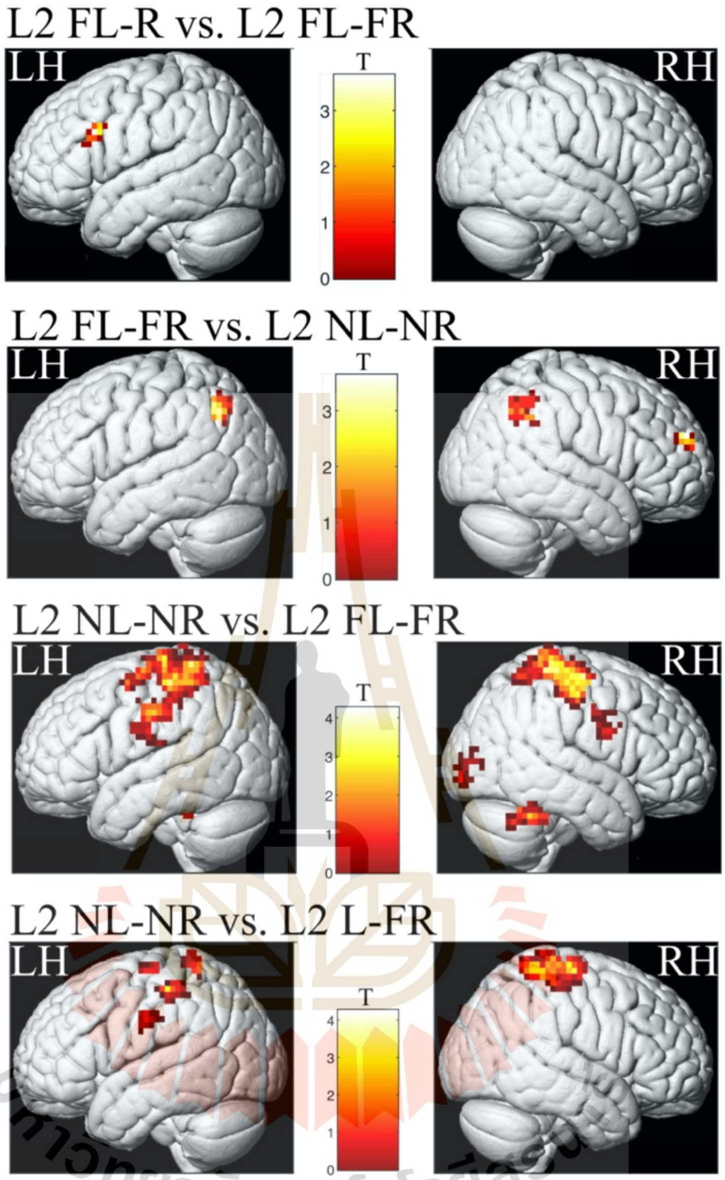


Figure 4.8 Comparisons of brain activation patterns within the L2 stimuli

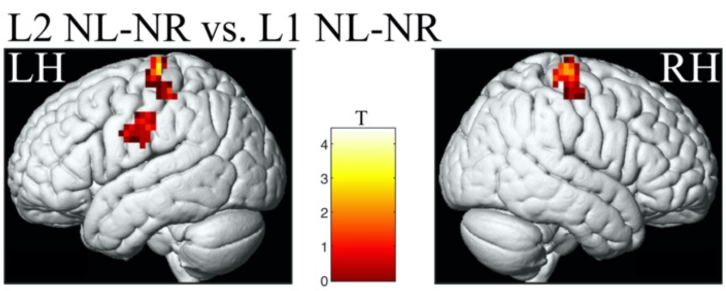


Figure 4.9 Comparisons of brain activation patterns between the L1 and L2 stimuli

(1) Comparisons of brain activation patterns within the L1 stimuli

In the comparisons within the L1 stimuli, FL-R vs. FL-FR, L-FR vs. FL-FR, and FL-FR vs. NL-NR showed significant differences of the brain regions while processing the signals ($p < .006$, uncorrected), but other contrasts did not present any statistical differences between the stimuli at $p < .006$ (uncorrected).

The contrast of FL-R vs. FL-FR reflected a strong involvement of the left PrCG and PoCG ($p < .006$, uncorrected). When compared to FL-FR, FL-R induced more activity in the areas of the left PrCG and PoCG.

L-FR vs. FL-FR revealed that significant differences were found in the left PrCG, inferior parietal lobule (IPL), and superior parietal lobule (SPL) ($p < .006$, uncorrected). It indicated that L-FR induced more activations in the regions of the left PrCG, IPL, and SPL as compared to FL-FR.

When contrasting FL-FR vs. NL-NR, it demonstrated significant differences in the right MFG and PrCG ($p < .006$, uncorrected), suggesting that FL-FR induced more increased activity in the right MFG and PrCG as compared to NL-NR.

(2) Comparisons of brain activation patterns within the L2 stimuli

In the contrasts of the L2 stimuli, FL-R vs. FL-FR, FL-FR vs. NL-NR, NL-NR vs. FL-FR, and NL-NR vs. L-FR presented significant differences ($p < .006$, uncorrected), but other contrasts did not show statistical differences when the threshold was at $p < .006$ (uncorrected).

The FL-R vs. FL-FR contrast reflected that significant activations were found in the brain area of the left MFG ($p < .006$, uncorrected), suggesting that more involvement of the left MFG was induced by FL-R relative to the FL-FR stimulus.

FL-FR vs. NL-NR revealed that more increased activity was observed in the regions of the bilateral IPL and the right MFG ($p < .006$, uncorrected). Compared to NL-NR, FL-FR resulted in more involvement of the bilateral IPL and the right MFG as processing the signal.

The two-sample *t*-test for NL-NR vs. FL-FR showed significant differences in the areas of the bilateral PoCG and PrCG, in addition, the right MFG at $p < .006$ (uncorrected). The results demonstrated that NL-NR led to more significant activations in the above-mentioned areas as compared to FL-FR.

The result of the NL-NR vs. L-FR contrast showed that the bilateral PrCG and the right PoCG were significantly different while processing the two stimuli ($p < .006$, uncorrected), which indicated that NL-NR induced more increased activity in the bilateral PrCG and the right PoCG when compared to L-FR.

(3) Comparisons of brain activation patterns between the L1 and L2 stimuli

Comparisons between the L1 and L2 stimuli of the same configuration presented the significant difference only in the contrast of L2 NL-NR vs. L1 NL-NR, but other contrasts did not show any statistical differences at $p < .006$ (uncorrected).

The L2 NL-NR vs. L1 NL-NR contrast revealed that the bilateral PrCG and the left PoCG had a significant difference in the processing of the two stimuli ($p < .006$, uncorrected). Compared to the processing of L1 NL-NR, the L2 NL-NR stimulus recruited more areas of the bilateral PrCG and the left PoCG.

4.1.2.3 Results of hemispheric lateralization

According to the ROIs determined in the individual analysis of the previous section of 4.1.2.1, hemispheric preference for processing the auditory stimuli of the current study was investigated by using the lateralization index (LI). The formula used for calculating the LI was introduced in section 3.6.2 in Chapter 3. The LI results of the ROIs are shown in Table 4.6.

Table 4.6 Hemispheric lateralization for the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions

Stimulus	ROI	V_L	V_R	LI	Hemispheric lateralization
L1					
FL-FR	STG	144	172	-0.08	-
	MTG	75	129	-0.26	R
	IFG	47	122	-0.44	R
	MFG	15	254	-0.88	R
FL-R	STG	144	115	0.11	-
	MTG	76	54	0.16	-

Table 4.6 Hemispheric lateralization for the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions (Cont.)

Stimulus	ROI	V_L	V_R	LI	Hemispheric lateralization
	IFG	41	17	0.41	L
	MFG	19	12	0.22	L
L-FR	STG	96	125	-0.13	-
	MTG	33	52	-0.22	R
	IFG	12	16	-0.14	-
	MFG	0	31	-1	R
NL-NR	STG	258	268	-0.01	-
	MTG	70	85	-0.09	-
L2					
FL-FR	STG	108	182	-0.25	R
	MTG	17	57	-0.54	R
	MFG	0	15	-1	R
FL-R	STG	242	144	0.25	L
	MTG	45	59	-0.13	-
	IFG	10	37	-0.57	R
	MFG	0	51	-1	R
L-FR	STG	251	254	-0.00	-
	MTG	31	89	-0.48	R
	IFG	0	17	-1	R
NL-NR	STG	250	261	-0.02	-
	MTG	26	58	-0.38	R
	PrCG	19	0	1	L
	PoCG	0	57	-1	R

Notes: LI = (V_L – V_R) / (V_L + V_R). V_L, the number of voxels activated for the left hemisphere; V_R, the number of voxels activated in the right hemisphere; LI > 0.20, left-sided dominance; LI < -0.20, right-sided dominance; -0.20 ≤ LI ≤ 0.20, symmetric / no clear hemispheric preference (Springer et al., 1999). STG, superior temporal gyrus; MTG, middle temporal gyrus; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; PrCG, precentral gyrus; PoCG, postcentral gyrus; L, left-sided; R, right-sided.

(1) Hemispheric lateralization for the L1 stimuli

In L1, the FL-FR signal induced right-sided dominance in the brain regions of the MTG, IFG, and MFG with the LI -0.26, -0.44, and -0.88 respectively. There was no clear hemispheric laterality in the STG.

For FL-R, it showed left-sided dominance in the IFG (LI = 0.41) and in the MFG (LI = 0.22). But no clear hemispheric laterality was observed in the STG and MTG.

L-FR resulted in right dominance in the MTG (LI = -0.22) and the MFG (LI = -1.00). But no hemispheric laterality was found in the STG and IFG.

In terms of the NL-NR stimulus, the ROIs of the STG and MTG did not exhibit hemispheric lateralization.

(2) Hemispheric lateralization for the L2 stimuli

Regarding the L2 stimuli, FL-FR led to right dominance in the areas of the STG (LI = -0.25), the MTG (LI = -0.54), and the MFG (LI = -1.00).

The FL-R stimulus brought about left-lateralized activations in the STG (LI = 0.25), but right-lateralized activations in the IFG (LI = -0.57) and the MFG (LI = -1.00). The MTG did not demonstrate hemispheric lateralization.

L-FR induced right-sided dominance in the MTG (LI = -0.48) and the IFG (LI = -1.00). There was no clear laterality in the STG.

As to NL-NR, right-lateralized activations were observed in the MTG with LI = -0.38 and the PoCG with LI = -1.00, but left-sided laterality was found in the PrCG (LI = 1.00). Hemispheric laterality in the STG was not clear.

4.2 Results of the semi-structured interviews

Semi-structured interviews were conducted to investigate students' feelings after listening to the auditory signals and their opinions of using the signals for language learning. Twelve participants who got various perspectives on the perception of the stimuli and were enthusiastic to share their views were invited as interviewees. Audio recordings, made with interviewees' permission, were used for transcription, coding, and interpretation by adopting content analysis (Creswell & Guetterman, 2019).

Students' opinions about the signals of the current study could be categorized into three themes: feelings after listening to the signals, understanding of the signals, and views on using the signals for language learning as shown in Figure 4.10.

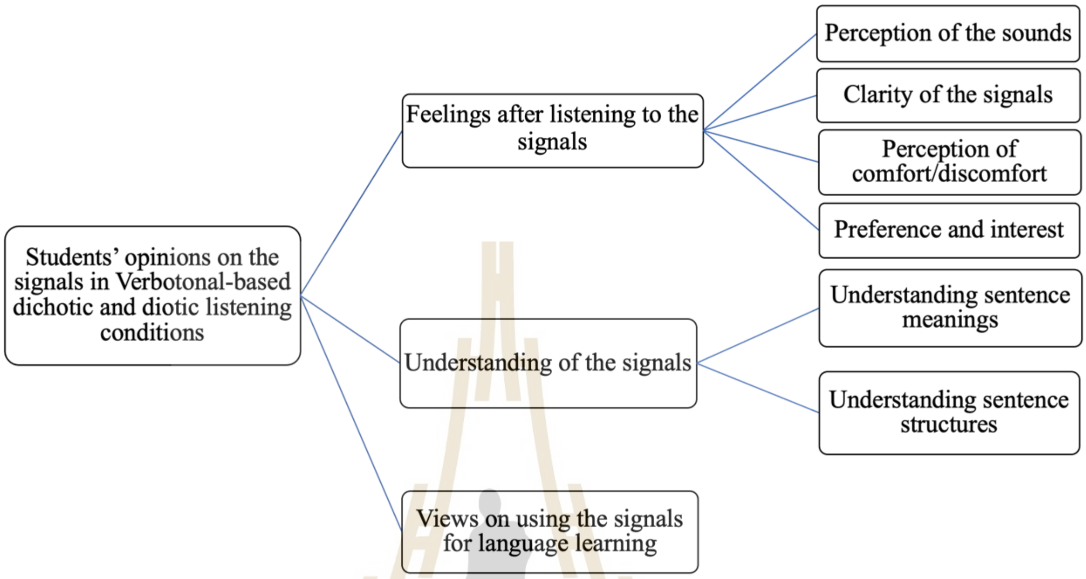


Figure 4.10 Students' opinions on the signals in Verbotonal-based dichotic and diotic listening conditions

4.2.1 Feelings after listening to the signals in Verbotonal-based dichotic and diotic listening conditions

4.2.1.1 Perception of the sounds

(1) L1 signals

The interviewees reported that the FL-FR signals were just humming sounds and unclear, but they could be recognized as Chinese sentences. The NL-NR signals were loud and like stereophonic sounds that both ears received the signals. One student described it as:

"......第四组信号感觉双耳都在共鸣一样，声音很大。(... the fourth group of signals (NL-NR) feels like both ears are resonating, and the sound is very loud.)" – Interviewee 06

For most of the students, the L1 signals in dichotic conditions were perceived differently in the left and right ears. Specifically, the speech sounds of the FL-R signals were clear in the right ear and the humming sounds in the left ear, and the situation was opposite in L-FR.

"第二组像是右边说话，第三组像左边说话。(The second group of signals (FL-R) is like speaking on the right, and the third group (L-FR) is like speaking on the left.)" – Interviewee 05

"第二组感觉右边清楚，左边会出现模糊声音，但整体清楚。第三组左边清楚，整体清晰。(The second group of signals (FL-R) feels clear on the right side and the muffled sounds were on the left side, but overall, it is clear. For the third group (L-FR), the left side is clear, and it is clear in general.)" – Interviewee 06

But for a few students, they did not feel the difference between the left and right ears during the presentation of the L1 dichotic signals.

Interestingly, two interviewees also claimed that there was a resemblance between FL-R and NL-NR regarding perception.

"第二组和第四组信号较为相似，且更为自然。 (The second (FL-R) and fourth (NL-NR) groups of signals are more similar and more natural.)" – Interviewee 04

"第二组和第四组感觉相差不大，就是刚开始听时感觉与第四组相比有些差异，但是不明显。(The second (FL-R) and fourth (NL-NR) groups of signals are not much different, that is, the feeling is somewhat different from that of the fourth group when you first listen, but it is not obvious.)" – Interviewee 11

(2) L2 signals

As for the L2 FL-FR stimulus, some respondents stated that it sounded similar to L1 FL-FR, which was humming sounds and unclear. And others thought the L2 FL-FR signals were harder to decipher and more unclear than that of Chinese. Even the students were not quite sure if the signals were in English.

"相对于中文更加模糊，任何词语都听不清楚，只能听到一些哼哼的声音。(Compared with Chinese, it is vaguer. I can't hear any words clearly. Only some humming sounds can be heard.)" – Interviewee 06

"听不懂，好像是英文。(I don't understand. It seems to be in English.)" – Interviewee 08

Similar to L1 NL-NR, the students felt balanced that both ears could perceive the same signals from the L2 NL-NR signal. One student mentioned:

“感觉整个声音都集中于脑袋中间。(It feels like the voice concentrates in the middle of my head.)” – Interviewee 06

Students’ perceptions of the L2 dichotic signals varied. Most of the respondents reported that the FL-R signal sounded like the humming sounds in the left ear and the speech sounds in the right ear, but a few described it oppositely that the humming sounds were in the right ear and the speech sounds in the left ear. The L-FR signal showed different perceptions as well.

“第二组左边耳朵有嗡嗡声，第三组右边耳朵有嗡嗡声。(The second group of signals (FL-R) has a humming sound in the left ear, and a humming sound was in the right ear for the third group of signals (L-FR).)” – Interviewee 10

“在听第三组信号时，感觉右耳听到了清楚的语音信号，没有感觉到嗡嗡声。第二组左边语音信号更清楚。(When listening to the third group of signals (L-FR), I felt a clear voice signal in the right ear without humming sounds. The voice signal is clearer on the left side for the second group of signals (FL-R).)” – Interviewee 05

In addition, several students responded that there was an obvious difference in auditory perception in the left and right ears for the dichotic conditions of L2 FL-R and L-FR, while some others reported little difference between the two.

Compared to the L1 dichotic signals, some students mentioned that the L2 dichotic signals they perceived were different, especially for FL-R.

“明显感觉第二组是右边发音，但是与中文明显不同。(It is obvious that the second group of signals (FL-R) is speaking on the right side, but it is apparently different from Chinese.)” – Interviewee 04

The Interviewee 04 further claimed that “感觉英文第一组和第二组声音在左边,非常相似。(It feels that the sounds of the first group (FL-FR) and the second group (FL-R) in English are on the left side, which is very similar)”.

4.2.1.2 Clarity of the signals

When asked about the feelings of listening to the signals during the experiment, all students talked about the signal clarity. As one interviewee put it:

“清晰是听感的主要关注点。(Clarity is my major concern for auditory perception.)” – Interviewee 05

All students responded that both L1 and L2 NL-NR stimuli were the clearest among all the signals.

“第四组能够清楚地听到，而且能正常理解意思。(For the fourth group of signals (NL-NR), I can hear clearly and understand the meaning.)” – Interviewee 04

“第四组很清晰，听得非常清楚。(The fourth group (NL-NR) is very clear, and I can hear it very clearly.)” – Interviewee 12

But there still was a difference of perception between L1 and L2 NL-NR due to language proficiency. The L2 NL-NR stimulus was reported as a clear signal at a faster speed, which was harder to comprehend than that of L1.

“英文第四组很清晰，但是语速过快，不能完全理解句子的意思，或者说能大致了解意思。中文第四组能够完全听懂。(The fourth group in English (L2 NL-NR) is very clear, but the speed of speaking is too fast to fully understand the meaning of the sentence. In other words, I can roughly understand the meaning. I can fully understand the fourth group in Chinese (L1 NL-NR).)” – Interviewee 02

For the clarity of the dichotic stimuli, FL-R was perceived as clear as L-FR in both L1 and L2, but the humming sounds were unusual and unfamiliar to the students.

“能够听清楚，但是总感觉怪怪的，总是有点听不清楚的感觉，但是又说不出来，这两组对于句子的理解倒是没有造成多大的困难。(I can hear clearly, but it always feels weird, always a little unclear. But I can’t explain why. These two groups (FL-R and L-FR) did not cause many difficulties in understanding the sentences.)” – Interviewee 01

Regarding the FL-FR signals, all respondents considered them as the most unclear sounds, further, L2 FL-FR was vaguer compared to L1 FL-FR:

“和中文第一组一样听不清，英文更模糊，更难理解句子。(Like the first group in Chinese (L1 FL-FR), it is not clear. And it (L2 FL-FR) is vaguer and more difficult to understand the sentences.)” – Interviewee 01

4.2.1.3 Perception of comfort/discomfort

Some students gained a perception of discomfort towards the FL-FR signals. They felt uncomfortable and unpleasant while listening to the FL-FR signal due to the lack of clarity. But several students held a neutral opinion about the perception of comfort and discomfort.

“不舒服，因为听不清楚。(I felt uncomfortable because I can’t hear clearly.)” – Interviewee 09

“听不清的信号，总是试图听清，给我造成了心情的不愉悦，尤其是第一组。(I always try to hear the unclear signals, which makes me feel unpleasant, especially the first group of signals (FL-FR).)” – Interviewee 06

“第一组只能听见哼哼的声音，不能说不舒服，只能说听不明白什么意思。(Only the sounds of humming can be heard in the first group (FL-FR). I can’t say that it is uncomfortable. I can only say that I don’t understand what it means.)” – Interviewee 02

As to the NL-NR signal, all the students responded that it was comfortable to listen to because the speech sounds could be heard clearly. For the dichotic FL-R signals in L1 and L2, most of the respondents considered them as comfortable signals as well.

“中文除了第四组，第二组更舒服。(Besides the fourth group in Chinese (L1 NL-NR), the second group (L1 FL-R) is more comfortable.)” – Interviewee 05

“英文第四组更舒服些，第二组还不错。(The fourth group in English (L2 NL-NR) is more comfortable, and the second group (L2 FL-R) is not bad.)” – Interviewee 02

But for the L-FR signal, some students reported the feeling of discomfort:

“中文第三组信号会让头脑感觉不舒服。(The third group of signals in Chinese (L1 L-FR) can make the brain feel uncomfortable.)” – Interviewee 05

“英文第三组感觉有点杂音，不舒服。(The third group in English (L2 L-FR) sounds a little noisy and uncomfortable.)” – Interviewee 03

“听英文第三组的时候脑袋会停止思考。(When listening to the third group in English (L2 L-FR), my brain stops thinking.)” – Interviewee 06

4.2.1.4 Preference and interest

Although the filtered signals were unusual and difficult for the students to understand, some expressed an interest in the FL-FR signals:

"对第一组信号更感兴趣，虽然句子理解有非常严重的困难。(I am more interested in the first group of signals (FL-FR), although there are very serious difficulties in sentence comprehension.)" – Interviewee 05

Others had a neutral position on the preference of the FL-FR signal:

"第一个听不清楚，没任何反感，只是觉得他在哼哼。(I cannot hear the first group (FL-FR) clearly. But no dislike for it. I just think he is humming.)" – Interviewee 04

When asked about the preferred signal, most students replied that the L1 and L2 NL-NR signals were their favorite because of the familiarity and the signal clarity.

"喜欢中文第四组，因为听的清楚，这是平常听到的，一听就清楚它的意思。(I like the fourth group in Chinese (L1 NL-NR) because I can hear it clearly. It is what I usually hear, and I know its meaning when I hear it.)" – Interviewee 01

"英文信号喜欢第四组，因为能够听得清楚些，所以会下意识地想把它听懂。(I like the fourth group in English (L2 NL-NR) because I can hear it clearly. I subconsciously want to understand it.)" – Interviewee 10

In terms of the dichotic signals, several interviewees expressed a preference for the FL-R signal over the NL-NR and L-FR signals, and they wished to hear more of this kind of speech sound in the future.

"第四组感觉没有第二组声音通透，更喜欢右耳听的清楚的感觉，因为第四组信号感觉双耳都在共鸣一样，声音很大。(I feel that the sounds of the fourth group (NL-NR) are not as transparent as the second group (FL-R). I prefer the clear sounds in the right ear because the fourth group of signals (NL-NR) feels like both ears are resonating, and the sound is very loud.)" – Interviewee 06

"第二组相较于第三组更清楚些，更喜欢第二组，说不清楚喜欢的理由。(The second group (FL-R) is clearer than the third group (L-FR). I prefer the second group (FL-R), but I can't explain why I like it.)" – Interviewee 04

“当听到第二组时，明显可以听清楚，心情会变得稍好。(When I hear the second group (FL-R), I can hear it clearly. My mood gets better.)” – Interviewee 09

“第二组语音非常喜欢，并且希望以后多听见这种声音。(I really like the voice in the second group (FL-R) and hope to hear more of it in the future.)” – Interviewee 02

In addition, some students showed a feeling of neither like nor dislike of the dichotic FL-R signal.

For the L-FR signal, one respondent showed a preference for it, especially in L1:

“中文第二、三组中更喜欢第三组，说不清楚喜欢的理由。 (Among the second (FL-R) and third groups (L-FR) in Chinese, I prefer the third group (L-FR), but I can’t explain why I like it.)” – Interviewee 01

But others showed dislike or held a neutral opinion of preference for the L-FR signal:

“第三组说不清感觉，就是不喜欢。(For the third group (L-FR), I cannot tell how it feels, I just don’t like it.)” – Interviewee 05

“第三组不喜欢也不讨厌。(I neither like nor hate the third group (L-FR).)” – Interviewee 10

4.2.2 Understanding of the signals in Verbotonal-based dichotic and diotic listening conditions

Students’ views on understanding the meanings and structures of the stimuli in Verbotonal-based dichotic and diotic listening conditions were further investigated. Most respondents stated that the FL-FR stimuli in both L1 and L2 hindered comprehension of the sentence meanings and structures:

“中文第一组非常模糊，造成理解的困难。虽然能听出是中文句子，但只能很模糊地听到个别的字，严重影响理解。尝试理解，非常吃力。(The first group of Chinese (L1 FL-FR) is very vague, which causes difficulty in understanding. Although it can be recognized as Chinese sentences, I can only hear individual words vaguely, which seriously affects understanding. Trying to understand the sentences entails strenuous effort.)” – Interviewee 07

“英文很难理解句子的含义, 直接不能听到句子结构。(It is very difficult to understand the meanings of the sentences in English (L2 FL-FR). I cannot hear the sentence structure at all.)” – Interviewee 06

As to the dichotic stimuli FL-R and L-FR, some students agreed that FL-R facilitated understanding of sentence meanings and structures but a few held an opposite view:

“第二组有助于理解句子的含义，语义理解没有造成困难，能听清楚语义信息。和第三组比，第二组更好。(The second group (FL-R) helps understand the meaning of sentences, which causes no difficulty in semantic comprehension. I can hear semantic information clearly. Compared to the third group (L-FR), the second group (FL-R) is better)” – Interviewee 02

“在结构理解上,第二组比第三组更好。(In terms of a structural understanding, the second group (FL-R) is better than the third group (L-FR).)” – Interviewee 04

Only a few stated that L1 L-FR could help understand the sentence meanings and structures:

“中文第三组有助于理解句子的含义，语义理解没有造成困难，能听清楚语义信息，能理解句子结构。(The third group in Chinese (L1 L-FR) helps understand the meaning of the sentence, and it does not cause difficulty in semantic understanding. I can hear the semantic information clearly and understand the sentence structure.)” – Interviewee 01

But for a few respondents like Interviewee 01, L2 L-FR was as difficult as L2 FL-FR to comprehend the meanings and structures of the sentences:

“英文第三组和第一组一样，很难理解句子意思和结构。(The third group in English (L2 L-FR), like the first group (L2 FL-FR), is difficult for me to understand the sentence meanings and structures.)” – Interviewee 01

Regarding the L1 and L2 NL-NR signals, the majority of students claimed that they could understand the sentence signals

“中文第四组更有助于理解句子的含义、结构，不费力，能够完全听懂。(The fourth group of Chinese (L1 NL-NR) is more helpful to understand the meanings and structures of the sentence. It is easy to understand and I can fully understand)” – Interviewee 02

“个人感觉句子的含义、结构都是建立在听清楚上的，那么感觉英文第四组更好。除了第四组外，第二、三组也可以。(Personally, I feel that the understanding of sentence meanings and structures is based on listening clearly, so I feel that the fourth group in English (L2 NL-NR) is better. Except for the fourth group (L2 NL-NR), the second (L2 FL-R) and third groups (L2 L-FR) are also fine.)” – Interviewee 04

When talking about the differences in understanding the sentences between the L1 and L2 signals, most of the students put it that they faced more challenges and had more difficulties in understanding the L2 signals due to the unfamiliarity of the second language.

“英文信号语速快，听起来更模糊。(The speed of English signals is fast and sounds vaguer.)” – Interviewee 08

“英文信号不能理解完整的意思，英文比较难懂。(I cannot understand the complete meanings of English signals, and English is difficult to understand.)” – Interviewee 11

“中文相对能获取到的内容多。中文更清晰，可能因为是母语，会更亲切。中文更容易让人专心。(I can get more information from Chinese signals. Chinese is clearer, perhaps because it is a native language. It sounds more friendly. Chinese makes me concentrate more.)” – Interviewee 07

“中文一旦听清楚就更容易稳定地产生联想，即使只有只言片语也能去联想句子的意思，英文即使听懂某几个单词也无法进行宽泛的联想，英文还需要去理解字词的意思和句子结构，比中文的困难。中文母语容易懂，英文不熟悉得思考。(Once you hear the Chinese signals clearly, it is easier to make associations stably. Even if you only hear a few words, you can associate the meaning of a sentence. But you cannot make broad associations if you hear and understand a few words in English. You also need to understand the word meanings and sentence structures in English, which is more difficult compared to Chinese. Chinese as a native language is easy to understand, but English is not familiar enough and I need to think more.)” – Interviewee 12

4.2.3 Views on using the signals for language learning

When asked about the views on using the signals in Verbotonal-based dichotic and diotic listening conditions for language learning, students expressed various opinions and their interests.

Some students expressed a preference for the unfiltered NL-NR signals as listening materials to learn English because the audios were clear enough to identify the words and structures:

"以后学习英语，还是希望继续使用第四组的声音，因为可以听清楚字词。(In the future, I still hope to continue to use the fourth group of sounds (NL-NR) to learn English, because I can hear the words clearly.)" – Interviewee 02

But several respondents accepted both of the unfiltered diotic signal NL-NR and the dichotic FL-R and L-FR signals as learning materials when studying a foreign language:

"学习英语想用第四组信号，第二组和第三组也可以，但是第一组不可以。 (When learning English, I would like to use the fourth group of signals (NL-NR). The second (FL-R) and third groups (L-FR) can also be used, but the first group (FL-FR) is not acceptable.)" – Interviewee 04

In addition, a number of students showed an interest and willingness to use all the signals and the order of signal presentation as in the experiment while learning English.

"如果在学习英文时，有这些语音信号，并按这种播放顺序是可以接受的。非常愿意进行尝试。(It is acceptable to have these voice signals and play them in this order when learning English. I am very willing to try.)" – Interviewee 05

To sum up, perceptions of the L1 and L2 signals in Verbotonal-based dichotic and diotic listening conditions varied from individual to individual, as well as their preferences for the signals. Students' perceptions and preferences were considerably influenced by the familiarity with the native language and the physical quality of the signals (the unfiltered and filtered sounds). But students still showed an interest in the low-pass filtered signals and were quite willing to learn English with the auditory signals in Verbotonal-based dichotic and diotic listening conditions.

4.3 Summary

This chapter presents the results of the current study. ERP findings illustrated the mental workload and latencies elicited by the L1 and L2 stimuli in Verbotonal-based dichotic and diotic listening conditions, as well as the differences in neural processing between the stimuli. Cerebral regions involved in the processing of the signals were reported in the fMRI results. In addition, comparisons of the regions activated by the stimuli were made, and hemispheric lateralization induced by the signals was investigated. Further, students' perceptions and preferences for the L1 and L2 signals were explored.

CHAPTER 5
DISCUSSION

This chapter discusses the findings reported previously. To answer the research questions raised in Chapter 1, this chapter is organized into three sections. The first section focuses on students' brain responses while listening to the L1 and L2 signals in the ERP and fMRI experiments. Students' opinions are explored in the second section. Taken together, the final section investigates if there exist optimal and non-optimal auditory language signals in Verbotonal-based dichotic and diotic listening conditions

5.1 Temporal and spatial neural signatures of the processing of the auditory language signals

5.1.1 Temporal and spatial neural signatures of the processing of the L1 signals

The current study identifies that the change of the physical quality of the auditory language signals leads to different activities in the brain. The ANOVA tests for the mean peak amplitudes of the elicited ERPs showed that the L1 stimuli in Verbotonal-based dichotic and diotic listening conditions were significantly different from each other ($p < .001$), indicating that the brain responded to the four configurations of the L1 stimuli variously. As VTA advocates, the perception of a language and the brain's neural connectivity could be restructured by the low-pass filtered (i.e., low-frequency) stimuli (Asp et al., 2012; Guberina & Asp, 1981, 2013; Kim & Asp, 2002). The results of the ANOVA tests demonstrated that the change of the physical quality of signals indeed impacted the processing of the auditory language signals, which could further influence the perception of the language being learned and the "meaning-making" process in language development/learning (Guberina & Asp, 1981; Lian, 2004).

5.1.1.1 Discussion on the ERP findings

Regarding the language-related ERP components, the L1 stimuli in Verbotonal-based dichotic and diotic listening conditions elicited the N400Pz, N400Cz, P400Pz, and P400Cz components. Specifically, L1 FL-FR, L-FR, and NL-NR elicited the N400Pz, suggesting the process of lexical access (Dien et al., 2010), but FL-R elicited the P400 for semantic integration. When compared the mean peak amplitudes, AUC, and reaction time of the components induced by the four stimuli, FL-R elicited the smallest peak amplitude and the smallest AUC with a longer reaction time, reflecting the lightest mental workload was required for semantic integration although the reaction time could be longer. The NL-NR, L-FR, and FL-FR stimuli resulted in a quite heavier mental load and a much shorter reaction time for lexical access. Compared to NL-NR and L-FR, FL-FR, as a both-ear-filtered prosodic signal, lacks the high frequencies to identify the words, so the mental workload is lowered for lexical access. The MMN analysis showed that the difference in the mental workload required for lexical access between FL-FR and FL-R was the smallest, but a huge difference between FL-FR and NL-NR.

As to the components at the central electrode site, FL-FR, FL-R, and NL-NR elicited the P400Cz while L-FR induced the N400Cz. Both of the components were related to semantic manipulation, but the P400Cz elicited by FL-FR, FL-R, and NL-NR demonstrated the semantic integration of sequential representations and an expectancy process for sequential probabilities (Dien, 2009; Dien et al., 2010; Liu et al., 2009). According to Cummings et al. (2006), the N400 reflected the earlier process of the established meaningful semantic representations. The L-FR signal induced the earlier process of semantic manipulation with the heaviest mental workload and a longer reaction time, but FL-R led to a quite light processing load and a short reaction time for the processes of semantic integration and expectancy. The results indicate that both the L-FR and FL-R induce semantic processing, but L-FR requires more neuronal resources and longer time for the processing of the established meaningful representations. For FL-R, semantic manipulation in the sequential representations and the expectancy process with a short reaction time is induced. From the MMN analysis, a great discrepancy in the mental load for the semantic integration and expectancy processes was found between FL-FR and L-FR, and the smallest difference was between NL-NR and FL-R. This great difference between FL-FR and L-FR indicates that FL-FR as an unusual and unfamiliar signal for the participants imposes a heavy cognitive load for the processing of semantic integration and the expectancy process, while L-FR results in the heaviest processing load for the earlier process of semantic manipulation in the established meanings. But the difference in cognitive load required for semantic integration of the NL-NR and FL-R stimuli is small.

Besides the earlier process of the established semantic representation, L-FR and NL-NR elicited the N600 at the central electrode site, indicating the later process of stimulus semantics and possibly the stimulus-general process (working memory, attention, maintaining task-demands, decision-making, etc.) (Cummings et al., 2006; Shahin et al., 2006). NL-NR induced the lowest mental workload for the stimulus semantics and stimulus-general processes with a quite long reaction time. It may be due to the fact that NL-NR is an unfiltered L1 signal that the participants are quite familiar with and all the frequencies are maintained to identify the words. But L-FR resulted in a relatively high load for stimulus semantics and stimulus-general processing. Meanwhile, FL-FR and FL-R elicited the P600 at the central electrode site, showing the process of syntactic manipulation. Notably, FL-R led to the heaviest mental workload and the shortest reaction time for syntactic processing at the central site, which thus made a huge difference in the processing load for syntactic manipulation between FL-R and L-FR according to the MMN analysis. This result indicates when FL-R involves more neuronal resources for syntactic processing, L-FR still requires mental workload for the later processing for stimulus semantics and cognitive processes such as working memory, maintaining task demands, decision-making, and attention (Cummings et al., 2006; Shahin et al., 2006).

Regarding the parietal electrode site, four configurations of the L1 stimuli elicited the P600 component, reflecting syntactic processing (Daltrozzo & Conway, 2014; Silva-Pereyra et al., 2007; Swaab et al., 2012). At the parietal site, FL-FR resulted in the heaviest processing load with a relatively short reaction time for syntactic processing. It is assumably due to the both-ear-filtered signals that make the comprehension of the sentence structures unrecognizable and difficult. But FL-R induced the lightest load and a rather short reaction time for syntactic processing at the parietal site. As a result, it is confirmed in the MMN analysis that the difference in the mental workload for syntactic manipulation between FL-FR and FL-R is greater. This result suggests that FL-FR without high frequencies to identify the words imposes more cognitive load and for syntactic processing, but the FL-R signal requires less neural resources and a shorter reaction time for the processing of the sentence structures.

5.1.1.2 Discussion on the fMRI findings

As for the brain activation patterns during listening to the L1 stimuli in Verbotonal-based dichotic and diotic listening conditions, fMRI results showed that overlapping and distinct brain regions were engaged in processing the four configurations of the signals.

Increased activation in the bilateral STG was detected after the presentation of the four stimuli in L1. The bilateral STG is identified as the area for spectrotemporal analysis of speech signals (Hickok & Poeppel, 2007), known as the primary auditory cortex processing the auditory signals from the contralateral ears (Mangold & Das, 2020). Moreover, the bilateral STG is involved in the acoustic analysis of the four configurations of the L1 stimuli since no clear laterality is found.

As to the MTG, four configurations of the L1 signals induced strong activations in the bilateral MTG. According to Hickok and Poeppel's model (Hickok & Poeppel, 2000, 2004, 2007) for speech perception, the MTG, connected to the STG (spectrotemporal analysis) and the STS (phonological access) in the ventral pathway of the temporal lobe, is identified to support speech comprehension, especially for lexical access and combinatorial processes with a weak left-hemisphere dominance. The current study confirms the results of the previous studies that early-stage speech perception and comprehension recruits both the STG and MTG in Chinese speakers. The right MTG, together with the bilateral STG and the left IFG, is identified as the area for the basic perception of Chinese linguistic pitches (Kwok et al., 2016). The left MTG, with the left IFG, is found to be involved in semantic processing in Mandarin Chinese (C.-Y. Wu et al., 2012; Zhang et al., 2019). The current results showed that right-lateralized activations were observed in the MTG after the FL-FR and L-FR signals were presented, but no clear hemispheric laterality was observed for FL-R and NL-NR. It indicates that the FL-R signal, eliciting activations in the bilateral STG and MTG with the left IFG, induces basic acoustic, phonological, and semantic processing. Similar to FL-R processing, the L1 NL-NR stimulus activates the bilateral STG and MTG without clear hemispheric laterality, indicating the acoustic, phonological, and semantic processing of the signal. Further, NL-NR induces activations only in the bilateral STG and MTG regions, which might result from the fact that L1 NL-NR is a both-ear-unfiltered signal in the first language that the participants are quite familiar with and proficient in. Thus, more intensive recruitment of the brain areas and a fewer number of the activated regions are found in the first language processing (Vingerhoets et al., 2003). The FL-FR signal is a both-ear-filtered signal containing sentence prosodic information, so bilateral activity in the MTG is detected for early-stage language perception, meanwhile, right-lateralized activity indicates basic acoustic and linguistic prosody processing. This processing pattern is similar to that of L-FR. For L-FR, the bilateral STG and MTG are involved in early-stage speech processing, however, it is lateralized to the right MTG for basic acoustic and linguistic processing rather than semantic processing in the left MTG. Possibly, for L-FR, the left hemisphere that is less specialized in processing prosodic information receives the filtered sounds from the right ear, and the right hemisphere gets the unfiltered signals. As a result, more involvement of the right hemisphere especially the right MTG is observed for acoustic and prosodic processing, but less engagement of the left MTG is detected for semantic processing.

When it comes to the frontal regions, the bilateral IFG was significantly activated by the L1 FL-FR, FL-R, and L-FR stimuli. Hemispheric laterality in the IFG showed that FL-FR resulted in right-hemisphere dominance, FL-R led to left-lateralized activations, and no clear laterality was found for L-FR. The IFG is linked in the dorsal stream supporting sensory-motor integration and semantic processing, which is left-dominant (Hickok & Poeppel, 2007; Newman et al., 2010; Obleser & Kotz, 2010). The left IFG is involved in tonal and semantic processing (e.g., Kwok et al., 2016), and functional connectivity from the left IFG to the left MTG is detected in semantic processing of Mandarin Chinese (Zhang et al., 2019). But the current study shows that the bilateral IFG is engaged in processing the both-ear-filtered prosodic signal FL-FR and the dichotic signals FL-R and L-FR. According to Gandour et al. (2004), the RH fundamentally manipulates the processing of prosody but language processing is more complicated than the processing of acoustic information such as tones and pitch, thus, the LH is involved in higher-order language processing. The current study highlights the prosody of the language signals by sending the filtered signals diotically and dichotically to the ears, which results in increased activations in the bilateral IFG for prosodic and linguistic processing. Although L1 FL-FR lacked the high frequencies to identify the words, participants in the interview claimed that the signals could be recognized as Chinese sentences. Assumably, FL-FR induces prosodic and linguistic processing in the bilateral IFG, in addition, the right-sided laterality in the IFG indicates more involvement for prosodic processing. Since L1 NL-NR is a both-ear-unfiltered signal in the native language, it is familiar to the participants who can be regarded as proficient language users. For proficient language users, they can manipulate the structure processing more automatically by recruiting the FOP and the aSTG (Friederici et al., 2003), but this process may be less automatized and the IFG (BA 44) is involved for less proficient users (Brauer & Friederici, 2007; Rüschemeyer et al., 2005). The Chinese students in the current study are proficient users to process L1 NL-NR, as a result, only the bilateral STG and MTG without any clear hemispheric lateralization are recruited for speech perception and comprehension. However, the dichotic signals FL-R and L-FR sounded somehow unusual and unfamiliar for the participants, therefore, involvement of the bilateral IFG for prosodic and linguistic processing was detected. It was left-lateralized in the IFG for FL-R processing, and laterality was not clear for L-FR. It may be due to the fact that FL-R sends the unfiltered signals to the left hemisphere that is more specialized in linguistic processing, as a result, the left IFG becomes more involved in linguistic and semantic processing. On the contrary, the L-FR sends the filtered signals to the left hemisphere that is less specialized in processing prosodic information. Thus, it results in the recruitment of the bilateral IFG for prosodic and linguistic processing without clear laterality, rather than the left-lateralized activations in the IFG for semantic processing.

Significant activations in the MFG were observed during the presentation of the L1 stimuli as well. FL-FR and FL-R induced robust activations in the bilateral MFG, and L-FR activated the right MFG. It is reported that Chinese speakers recruit the left IFG for complex linguistic processing and the left MFG for working memory during extracting phonological information from the characters (Siok et al., 2008; Tan et al., 2005), in addition, the left MFG is involved in Chinese morphological judgments (Zou et al., 2016). For Chinese prosodic processing, right-lateralized activations in the STS and MFG were detected for the processing of prosodic units (Gandour et al., 2004). Further, a rightward asymmetry in the MFG is observed in the processing of sentence-level linguistic prosody for Chinese speakers (Gandour et al., 2007). In the current study, FL-FR and FL-R elicited activations in the bilateral MFG, together with the bilateral STG, MTG, and IFG, indicating complex language processing with prosodic and linguistic processing at the sentence level in both hemispheres. But hemispheric laterality showed right-sided dominance in the MFG for FL-FR and left dominance in the MFG for processing FL-R. Since the participants can recognize it as a signal of Chinese sentences, linguistic processing is induced in the bilateral MFG for complex sounds, meanwhile, FL-FR is a both-ear-filtered prosodic signal that results in right-lateralized activations in the MFG for prosodic processing. As for FL-R, both hemispheres receive the signals that they are specialized in processing, i.e., the linguistic signals to the left hemisphere and the prosodic signals to the right hemisphere. As a result, the bilateral MFG, together with the bilateral STG, MTG, and IFG, get involved in complex linguistic and prosodic processing, especially the left MFG shows more engagement in phonological, semantic processing, and working memory (Siok et al., 2008; Tan et al., 2005; Tan et al., 2000; C.-Y. Wu et al., 2012). Different from FL-FR and FL-R, the L-FR stimulus only induced significant activity in the right MFG, which may also result from hemispheric specialization. To be specific, L-FR sent the unfiltered signals to the right hemisphere and the filtered signals are sent to the left hemisphere. Then, the right MFG was strongly activated for sentence-level linguistic prosody, but the filtered prosodic signals from the right ear to the left hemisphere did not significantly activate the left MFG for phonological and semantic processing (Tan et al., 2000; C.-Y. Wu et al., 2012). Unlike the processing of FL-FR and FL-R recruiting the bilateral MFG for complex language processing, the L-FR stimulus with the right-ear-filtered signals sounds unusual to the left hemisphere that is less specialized in prosodic processing. As a result, it seems that the left MFG is reluctant to process the filtered prosodic signals coming from the right ear.

5.1.1.3 Discussion on the comparisons within the L1 stimuli

To discuss the neural processing differences between the L1 stimuli in Verbotonal-based dichotic and diotic listening conditions, the results of ERP MMN differences and the results of the fMRI group analysis are taken into account. The MMN results showed that greater differences in semantic integration and lexical access were found in FL-FR – L-FR and L-FR – FL-FR, respectively, at the Cz site in the 300-500 ms time window. Meanwhile, the results of the fMRI group analysis are consistent with the MMN results, demonstrating that L-FR induced more activations in the regions of the left PrCG, IPL, and SPL as compared to FL-FR. A meta-analysis shows that the left PrCG is involved in phoneme, lexical tone, and prosody perception, especially for tonal language speakers (Liang & Du, 2018). The anterior part of the left IPL along with the STG forms the Wernicke area, which is responsible for the phonological analysis of words and merging syllables into words (Prpić, 2015). In addition, the posterior part of the left IPL is the site for semantic analysis and multisensory synthesis (Prpić, 2015). For Chinese speakers, the left SPL is involved in Chinese character reading (Fu et al., 2002). Thus, both of the MMN and fMRI results suggest a huge difference in lexical/semantic processing between the neural processing of L-FR and FL-FR, indicating more neuronal resources and brain regions are required for the lexical/semantic processing of the L-FR signal as compared to FL-FR. On the contrary, a small difference in the neural resources/processing load required for semantic integration and lexical access between FL-R and NL-NR is found in the MMN results, which is not statistically significant in the comparison of activated brain regions in the fMRI group analysis. It indicates that the brain regions for processing FL-R and NL-NR are not significantly different.

At the parietal site in the 300-500 ms time window, the MMN results demonstrated that a greater difference existed between FL-FR and NL-NR for lexical access, which was consistently detected in the fMRI group analysis that FL-FR significantly induced more involvement of the right MFG and PrCG than that of NL-NR. As mentioned above, the MFG is responsible for processing sentence-level linguistic prosody in Chinese speakers (Gandour et al., 2007). In addition, the right PrCG is involved in the sensorimotor processing of Chinese syllables in the word rhyming judgment task (Cao et al., 2014), as well as the involvement in cognitive control of language switching in bilinguals (Luk et al., 2012). The both-ear-filtered signal FL-FR is a prosodic signal that is unusual for the participants, but the signal still can be recognized as the Chinese sentence signal. Thus, this unfamiliar and prosodic FL-FR signal recruits more neuronal resources and more brain regions for prosodic processing and lexical access, and even sensorimotor processing and cognitive control are involved in FL-FR processing, especially when compared to the unfiltered sentence signal NL-NR. On the other hand, a quite small discrepancy in the processing load for lexical access between FL-FR and FL-R was observed in the MMN analysis, which might result from both of the signals elicited the smaller peak amplitudes and AUCs during lexical access and semantic integration processes respectively at the parietal site. In addition, the difference in the brain regions for processing the two stimuli was not significant in the fMRI group analysis.

As to the MMN findings at the central electrode site in the 500-900 ms time window, the largest differences in the processing load were observed between FL-R and L-FR. Compared to L-FR, more neuronal resources are required for the syntactic processing of FL-R. But processing L-FR needs more mental workload for stimulus semantics and stimulus-general processes (i.e., cognitive processes such as working memory, decision-making, maintain task demands, and attention) (Cummings et al., 2006; Shahin et al., 2006). Although the mental workload for processing FL-R and L-FR differs in the MMN results, brain regions for processing the two stimuli are not significantly different from the fMRI group analysis.

The greater difference in the processing load for the later processes of stimulus semantics and cognitive processes was also found in the comparison between FL-R and FL-FR at the parietal site in the 500-900 ms time window. It indicates that FL-R requires more neuronal resources for stimulus semantics and cognitive processes as compared to FL-FR. It is confirmed in the fMRI group analysis that the processing of FL-R induces more involvement of the left PrCG and PoCG as compared to the processing of FL-FR. The left PrCG, as mentioned before, is responsible for phoneme, lexical tone, and prosody perception in tonal language speakers (Liang & Du, 2018), and the PoCG is involved in a more complex sensory-motor network for auditory feedback related to speech production and motor control (Parkinson et al., 2012), especially the left PoCG is considered to be involved in speaking (Behroozmand et al., 2015; Zheng et al., 2010). FL-R with linguistic and prosodic signals in the dichotic listening condition involves more neuronal resources and brain regions for stimulus semantics and cognitive processes in a complex sensory-motor network as compared to the both-ear-filtered prosodic signal FL-FR. In addition, the difference in the processing load between FL-FR and NL-NR was also great, indicating that FL-FR required more neural resources for syntactic processing when compared to NL-NR at the parietal site in the 500-900 ms time window. Consistently, the fMRI group analysis showed a significant difference in the right MFG and PrCG between FL-FR and NL-NR. L1 NL-NR is the L1 unfiltered language signal that the students are familiar with and proficient at processing, while FL-FR is a prosodic signal that is unusual for the participants. As a result, the right MFG becomes more involved in the processing of sentence-level linguistic prosody in FL-FR (Gandour et al., 2007; Gandour et al., 2004), and more activity is detected in the right PrCG for the sensorimotor processing of Chinese signals (Cao et al., 2014). The results indicate that the both-ear-filtered FL-FR signal, lacking the high frequencies to identify the words, imposes more cognitive load for the processing of the sentence structures, meanwhile, FL-FR activates more brain areas for the sensorimotor processing and the processing of sentence prosody as compared to NL-NR.

5.1.2 Temporal and spatial neural signatures of the processing of the L2 signals

5.1.2.1 Discussion on the ERP findings

At the parietal electrode site in the 300-500 ms time window, the four configurations of the L2 signals (L2 FL-FR, FL-R, L-FR, and NL-NR) elicited lexical access (Dien, 2009; Dien et al., 2010; Liu et al., 2009), indicating earlier semantic processing of the established meaningful representations (Cummings et al., 2006).

Among the four configurations, the least neuronal resources over the processing time were required for lexical access of the FL-R signal. As a result, the MMN analysis showed a larger difference in the processing load of lexical access between FL-R and FL-FR, in addition, greater differences were found in the mental workload of semantic integration in the comparisons of L-FR – FL-R and NL-NR – FL-R. This result suggests that the processing load for semantic processing and lexical access of the FL-R signal is quite small so that the differences of the neuronal resources involved in lexical and semantic processing between FL-R and the other three stimuli become larger.

In the findings at the central site in the 300-500 ms time window, L2 FL-FR, FL-R, and NL-NR induced the earlier semantic processing of the established meaningful representations (Cummings et al., 2006), but only L-FR involved more neural resources for semantic integration of the sequential representations and expectancy processes (Dien, 2009; Dien et al., 2010; Liu et al., 2009). From the peak amplitude and AUC for the earlier semantic processing, FL-FR induced the smallest processing load as compared to NL-NR and FL-R. The processing of L2 FL-FR at the central site is different from L1 FL-FR. According to the interviewees, L1 FL-FR can be recognized as Chinese sentences so the mental workload for semantic integration and the expectancy process is relatively higher. However, L2 FL-FR, as a prosodic signal in L2, cannot affirmatively be recognized as English sentences (from the interview data). Thus, the smallest load for processing L2 FL-FR may result from the factor of L2 language proficiency and the lack of high frequencies to identify the words. In other words, the processing of the L2 both-ear-filtered signal could be considered as prosody processing with the smallest load for the semantic processing for Chinese EFL learners. On the contrary, NL-NR with unfiltered linguistic and prosodic information induced a heavier load for semantic processing. Therefore, a huge difference for semantic manipulation between FL-FR and NL-NR was found. Further, the L-FR signal elicited the heaviest load for sematic integration and expectancy processes, and NL-NR involved a higher load for the earlier semantic processing, thus making the difference in semantic processing between NL-NR and L-FR greater.

As to the results at the central electrode site in the 500-900 ms time window, only the FL-R elicited the later semantic processing and stimulus-general processes with the smallest processing load. In the meantime, FL-FR, L-FR, and NL-NR induce syntactic processing. FL-FR, as the prosodic signal lacking information of the sentence structures, resulted in the highest mental load for syntactic processing at the central site. It is different from the smallest load for L2 FL-FR semantic processing discussed above. This result indicates that the both-ear-filtered FL-FR signal lacking linguistic information lowers the load for semantic processing but more neuronal resources are involved for syntactic manipulation. Therefore, it is not surprising that the greatest difference in neural processing is observed between FL-R and FL-FR.

Four configurations of the L2 stimuli elicited syntactic processing at the parietal electrode site in the 500-900 ms time window. The syntactic manipulation of L-FR required the highest mental workload with the longest reaction time of all, and the FL-R signal induced the second-highest load for syntactic processing. But the FL-FR and NL-NR signals resulted in a relatively lighter processing load at the parietal site. Unlike the syntactic processing of FL-FR with a heavy load at the central site, fewer neural resources are involved in syntactic processing for FL-FR at the parietal site as compared to the other three stimuli. In the comparisons of the neural processing between the stimuli, the greater difference in syntactic processing was between FL-R and FL-FR. This result indicates that the FL-R signal involves more neuronal resources for syntactic processing as compared to the prosodic signal FL-FR at the parietal site. Meanwhile, another great difference in stimulus-general processes was observed between FL-FR and NL-NR. It indicates that processing the both-ear-filtered signal FL-FR requires more neuronal resources for the cognitive processes (such as working memory, attention, maintaining task demands), especially when compared to the unfiltered NL-NR at the parietal site. However, the difference in the cognitive processes is quite small between FL-FR and L-FR. It suggests that the involvement of the neuronal resources for the cognitive processes is at a similar level between FL-FR and L-FR. In addition, L-FR and FL-R showed a small difference in syntactic processing. It indicates that the neural resources involved in syntactic manipulation between L-FR and FL-R are at a similar level, thus making the difference slight.

5.1.2.2 Discussion on the fMRI findings

The brain activation patterns for the L1 and L2 stimuli share some similarities, but distinct regions and lateralization patterns are detected while processing the L2 stimuli in Verbotonal-based dichotic and diotic listening conditions.

Increased activity in the bilateral STG was detected for processing the four configurations of the L2 stimuli (i.e., L2 FL-FR, FL-R, L-FR, and NL-NR). Activations in the bilateral STG indicate the spectrotemporal analysis of the signals in the primary auditory cortex (Hickok & Poeppel, 2007; Mangold & Das, 2020). But hemispheric laterality results showed that FL-FR induced right-sided dominance, FL-R resulted in left-sided dominance, and L-FR and NL-NR led to unclear hemispheric dominance in the STG. The L2 FL-FR signal is a both-ear-filtered prosodic signal with unidentifiable words and sentence structures in English. For the participants, FL-FR is so unusual and unfamiliar that they cannot affirmatively recognize it as English sentences (from the interview data). Thus, the STG is bilaterally involved in the acoustic analysis of FL-FR, especially more right-lateralized activations in the STG for decoding prosody vocalization containing unintelligible speech (Grandjean et al., 2005; Sander et al., 2005). As for the processing of the FL-R signal, bilateral STG activity with left-sided dominance appears to be involved in foreign language processing since activity in the left STG increases more with higher foreign-language proficiency (Mårtensson et al., 2012) and involves in cross-language switch (Hosoda et al., 2012). Different from the processing of the four configurations of the L1 signals without clear laterality in the STG, the L2 signals induce different lateralization patterns in the STG. The results indicate that the bilateral STG is involved in the primary auditory processing of the four configurations of the L2 stimuli. But the right STG is more engaged for prosody decoding of unintelligible speech in L2 FL-FR, and left-sided dominance of the STG for foreign language processing is observed in L2 FL-R. Further, no clear hemispheric laterality in the STG is found while processing L2 L-FR and NL-NR.

The MTG was bilaterally involved in the processing of the L2 stimuli, in addition, right-lateralized activations were detected while processing FL-FR, L-FR, and NL-NR, but FL-R did not induce clear hemispheric laterality in the MTG. Activations in the bilateral MTG with the STG induced by the L2 signals confirm that the MTG is involved in early-stage speech perception and comprehension, including lexical access and combinatorial processes (Hickok & Poeppel, 2000, 2004, 2007). Like the processing of L1 FL-FR as discussed before, bilateral activity in the MTG with the STG is observed for speech perception, in addition, right-lateralized activity in the MTG suggests the acoustic and prosodic processing of the signal. The L2 FL-FR signal is a both-ear-filtered prosodic signal in English, so it is an unfamiliar prosodic signal for the participants to process. As a result, early-stage language perception of L2 FL-FR is detected in the bilateral MTG with the bilateral STG, further, right-sided MTG activity, especially for Chinese speakers, is responsible for the basic perception of acoustic information such as tones and pitch (Kwok et al., 2016). L-FR sends the unfiltered language signals to the right hemisphere and the filtered prosodic signals to the left hemisphere, which seems to be a signal that both hemispheres are less specialized for processing. Except for activations in the bilateral STG and MTG for early-stage language perception and comprehension, L-FR leads to more involvement of the right MTG for acoustic and prosodic processing and less engagement of the left MTG for semantic processing (C.-Y. Wu et al., 2012; Zhang et al., 2019). This result is consistent with the findings of L1 L-FR. But L2 NL-NR induced activations in the bilateral MTG and STG with right-sided laterality in the MTG, which is different from the L1 NL-NR processing. It may be due to the fact the L2 NL-NR is a foreign-language signal that word identification relies heavily on the right MTG for extraction of prelexical acoustic and phonetic information when the late bilinguals lack semantic context in L2 (Hervais-Adelman et al., 2014). However, the bilateral MTG without clear laterality, together with left-lateralized STG, was found for processing the FL-R signal. It indicates that FL-R, unlike the other three L2 stimuli, induces bilateral MTG involvement for acoustic, phonological, and semantic processing (C.-Y. Wu et al., 2012; Zhang et al., 2019) in L2 without clear hemispheric laterality.

Regarding the frontal areas, bilateral IFG activity was detected as processing L2 FL-R with right-sided laterality, and L2 L-FR only induced increased activation in the right IFG. In addition, L2 FL-FR and NL-NR did not significantly activate the IFG. The left-lateralized IFG activity is generally considered to be involved in sensory-motor integration and semantic processing (Hickok & Poeppel, 2007; Newman et al., 2010; Obleser & Kotz, 2010), but the right IFG is related to bilingual/multilingual cognitive control in language switching (Bialystok et al., 2012; de Bruin et al., 2014; Ma et al., 2014; Yang et al., 2018; Zhu et al., 2020). More and more studies identify that activity in the right IFG increases during the naming task in the non-dominant language relative to the dominant language (Zhu et al., 2020), and during switches with larger costs compared to smaller switch costs (Bialystok et al., 2012), as with switches to the L2 and L3 compared to non-switch trials (de Bruin et al., 2014). Moreover, increased activity in the right IFG is detected in tonal bilinguals compared to tonal monolinguals (Gao et al., 2020). Thus, the concept of inhibitory control is put forward to indicate that bilinguals/multilinguals use inhibition to switch between languages, and the right IFG as a region in the domain-general inhibition areas plays a role in inhibiting competitors in the dominant language (Bialystok et al., 2012; de Bruin et al., 2014; Ma et al., 2014; Yang et al., 2018; Zhu et al., 2020). In addition, right-sided laterality in the IFG may account for acoustic processing in tonal bilinguals (Gao et al., 2020). The results of the current study indicate that L2 FL-R, sending the unfiltered L2 signals to the left hemisphere and the filtered signals to the right hemisphere, induces sensory-motor integration and semantic processing in the left IFG, but right-lateralized activations in the IFG suggest greater cognitive control in language switching and acoustic analysis of the filtered L2 signals from the left ear. Unlike the L1 FL-R signal induces left-lateralized activations in the IFG for linguistic and semantic processing, processing L2 FL-R exerts more controlled processing resources for language switching. But for L2 L-FR, only the right IFG was activated. This result suggests that the left-ear-unfiltered and right-ear-filtered L2 signals do not significantly induce sensory-motor integration and semantic processing in the left IFG, but recruit more controlled processing resources during language switching and acoustic processing.

In addition, the L2 FL-FR and FL-R stimuli resulted in significant right MFG activity. As discussed before, the right MFG is involved in Chinese prosodic processing, especially the processing of sentence-level linguistic prosody for Chinese speakers (Gandour et al., 2007; Gandour et al., 2004). As a key component in the cognitive control network, the MFG, connected to the IFG, is related to executive control functions such as working memory, maintaining task demands, inhibition, and interference suppression (Chen et al., 2020; Elliott, 2003). Thus, right MFG activity during language switching is associated with domain-general cognitive control, suggesting inhibiting the non-target language to choose the target language in the bilinguals (e.g., Abutalebi et al., 2007; Chen et al., 2020; de Bruin et al., 2014). Unlike L1 FL-FR and FL-R activate the bilateral MFG for complex linguistic and prosodic processing, the L2 FL-FR and FL-R signals only induce the right MFG. The results indicate that the both-ear-filtered prosodic FL-FR and the left-ear-filtered FL-R (sending the filtered signals to the right hemisphere) signals significantly induce prosodic processing in the right MFG. In the meantime, as the L2 signals, FL-FR and FL-R exert more controlled processing resources for language switching.

The L2 NL-NR also induced significant activations in the left PrCG and the right PoCG. The left PrCG, as discussed before, is involved in phoneme, lexical tone, and prosody perception in tonal language speakers (Liang & Du, 2018). But different from L1 processing, stronger left PrCG activations are reported in bilingual language switching and speech motor control (e.g., Guo et al., 2011; Hernandez, 2009; Luk et al., 2012). The result of the current study is consistent with the previous findings in the meta-analysis studies that more involvement of the left PrCG is found in L2 processing than that of L1 (Indefrey, 2006; Liu & Cao, 2016), indicating an L2 effect that requires more language control (de Bruin et al., 2014; Ma et al., 2014) and sensory-motor coordination in L2 processing (Martin et al., 2015). Meanwhile, the right PoCG was recruited in the processing of L2 NL-NR. As mentioned before, the PoCG is a primary somatosensory region (Hernandez, 2009), which is involved in a more complex sensory-motor network for auditory feedback related to speech production and motor control (Parkinson et al., 2012). The right PoCG is identified to be involved in cognitive control and speech motor control in the bilinguals (e.g., de Bruin et al., 2014; Guo et al., 2011; Marangolo et al., 2009). Thus, the processing of L2 NL-NR not only recruits the bilateral STG and MTG for speech perception and comprehension but a more complex sensory-motor network for cognitive control in L2.

5.1.2.3 Discussion on the comparisons within the L2 stimuli

Taking the results of the ERP MMN analysis and the fMRI group analysis together, the differences in the neural processing between the L2 stimuli in Verbotonal-based dichotic and diotic listening conditions could be investigated.

The ERP MMN analysis showed that the greatest difference in semantic manipulation was found between FL-FR and NL-NR at the central electrode site in the 300-500 ms time window. Consistently, the contrast of FL-FR vs. NL-NR demonstrated significant differences in the regions of the bilateral IPL and the right MFG in the fMRI group analysis. It indicates that the processing of FL-FR recruits more brain regions of the bilateral IPL and the right MFG as compared to NL-NR. As discussed before, the left IPL, including the Wernicke's area, is the classic language area for comprehension (Hickok & Poeppel, 2007; Indefrey & Levelt, 2000; Prpić, 2015). But more evidence shows that the bilateral IPL regions are involved in phonological working memory, lexical-phonological representation, semantic integration, and vocabulary learning in a second language (e.g., Abutalebi et al., 2013; Baddeley, 2003; Della Rosa et al., 2013; Li et al., 2014; Mechelli et al., 2004; J. Wu et al., 2012; Yang et al., 2015), which is essential in forming and maintaining lexical representations to further facilitate learning novel words in a foreign language (Baddeley et al., 1998). In addition, the right MFG is involved in the processing of sentence-level linguistic prosody in Chinese speakers (Gandour et al., 2007; Gandour et al., 2004). The L2 FL-FR signal, as a both-ear-filtered prosodic signal in L2, sounds novel and unusual to the participants. As a result, more involvement of the bilateral IPL, together with the right MFG, is found for phonological working memory (e.g., Abutalebi et al., 2013; Baddeley, 2003; Della Rosa et al., 2013; Li et al., 2014; Mechelli et al., 2004; J. Wu et al., 2012; Yang et al., 2015) and sentence-level prosodic processing (Gandour et al., 2007; Gandour et al., 2004) when compared to L2 NL-NR. The second-largest difference in the neural resources involved in semantic manipulation was observed between NL-NR and L-FR in the MMN analysis. The results of the fMRI group analysis also identified this difference that significant differences were found in the bilateral PrCG and the right PoCG in the contrast of NL-NR vs. L-FR. It indicates that the processing of L2 NL-NR recruits more regions of the bilateral PrCG and the right PoCG as compared to that of

L2 L-FR. Specifically, more involvement of the left PrCG is detected in L2 NL-NR processing for phoneme, lexical, and prosody perception (Liang & Du, 2018), more activity in the right PrCG is found for sensorimotor processing (Cao et al., 2014) and cognitive control of language switching (Luk et al., 2012), and more engagement of the right PoCG is observed for cognitive control in L2 (e.g., de Bruin et al., 2014; Guo et al., 2011; Marangolo et al., 2009) relative to the processing of L2 L-FR. Additionally, the difference in neural processing between L-FR and FL-R was small in the MMN analysis, and the brain regions involved in processing the two signals were not significantly different in the fMRI group analysis. It indicates that the discrepancy in the neuronal resources involved for the semantic processing of L-FR and FL-R is small, and the brain activation patterns for processing the two signals are similar.

At the parietal electrode site in the 300-500 ms time window, a greater difference in lexical access between FL-R and FL-FR was observed in the MMN analysis. Meanwhile, the fMRI group analysis supported the difference that the left MFG was more involved in FL-R processing for phonological and semantic processing (Tan et al., 2000; C.-Y. Wu et al., 2012) as compared to FL-FR. It may be due to the fact that the FL-R signal sends unfiltered signals to the left hemisphere, which induces more involvement of the left MFG for phonological and semantic processing. But FL-FR sends the filtered prosodic signals to both hemispheres, which results in less semantic processing in this area. Besides, greater differences in the neural resources involved for semantic integration were found in L-FR – FL-R and NL-NR – FL-R, but the differences in the activated brain regions between the signals were not significant. The results indicate that L-FR and NL-NR require more neuronal resources for semantic integration and expectancy processes as compared to the processing of FL-R although the brain activation patterns between L-FR and FL-R, and between NL-NR and FL-R are not significantly different.

For the central electrode site in the 500-900 ms time window, a great difference in the neuronal resources involved for syntactic processing was found between FL-FR and FL-R. On the contrary, FL-R – FL-FR induced a huge discrepancy in the neural resources involved for later stimulus semantics and stimulus-general processes. The results of the fMRI group analysis also showed that more robust activations in the left MFG were detected during FL-R processing as compared to FL-FR. This result indicates that the L2 FL-FR signal, as a both-ear-filtered prosodic signal in L2, lacks high frequencies to identify the words and structures, which imposes more cognitive load for syntactic and semantic processing as compared to L2 FL-R. Further, L2 FL-R sends the unfiltered signals to the left hemisphere, which helps phonological and semantic processing in the left MFG (Tan et al., 2000; C.-Y. Wu et al., 2012). As a result, L2 FL-R induces more involvement of the neural resources for stimulus semantics and stimulus-general processes (such as working memory, attention, and decision making) as compared to L2 FL-FR. Another great difference was observed between L2 FL-FR and NL-NR, indicating that the processing of FL-FR involved more neural resources for syntactic processing when compared to NL-NR. In the meantime, the fMRI group analysis demonstrated that more regions of the bilateral IPL and the right MFG were recruited for L2 FL-FR as compared to L2 NL-NR. Due to the lack of syntactic and semantic information, the prosodic signal L2 FL-FR imposes more cognitive load for signal processing, which recruits more brain areas for phonological working memory (e.g., Abutalebi et al., 2013; Baddeley, 2003; Della Rosa et al., 2013; Li et al., 2014; Mechelli et al., 2004; J. Wu et al., 2012; Yang et al., 2015) and sentence-level prosodic processing (Gandour et al., 2007; Gandour et al., 2004) relative to the processing of L2 NL-NR. Conversely, the brain areas of the bilateral PrCG, PoCG and the right MFG were more involved in the processing of L2 NL-NR when compared to L2 FL-FR in the fMRI group analysis. The result suggests that the processing of L2 NL-NR, an unfiltered L2 signal, recruits more brain regions such as the left PrCG for phoneme, lexical, and prosody perception (Liang & Du, 2018), the right PrCG for the cognitive control of language switching (Luk et al., 2012), the bilateral PoCG for sensorimotor processing and cognitive control (de Bruin et al., 2014; Guo et al., 2011; Marangolo et al., 2009; Parkinson et al., 2012), and the right MFG for sentence-level linguistic prosody processing (Gandour et al., 2007; Gandour et al., 2004). Compared to the prosodic signal L2 FL-FR without adequate semantic and syntactic information, the processing of the unfiltered NL-NR signals in L2 involves not only the linguistic processing of acoustic, phoneme, lexical, and prosodic perception but a more complex network including high-level cognitive control of language switching and sensorimotor processing.

Regarding the parietal site in the 500-900 ms time window, the MMN analysis revealed the great difference in the neural resources involved for syntactic processing between FL-R and FL-FR. The finding in the fMRI group analysis supported the MMN result that more recruitment of the left MFG was detected for FL-R processing relative to FL-FR processing. Compared to the prosodic signal FL-FR, FL-R sends the unfiltered and filtered signals to the left and right hemispheres respectively, which requires more neuronal resources for syntactic processing, meanwhile, the left MFG is more involved for working memory, phonological, and semantic processing (Siok et al., 2008; Tan et al., 2005; Tan et al., 2000; C.-Y. Wu et al., 2012). In addition, the huge difference in the neural resources involved for the later stimulus semantics and stimulus-general processes were found in the comparison of FL-FR and NL-NR. The fMRI group analysis identified this result that the processing of FL-FR induced more recruitment of the bilateral IPL and the right MFG as compared to NL-NR. The L2 FL-FR is a both-ear-filtered prosodic signal, which is novel and unusual to the participants. As a result, it requires more non-specific cognitive processes such as attention, working memory, and maintaining task demands (Cummings et al., 2006; Shahin et al., 2006) as compared to the processing of the unfiltered NL-NR signal. Further, FL-FR processing induces more involvement of the bilateral IPL and the right MFG for phonological working memory (e.g., Abutalebi et al., 2013; Baddeley, 2003; Della Rosa et al., 2013; Li et al., 2014; Mechelli et al., 2004; J. Wu et al., 2012; Yang et al., 2015) and sentence-level prosodic processing (Gandour et al., 2007; Gandour et al., 2004).

5.1.3 Comparisons between the L1 and L2 signals

From the fMRI group analysis, only the contrast of L2 NL-NR vs. L1 NL-NR showed significant differences in the bilateral PrCG and the left PoCG regions, but other comparisons between the brain regions involved in processing the L1 and L2 stimuli did not demonstrate differences. On one hand, the L2 signals in Verbotonal-based dichotic and diotic listening conditions are processed through the neural structures underlying the processing of the L1 signals. On the other hand, L2 processing, especially the unfiltered L2 NL-NR, requires more areas involved for linguistic processing and high-level cognitive processes, including phoneme, lexical, and prosody perception in the left PrCG (Liang & Du, 2018), cognitive control of language switching in the right PrCG (Luk et al., 2012), and sensorimotor processing and cognitive control in the bilateral PoCG (de Bruin et al., 2014; Guo et al., 2011; Marangolo et al., 2009; Parkinson et al., 2012) as compared to L1 NL-NR processing. The results confirm the previous findings that L2 is processed and acquired through the same neural mechanism and structures that are responsible for L1 processing and acquisition (Abutalebi, 2008). Further, more controlled processing is found in less proficient L2 users (i.e., in a weak L2 system), indicating competition and conflict between languages (Abutalebi, 2008; Abutalebi & Green, 2007). As the speakers become more “native-like” proficient in L2, this difference in cognitive control involved in L1 and L2 processing will disappear and L2 processing will be more automatic as L1 processing (Abutalebi, 2008). Interestingly, among the four configurations of the L1 and L2 stimuli (i.e., FL-FR, FL-R, L-FR, and NL-NR), the difference in the activated brain regions is only observed in the contrast between L2 NL-NR and L1 NL-NR. It means that only the processing of the both-ear-unfiltered L2 signals shows an L2 effect involving more controlled processing resources for language switching (Abutalebi, 2008; de Bruin et al., 2014; Ma et al., 2014) when compared to the both-ear-unfiltered L1 signals. However, the other three configurations of the L1 and L2 stimuli (i.e., FL-FR, FL-R, and L-FR) are processed through similar neural structures or brain regions that are not statistically different. It indicates that the L2 signals with low-pass filtered speech sounds in dichotic and diotic listening conditions could reduce the L2 effect, which means the processing of L2 FL-FR, FL-R, and L-FR would recruit similar brain areas as in L1 rather than involving more regions for cognitive control.

Although the brain regions involved in processing the L2 FL-FR, FL-R, and L-FR signals are not significantly different from that of L1, the neuronal resources involved in semantic and syntactic processing show some differences in the ERP MMN analysis. The differences in the neural resources involved in semantic processing were great between L2 L-FR and L1 L-FR at both the central and parietal electrode sites in the time window of 300-500 ms. This result suggests when compared to L1 L-FR processing, more neuronal resources are involved for semantic processing / lexical access in L2 L-FR even though the brain regions recruited for processing the two signals are similar. In addition, the greater differences in the neural resources involved for semantic integration and expectancy processes were observed between L1 NL-NR and L2 NL-NR. Processing the both-ear-unfiltered L1 signals involves more neural resources for semantic integration of sequential representations and the expectancy process as compared to L2 NL-NR processing. Due to L1 processing, more manipulations of integrating the words of the sentences into the context (i.e., sequential representations) are observed (Dien et al., 2010). Moreover, the expectancy processes for sequential probabilities are more involved in the processing of the L1 NL-NR relative to L2 NL-NR processing (Dien et al., 2010).

At both the central and parietal electrodes in the 500-900 ms time window, the greater differences in the neural resources involved for syntactic processing and cognitive processes were found between the L1 and L2 FL-FR signals. Specifically, compared to L2 FL-FR, processing L1 FL-FR requires more neuronal resources for syntactic processing. As the interviewees mentioned in the semi-structured interviews, L1 FL-FR still could be recognized as Chinese sentences but L2 FL-FR could not affirmatively be identified as English sentences. As a result, more neural resources are involved in L1 FL-FR syntactic processing. On the contrary, more neural resources are involved for the non-specific cognitive processes such as working memory, decision-making, maintaining task demands, and attention (Cummings et al., 2006; Shahin et al., 2006) in L2 FL-FR processing when compared to L1 FL-FR, even though the brain regions recruited for processing the two signals are not significantly different.

5.2 Students' opinions on the L1 and L2 signals in Verbotonal-based dichotic and diotic listening conditions

5.2.1 Auditory perceptions of the signals

Inter-individual variability in auditory perceptions of language signals always exists, which may be due to the ways that we perceive speech are constrained and constructed by the operational histories (Lian, 2004; Lian & Pineda, 2014), including individuals' various language experiences (e.g., Gandour et al., 2004; Lotto & Holt, 2016), cognitive and metacognitive abilities (e.g., Vandergrift, 2003a; Vandergrift, 2003b, 2007), neurophysiological bases for auditory processing (e.g., Pernet et al., 2015), and so on. Thus, different auditory perceptions of the L1 and L2 signals in Verbotonal-based dichotic and diotic listening conditions were found during the semi-structured interviews.

For the dichotic signals FL-R and L-FR in L1, most of the interviewees reported that the humming sounds were heard in the left ear for FL-R but the humming sounds came from the right ear for the L-FR signal. A few students could not feel the difference between the two dichotic signals and even felt FL-R and the unfiltered signal NL-NR were similar to each other. Interestingly, the L2 dichotic signals were perceived differently. Some respondents described FL-R as filtered in the left ear but others said it was filtered in the right ear, so was L-FR. As Pressnitzer et al. (2018) explain, listeners perceptually emphasize or focus on different parts of the frequency range. Thus, inter-individual differences in auditory perception of speech could demonstrate idiosyncratic frequency weighting that might be a factor that influences covert attention (Pressnitzer et al., 2018). In addition, listeners' covert attention is initially attracted by the sounds in low frequencies and may remain there due to spectral continuity or auditory binding (Chambers et al., 2017; Moore & Gockel, 2012). The current study diotically and dichotically sends the low-pass filtered sentence signals with frequencies under 320 Hz and the unfiltered signals to the left and right ears, which may lead to a greater frequency weight of the auditory cues for the participants. As a result, a possible explanation of different auditory perceptions of the dichotic signals (especially in L2) is that participants' attention is influenced by the frequency weight of the filtered and unfiltered speech sounds in the dichotic listening conditions. Varying perception and sensitivity to the signals with different frequencies in both ears may affect or shift students' attention while listening, thus making their descriptions of auditory perception different.

Regarding the both-ear-filtered signals, L1 FL-FR could still be recognized as Chinese sentences that required more effort to process/understand, but L2 FL-FR sounded vaguer than L1 FL-FR so that it could not affirmatively be identified as English sentences. One of the differences between L1 and L2 processing is how attentional resources are allocated to what linguistic features (Mora & Mora-Plaza, 2019). Unlike highly-automatic attention control in L1 processing, the lower proficiency level of L2 may cause perceptual difficulties in phonologically encoding the L2 signals by using the L2-specific cue-weighting (Bohn, 1995; Flege, 1995; Mora & Mora-Plaza, 2019). Thus, perceptual sensitivity to L1 allows the participants to recognize the L1 FL-FR signals as sentences in Chinese even though they cannot hear the words clearly.

Another issue that the respondents mentioned a lot regarding auditory perception is signal clarity. Due to the physical quality of the low-pass filtered stimuli, filtering removed the frequencies above 320 Hz that could help identify the words. Further, the low-pass filtered and unfiltered signals were diotically and dichotically sent to the participants. Compared to the unfiltered signals in both ears or in one ear, the filtered signals were not clear enough to identify the words in the sentences. The interviewees admitted that they could not hear the filtered signals clearly in both L1 and L2, which affected their understanding. This finding supports the results of Eisenberg et al. (1998) that the stimulus bandwidth influences listeners' perceived clarity in the subjective evaluation since speech information is largely encoded in frequencies above 500 Hz (Griesinger, 2013). As a result, the participants of the current study perceived the 320 Hz low-pass filtered signals as unclear signals when compared to the full-bandwidth signals (i.e., the unfiltered signals). But perceived clarity is the internal reference for clearness of the auditory signals, which is different from stimulus intelligibility (Eisenberg et al., 1998). Understanding of the stimulus meanings and structures will be discussed in the following section 5.2.3.

As to the perception of discomfort, the students reported that FL-FR in both L1 and L2 made them feel uncomfortable and unpleasant while listening because they were not clear. It is the first time for the students to listen to speech signals of this kind. In addition, due to the lack of the high frequencies in FL-FR, speech information except for the prosodic features is not audible. Thus, it is reasonable that the participants have to struggle with the processing of these unfamiliar and unusual signals. This would be the reason that they felt uncomfortable while listening to L1 and L2 FL-FR. On the contrary, the full-bandwidth signal NL-NR sounds clearer. It is what the students hear in daily speech/communication, so they get used to it.

Therefore, the NL-NR signal is reported as a comfortable signal to listen to. However, the dichotic signal L-FR was perceived as an uncomfortable signal. A possible explanation is that L-FR sends the filtered signals to the left hemisphere and the unfiltered signals to the right hemisphere, which violates hemispheric specialization for linguistic processing in the left hemisphere and prosodic perception in the right hemisphere (Gandour et al., 2004; Meyer et al., 2002; Sammler et al., 2015; Tervaniemi & Hugdahl, 2003; Vigneau et al., 2006). Both of the hemispheres are less specialized in processing the signals that L-FR dichotically sends, thus making both hemispheres struggle. This might be a reason that the interviewees reported the feeling of discomfort. Interestingly, the other dichotic signal FL-R was thought to be a comfortable signal besides the unfiltered signal NL-NR. It could also be explained by hemispheric specialization that both hemispheres received the signals they were more specialized in processing. As a result, it is less effortful for both hemispheres to process, indicating a reason that FL-R is perceived as a comfortable signal.

5.2.2 Preference and interest in the signals

Although the both-ear-filtered signal FL-FR was unusual to the participants, some students expressed their interest and preference for it and others had a neutral opinion. This result is consistent with a study by Luu (2021), which uses the 320 Hz low-pass filtered signals in a self-regulated online system to improve Vietnamese EFL students' English listening ability. The majority of participants in Luu's study also showed preference or a neutral position for the filtered signals when compared to the traditional method using the unfiltered signals as listening materials. It is observable that most of the students are interested in or willing to accept the "novel" speech sounds during the language-learning process.

The unfiltered signals NL-NR were favored by most of the students. The respondents gave the reasons that speech information of NL-NR could be heard clearly, and it was the sound that they heard in daily speech. For the dichotic signals, most participants expressed a preference for FL-R, or no obvious preference was reported. It was hard for the participants to explain the reasons that they preferred FL-R, but some mentioned that FL-R was clearer compared to other stimuli. L-FR was not favored by some of the interviewees since the feeling of discomfort as discussed above.

Only one student showed a preference for L-FR without any reason. As discussed before, auditory perception is different from individual to individual due to the listeners' various operational histories (Lian, 2004; Lian & Pineda, 2014). Participants' subjective opinions of preferences and dislikes help us understand how they really feel during listening to the signals in Verbotonal-based dichotic and diotic listening conditions.

5.2.3 Understanding of the sentence meanings and structures

For the understanding of the signals, the participants expressed the differences and difficulties in understanding the meanings and structures of the sentences. The both-ear-filtered signals were difficult for the participants to understand since the frequencies above 320 Hz that contained acoustic information to identify words were removed. The participants could not understand the meanings of the signals let alone the structures. The qualitative results are supported by the ERP and fMRI results. In the ERP MMN analysis of L1, greater differences were found between FL-FR and FL-R in the neural resources involved in syntactic processing, between FL-FR and L-FR in semantic integration, and between FL-FR and NL-NR in semantic and syntactic manipulations. Consistently, the fMRI group analysis revealed the differences of the activated brain regions in the left PrCG and PoCG between FL-R and FL-FR, the left PrCG, IPL, and SPL between L-FR and FL-FR, and the right MFG and PrCG between FL-FR and NL-NR. Compared to the processing of the other three stimuli, FL-FR processing leads to differences in the neuronal resources involved for semantic and syntactic manipulations, and brain regions recruited for processing. As to L2 processing, huge differences were observed between FL-FR and FL-R in the neural resources involved in syntactic processing, and between FL-FR and NL-NR in semantic and syntactic manipulations. The recruited brain regions were also different between FL-R and FL-FR in the left MFG, and between FL-FR and NL-NR in the right IPL, MFG, and the left IPL. Interestingly, L2 FL-FR and L2 L-FR did not show great differences in the neural processing of semantic and syntactic information, and the brain regions recruited for processing the two signals were not significantly different. This finding may answer why a few participants state that L2 L-FR is as difficult as L2 FL-FR in understanding the meanings and structures.

On the contrary, the participants reported that the unfiltered signals NL-NR in L1 and L2 were semantically and syntactically easier to comprehend because of the perceived clarity. But the difficulty they faced was that the lower proficiency of L2 led to more processing efforts as compared to L1 processing. This finding is consistent with the fMRI results that L2 NL-NR recruits more regions of the bilateral PrCG and the left PoCG for linguistic processing and high-level cognitive processes (de Bruin et al., 2014; Guo et al., 2011; Liang & Du, 2018; Luk et al., 2012; Marangolo et al., 2009; Parkinson et al., 2012) as compared to the processing of L1 NL-NR. But the L1 processing pattern is more automatic (Abutalebi, 2008), which requires a more intensive recruitment of the brain areas and a fewer number of the activated regions (Vingerhoets et al., 2003).

Views on the understanding of the dichotic signals FL-R and L-FR varied. Some held that FL-R helped comprehend sentence meanings and structures, but a few claimed that L-FR did. The reasons that it is easy to understand the FL-R signal may be hemispheric specialization for linguistic and prosodic processing in the left and right hemispheres (Gandour et al., 2004; Meyer et al., 2002; Sammler et al., 2015; Tervaniemi & Hugdahl, 2003; Vigneau et al., 2006). Specifically, FL-R sends the filtered signal (i.e., prosody) to the right hemisphere and the unfiltered signal to the left hemisphere so that both hemispheres receive the signals that they are more specialized in processing. In addition, the ERP individual analysis suggested that FL-R in both L1 and L2 induced a lower processing load in semantic and syntactic manipulations as compared to other stimuli, especially the L-FR signal. But for a few participants, it was easy to understand semantic and syntactic information of L1 L-FR. It is reasonable that the perceived difficulty is subjective and varies from person to person, which reminds us that listening as a cognitive activity is personalized and individual due to diverse operational histories (Lian, 2004; Lian & Pineda, 2014).

We have to admit that the current study does not adopt a comprehension task during the experiment, thus, it is just students' subjective evaluation of their understanding of the stimuli rather than an objective assessment of comprehension. But it provides us students' perceptions of their comprehension and their views on the signals.

5.2.4 The signals for language learning

NL-NR contains all frequencies so that speech information can be heard clearly. Therefore, it is not surprising that most of the students suggest using the unfiltered signals NL-NR when learning a language. The participants mentioned that NL-NR was the signal they usually hear in daily speech/communication. It is understandable that the students prefer the most familiar signal, in addition, this familiar signal is perceived as the clearest signal. However, the most familiar signal does not mean it is the optimal one for the brain to process. As discussed above, the processing load or the neuronal resources involved in semantic and syntactic processing, and the brain regions recruited for processing differed in the different signals. Indeed, the quantitative results reveal that the auditory signals could be optimized and there could exist the optimal and the non-optimal signal regarding the neural processing. The optimized auditory signal could be applied in learning English for Chinese students, which will be discussed in the next section.

Noteworthily, a lot of participants expressed an interest and willingness to listen to the signals in Verbotonal-based dichotic and diotic listening conditions when learning English. They even stated that they were willing to learn English with all the four configurations of the L2 signals presented in the order of FL-FR, FL-R, L-FR, and NL-NR as in the experiment. Although the students could not explicitly explain how the signals would influence their learning and why they preferred using the signals in learning, most of them expressed a positive attitude toward the signals to English learning. Importantly, the students were interested in these signals.

5.3 The optimal and non-optimal signals in L1 and L2 in Verbotonal-based dichotic and diotic listening conditions

According to cognitive load theory (as reviewed in Chapter 2), the first concern of the instructional design is to optimize students' level of cognitive load in the learning process, as a result, to avoid cognitive overload that hinders students' learning performance and outcomes (Antonenko et al., 2014; de Jong et al., 2009; Sweller et al., 2011). Students would benefit from the optimized cognitive load in different learning contexts because they could maintain an optimal level in mental processing load, which facilitates knowledge acquisition by transforming information from working memory to long-term memory effectively (Antonenko et al., 2014; Sweller et al., 2011). Previous studies attempted to optimize EFL students' cognitive load in listening by means of changing the listening modalities in multimedia environments, however, the results were controversial that the changes of listening modalities might impose additional cognitive load for the students rather than lowering the load (Aldera & Mohsen, 2013; Diao et al., 2007; Mohsen, 2016). Thus, the current study endeavors to optimize Chinese EFL students' cognitive load in listening to L1 and L2 sentence signals by changing the physical quality of the auditory signals on the principles of verbotonalism (Guberina & Asp, 1981, 2013) and hemispheric specialization for linguistic and prosodic perception (Gandour et al., 2004; Meyer et al., 2002; Sammler et al., 2015; Tervaniemi & Hugdahl, 2003; Vigneau et al., 2006). This section discusses the optimal and non-optimal auditory language signals in Verbotonal-based dichotic and diotic listening conditions based on the ERP findings of the cognitive load / mental workload in semantic and syntactic processing, the fMRI results of the brain regions recruited for processing, and students' opinions on the signals.

5.3.1 The optimal L1 auditory signal

For the L1 signals, FL-R seems to be an optimal auditory language signal based on the reduction of cognitive load for semantic and syntactic (to some extent) processing and students' auditory perceptions. Compared to the other three stimuli in L1 (i.e., L1 FL-FR, L-FR, and NL-NR), the L1 FL-R signal elicited a lower processing load for semantic manipulation. Specifically, FL-R induced the lowest mental workload for semantic integration and sequential expectancy processes with a shorter reaction time at the central site and a longer reaction time at the parietal areas though (Dien, 2009; Dien et al., 2010; Liu et al., 2009). With regard to syntactic processing, FL-R elicited the lowest load with a short reaction time for the late sentence structure-specific processes at the parietal site (Schacht et al., 2014) among the four stimuli. However, the FL-R resulted in the heaviest load with a short reaction time for structural processing at the central site. Thus, it could be estimated that FL-R lowers the mental workload for semantic manipulation, including the processes for semantic integration and sequential expectancy. For the syntax-related processes, studies reported that a more parietal P600 was observed in native speakers while a more central P600 was found in L2 speakers (Nickels et al., 2013; White et al., 2012). The current study supports this finding that L1 NL-NR indeed results in a larger P600 effect at the parietal site than that at the central site. Further, the parietal P600 may be related to the late sentence structure-specific processes, which are more task-sensitive (Schacht et al., 2014). It is clear that FL-R lowers the load for the late structure-specific processes regarding syntactic manipulation compared to the other three signals. But it has to be acknowledged that FL-R imposes a heavy load for structural processing at the central areas. A possible explanation could be that the dichotic signal FL-R is somewhat unusual to the listeners, which requires allocating more resources for the earlier process of structure analyzing. The delay of shifting from the earlier structural processing to the late structure-specific processes indicates the unfamiliarity of the signal, which is similar to the description of the delay in L2 learners (White et al., 2012). But more studies are required to further elucidate this speculation. Here, it could be indicated that FL-R lowers the mental load for the late structure-specific processing rather than the earlier structural processing.

From the group analysis, both of the ERP MMN and fMRI results revealed the difference between FL-R and FL-FR. As a language signal that is left-ear-filtered and right-ear-unfiltered, FL-R involves more neural resources for stimulus semantics and stimulus-general processes in a complex sensory-motor network as compared to the both-ear-filtered prosodic signal FL-FR. Compared to L-FR, the MMN results indicate that processing FL-R involves more neuronal resources for syntactic processing, but L-FR requires more mental workload for the later processing of stimulus semantics and cognitive processes such as working memory, maintain task demands, decision-making, and attention (Cummings et al., 2006; Shahin et al., 2006). Apart from the MMN difference, the brain regions recruited for FL-R and L-FR processing are not significantly different. As to the comparison between FL-R and NL-NR, the MMN results showed the smallest difference of the neuronal resource involved in semantic processing, and the brain regions recruited for processing the two were not significantly different. This result suggests that the neural processing patterns for the FL-R and NL-NR signals are quite similar, but, as discussed above, FL-R lowers the mental workload for semantic processing and the later structure-specific processing as compared to NL-NR.

As for students' opinions of the FL-R signal, positive or at least neutral opinions were reported. Except for NL-NR, some students expressed their preferences for FL-R. Others showed no obvious preference, holding a neutral opinion. As an unfiltered signal, NL-NR is more familiar to the students. Thus, it is reasonable that students would prefer NL-NR over the other three signals, but, as mentioned above, it does not mean that NL-NR is the optimal signal for the brain to process. The dichotic signal FL-R is thought to be as natural as or sounds similar to NL-NR. It suggests that FL-R, at least, does not arouse the feeling of novelty. Compared to L-FR, FL-R was considered as a clearer signal. More importantly, no one reported the feeling of discomfort while listening to the FL-R signal.

Taken together, the L1 FL-R signal, in line with hemispheric specialization, could be regarded as an optimal L1 auditory signal for processing.

5.3.2 The optimal L2 auditory signal

The current findings identify that the L2 FL-R signal could be an optimal L2 auditory signal for the students to process. From the individual analysis of the ERP components, FL-R induced a lower mental workload for semantic and syntactic processing relative to the other stimuli. To be specific, FL-R resulted in the early process for the established meaningful semantic representations at the central and parietal areas (Cummings et al., 2006). At the central site, the processing load induced by FL-R for semantic manipulation was lower than L-FR and NL-NR, further, the load elicited by FL-R for lexical access at the parietal site (Dien, 2009; Dien et al., 2010; Liu et al., 2009) was the lightest among the four. In addition, FL-R elicited the later processing for stimulus semantics and stimulus-general processes (Cummings et al., 2006; Itoh et al., 2005; King & Kutas, 1995; Koelsch et al., 2003) with the lightest mental load and shortest reaction time of all. Though FL-R induced a heavier load and a longer reaction time for syntactic manipulation than that of NL-NR at the parietal areas, the load and the reaction time elicited by FL-R was lower and shorter than that of L-FR. As a result, FL-R could lower the mental load for the early semantic manipulation and the later processes for stimulus semantics and cognitive processes, in the meantime, FL-R induces the lower load for syntactic processing as compared to L-FR.

The ERP MMN and fMRI group analyses revealed only when compared to FL-FR, FL-R demonstrated differences in the neural resources involved for the early, the later processes of semantic manipulation, and syntactic processing in the brain region of the left MFG for working memory, phonological, and semantic processing (Siok et al., 2008; Tan et al., 2005; Tan et al., 2000; C.-Y. Wu et al., 2012). But as FL-R was compared to L-FR and NL-NR, no such differences in the neuronal resources involved for semantic or syntactic manipulation, and brain regions were found. It indicates that FL-R does not involve additional mental workload or brain areas for semantic or syntactic processing as compared to L-FR and NL-NR.

Similar to L1, students' opinions of L2 FL-R are positive or neutral. Except for the L2 NL-NR signal, some students showed preferences for L2 FL-R or no obvious preference. L2 NL-NR is a familiar audio signal without any "novel" sounds for the students, which could explain why most students tend to prefer it. Noteworthily, several interviewees expressed that FL-R helped understand the sentences and they enjoyed listening to FL-R. Also, no one reported the feeling of discomfort while listening to L2 FL-R. In terms of learning English with the signal, the students showed an interest and willingness to use L2 FL-R, even with the other three signals.

Therefore, the L2 FL-R signal, in line with hemispheric specialization, could be taken as an optimal L2 auditory signal for Chinese EFL students to process and learn with.

5.3.3 The non-optimal auditory signal

There might also exist the non-optimal signals in Verbotonal-based dichotic and diotic listening conditions, which imposes additional cognitive load for semantic and syntactic processing.

Among the L1 signals, it is not clear which signal is non-optimal for processing. According to hemispheric specialization, L-FR might be thought of as a non-optimal signal. From the ERP individual analysis, the L-FR signal only induced the highest processing load with a longer reaction time for the early semantic manipulation at the central areas. However, L-FR led to a lower mental load for lexical access and syntactic processing at the parietal areas as compared to NL-NR. As for the group analysis, only the contrast of L-FR and FL-FR demonstrated great differences in both ERP MMN and fMRI results. Specifically, L1 L-FR involved more neural resources for semantic manipulation and recruited more areas of the left PrCG, IPL, and SPL for lexical/semantic processing when compared to L1 FL-FR. The great difference between L-FR and FL-R indicates that more involvement of the neural resources for the later processing of L-FR stimulus semantics and cognitive processes (such as working memory, maintain task demands, decision-making, and attention) (Cummings et al., 2006; Shahin et al., 2006) as compared FL-R. But the brain areas recruited for processing L-FR and FL-R are not significantly different. In terms of students' perceptions of the L1 L-FR, opinions varied that some showed feelings of dislike or discomfort, but a few expressed preferences and reported that L1 L-FR helped understand the L1 sentences. So far, it is unclear that L1 L-FR is the non-optimal L1 auditory signal for processing.

As for the L2 signals, L-FR seems to be the non-optimal auditory signal to process. L2 L-FR induced the heaviest processing load for the processes of lexical access, semantic integration, and sequential expectancy at the central and parietal areas (Dien, 2009; Dien et al., 2010; Liu et al., 2009). Further, the heaviest load was elicited by the L2 L-FR for syntactic processing at the parietal areas among the four signals. In the group analysis, the MMN differences showed that L-FR involved more neural resources for semantic integration and sequential expectancy processes as compared to FL-R and NL-NR respectively, though the brain regions did not exhibit significant differences in processing them. As mentioned above, students' perceptions of L-FR varied. But it is worth pointing out that L-FR made some participants feel uncomfortable while listening. Even though a few interviewees showed preferences for L-FR, some students expressed dislike for it. The statement of dislike was referred to L-FR and FL-FR during the interviews. As a result, the L2 L-FR signal, imposing more mental processing load for semantic and syntactic manipulations, may violate hemispheric specialization, thus making it the non-optimal L2 auditory signal for Chinese EFL students to process.

The reason that L-FR might be non-optimal for L2 processing but not for L1 could be the proficiency in the native language. The neural mechanisms for L1 processing are more automatic (e.g., Friederici et al., 2003) while the processing for L2 learners or less proficient users may be less automatized (e.g., Brauer & Friederici, 2007; Rüschemeyer et al., 2005). Thus, no matter the speech sounds of L1 are filtered in either the left or right ear, the signals would be processed in the highly-automatic neural mechanisms. However, due to the lower level of L2 proficiency, whether the signals are left-ear- or right-ear-filtered matters, which would lead to the changes of cognitive load for semantic and syntactic processing while listening to the L2 signals.

5.4 Summary

To sum, this chapter discusses the results of the individual and group analyses in the combined ERP and fMRI experiment, and students' perceptions of the signals from the semi-structured interviews. Then, the optimal L1 and L2 auditory signals, as well as the non-optimal signal in L2 are discussed.

CHAPTER 6

IMPLICATIONS, RECOMMENDATIONS, AND CONCLUSION

In this concluding chapter, a summary of the findings and contributions is made in the first section. Then, implications of the current study are provided in the second section. Strengths and limitations are discussed in the following section. Finally, suggestions for future studies are put forward.

6.1 Summary of the study

The current study explored how the physical quality of the language signals sent to learners either dichotically or diotically influenced brain activity leading to language perception. Based on theories of verbotonalism, hemispheric specialization, and cognitive load, this study created the L1 and L2 signals in Verbotonal-based dichotic and diotic listening conditions, which attempted to provide insight into Chinese EFL university students' neurobiological bases of auditory language perception and to identify an optimal auditory language input. A combined ERP and fMRI experiment was performed to unravel the temporal and spatial neural signatures of the processing of the L1 and L2 auditory signals. Further, students' views on the signals were investigated through semi-structured interviews. Thirty students from Kunming Medical University, China took part in the study.

As the physical features of the auditory language signals changed, the brain could differentiate the signals in different configurations. Specifically, the filtered prosodic and unfiltered speech sounds under various diotic and dichotic listening conditions induced different neural processing patterns and auditory perceptions, and the brain could distinguish these different signals. In addition, the processing of L2 signals, on one hand, recruited similar brain regions employed by L1, indicating the accommodation process (Abutalebi & Green, 2007; Chee et al., 1999); on the other hand, the L2 effect revealed that more areas for high-order cognitive control and sensory-motor coordination were involved in L2 processing relative to L1 processing (de Bruin et al., 2014; Ma et al., 2014; Martin et al., 2015). The current study confirmed that language perception was an active cognitive process determined by both context-driven top-down and stimulus-driven bottom-up processes (Heald et al., 2016). Since the low-frequency stimulation based on VTA could restructure the ways that listeners perceive the auditory language signals (Lian & Sussex, 2018), the optimized auditory language signal with filtered and unfiltered sounds in the dichotic listening condition would be in line with hemispheric specialization and would better raise students' awareness, modify students' auditory perceptions, and facilitate language learning/development and, in addition, induce maximal neuroplasticity in the individual brain (Guberina & Asp, 2013; Kim & Asp, 2002; Lian & Sussex, 2018).

Based on the mental workload, brain regions involved in the processing, and students' views on the signals, the current study identified that FL-R was an optimal auditory signal in both L1 and L2. In L1 processing, FL-R could reduce the mental load for the processes of semantic integration and sequential expectancy. In addition, FL-R lowered the mental load for the late structure-specific processing of the L1 sentences.

As for L2, FL-R could lower the mental load for the early semantic manipulation, and the later processes for stimulus semantics and cognitive processes. In the interviews, most students stated that they preferred listening to the FL-R signal since it was clearer compared to FL-FR and L-FR, and they did not report any feeling of discomfort. Besides, some respondents claimed that FL-R could help understand sentence meanings and structures. Considering hemispheric specialization for linguistic and prosodic perception in the left and right hemispheres respectively (Gandour et al., 2004; Meyer et al., 2002; Sammler et al., 2015; Tervaniemi & Hugdahl, 2003; Vigneau et al., 2006), the L1 and L2 FL-R signals, sending linguistic and prosodic signals dichotically to the left and right hemispheres that were more specialized in processing, appeared best suited for optimal processing. Meanwhile, a non-optimal signal in L2 was identified. L-FR, violating hemispheric specialization, could impose a larger mental processing load for semantic and syntactic manipulations. Moreover, L2 L-FR made some students feel uncomfortable while listening. But it was unclear whether L1 L-FR was non-optimal for students to process.

Compared to L1 processing, a more complex sensory-motor network for cognitive control was required in L2 processing (e.g., Indefrey, 2006; Liu & Cao, 2016), especially when processing the NL-NR signals in the current study. However, the L2 signals with low-pass filtered speech sounds in either dichotic or diotic listening condition could reduce the L2 effect. The current study identified that the processing of L2 FL-FR, FL-R, and L-FR would recruit similar brain regions as in L1 processing instead of involving more areas for cognitive control.

In terms of students' opinions on using the signals for learning English, the respondents showed an interest in all the filtered and unfiltered signals in either dichotic or diotic listening condition, further, they expressed willingness to listen to the signals during the learning process.

6.2 Implications

The outcomes of the current study carry profound implications for educators, policymakers, teachers, students, and researchers in the field of foreign language learning and teaching, perhaps in other fields too. The broader implication for each stakeholder mentioned above is that we need to turn our attention to the physical aspects of language input sent to the students. The current study revealed that changing the physical language signals results in changes in neural processing patterns. It is an example that the physical language signals influence students' brain activity and perception, and may affect learning outcomes in this way or other ways. Thus, we should be more aware of the significance of the physical language signals.

In addition, more experiments and investigations are needed to further understand the impacts of the optimal physical language signals (perhaps in other forms) on brain activity and neural processing. Wen et al. (2020) found that each student's optimal frequency bands for perception, production, and correction of English pronunciation varied, which reminded us that the optimal physical language signals might be in different forms for different individuals. But our understanding is limited currently. Thus, educators, researchers, and teachers should further explore the optimal physical language signals that help enhance the learning of a second language.

6.2.1 For educators and policymakers

The current study is a pioneering attempt to restructure students' language perception and raise their awareness of the language by optimizing physical language input, showing great potential in facilitating language learning. As this study has recognized the importance of the physical language signals in processing and learning a language, educators and policymakers should be aware of the importance and establish policies on what truly benefits students and facilitates learning.

Secondly, educators and policymakers should support new development and research as such. The current study is an interdisciplinary study involving language and neuroscience studies, which yields successful outcomes and has far-reaching influences on each stakeholder in these areas. Thus, policies should be formulated to encourage and support new development, technologies, and research in the field of language teaching/learning from diverse perspectives and subjects.

Thirdly, educators and policymakers should extend the academic discourse in the field of Teaching English to Speakers of Other Languages (TESOL) by allowing and encouraging teachers and researchers to communicate and develop further understanding of interdisciplinary research areas and topics. Such investigations should be encouraged with awareness of the influence of the physical language signals on students' brain responses. More importantly, more voices of such studies should be heard in academia.

6.2.2 For teachers and students

First and foremost, the students' biological basis – the body, especially the brain, should be valued during the process of foreign language teaching/learning. We are embodied so that our brains, bodies, and bodily experiences constrain, regulate, and structure the mind by allowing sensorimotor systems to receive sensory input and produce behavioral output (Foglia & Wilson, 2013). Thus, students do not just think with their minds, but with and through their bodies. Likewise, students do not just learn languages with their minds, but with and through their bodies. It is clear from the current study that different auditory language signals lead the brain to respond differently, which might further result in different learning outcomes. Teachers and students should be more aware of and turn attention to the input signal as changes in physical language signal input induce changes in neural processing patterns.

Both teachers and Chinese EFL students would benefit from the findings of the present study. The signal of L2 FL-R was identified to be an optimal signal that could lower the mental workload for the early semantic manipulation, and the later processes for stimulus semantics and cognitive processes. Thus, teachers and students could use this signal as listening materials when teaching and learning English. Teachers can prepare the materials by employing audio editing software such as Audacity and Audition to low-pass filter the speech sounds in the left ear/channel. Also, students are encouraged to edit the audios on their own.

In addition, the neuroscience of how the brain learns and how the physical language signal influences students' neural processing should be included in teacher education programs. Meanwhile, further investigations and applications of the impact of the physical language signals on the brain are needed in teaching languages.

6.2.3 For researchers

In the field of TESOL, researchers usually neglect the student's "body" while paying more attention to studies on developing students' higher-order cognitive skills such as memorizing words and structures and communicating in oral and written forms. As a result, how the brain processes the incoming signal and how the physical features of auditory signals influence language processing are under-valued and under-researched. Scholars should bear in mind that students' perceptions of the language signals derive from their neurobiological bases and the physical features of the incoming signals. Thus, researchers should turn attention to neuroscience studies, including how the brain processes language signals and how the physical language signals affect language processing and learning.

The current study provided evidence that neural processing patterns are impacted by the incoming auditory signals, and more importantly, the way of changing the physical language signals has enormous potential for language processing and learning. Further, the FL-R signal was identified as an optimal auditory signal that could lower the processing load and could be best suited for the students to process. If the efficacy and effectiveness of the optimal auditory signal FL-R in improving English learning could be further evaluated and confirmed by researchers in different EFL contexts, the implications of the current study are clear that FL-R could be used in the EFL/ESL course design. Therefore, this study makes notable contributions to the principles and approaches in language education. Researchers in EFL should recognize and capitalize on the role of the physical language signals in language learning. Based on the outcomes of the current study and subsequent studies, new teaching methods or approaches could be researched, which would lead to a positive future for language learning and language education enterprise.

Finally, as this study also provided the neural processing patterns for L1, the results could also be applied to Chinese speakers with pathological conditions relating to language. Researchers or clinical practitioners might be able to adopt the L1 optimal physical language signals as materials for speech therapy. This might be a new area of research.

6.3 Strengths and limitations of the study

The current study employed a mixed methods research design to investigate students' brain responses and views of the low-pass filtered and unfiltered language signals in the dichotic and diotic listening conditions. Quantitative data derived from the ERP and fMRI experiments and qualitative data from the semi-structured interviews corroborated and validated the results of the present study. Quantitative results revealed the temporal and spatial patterns of neural activity while processing both L1 and L2 auditory signals. The qualitative method assisted us in better understanding students' perceptions, preferences, interests, and views after listening to the signals. As a result, this study yielded encouraging outcomes related to the optimal and non-optimal auditory signals for L2 perception and processing.

Despite the promising results, there are some limitations. The first weakness of the current study, maybe for most cognitive neuroscience studies, is that only averaged data is reported rather than individual results. As data is pooled, we will lose data in many cases (Uttal, 2011).

Secondly, the current study like most of the cognitive neuroscience studies employed the techniques of EEG and brain images, and analyzed the data at the gross anatomical level, i.e., the macroscopic level. But the microscopic level, manifesting the neuronal networks/interactions, may explain how the mind is instantiated, which is germane to cognition. Unfortunately, we currently do not have the means to detect microscopic networks. Due to the complexity and numerousness of the neurons and neuronal networks, it is intractable to analyze and explain the bottom-up and top-down cognitive processes at the microscopic level (Uttal, 2011).

Thirdly, the current study strictly controlled the variables that could influence the neural processing of language such as handedness. We have to acknowledge that only strongly right-handed students are recruited as participants in the current study, which, to some extent, is not a fully randomized trial. Thus, how left-handed students process the auditory language signals is still unknown. When applying the current results more broadly, we need to be more careful, especially with left-handed students, because the neural processing patterns for language in left-handed students are not investigated in this study.

Last but not the least, the sample size of the current study was small, and the participants were Chinese university students with an intermediate English proficiency level. Thus, the results should be generalized with great care to the population of Chinese EFL learners. In addition, since the participants' native language – Chinese is a tonal language, it may affect the generalization of the outcomes to EFL students with different language backgrounds.

6.4 Suggestions for future research

Future studies can involve more participants with different English proficiency levels at different ages to fully recognize the neural processing patterns for the signals in Verbotonal-based dichotic and diotic listening conditions, further, to validate the current results of the optimal and non-optimal signals for Chinese EFL students. In addition, future studies can recruit non-tonal language speakers as participants to explore the difference or similarity of the neural processing of the language signals between the tonal and non-tonal language speakers.

The current study investigated the temporal and spatial patterns of neural activity as processing the auditory signals, but the connectivity between brain regions was still unknown. Future research can use a psycho-physiological interaction (PPI) analysis to estimate an interaction between the given psychological context (the task) and the physiological state (brain activity in a seed ROI), indicating the task-specific changes between activity in different cerebral regions (Friston, 2011; Friston et al., 1997; O'Reilly et al., 2012).

As mentioned above, the present study focused on the neural processing of the signals in Verbotonal-based dichotic and diotic listening conditions, but it did not touch upon comprehension of the auditory signals. Future research can address this issue by performing a listening comprehension task to further discuss how listening comprehension is influenced by changing the physical features of the auditory signals, in addition, to figure out the relationship between the neural processing patterns and the listening performance.

In order to make a broader generalization, future studies could conduct a fully randomized trial by involving both right- and left-handed students as participants. As the neural processing patterns for language in left-handed subjects are not fully understood, more research is needed to determine their "language site." Further, studies could expand the project to identify the optimal auditory language signal for left-handed EFL students, which would be worth investigating and open new possibilities for personalized learning.

Individual differences in brain activity during processing the signals should also be taken into consideration in the future. The neural processing patterns reported in this study are derived from averaged data, thus, individual differences are ignored. As each individual uses different cognitive processes and strategies involving distinct brain circuits, future studies could adopt data from resting-state scanning to evaluate individual brain connectivity and networks under the task-free condition, and relate the individual brain networks to individual task-evoked responses (Tavor et al., 2016).

In this way, individual differences in brain activity and individual task activation patterns for processing the signals could be unraveled.

To conclude, the current study has provided insight into Chinese students' neural processing of auditory language signals. As a primary connector of the "mind-body" problem, the auditory input signals play a vital role in raising students' awareness and restructuring their perceptions of the foreign language signals during the learning process. This study created the L1 and L2 signals in Verbotonal-based dichotic and diotic listening conditions to identify an optimal auditory language signal based on theories of verbotonalism, hemispheric specialization, and cognitive load. By implementing a combined ERP and fMRI experiment and the semi-structured interviews, the temporal and spatial patterns of neural activity while listening to the L1 and L2 signals were detected, and students' views after listening were investigated. Findings revealed the optimal auditory signal without imposing additional mental workload and the non-optimal signal imposing more load for processing. The present study has addressed the value of recognizing the neurobiological bases for language processing and the impact of the physical language signals on neural processing. Ideally, students would receive the optimal input and avoid the non-optimal signal, which could get their brains actively and efficiently involved in the language processing and learning processes. All in all, this study represents only a small step in the iterative process of progressive refinement of research into the nature of the physical signal as it affects L2 learning. Once general processing patterns have been established, future studies will delve into the details of how else to manipulate the physical language signal so as to enhance the L2 learning process.

REFERENCES

Abutalebi, J. (2008). Neural aspects of second language representation and language control. *Acta Psychologica, 128*(3), 466-478. https://doi.org/10.1016/j.actpsy.2008.03.014

Abutalebi, J., Annoni, J.-M., Zimine, I., Pegna, A. J., Seghier, M. L., Lee-Jahnke, H., Lazeyras, F., Cappa, S. F., & Khateb, A. (2007). Language Control and Lexical Competition in Bilinguals: An Event-Related fMRI Study. *Cerebral Cortex, 18*(7), 1496-1505. https://doi.org/10.1093/cercor/bhm182

Abutalebi, J., Della Rosa, P. A., Ding, G., Weekes, B., Costa, A., & Green, D. W. (2013). Language proficiency modulates the engagement of cognitive control areas in multilinguals. *Cortex, 49*(3), 905-911. https://doi.org/10.1016/j.cortex.2012.08.018

Abutalebi, J., & Green, D. (2007). Bilingual language production: The neurocognition of language representation and control. *Journal of Neurolinguistics, 20*(3), 242-275. https://doi.org/10.1016/j.jneuroling.2006.10.003

Alazard, C., Astesano, C., & BilliÈres, M. (2011). Multiphonia: A multimodal database of phonetics teaching methods in classroom interactions. http://www.lreccnf.org/proceedings/lrec2012/pdf/254_Paper.pdf

Aldera, A. S., & Mohsen, M. A. (2013). Annotations in captioned animation: Effects on vocabulary learning and listening skills. *Computers & Education, 68*, 60-75.

Alho, K., Salonen, J., Rinne, T., Medvedev, S. V., Hugdahl, K., & Hämäläinen, H. (2012). Attention-related modulation of auditory-cortex responses to speech sounds during dichotic listening. *Brain research, 1442*, 47-54.

Antonenko, P., & Niederhauser, D. S. (2010). The influence of leads on cognitive load and learning in a hypertext environment. *Computers in Human Behavior, 26*(2), 140-150.

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational psychology review, 22*(4), 425-438.

Antonenko, P., van Gog, T., & Paas, F. (2014). Implications of Neuroimaging for Educational Research. In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology* (pp. 51-63). Springer. https://doi.org/10.1007/978-1-4614-3185-5_5

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. A. (2018). *Introduction to research in education* (10th ed.). Cengage.

Asbjornsen, A. E., & Helland, T. (2006). Dichotic listening performance predicts language comprehension. *Laterality, 11*(3), 251-262. https://doi.org/10.1080/13576500500489360

Asp, C. W. (2006). *Verbotonal speech treatment*. Plural Publishing.

Asp, C. W., Kline, M., & Koike, K. J. (2012). Verbotonal worldwide. In R. Goldfarb (Ed.), *Translational speech-language pathology and audiology: Essays in honor of Dr. Sadanand Singh* (pp. 319–326). Plural Publishing.

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research, 18*(2), 172-184. https://doi.org/10.1016/j.cogbrainres.2003.10.002

Ayres, P. (2006). Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology, 20*, 287-298.

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *The American Journal of Psychology, 103*, 167-193.

Babbie, E. R. (2016). *The practice of social research* (14th ed.). Cengage.

Baddeley, A. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189-208. https://doi.org/10.1016/S0021-9924(03)00019-4

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173.

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., & Hyde, J. S. (1993). Processing strategies for time-course data sets in functional MRI of the human brain. *Magnetic Resonance in Medicine, 30*(2), 161-173. https://doi.org/10.1002/mrm.1910300204

Beauchemin, M., & De Beaumont, L. (2005). Statistical analysis of the mismatch negativity: To a dilemma, an answer. *Tutorials in Quantitative Methods for Psychology, 1*(1), 18-24.

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard, M. A., & Greenlee, J. D. W. (2015). Sensory–motor networks involved in speech production and motor control: An fMRI study. *Neuroimage, 109*, 418-428. https://doi.org/10.1016/j.neuroimage.2015.01.040

Berl, M. M., Vaidya, C. J., & Gaillard, W. D. (2006). Functional imaging of developmental and adaptive changes in neurocognition. *Neuroimage, 30*(3), 679-691.

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences, 16*(4), 240-250. https://doi.org/10.1016/j.tics.2012.03.001

Bitan, T., Cheon, J., Lu, D., Burman, D. D., & Booth, J. R. (2009). Developmental increase in top-down and bottom-up processing in a phonological task: An effective connectivity, fMRI Study. *Journal of Cognitive Neuroscience, 21*(6), 1135-1145.

Bitan, T., Cheon, J., Lu, D., Burman, D. D., Gitelman, D. R., Mesulam, M.-M., & Booth, J. R. (2007). Developmental changes in activation and effective connectivity in phonological processing. *Neuroimage, 38*(3), 564-575.

Böcker, K. B. E., Bastiaansen, M. C. M., Vroomen, J., Brunia, C. H. M., & De Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology, 36*(6), 706-720. https://doi.org/10.1017/S0048577299971767

Bohn, O.-S. (1995). Cross-language perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). York Press.

Booth, J. R. (2010). Development and language. In G. F. Koob, M. Le Moal, & R. F. Thompson (Eds.), *Encyclopaedia of behavioral neuroscience* (Vol. 1, pp. 387-395). Academic Press.

Born, J., & Wilhelm, I. (2012). System consolidation of memory during sleep. *Psychological research, 76*(2), 192-203.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: a neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113*, 787-821.

Bornkessel, I., Zysset, S., Friederici, A. D., von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *Neuroimage, 26*(1), 221-233. https://doi.org/10.1016/j.neuroimage.2005.01.032

Brådvik, B., Dravins, C., Holtås, S., Rosén, I., Ryding, E., & Ingvar, D. H. (1991). Disturbances of speech prosody following right hemisphere infarcts. *Acta Neurologica Scandinavica, 84*(2), 114-126. https://doi.org/10.1111/j.1600-0404.1991.tb04919.x

Brauer, J., & Friederici, A. D. (2007). Functional Neural Networks of Semantic and Syntactic Processes in the Developing Brain. *Journal of Cognitive Neuroscience, 19*(10), 1609-1623. https://doi.org/10.1162/jocn.2007.19.10.1609

Britton, B. K., & Tesser, A. (1982). Effects of prior knowledge on use of cognitive capacity in three complex cognitive tasks. *Journal of verbal learning and verbal behavior, 21*(4), 421-436.

Broca, P. P. (1861a). Nouvelle observation d'aphémie produite par une lésion de la moitié postérieure des deuxième et troisième circonvolutions frontales. *Bulletins de la Société d'Anatomie Paris, 6*, 398-407.

Broca, P. P. (1861b). Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletins de la Sociètè d'Anatomie (Paris), 6*(2e serie), 330-357.

Brockway, J. P. (2000). Two functional magnetic resonance imaging f(MRI) tasks that may replace the gold standard, Wada testing, for language lateralization while giving additional localization information. *Brain and Cognition, 43*(1-3), 57-59.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the role of the P600 in language comprehension. *Brain research, 1446*, 127–143.

Brouwer, H., & Hoeks, J. (2013). A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Frontiers in human neuroscience, 7*, 1–12.

Brumfit, C. J., & Johnson, K. (1979). *The communicative approach to language teaching* (Vol. 308). Oxford University Press.

Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current opinion in neurobiology, 17*(2), 243-250.

Buxton, R. B., Wong, E. C., & Frank, L. R. (1998). Dynamics of blood flow and oxygenation changes during brain activation: The balloon model. *Magnetic Resonance in Medicine, 39*, 855-864. https://doi.org/10.1002/mrm.1910390602

Cancelliere, A. E. B., & Kertesz, A. (1990). Lesion localization in acquired deficits of emotional expression and comprehension. *Brain and Cognition, 13*(2), 133-147. https://doi.org/10.1016/0278-2626(90)90046-Q

Cao, F., Young Kim, S., Liu, Y., & Liu, L. (2014). Similarities and differences in brain activation and functional connectivity in first and second language reading: Evidence from Chinese learners of English. *Neuropsychologia, 63*, 275-284. https://doi.org/10.1016/j.neuropsychologia.2014.09.001

Cappa, S. F., & Vignolo, L. A. (1983). CT scan studies of aphasia. *Human neurobiology, 2*(3), 129-134.

Caramazza, A., & Berndt, R. S. (1978). Semantic and syntactic processes in aphasia: A review of the literature. *Psychological Bulletin, 85*(4), 898-918. https://doi.org/10.1037/0033-2909.85.4.898

Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer, 21*(3), 77-88.

Casey, B., Galvan, A., & Hare, T. A. (2005). Changes in cerebral functional organization during cognitive development. *Current opinion in neurobiology, 15*(2), 239-244.

Chambers, C., Akram, S., Adam, V., Pelofi, C., Sahani, M., Shamma, S., & Pressnitzer, D. (2017). Prior context in audition informs binding and shapes simple features. *Nature Communications, 8*(1), 15027. https://doi.org/10.1038/ncomms15027

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and instruction, 8*(4), 293-332.

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*(2), 233-246.

Checa, P., & Fernández-Berrocal, P. (2015). The role of intelligence quotient and emotional intelligence in cognitive control processes. *Frontiers in Psychology, 6*, 1853. https://doi.org/10.3389/fpsyg.2015.01853

Chee, M. W. L., Tan, E. W. L., & Thiel, T. (1999). Mandarin and English Single Word Processing Studied with Functional Magnetic Resonance Imaging. *The Journal of Neuroscience, 19*(8), 3050. https://doi.org/10.1523/JNEUROSCI.19-08-03050.1999

Chen, H.-J. H. (2011). Developing and evaluating SynctoLearn, a fully automatic video and transcript synchronization tool for EFL learners. *Computer Assisted Language Learning, 24*(2), 117-130.

Chen, M., Ma, F., Wu, J., Li, S., Zhang, Z., Fu, Y., Lu, C., & Guo, T. (2020). Individual differences in language proficiency shape the neural plasticity of language control in bilingual language production. *Journal of Neurolinguistics, 54*, 100887. https://doi.org/10.1016/ j.jneuroling.2020.100887

Chien, P. J., Friederici, A. D., Hartwigsen, G., & Sammler, D. (2021). Intonation processing increases task-specific fronto-temporal connectivity in tonal language speakers. *Human brain mapping, 42*, 161-174. https://doi.org/10.1002/hbm.25214

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and cognitive processes, 27*(2), 231-256.

Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education* (8th ed.). Routledge.

Confucius Institute Headquarters. (2009). *Chinese Proficiency Test Syllabus Level 4*. The Commercial Press.

Copland, D. (2003). The basal ganglia and semantic engagement: Potential insights from semantic priming in individuals with subcortical vascular lesions, Parkinson's disease, and cortical lesions. *Journal of the International Neuropsychological Society, 9*(7), 1041-1052.

Cortiella, C., & Horowitz, S. H. (2014). *The state of learning disabilities: Facts, trends and emerging issues*. National Center for Learning Disabilities.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences, 24*(1), 87-114.

Cowan, N., Winkler, I., Teder, W., & Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *Journal of Experimental Psychology: Learning, Memory and Cognition, 19*(4), 909-921. https://doi.org/10.1037/0278-7393.19.4.909

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage.

Creswell, J. W., & Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

Cummings, A., Čeponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain research, 1115*(1), 92-107.

Cummings, J. L. (1993). Frontal-subcortical circuits and human behavior. *Archives of Neurology, 50*(8), 873-880.

Dale, A. M., & Buckner, R. L. (1997). Selective averaging of rapidly presented individual trials using fMRI. *Human brain mapping, 5*(5), 329-340. https://doi.org/10.1002/(SICI)1097-0193(1997)5:5<329::AID-HBM1>3.0.CO;2-5

Daltrozzo, J., & Conway, C. M. (2014). Neurocognitive mechanisms of statistical-sequential learning: what do event-related potentials tell us? *Frontiers in human neuroscience, 8*, 437.

Damiani, A. J. (2003). *The grammar translation method of language teaching*. Longman.

Dapretto, M., & Bookheimer, S. Y. (1999). Form and Content: Dissociating Syntax and Semantics in Sentence Comprehension. *Neuron, 24*(2), 427-432. https://doi.org/10.1016/S0896-6273(00)80855-7

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: Top-down influences on the interface between audition and speech perception. *Hearing research, 229*(1), 132-147. https://doi.org/10.1016/j.heares.2007.01.014

de Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. *Neuroimage, 90*, 348-359. https://doi.org/10.1016/j.neuroimage.2013.12.049

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science, 38*(2), 105-134.

de Jong, T., Van Gog, T., Jenks, K., Manlove, S., Van Hell, J., Jolles, J., Van Merriënboer, J., Van Leeuwen, T., & Boschloo, A. (2009). *Explorations in learning and the brain: On the potential of cognitive neuroscience for educational science*. Springer.

Declerck, M., & Kormos, J. (2012). The effect of dual task demands and proficiency on second language speech production. *Bilingualism: Language and Cognition, 15*(4), 782-796.

DeKeyser, R. M. (1997). Beyond explicit rule learning: Automatizing second language morphosyntax. *Studies in Second Language Acquisition, 19*(2), 195-221.

Della Rosa, P. A., Videsott, G., Borsa, V. M., Canini, M., Weekes, B. S., Franceschini, R., & Abutalebi, J. (2013). A neural interactive location for multilingual talent. *Cortex, 49*(2), 605-608. https://doi.org/10.1016/j.cortex.2012.12.001

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*, 237-261.

Diehl, R. L., Lotto, A. J., & Holt, L. L. (2004). Speech Perception. *Annual review of psychology, 55*(1), 149-179. https://doi.org/10.1146/annurev.psych.55.090902.142028

Dien, J. (2009). The neurocognitive basis of reading single words as seen through early latency ERPs: a model of converging pathways. *Biological Psychology, 80*(1), 10-22.

Dien, J., Michelson, C. A., & Franklin, M. S. (2010). Separating the visual sentence N400 effect from the P400 sequential expectancy effect: Cognitive and neuroanatomical implications. *Brain research, 1355*, 126-140.

Donaldson, D. I. (2004). Parsing brain activity with fMRI and mixed designs: What kind of a state is neuroimaging in? *Trends in Neurosciences, 27*, 442-444. https://doi.org/10.1016/j.tins.2004.06.001

Dronkers, N. F., Wilkins, D. P., Van Valin, R. D., Jr., Redfern, B. B., & Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition, 92*(1-2), 145-177.

Eggemeier, F. T. (1988). Properties of workload assessment techniques. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 41-62). North Holland, Elsevier.

Eisenberg, L. S., Dirks, D. D., Takayanagi, S., & Martinez, A. S. (1998). Subjective Judgments of Clarity and Intelligibility for Filtered Stimuli With Equivalent Speech Intelligibility Index Predictions. *Journal of Speech, Language, and Hearing Research, 41*(2), 327-339. https://doi.org/10.1044/jslhr.4102.327

Elliott, R. (2003). Executive functions and their disorders: Imaging in clinical neuroscience. *British Medical Bulletin, 65*(1), 49-59. https://doi.org/10.1093/bmb/65.1.49

Ferreira, F., & Tanenhaus, M. K. (2007). Introduction to the special issue on language–vision interactions. *Journal of Memory and Language, 57*(4), 455-459. https://doi.org/10.1016/j.jml.2007.08.002

Flege, J. E. (1995). Second-language Speech Learning: Theory, Findings, and Problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). York Press.

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(3), 319-325.

Fowler, C. A. (2008). The FLMP STMPed. *Psychonomic Bulletin & Review, 15*(2), 458-462.

Fowler, F. J. (2009). *Survey research methods* (4th ed.). Sage.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Francis, A. L., Baldwin, K., & Nusbaum, H. (2000). Effects of training on attention to acoustic cues. *Perception & psychophysics, 62*(8), 1668-1680.

Francis, A. L., & Nusbaum, H. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human perception and performance, 28*(2), 349-366.

Francis, A. L., & Nusbaum, H. (2009). Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics, 71*(6), 1360-1374.

Freeman, Y. S., & Freeman, D. E. (1992). *Whole language for second language learners*. Heinemann.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: from syllables to sentences. *Trends in Cognitive Sciences, 9*(10), 481-488.

Friederici, A. D. (2012). The cortical language circuit: from auditory perception to sentence comprehension. *Trends in Cognitive Sciences, 16*(5), 262-268. https://doi.org/10.1016/j.tics.2012.04.001

Friederici, A. D. (2016). The neuroanatomical pathway model of language: Syntactic and semantic networks. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 349-356). Academic Press.

Friederici, A. D., Fiebach, C. J., Schlesewsky, M., Bornkessel, I. D., & Von Cramon, D. Y. (2006). Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex, 16*(12), 1709-1717. https://doi.org/10.1093/cercor/bhj106

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current opinion in neurobiology, 23*(2), 250-254. https://doi.org/10.1016/j.conb.2012.10.002

Friederici, A. D., Rüschemeyer, S. A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170-177. https://doi.org/10.1093/cercor/13.2.170

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience and Biobehavioral Reviews, 25*(4), 355-373.

Friston, K. J. (2011). Functional and Effective Connectivity: A Review. *Brain Connectivity, 1*(1), 13-36. https://doi.org/10.1089/brain.2011.0008

Friston, K. J., Ashburner, J., Frith, C. D., Poline, J. B., Heather, J. D., & Frackowiak, R. S. (1995). Spatial registration and normalization of images. *Human brain mapping, 3*(3), 165-189.

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and Modulatory Interactions in Neuroimaging. *Neuroimage, 6*(3), 218-229. https://doi.org/10.1006/nimg.1997.0291

Friston, K. J., Holmes, A. P., Price, C. J., Büchel, C., & Worsley, K. J. (1999). Multisubject fMRI studies and conjunction analyses. *Neuroimage, 10*(4), 385-396. https://doi.org/10.1006/nimg.1999.0484

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. (1994). Statistical parametric maps in functional imaging: a general linear approach. *Human brain mapping, 2*(4), 189-210.

Fu, S., Chen, Y., Smith, S., Iversen, S., & Matthews, P. M. (2002). Effects of Word Form on Brain Processing of Written Chinese. *Neuroimage, 17*(3), 1538-1548. https://doi.org/10.1006/nimg.2002.1155

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *Neuroreport, 14*(5), 735-738.

Galotti, K. M. (2017). *Cognitive development: Infancy through adolescence* (2nd ed.). Sage.

Gandour, J., Tong, Y., Talavage, T., Wong, D., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2007). Neural basis of first and second language processing of sentence-level linguistic prosody. *Human brain mapping, 28*(2), 94-108. https://doi.org/10.1002/hbm.20255

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage, 23*(1), 344-357.

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., & Hutchins, G. D. (2000). A crosslinguistic PET study of tone perception. *Journal of Cognitive Neuroscience, 12*(1), 207-222. https://doi.org/10.1162/089892900561841

Gao, Z., Guo, X., Liu, C., Mo, Y., & Wang, J. (2020). Right inferior frontal gyrus: An integrative hub in tonal bilinguals. *Human brain mapping, 41*(8), 2152-2159. https://doi.org/10.1002/hbm.24936

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics, 39*(4), 546-557. https://doi.org/10.1016/j.wocn.2011.07.002

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging studies of language production and comprehension. *Annual review of psychology, 54*(1), 91-114.

Giard, M., Collet, L., Bouchet, P., & Pernier, J. (1994). Auditory selective attention in the human cochlea. *Brain research, 633*(1), 353-356. https://doi.org/10.1016/0006-8993(94)91561-X

Gilabert, R., & Barón, J. (2013). The impact of increasing task complexity on L2 pragmatic moves. In A. Mackey & K. McDonough (Eds.), *Second language interaction in diverse educational contexts* (pp. 45-69). John Benjamins.

Gilabert, R., Barón, J., & Levkina, M. (2011). Manipulating task complexity across task types and modes. In P. Robinson (Ed.), *Second language task complexity: Researching the cognition hypothesis of language learning and performance* (pp. 105-140). John Benjamins.

Goswami, U. (2004). Neuroscience and education. *British Journal of Educational Psychology, 74*, 1–14.

Gouvea, A. C., Phillips, C., Kazanina, N., & Poeppel, D. (2010). The linguistic processes underlying the P600. *Language and cognitive processes, 25*(2), 149-188.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: brain responses to angry prosody in meaningless speech. *Nature neuroscience, 8*(2), 145-146. https://doi.org/10.1038/nn1392

Greenbaum, S., & Nelson, G. (2002). *An introduction to English grammar* (2nd ed.). Pearson Education.

Greenbaum, S., & Nelson, G. (2009). *An introduction to English grammar* (3rd ed.). Routledge.

Greenspan, S. L., Nusbaum, H., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 421-433.

Grewe, T., Bornkessel, I., Zysset, S., Wiese, R., von Cramon, D. Y., & Schlesewsky, M. (2005). The emergence of the unmarked: A new perspective on the language-specific function of Broca's area. *Human brain mapping, 26*(3), 178-190. https://doi.org/10.1002/hbm.20154

Griesinger, D. H. (2013). What is "clarity", and how it can be measured? *Proceedings of Meetings on Acoustics, 19*(1), 015003. https://doi.org/10.1121/1.4799418

Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., & Tyler, L. K. (2013). Functional Organization of the Neural Language System: Dorsal and Ventral Pathways Are Critical for Syntax. *Cerebral Cortex, 23*(1), 139-147. https://doi.org/10.1093/cercor/bhr386

Grodzinsky, Y., & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current opinion in neurobiology, 16*(2), 240-246. https://doi.org/10.1016/j.conb.2006.03.007

Guberina, P. (1972). *Restricted bands of frequencies in auditory rehabilitation of deaf*. Institute of Phonetics, Faculty of Arts, University of Zagreb.

Guberina, P., & Asp, C. W. (1981). *The verbo-tonal method for rehabilitating people with communication problems*. World Rehabilitation Fund.

Guberina, P., & Asp, C. W. (2013). *The Verbotonal Method*. Artresor Naklada.

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese–English bilinguals. *Neuroimage, 56*(4), 2300-2309. https://doi.org/10.1016/j.neuroimage.2011.03.049

Hagoort, P. (2008). The fractionation of spoken language understanding by measuring electrical and magnetic brain signals. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1055-1069.

Hain, T. C. (2006). *Hearing loss*. Retrieved May 8, 2020 from http://www.dizziness-and-balance.com/disorders/hearing/hearing.html

Halgren, E., Baudena, P., Heit, G., Clarke, M., & Marinkovic, K. (1994). Spatio-temporal stages in face and word processing. 1. Depth recorded potentials in the human occipital and parietal lobes. *Journal of Physiology-Paris, 88*(1), 1-50.

Harris, C. L. (2006). *Language and cognition*. MacMillan.

Hashemi, L., & Thomas, B. (2013). *Objective PET Teacher's Book* (4th, Ed.). Cambridge University Press.

He, B. (2014). *Improving the English Pronunciation of Chinese EFL Learners through the Integration of CALL and Verbotonalism* [Doctoral dissertation, Suranaree University of Technology]. Thailand.

He, B., Sangarun, P., & Lian, A. P. (2015). Improving the English pronunciation of Chinese EFL university students through the integration of CALL and verbotonalism. Seventeenth international CALL research conference: Task design and CALL, Tarragona, Spain: University of Antwerp.

Heald, S., Klos, S., & Nusbaum, H. (2016). Understanding Speech in the Context of Variability. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 195-208). Academic Press. https://doi.org/10.1016/B978-0-12-407794-2.00017-1

Heald, S., & Nusbaum, H. (2014). Speech perception as an active cognitive process. *Frontiers in systems neuroscience, 8*(35). https://doi.org/10.3389/fnsys.2014.00035

Heigham, J., & Croker, R. (2009). *Qualitative Research in Applied Linguistics: A Practical Introduction*. Palgrave Macmillan.

Hernandez, A. E. (2009). Language switching in the bilingual brain: What's next? *Brain and Language, 109*(2), 133-140. https://doi.org/10.1016/j.bandl.2008.12.005

Hervais-Adelman, A., Pefkou, M., & Golestani, N. (2014). Bilingual speech-in-noise: Neural bases of semantic context use in the native language. *Brain and Language, 132*, 1-6. https://doi.org/10.1016/j.bandl.2014.01.009

Hickok, G. (2012). The cortical organization of speech processing: Feedback control and predictive coding the context of a dual-stream model. *Journal of Communication Disorders, 45*(6), 393-402. https://doi.org/10.1016/j.jcomdis.2012.06.004

Hickok, G., & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences, 4*(4), 131-138.

Hickok, G., & Poeppel, D. (2004). Dorsal and ventral streams: A framework for understanding aspects of the functional anatomy of language. *Cognition, 92*(1), 67-99. https://doi.org/10.1016/j.cognition.2003.10.011

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*(5), 393-402. https://doi.org/10.1038/nrn2113

Hickok, G., & Poeppel, D. (2016). Neural basis of speech perception. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 299-310). Academic Press.

Hosoda, C., Hanakawa, T., Nariai, T., Ohno, K., & Honda, M. (2012). Neural mechanisms of language switch. *Journal of Neurolinguistics, 25*(1), 44-61. https://doi.org/10.1016/j.jneuroling.2011.08.007

Hu, Y., & Uno, S. (2005). Effectiveness of a new teaching method based on the verbo-tonal method for Japanese beginners' learning of Chinese voice tones. *Japanese Journal of Educational Psychology, 53*(4), 541-550.

Huang, B., & Liao, X. (2017). *Modern Chinese* (Vol. 2). Higher Education Press.

Huffman, R. F., & Henson, O. W. (1990). The descending auditory pathway and acousticomotor systems: Connections with the inferior colliculus. *Brain Research Reviews, 15*(3), 295-323. https://doi.org/10.1016/0165-0173(90)90005-9

Hugdahl, K. (2003). Dichotic listening in the study of auditory laterality. In K. Hugdahl & R. J. Davidson (Eds.), *The Asymmetrical Brain* (pp. 441-476). MIT Press.

Indefrey, P. (2006). A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean? *Language Learning, 56*, 279-304. https://doi.org/10.1111/j.1467-9922.2006.00365.x

Indefrey, P. (2012). Hemodynamic Studies of Syntactic Processing. In M. Faust (Ed.), *The Handbook of Neuropsychology of Language* (pp. 209-228). Wiley-Blackwell.

Indefrey, P., & Levelt, W. J. M. (2000). The Neural Correlates of Language Production. In M. S. Gazzaniga (Ed.), *The New Cognitive Neurosciences* (2nd ed., pp. 845-865). MIT Press.

Institute for Quality and Efficiency in Health Care. (2008). *Hearing loss and deafness: Normal hearing and impaired hearing*. Retrieved May 8, 2020 from https://www.ncbi.nlm.nih.gov/books/NBK390300/

Itoh, K., Suwazono, S., Arao, H., Miyazaki, K. i., & Nakada, T. (2005). Electrophysiological correlates of absolute pitch and relative pitch. *Cerebral Cortex, 15*(6), 760-769.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative Science Quarterly, 24*(4), 602-611.

Kablan, Z., & Erden, M. (2008). Instructional efficiency of integrated and separated text with animated presentations in computer-based science instruction. *Computers & Education, 51*(2), 660-668.

Kadosh, K. C., & Johnson, M. H. (2007). Developing a cortex specialized for face perception. *Trends in Cognitive Sciences, 11*(9), 367-369.

Kappenman, E. S., & Luck, S. J. (2011). ERP components: The ups and downs of brainwave recordings. In E. S. Kappenman & S. J. Luck (Eds.), *The Oxford Handbook of Event-Related Potential Components* (pp. 3-30). Oxford University Press.

Kaufman, A., S., & Lichtenberger, E. O. (2005). *Assessing Adolescent and Adult Intelligence* (3rd ed.). John Wiley & Sons, Inc.

Kerr, B. (1973). Processing demands during mental operations. *Memory and Cognition, 1*(4), 401-412.

Kester, L., Kirschner, P. A., & van Merriënboer, J. J. G. (2006). Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology, 31*(2), 167-185.

Kim, Y., & Asp, C. W. (2002). Low Frequency Perception of Rhythm and Intonation Speech Patterns by Normal Hearing Adults. *Speech Sciences, 9*(1), 7-16.

King, J. W., & Kutas, M. (1995). Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience, 7*(3), 376-395.

Kinno, R., Kawamura, M., Shioda, S., & Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Human brain mapping, 29*(9), 1015-1027. https://doi.org/10.1002/hbm.20441

Klem, G. H., Lüders, H. O., Jasper, H., & Elger, C. (1999). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology. Supplement, 52*(3), 3-6.

Knecht, S., Dräger, B., Deppe, M., Bobe, L., Lohmann, H., Flöel, A., Ringelstein, E.-B., & Henningsen, H. (2000). Handedness and hemispheric language dominance in healthy humans. *Brain, 123*(12), 2512-2518. https://doi.org/10.1093/brain/123.12.2512

Koelsch, S., Gunter, T., Schröger, E., & Friederici, A. D. (2003). Processing tonal modulations: An ERP study. *Journal of Cognitive Neuroscience, 15*(8), 1149-1159.

Kolb, B., & Whishaw, I. Q. (2015). *Fundamentals of human neuropsychology* (7th ed.). Worth Publishers.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon Press.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman.

Krashen, S. D. (2003). *Explorations in language acquisition and use*. Heinemann.

Kraus, N., & Anderson, S. (2015). Identifying neural signatures of auditory function. *The Hearing Journal, 68*(1), 38-40.

Kraus, N., & Nicol, T. (2014). The Cognitive Auditory System: The Role of Learning in Shaping the Biology of the Auditory System. In A. N. Popper & R. R. Fay (Eds.), *Perspectives on Auditory Research* (pp. 299-319). Springer.

Krippendorf, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed.). Sage.

Krishnan, A. (2002). Human frequency-following responses: representation of steady-state synthetic vowels. *Hearing research, 166*, 192-201.

Krishnan, A. (2007). Frequency-following response. In R. F. Burkard, M. Don, & J. J. Eggermont (Eds.), *Auditory evoked potentials: Basic Principles and clinical application* (pp. 313-333). Lippincott, Williams & Wilkins.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161-168. https://doi.org/10.1016/j.cogbrainres.2005.05.004

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831-843.

Kuhl, P. K., & Rivera-Gaxiola, M. (2008). Neural substrates of language acquisition. *Annu. Rev. Neurosci., 31*, 511-534.

Kuperberg, G. R., McGuire, P. K., Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Wright, I. C., Lythgoe, D. J., Williams, S. C. R., & David, A. S. (2000). Common and distinct neural substrates for pragmatic, semantic, and syntactic processing of spoken sentences: An fMRI study. *Journal of Cognitive Neuroscience, 12*(2), 321-341. https://doi.org/10.1162/089892900562138

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470. https://doi.org/10.1016/S1364-6613(00)01560-6

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual review of psychology, 62*, 621-647.

Kutas, M., Kiang, M., & Sweeney, K. (2012). Potentials and paradigms: event-related brain potentials and neuropsychology. In M. Faust (Ed.), *The Handbook of Neuropsychology of Language* (pp. 545-564). Wiley-Blackwell.

Kwok, V. P. Y., Dan, G., Yakpo, K., Matthews, S., & Tan, L. H. (2016). Neural systems for auditory perception of lexical tones. *Journal of Neurolinguistics, 37*, 34-40. https://doi.org/10.1016/j.jneuroling.2015.08.003

Lambon Ralph, M. A., & Patterson, K. (2008). Generalization and Differentiation in Semantic Memory. *Annals of the New York Academy of Sciences, 1124*(1), 61-76. https://doi.org/10.1196/annals.1440.006

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review, 72*(4), 275-309.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching*. Oxford University Press.

Lee, D. J., Pouratian, N., Bookheimer, S. Y., & Martin, N. A. (2010). Factors predicting language lateralization in patients with perisylvian vascular malformations. *Journal of neurosurgery, 113*(4), 723-730.

Leitan, N. D., & Chaffey, L. (2014). Embodied cognition and its applications: A brief review. *Sensoria: A Journal of Mind, Brain & Culture, 10*(1), 3-10.

Levine, M. D. (2009). Differences in Learning and Neurodevelopmental Function in School-age Children. In W. B. Carey, A. C. Crocker, W. L. Coleman, E. R. Elias, & H. M. Feldman (Eds.), *Developmental-Behavioral Pediatrics* (4th ed., pp. 535-546). W.B. Saunders. https://doi.org/10.1016/B978-1-4160-3370-7.00055-9

Levkina, M., & Gilabert, R. (2012). The effects of cognitive task complexity on L2 oral production. In A. Housen, I. Vedder, & F. Kuiken (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp. 171-198). John Benjamins.

Levkina, M., & Gilabert, R. (2014). Task sequencing in the L2 development of spatial expressions. In M. Baralt, R. Gilabert, & P. Robinson (Eds.), *Task sequencing and instructed second language learning* (pp. 37-70). Bloomsbury.

Li, P., Legault, J., & Litcofsky, K. A. (2014). Neuroplasticity as a function of second language learning: Anatomical changes in the human brain. *Cortex, 58*, 301-324. https://doi.org/10.1016/j.cortex.2014.05.001

Li, X., Chen, Y., & Yang, Y. (2011). Immediate integration of different types of prosodic information during on-line spoken language comprehension: An ERP study. *Brain research, 1386*, 139-152. https://doi.org/10.1016/j.brainres.2011.02.051

Lian, A. P. (1980). *Intonation patterns of French (Teacher's book)*. River Seine Publications Pty Ltd.

Lian, A. P. (2004). Technology-enhanced language learning environments: A rhizomatic approach. In J.-B. Son (Ed.), *Computer-assisted language learning: Concepts, contexts and practices* (pp. 1-20). iUniverse.

Lian, A. P., & Lian, A. (1997). The Secret of the Shao-Lin Monk: Contribution to an intellectual framework for language-learning. *On-CALL, 11*(2), 2-18.

Lian, A. P., & Pineda, M. V. (2014). Rhizomatic learning: "As… when… and if…" A strategy for the ASEAN community in the 21st century. *Beyond Words, 2*(1), 1-28.

Lian, A. P., & Sussex, R. (2018). Toward a Critical Epistemology for Learning Languages and Cultures in Twenty-First Century Asia. In A. Curtis & R. Sussex (Eds.), *Intercultural*

Communication in Asia: Education, Language and Values (pp. 37-54). Springer International Publishing. https://doi.org/10.1007/978-3-319-69995-0_3

Liang, B., & Du, Y. (2018). The Functional Neuroanatomy of Lexical Tone Perception: An Activation Likelihood Estimation Meta-Analysis. *Frontiers in Neuroscience, 12*. https://doi.org/10.3389/fnins.2018.00495

Liberman, A. M., Cooper, F. S., Harris, K. S., MacNeilage, P. F., & Studdert-Kennedy, M. (1964). Some observations on a model for speech perception. The AFCRL symposium on models for the perception of speech and visual form, Boston, MA.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Liu, B., Wang, Z., & Jin, Z. (2009). The integration processing of the visual and auditory information in videos of real-world events: an ERP study. *Neuroscience Letters, 461*(1), 7-11.

Liu, H., & Cao, F. (2016). L1 and L2 processing in the bilingual brain: A meta-analysis of neuroimaging studies. *Brain and Language, 159*, 60-73. https://doi.org/10.1016/j.bandl.2016.05.013

Logothetis, N. K. (2012). What We Can and What We Can't Do with fMRI. Society for Neuroscience: 2012 Short Course II MRI and Advanced Imaging in Animals and Humans, New Orleans, LA.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences, 13*(3), 110-114.

Lotto, A. J., & Holt, L. L. (2016). Speech perception: The view from the auditory system. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 185-194). Academic Press.

Luck, S. J. (2014). *An introduction to the event-related potential technique* (2nd ed.). MIT Press.

Ludovic, K. (2010). Phonetic correction in class with verbo-tonal method. *Studies in Language and Literature, 30*(1), 35-56.

Luk, G., Green, D. W., Abutalebi, J., & Grady, C. (2012). Cognitive control for language switching in bilinguals: A quantitative meta-analysis of functional neuroimaging studies. *Language and cognitive processes, 27*(10), 1479-1488. https://doi.org/10.1080/01690965.2011.613209

Luu, T. M. V. (2021). *Improving Vietnamese EFL Learners' Listening Ability through an Optimized Prosodic Approach* [Doctoral dissertation, Suranaree University of Technology]. Thailand.

Lyketsos, C. G., Kozauer, N., & Rabins, P. V. (2007). Psychiatric manifestations of neurologic disease: Where are we headed? *Dialogues in Clinical Neuroscience, 9*(2), 111-124.

Ma, H., Hu, J., Xi, J., Shen, W., Ge, J., Geng, F., Wu, Y., Guo, J., & Yao, D. (2014). Bilingual cognitive control in language switching: An fMRI study of English-Chinese late bilinguals. *PLoS One, 9*(9), e106468. https://doi.org/10.1371/journal.pone.0106468

Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). Oxford University Press.

Mahmoudian, S., Farhadi, M., Najafi-Koopaie, M., Darestani-Farahani, E., Mohebbi, M., Dengler, R., Esser, K.-H., Sadjedi, H., Salamat, B., Danesh, A. A., & Lenarz, T. (2013). Central auditory processing during chronic tinnitus as indexed by topographical maps of the mismatch negativity obtained with the multi-feature paradigm. *Brain research, 1527*, 161-173. https://doi.org/10.1016/j.brainres.2013.06.019

Maison, S., Micheyl, C., & Collet, L. (2001). Influence of focused auditory attention on cochlear activity in humans. *Psychophysiology, 38*(1), 35-40.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences, 106*(20), 8362-8367. https://doi.org/10.1073/pnas.0810928106

Malicka, A., & Levkina, M. (2012). Measuring task complexity: Does EFL proficiency matter? In A. Shehadeh & C. Coombe (Eds.), *Task-based language teaching in foreign language contexts: Research and implementation* (pp. 43-66). John Benjamins.

Mangold, S. A., & Das, J. M. (2020). *Neuroanatomy, cortical primary auditory area*. StatPearls.

Marangolo, P., Rizzi, C., Peran, P., Piras, F., & Sabatini, U. (2009). Parallel recovery in a bilingual aphasic: a neurolinguistic and fMRI study. *Neuropsychology, 23*(3), 405–409. https://doi.org/10.1037/a0014824

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of educational psychology, 88*, 49-63.

Mårtensson, J., Eriksson, J., Bodammer, N. C., Lindgren, M., Johansson, M., Nyberg, L., & Lövdén, M. (2012). Growth of language-related brain areas after foreign language learning. *Neuroimage, 63*(1), 240-244. https://doi.org/10.1016/j.neuroimage.2012.06.043

Martin, A., Schurz, M., Kronbichler, M., & Richlan, F. (2015). Reading in the brain of children and adults: A meta-analysis of 40 functional magnetic resonance imaging studies. *Human Brain Mapping, 36*(5), 1963-1981. https://doi.org/10.1002/hbm.22749

Massaro, D. W., & Chen, T. H. (2008). The motor theory of speech perception revisited. *Psychonomic Bulletin & Review, 15*(2), 453-457.

Mayring, P. (2004). Qualitative content analysis. In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *A Companion to Qualitative Research* (pp. 266-269). Sage.

Mazoyer, B., Zago, L., Jobard, G., Crivello, F., Joliot, M., Perchey, G., Mellet, E., Petit, L., & Tzourio-Mazoyer, N. (2014). Gaussian mixture modeling of hemispheric lateralization for language in a large sample of healthy individuals balanced for handedness. *PLoS One, 9*(6), e101165. https://doi.org/10.1371/journal.pone.0101165

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive psychology, 18*(1), 1-86.

McDowell, K., Kerick, S., Santa Maria, D., & Hatfield, B. (2003). Aging, physical activity, and cognitive processing: an examination of P300. *Neurobiology of aging, 24*(4), 597-606.

McLaughlin, B. (1987). *Theories of second-language learning*. Edward Arnold.

McLaughlin, J., Osterhout, L., & Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature neuroscience, 7*(7), 703-704.

McNerney, S. (2011). A brief guide to embodied cognition: Why you are not your brain. *Scientific American*. https://blogs.scientificamerican.com/guest-blog/a-brief-guide-to-embodied-cognition-why-you-are-not-your-brain/

Mechelli, A., Crinion, J. T., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R. S., & Price, C. J. (2004). Structural plasticity in the bilingual brain. *Nature, 431*(7010), 757-757. https://doi.org/10.1038/431757a

Mellott, J. G., Bickford, M. E., & Schofield, B. R. (2014). Descending projections from auditory cortex to excitatory and inhibitory cells in the nucleus of the brachium of the inferior colliculus. *Frontiers in systems neuroscience, 8*(188). https://doi.org/10.3389/fnsys.2014.00188

Metter, E. J., Riege, W. H., Hanson, W. R., Camras, L. R., Phelps, M. E., & Kuhl, D. E. (1984). Correlations of glucose metabolism and structural damage to language function in aphasia. *Brain and Language, 21*(2), 187-207. https://doi.org/10.1016/0093-934X(84)90046-4

Meyer, L., Obleser, J., Anwander, A., & Friederici, A. D. (2012). Linking ordering in Broca's area to storage in left temporo-parietal regions: The case of sentence processing. *Neuroimage, 62*(3), 1987-1998. https://doi.org/10.1016/j.neuroimage.2012.05.052

Meyer, M., Alter, K., Friederici, A. D., Lohmann, G., & von Cramon, D. Y. (2002). fMRI reveals brain regions mediating slow prosodic modulations in spoken sentences. *Human brain mapping, 17*(2), 73-88. https://doi.org/10.1002/hbm.10042

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mohsen, M. A. (2016). The use of help options in multimedia listening environments to aid language learning: A review. *British Journal of Educational Technology, 47*(6), 1232-1242.

Montoya-Martínez, J., Bertrand, A., & Francart, T. (2019). Optimal number and placement of EEG electrodes for measurement of neural tracking of speech. *bioRxiv*, 800979. https://doi.org/10.1101/800979

Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1591), 919-931. https://doi.org/doi:10.1098/rstb.2011.0355

Mora, J. C., & Mora-Plaza, I. (2019). Contributions of cognitive attention control to L2 speech learning. In A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, & M. H. Sørensen (Eds.), *A Sound Approach to Language Matters - In Honor of Ocke-Schwen Bohn* (pp. 477–499). Department of English, School of Communication & Culture, Aarhus University.

Morgan-Short, K., & Tanner, D. (2014). Event-related potentials (ERPs). In J. Jegerski & B. VanPatten (Eds.), *Research methods in second language psycholinguistics* (pp. 127-152). Routledge.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences, 104*(40), 15894-15898.

Neisser, U. (1967). *Cognitive psychology*. Appleton-Century-Crofts.

Newman, S. D., Ikuta, T., & Burns, T. (2010). The effect of semantic relatedness on syntactic analysis: An fMRI study. *Brain and Language, 113*(2), 51-58. https://doi.org/10.1016/j.bandl.2010.02.001

Nickel, P., & Nachreiner, F. (2003). Sensitivity and diagnosticity of the 0.1-Hz component of heart rate variability as an indicator of mental workload. *Human factors, 45*(4), 575-590.

Nickels, S., Opitz, B., & Steinhauer, K. (2013). ERPs show that classroom-instructed late second language learners rely on the same prosodic cues in syntactic parsing as native speakers. *Neuroscience Letters, 557*, 107-111. https://doi.org/10.1016/j.neulet .2013.10.019

Nusbaum, H., & Henly, A. S. (1992). Listening to speech through an adaptive window of analysis. In B. Schouten (Ed.), *The processing of speech: From the auditory periphery to word recognition* (pp. 339-348). Mouton-De Gruyter.

Nusbaum, H., & Magnuson, J. (1997). Talker normalization: Phonetic constancy as a cognitive process. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 109-132). Academic Press.

Nusbaum, H., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.), *Speech perception, production, and linguistic structure* (pp. 113-134). OHM Publishing Company.

Nusbaum, H., & Schwab, E. C. (1986). The Role of Attention and Active Processing in Speech Perception. In E. C. Schwab & H. C. Nusbaum (Eds.), *Pattern Recognition by Humans and Machines: Vol. 1. Speech perception* (pp. 113-157). Academic Press. https://doi.org/10.1016/B978-0-12-631403-8.50009-6

O'Reilly, J. X., Woolrich, M. W., Behrens, T. E. J., Smith, S. M., & Johansen-Berg, H. (2012). Tools of the trade: Psychophysiological interactions and functional connectivity. *Social cognitive and affective neuroscience, 7*(5), 604-609. https://doi.org/10.1093/scan/nss055

Obleser, J., & Kotz, S. A. (2010). Expectancy Constraints in Degraded Speech Modulate the Language Comprehension Network. *Cerebral Cortex, 20*(3), 633-640. https://doi.org/10.1093/cercor/bhp128

Obretenova, S., Halko, M. A., Plow, E. B., Pascual-Leone, A., & Merabet, L. B. (2010). Neuroplasticity associated with tactile language communication in a deaf-blind subject. *Frontiers in human neuroscience, 3*. https://doi.org/10.3389/neuro.09 .060.2009

Ohta, S., Fukui, N., & Sakai, K. L. (2013). Syntactic computation in the human brain: the degree of merger as a key factor. *PLoS One, 8*(2), e56230.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology, 84*, 429-434.

Paas, F., Ayres, P., & Pachman, M. (2008). Assessment of cognitive load in multimedia learning: Theory, methods and applications. In D. H. Robinson & G. Schraw (Eds.), *Recent innovations in educational technology that facilitate student learning* (pp. 11-35). Information Age Publishing.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational psychologist, 38*(1), 63-71.

Paas, F., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology, 86*(1), 122-133.

Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and motor skills, 79*(1), 419-430.

Parker, G. J. M., Luzzi, S., Alexander, D. C., Wheeler-Kingshott, C. A. M., Ciccarelli, O., & Lambon Ralph, M. A. (2005). Lateralization of ventral and dorsal auditory-language pathways in the human brain. *Neuroimage, 24*(3), 656-666. https://doi.org/10.1016/j.neuroimage.2004.08.047

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *Neuroimage, 61*(1), 314-322. https://doi.org/10.1016/j.neuroimage.2012.02.068

Patterson, K., Nestor, P. J., & Rogers, T. T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience, 8*(12), 976-987. https://doi.org/10.1038/nrn2277

Paulmann, S. (2016). The neurocognition of prosody. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 1109-1120). Academic Press.

Paulmann, S., Jessen, S., & Kotz, S. A. (2012). It's special the way you say it: An ERP investigation on the temporal dynamics of two types of prosody. *Neuropsychologia, 50*(7), 1609-1620. https://doi.org/10.1016/j.neuropsychologia.2012.03.014

Peelle, J. E. (2019). The neural basis for auditory and audiovisual speech perception. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge Handbook of Phonetics* (pp. 193-216). Routledge. https://doi.org/10.4324/9780429056253-9

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: The analysis of functional brain images*. Elsevier.

Pernet, C. R., McAleer, P., Latinus, M., Gorgolewski, K. J., Charest, I., Bestelmeyer, P. E. G., Watson, R. H., Fleming, D., Crabbe, F., Valdes-Sosa, M., & Belin, P. (2015). The human voice areas: Spatial organization and inter-individual variability in temporal and extra-temporal cortices. *Neuroimage, 119*, 164-174. https://doi.org/10.1016/j.neuroimage .2015.06.050

Petersen, S. E., & Dubis, J. W. (2012). The mixed block/event-related design. *Neuroimage, 62*(2), 1177-1184. https://doi.org/10.1016/j.neuroimage.2011.09.084

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331*(6157), 585-589. https://doi.org/10.1038/331585a0

Petersen, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1989). Positron Emission Tomographic Studies of the Processing of Singe Words. *Journal of Cognitive Neuroscience, 1*(2), 153-170. https://doi.org/10.1162/jocn.1989.1.2.153

Plante, E., Creusere, M., & Sabin, C. (2002). Dissociating Sentential Prosody from Sentence Processing: Activation Interacts with Task Demands. *Neuroimage, 17*(1), 401-410. https://doi.org/10.1006/nimg.2002.1182

Poldrack, R. A. (2018). *The new mind readers: What neuroimaging can and cannot reveal about our thoughts*. Princeton University Press.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and instruction, 12*(1), 61-86.

Pressnitzer, D., Graves, J., Chambers, C., de Gardelle, V., & Egré, P. (2018). Auditory Perception: Laurel and Yanny Together at Last. *Current Biology, 28*(13), R739-R741. https://doi.org/10.1016/j.cub.2018.06.002

Prpić, N. (2015). Language processing – Role of inferior parietal lobule. *Gyrus, 3*(3), 173-175. https://doi.org/10.17486/gyr.3.1037

Rauschecker, J. P., & Scott, S. K. (2009). Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature neuroscience, 12*(6), 718-724. https://doi.org/10.1038/nn.2331

Renard, R., & Van Vlasselaer, J. J. (1976). *Foreign language teaching with an integrated methodology: The SGAV (Structuro-Global Audio-Visual) methodology*. Didier.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgments: A validation study. *Studies in Second Language Acquisition, 38*(4), 703-737.

Ridge, E. (1992). Communicative language teaching: Time for review? *Stellenbosch Papers in Linguistics Plus, 21*, 95-108.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied linguistics, 22*(1), 27-57.

Robinson, P. (2007). Task complexity, theory of mind, and intentional reasoning: Effects on L2 speech production, interaction, uptake and perceptions of task difficulty. *International Review of Applied Linguistics, 45*, 193-213.

Röder, B., Stock, O., Neville, H., Bien, S., & Rösler, F. (2002). Brain Activation Modulated by the Comprehension of Normal and Pseudo-word Sentences of Different Processing Demands: A Functional Magnetic Resonance Imaging Study. *Neuroimage, 15*(4), 1003-1014. https://doi.org/10.1006/nimg.2001.1026

Rolheiser, T., Stamatakis, E. A., & Tyler, L. K. (2011). Dynamic Processing in the Human Language System: Synergy between the Arcuate Fascicle and Extreme Capsule. *The*

Journal of Neuroscience, 31(47), 16949. https://doi.org/10.1523/JNEUROSCI.2725-11.2011

Rombouts, S. A., Barkhof, F., Hoogenraad, F. G., Sprenger, M., Valk, J., & Scheltens, P. (1997). Test-retest analysis with functional MR of the activated area in the human visual cortex. *American Journal of Neuroradiology, 18*(7), 1317-1322.

Ross, A., & Willson, V. L. (2018). *Basic and advanced statistical tests: Writing results sections and creating tables and figures*. Springer.

Rüschemeyer, S. A., Fiebach, C. J., Kempe, V., & Friederici, A. D. (2005). Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian. *Human brain mapping, 25*(2), 266-286.

Rymarczyk, K., & Grabowska, A. (2007). Sex differences in brain control of prosody. *Neuropsychologia, 45*(5), 921-930. https://doi.org/10.1016/j.neuropsychologia.2006.08.021

Sammler, D., Grosbras, M. H., Anwander, A., Bestelmeyer, P. E., & Belin, P. (2015). Dorsal and ventral pathways for prosody. *Current Biology, 25*(23), 3079-3085.

Sander, D., Grandjean, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). Emotion and attention interactions in social cognition: Brain regions involved in processing anger prosody. *Neuroimage, 28*(4), 848-858. https://doi.org/10.1016/j.neuroimage.2005.06.023

Santi, A., & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *Neuroimage, 51*(4), 1285-1293. https://doi.org/10.1016/j.neuroimage .2010.03.034

Sasayama, S. (2013). Is a "complex" task really complex? Measuring task complexity independently from linguistic production. the 5th Biennial International Conference on Task-Based Language Teaching, Banff, Alberta, Canada,

Sattler, J. M. (2001). Issues related to the measurement and change of intelligence. In J. M. Sattler (Ed.), *Assessment of children: Cognitive applications* (4th ed., pp. 160-182). Jerome M. Sattler, Publisher, Inc.

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences, 105*(46), 18035-18040. https://doi.org/10.1073/pnas.0805234105

Schacht, A., Sommer, W., Shmuilovich, O., Martíenz, P. C., & Martín-Loeches, M. (2014). Differential Task Effects on N400 and P600 Elicited by Semantic and Syntactic Violations. *PLoS One, 9*(3), e91226. https://doi.org/10.1371/journal.pone.0091226

Schmutz, P., Heinz, S., Métrailler, Y., & Opwis, K. (2009). Cognitive Load in eCommerce Applications—Measurement and Effects on User Satisfaction. *Advances in Human-Computer Interaction, 2009*. https://doi.org/10.1155/2009/121494

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1-66.

Schultheis, H., & Jameson, A. (2004). Assessing cognitive load in adaptive hypermedia systems: Physiological and behavioral methods. In P. DeBra & W. Nejdl (Eds.), *Adaptive hypermedia and adaptive web-based systems* (Vol. 3137, pp. 225-234). Springer.

Schulz, P., & Steimer, T. (2000). Psychotropic medication, psychiatric disorders, and higher brain functions. *Dialogues in Clinical Neuroscience, 2*(3), 177-182.

Schwab, E. C., Nusbaum, H., & Pisoni, D. B. (1985). Some Effects of Training on the Perception of Synthetic Speech. *Human factors, 27*(4), 395-408. https://doi.org/10.1177/001872088502700404

Seeber, K. G. (2011). Cognitive load in simultaneous interpreting: Existing theories - New models. *Interpreting, 13*(2), 176-204.

Shahin, A. J., Alain, C., & Picton, T. W. (2006). Scalp topography and intracerebral sources for ERPs recorded during auditory target detection. *Brain Topography, 19*(1), 89-105.

Shi, J., Zhang, Z., Yao, Z., Hao, G., & Chen, N. (2006). Cognitive neural substrates for Chinese-English languages studied with functional MRI. *Chinese Journal of Behavioral Medicine and Brain Science, 15*(4), 349-351.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*, 127-190.

Silva-Pereyra, J., Conboy, B. T., Klarman, L., & Kuhl, P. K. (2007). Grammatical processing without semantics? An event-related brain potential study of preschoolers using jabberwocky sentences. *Journal of Cognitive Neuroscience, 19*(6), 1050-1065.

Silverman, D. (2018). *Doing qualitative research* (5th ed.). Sage.

Siok, W. T., Niu, Z., Jin, Z., Perfetti, C. A., & Tan, L. H. (2008). A structural–functional basis for dyslexia in the cortex of Chinese readers. *Proceedings of the National Academy of Sciences, 105*(14), 5561-5566. https://doi.org/10.1073/pnas.0801750105

Small, S. L., & Hickok, G. (2016). The Neurobiology of Language. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 3-9). Academic Press.

Springer, J. A., Binder, J. R., Hammeke, T. A., Swanson, S. J., Frost, J. A., Bellgowan, P. S., Brewer, C. C., Perry, H. M., Morris, G. L., & Mueller, W. M. (1999). Language dominance in neurologically normal and epilepsy subjects: a functional MRI study. *Brain, 122*(11), 2033-2046.

Starkstein, S. E., Federoff, J. P., Price, T. R., Leiguarda, R. C., & Robinson, R. G. (1994). Neuropsychological and neuroradiologic correlates of emotional prosody comprehension. *Neurology, 44*(3 Part 1), 515-515.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature neuroscience, 2*(2), 191-196. https://doi.org/10.1038/5757

Stevens, R. H., Galloway, T., & Berka, C. (2007). Allocation of time, workload, engagement and distraction as students acquire problem solving skills. In D. Schmorrow, D. Nicholson, J. Drexler, & L. Reeves (Eds.), *Foundations of augmented cognition* (pp. 128-137). Strategic Analysis.

Sundheim, S. T., & Voeller, K. K. (2004). Psychiatric implications of language disorders and learning disabilities: Risks and management. *Journal of Child Neurology, 19*(10), 814-826. https://doi.org/10.1177/08830738040190101001

Swaab, T. Y., Ledoux, K., Camblin, C. C., & Boudewyn, M. A. (2012). Language-related ERP components. In S. J. Luck & E. S. Kappenman (Eds.), *Oxford Handbook of Event-Related Potential Components* (pp. 397-440). Oxford University Press.

Swan, M. (1985). A critical look at the communicative approach (2). *ELT Journal, 39*(2), 76-87.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science, 12*(2), 257-285.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review, 22*(2), 123-138.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. Springer.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and instruction, 1*, 59-89.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational psychology review, 10*(3), 251-296.

Szaflarski, J. P., Rajagopal, A., Altaye, M., Byars, A. W., Jacola, L., Schmithorst, V. J., Schapiro, M. B., Plante, E., & Holland, S. K. (2012). Left-handedness and language lateralization in children. *Brain research, 1433*, 85-97. https://doi.org/10.1016/j.brainres .2011.11.026

Talairach, J., & Tournoux, P. (1988). *Co-Planar Stereotactic Atlas of the Human Brain*. Thieme.

Tan, L. H., Laird, A. R., Li, K., & Fox, P. T. (2005). Neuroanatomical correlates of phonological processing of Chinese characters and alphabetic words: A meta-analysis. *Human brain mapping, 25*(1), 83-91.

Tan, L. H., Spinks, J. A., Gao, J. H., Liu, H. L., Perfetti, C. A., Xiong, J., Stofer, K. A., Pu, Y., Liu, Y., & Fox, P. T. (2000). Brain activation in the processing of Chinese characters and words: A functional MRI study. *Human brain mapping, 10*(1), 16-27. https://doi.org/10.1002/(SICI)1097-0193(200005)10:1<16::AID-HBM30>3.0.CO;2-M

Tavor, I., Jones, O. P., Mars, R. B., Smith, S. M., Behrens, T. E., & Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science, 352*(6282), 216-220. https://doi.org/10.1126/science.aad8127

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews, 43*(3), 231-246.

Tie, Y., Suarez, R. O., Whalen, S., Radmanesh, A., Norton, I. H., & Golby, A. J. (2009). Comparison of blocked and event-related fMRI designs for pre-surgical language mapping. *Neuroimage, 47*(Suppl 2), T107-T115. https://doi.org/10.1016/j.neuroimage. 2008.11.020

Tokowicz, N., & MacWhinney, B. (2005). Implicit and Explicit Measures of Sensitivity to Violations in Second Language Grammar: An Event-Related Potential Investigation. *Studies in Second Language Acquisition, 27*(2), 173-204. https://doi.org/10.1017/ S0272263105050102

Tracy, D. K., Ho, D. K., O'Daly, O., Michalopoulou, P., Lloyd, L. C., Dimond, E., Matsumoto, K., & Shergill, S. S. (2011). It's not what you say but the way that you say it: An fMRI study of differential lexical and non-lexical prosodic pitch processing. *BMC neuroscience, 12*(1), 128.

Tu, L., Wang, J. J., Abutalebi, J., Jiang, B., Pan, X. M., Li, M., Gao, W., Yang, Y. C., Liang, B. S., Lu, Z., & Huang, R. W. (2015). Language exposure induced neuroplasticity in the bilingual brain: A follow-up fMRI study. *Cortex, 64*, 8-19. https://doi.org/10.1016/j.cortex.2014.09.019

Ullsperger, M., & Debener, S. (2010). *Simultaneous EEG and fMRI: recording, analysis, and application*. Oxford University Press.

Uttal, W. R. (2011). *Mind and Brain: A Critical Appraisal of Cognitive Neuroscience*. The MIT Press.

van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., & Schmidt, H. G. (2004). Memory load and the cognitive pupillary response in aging. *Psychophysiology, 41*(2), 167-174.

van Gog, T., Kester, L., Nievelstein, F., Giesbers, B., & Paas, F. (2009). Uncovering cognitive processes: Different techniques that can contribute to cognitive load research and instruction. *Computers in Human Behavior, 25*(2), 325-331.

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational psychologist, 43*(1), 16-26.

van Herten, M., Kolk, H. H. J., & Chwilla, D. J. (2005). An ERP study of P600 effects elicited by semantic anomalies. *Cognitive Brain Research, 22*(2), 241–255.

Vandergrift, L. (2003a). From Prediction Through Reflection: Guiding Students: Through the Process of L2 Listening. *The Canadian Modern Language Review, 59*(3), 425-440. https://doi.org/10.3138/cmlr.59.3.425

Vandergrift, L. (2003b). Orchestrating Strategy Use: Toward a Model of the Skilled Second Language Listener. *Language Learning, 53*(3), 463-496. https://doi.org/10.1111/1467-9922.00232

Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching, 40*(3), 191-210. https://doi.org/10.1017/S0261444807004338

Vanhoucke, E., Cousin, E., & Baciu, M. (2013). Hemispheric asymmetry modulation for language processing in aging: meta-analysis of studies using the dichotic listening test. *Geriatrie et psychologie neuropsychiatrie du vieillissement, 11*(1), 57-64.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., Mazoyer, B., & Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *Neuroimage, 30*(4), 1414-1432.

Vingerhoets, G., Borsel, J. V., Tesink, C., van den Noort, M., Deblaere, K., Seurinck, R., Vandemaele, P., & Achten, E. (2003). Multilingualism: an fMRI study. *Neuroimage, 20*(4), 2181-2196. https://doi.org/10.1016/j.neuroimage.2003.07.029

Welman, C., Kruger, S. J., & Mitchell, B. (2005). *Research methodology* (3rd ed.). Oxford University Press.

Wen, F., Lian, A. P., & Sangarun, P. (2020). Determination of corrective optimals for Chinese university learners of English. *Govor, 37*(1), 3-29.

Wernicke, C. (1874). *Der aphasische symptomenkomplex*. Cohn & Weigert.

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1), 1-12. https://doi.org/10.1016/j.edurev.2006.11.001

White, E. J., Genesee, F., & Steinhauer, K. (2012). Brain Responses before and after Intensive Second Language Learning: Proficiency Based Changes and First Language Background Effects in Adult Learners. *PLoS One, 7*(12), e52318. https://doi.org/10.1371/journal.pone.0052318

Wilson, Stephen M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., Ogar, J. M., DeLeon, J., Miller, B. L., & Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron, 72*(2), 397-403. https://doi.org/10.1016/j.neuron.2011.09.014

Witteman, J., van Ijzendoorn, M. H., van de Velde, D., van Heuven, V. J. J. P., & Schiller, N. O. (2011). The nature of hemispheric specialization for linguistic and emotional prosodic perception: A meta-analysis of the lesion literature. *Neuropsychologia, 49*(13), 3722-3738. https://doi.org/10.1016/j.neuropsychologia.2011.09.028

Wong, P. C. M., Ciocca, V., Chan, A. H. D., Ha, L. Y. Y., Tan, L. H., & Peretz, I. (2012). Effects of culture on musical pitch perception. *PLoS One, 7*(4), e33424. https://doi.org/10.1371/journal.pone.0033424

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature neuroscience, 10*(4), 420-422. https://doi.org/10.1038/nn1872

World Medical Association. (2013). World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *The Journal of the American Medical Association, 310*(20), 2191-2194. https://doi.org/10.1001/jama.2013.281053

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited—Again. *Neuroimage, 2*(3), 173-181. https://doi.org/10.1006/nimg.1995.1023

Wu, C.-Y., Ho, M.-H. R., & Chen, S.-H. A. (2012). A meta-analysis of fMRI studies on Chinese orthographic, phonological, and semantic processing. *Neuroimage, 63*(1), 381-391. https://doi.org/10.1016/j.neuroimage.2012.06.047

Wu, J., Li, X., Yang, J., Cai, C., Sun, H., & Guo, Q. (2012). Prominent activation of the bilateral inferior parietal lobule of literate compared with illiterate subjects during Chinese logographic processing. *Experimental Brain Research, 219*(3), 327-337. https://doi.org/10.1007/s00221-012-3094-8

Xing, F. (2016). *Modern Chinese Grammar: A Clause-Pivot Approach*. Routledge.

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics, 33*, 29-49. https://doi.org/10.1016/j.jneuroling.2014.09.004

Yang, J., Ye, J., Wang, R., Zhou, K., & Wu, Y. J. (2018). Bilingual Contexts Modulate the Inhibitory Control Network. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.00395

Yang, Y. (2016). *Improving the Phonological Working Memory and English Speaking Skills of Chinese Primary EFL Learners with a Verbotonal-Based Approach* [Doctoral dissertation, Suranaree University of Technology]. Thailand.

Yang, Y., Wannaruk, A., & Lian, A. P. (2017). Improving the English-speaking skills of Chinese primary EFL learners with a verbotonal approach. *Rangsit Journal of Arts and Sciences, 7*(2), 141-156.

Zhang, F. Z. (2006). *The teaching of Mandarin prosody: A somatically-enhanced approach for second language learners* [Doctoral dissertation, University of Canberra]. Australia.

Zhang, F. Z. (2012). Teaching Mandarin prosody: A Somatically-Enhanced Approach for second language learners. *Taiwan Journal of Chinese as a Second Language, 4*(1), 95-130.

Zhang, Q., Wang, H., Luo, C., Zhang, J., Jin, Z., & Li, L. (2019). The neural basis of semantic cognition in Mandarin Chinese: A combined fMRI and TMS study. *Human brain mapping, 40*(18), 5412-5423.

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional Overlap between Regions Involved in Speech Perception and in Monitoring One's Own Voice during Speech Production. *Journal of Cognitive Neuroscience, 22*(8), 1770-1781. https://doi.org/10.1162/jocn.2009.21324

Zhu, J. D., Seymour, R. A., Szakay, A., & Sowman, P. F. (2020). Neuro-dynamics of executive control in bilingual language switching: An MEG study. *Cognition, 199*, 104247. https://doi.org/10.1016/j.cognition.2020.104247

Zou, L., Packard, J. L., Xia, Z., Liu, Y., & Shu, H. (2016). Neural Correlates of Morphological Processing: Evidence from Chinese. *Frontiers in human neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00714

APPENDICES

APPENDIX A

A questionnaire on participants' personal information and handedness

English version

Directions:

This questionnaire is designed to gather your personal information. There is no right or wrong answer. Please read each statement carefully, and complete the questions by filling out or ticking (√) your response/opinions. Your answers will be used confidentially. Thank you for your participation!

Part 1 Personal Information

1. Age:
2. Gender:
3. Do you have difficulty hearing what teachers say in the classroom? Yes☐ No☐
4. Do you have difficulty hearing a whispered voice, a finger rub, or a clock tick?
   Yes☐ No☐
5. Have you ever worn a hearing aid? Yes☐ No☐
6. Without wearing glasses, can you see the regular print in books/newspapers?
   Yes☐ No☐ If no, can you read books/newspapers with glasses? Yes☐ No☐
7. Without wearing glasses, can you see street signs while walking on the street?
   Yes☐ No☐ If no, can you see them with glasses? Yes☐ No☐
8. Do you have neurological, psychiatric, or vascular diseases? Yes☐ No☐
9. Are you taking psychotropic or anti-hypertensive medications now? Yes☐ No☐
10. Your first language is ___________.
11. How many years have you been learning English? ___________.
12. Have you passed CET-4? Yes☐ No☐ The score is ___________.
13. Your self-evaluation of English proficiency: Elementary☐ Pre-intermediate☐
    Intermediate☐ Upper intermediate☐ Advance☐
14. Have you ever received professional musical training before? Yes☐ No☐

Part 2 Handedness Questionnaire (Adopted from Oldfield, 1971)

Please read each statement carefully and tick (√) the response that represents your opinions.

Which hand do you prefer to use when:

	Which hand you prefer for that activity?			Do you ever use the other hand?
Writing:	Left☐	No preference☐	Right☐	Yes☐
Drawing:	Left☐	No preference☐	Right☐	Yes☐
Throwing:	Left☐	No preference☐	Right☐	Yes☐
Using scissors:	Left☐	No preference☐	Right☐	Yes☐
Using a toothbrush:	Left☐	No preference☐	Right☐	Yes☐
Using a knife (without a fork):	Left☐	No preference☐	Right☐	Yes☐
Using a spoon:	Left☐	No preference☐	Right☐	Yes☐
Using a broom (upper hand):	Left☐	No preference☐	Right☐	Yes☐
Striking a match:	Left☐	No preference☐	Right☐	Yes☐
Opening a box (holding the lid):	Left☐	No preference☐	Right☐	Yes☐
Holding a computer mouse:	Left☐	No preference☐	Right☐	Yes☐
Using a key to unlock a door:	Left☐	No preference☐	Right☐	Yes☐
Holding a hammer:	Left☐	No preference☐	Right☐	Yes☐
Holding a brush or comb:	Left☐	No preference☐	Right☐	Yes☐
Holding a cup while drinking:	Left☐	No preference☐	Right☐	Yes☐

Chinese version

说明：
本次问卷是用于收集您的个人信息。问卷答案没有对错之分。请认真阅读每一个问题，根据自己的实际情况填写或在方框内打钩（√）。您的答案将作保密处理后再使用。感谢您的参与。

第一部分 个人信息

1. 年龄：
2. 性别：
3. 您在课堂上是否感到听老师说话有困难？是□ 否□
4. 您是否难以听清低声的耳语，手指的摩擦或时钟的滴答声？是□ 否□
5. 您是否戴过助听器？是□ 否□
6. 不戴眼镜，您是否能看清书籍/报纸上常规大小的字吗？是□ 否□ 如果不能，戴了眼镜后是否能看清？是□ 否□
7. 不戴眼镜，您走在路上是否能看清路牌？是□ 否□ 如果不能，戴了眼镜后是否能看清？是□ 否□
8. 您是否患有神经性疾病，精神疾病或血管疾病？是□ 否□
9. 您是否正在服用精神类或抗高血压类药物？是□ 否□
10. 您的母语是 ________.
11. 您学习英语几年了？ ________.
12. 您是否通过了大学英语四级（CET 4）？是□ 否□ 分数是 ________.
13. 您对英语水平的自我评价: 初级□ 中初级□ 中等□ 中高级□ 高级□
14. 您是否接受过专业的音乐训练？是□ 否□

第二部分 左右利手调查问卷 (Oldfield, 1971)

请仔细阅读下列每一条陈述，根据自己的实际情况在方框内打钩（√）。

在以下情形中，你更喜欢使用哪只手:

	您喜欢使用哪只手进行以下活动？			**您曾经用过另一只手吗？**
写字：	左□	没有偏好□	右□	是□
绘画：	左□	没有偏好□	右□	是□
投掷：	左□	没有偏好□	右□	是□
用剪刀：	左□	没有偏好□	右□	是□
刷牙：	左□	没有偏好□	右□	是□
用刀：	左□	没有偏好□	右□	是□
用汤匙：	左□	没有偏好□	右□	是□
用扫帚（在上的手）：	左□	没有偏好□	右□	是□
划火柴：	左□	没有偏好□	右□	是□
打开盒子（握盖子的手）：	左□	没有偏好□	右□	是□
拿鼠标：	左□	没有偏好□	右□	是□
拿钥匙开门：	左□	没有偏好□	右□	是□
拿锤子：	左□	没有偏好□	右□	是□
拿刷子或梳子:	左□	没有偏好□	右□	是□
喝水时拿杯子：	左□	没有偏好□	右□	是□

APPENDIX B

Informed consent form for participating in the experiment

English version

Part 1 Introduction

You are invited to participate in a study on measuring brain activity in response to different auditory language signals. This study adopts a combined ERP and fMRI experiment and a semi-structured interview. The study aims to explore an optimal language input for English-as-a-Second-Language students. Since you are qualified for the requirements, you are invited to be involved in the study.

This Informed Consent offers you some information to help you decide whether you are willing to join in the research. Participating in the study is voluntary. This project is approved by the Ethics Committee of Kunming Medical University. If you agree to join in the research, please read the following part.

1. Introduction

In the study, we will perform a combined ERP and fMRI experiment, which will last 30-40 minutes. Eight runs of the experiment will be performed with different auditory signals. The signals are Chinese and English conversations and the sounds that are totally different from Chinese/English conversations. During the experiment, please close your eyes and listen to each signal carefully. After the experiment, a semi-structured interview will be conducted. The interview will last for around 15 minutes. You will not be asked to share personal beliefs or stories, and you do not have to share any knowledge that you are not comfortable sharing. The entire discussion will be recorded, but no one will be identified by name from the recording.

2. Discomfort and risks that may occur to participants

During the experiment, you cannot move for about 40 minutes, which may make you feel discomfort.

3. Confidentiality

Your record will be kept properly and handled in strict confidentiality. It may be monitored by the Ethics Committee, but your information will not be disclosed to the public.

4. Contact person during the research

Mr. CAI Xirui, Tel: +86 13888574715, Email: caixirui@kmmu.edu.cn

5. You have the following rights: you are informed and agree to take part in the project. Participating or quitting the study is voluntary. Even if you quit the research during the trial, you will not have the loss of interest or any punishment.

6. If the unexpected experimental influence were discovered during the experiment, we would change the related content of this informed consent if necessary. Please sign again then.

7. If the experimental procedures do harm to you, appropriate compensation will be given to the participants, including transportation expenses, lost wages and medical expenses.

8. This program was approved by the Ethics Committee of Suranaree University of Technology, Thailand. If any procedure is against the scheme, participants can make a complaint to the Ethics Committee.

Part 2 Consent

1. I have already read the Informed Consent. The researchers have explained the study plan to me clearly. I totally understand the aim, method, my rights and risks of taking part in this study. I have been told that my personal information is confidential, and my privacy is protected.

I volunteer to take part in this study, cooperate with the researchers in accordance with the research method and the content in this Informed Consent, and complete the tasks seriously.

Participant's signature:

Date:

2. I have clearly explained the research aim, procedures and risks that may occur to the participants, and answered the participant's questions satisfactorily.

Researcher's signature:

Date:

Chinese version

知情同意书

第一部分 知情部分

您将被邀请参加一项测量大脑对不同听觉语言信号的活动研究。该研究采用 ERP 和 fMRI 实验，以及半结构式采访。本研究旨在探索英语作为二语学习的最优语音输入信号。由于您符合入组条件而被邀请加入此项研究。

本知情同意书提供给您一些信息以帮助您决定是否参加此项研究。您参加本项研究是自愿的。本次研究已通过昆明医科大学医学伦理委员会审查。如果您同意加入此项研究，请看下列说明。

请您仔细阅读，如有任何疑问请向该项研究的研究者提出。

1. 项目介绍

在本研究中，我们将分别进行ERP和fMRI采集，历时30-40分钟，共有八组语音信号。在每一组语音信号中，将有不同的中文和英语对话和完全区别于对话的音频。整个实验过程请您闭上双眼，认真聆听每一个音频。之后，是一个半结构式的采访。采访将持续约15分钟。您不会被要求分享个人信念或故事，也不必分享任何您不愿意分享的内容。整个采访将被录音，但是录音不会以姓名标识。

2. 研究给受试者可能带来的不适和风险

在实验过程中，您需要坚持约40分钟左右不能动，可能会让您感到不适。

3. 研究的保密性

您的记录将被妥善保管，作保密处理，有可能会接受有关部门(伦理委员会)的监察，但不会对外披露其内容。

4. 试验过程中联系研究者的姓名和联系办法

蔡希睿，电话：+86 13888574715，Email：caixirui@kmmu.edu.cn

5. 您拥有以下权利：自由参加和退出、知情、同意，参加试验是自愿的，即使中途退出试验也不会有权益上的损失或任何惩罚。

6. 在研究中若发现预期以外的试验影响，必要时可能对知情同意书相关内容进行修改，届时将请您重新签名确认。

7. 在研究中若因试验步骤的原因造成损害，我们将给予受试者相应的补偿，例如交通费、误工费、医疗费用。

8. 该试验方案经泰国苏南拉里理工大学医学伦理委员会批准实施，试验过程中有任何违反研究方案的情况，受试者可以直接向伦理委员会投诉。

第二部分 同意部分

1. 我已详细阅读了知情同意书，研究者已向我作了详尽的研究方案说明，我完全了解参加本次研究的目的、性质、方法及我的权益和风险，得知我的个人资料是保密的，隐私权也得到保护。

我自愿参加本次研究，并同意按照研究方法和知情同意书的内容配合研究者操作，认真完成本次研究。

受试者签名：

日期：

2. 我已向该受试者充分解释和说明了本研究的目的、操作过程以及受试者参加该试验可能存在的风险，并满意地回答了受试者的所有有关问题。

研究者签名：

日期：

APPENDIX C

Guiding questions in a semi-structured interview

English version

1. What do you feel when listening to the signals in the experiment?
2. Do you feel comfortable/uncomfortable when listening to a/some certain kind/kinds of signal/signals? Please specify which one/ones makes/make you feel comfortable/ uncomfortable?
3. Do you think you can hear more clearly/unclearly with a/some certain kind/kinds of signal/signals? Please specify the signal/signals.
4. Is there a signal that helps you understand the meaning of the sentences better? Please specify the signal.
5. Is there a signal that helps you understand the structure of the sentences better? Please specify the signal.
6. Is there a signal that you think is easier to understand? Please specify the signal.
7. Is there a signal that makes you take less/more effort to understand the sentences? Which kind of signal is it?
8. Do you prefer any kind of signal to other signals in the experiment?
9. What do you think of using the signals for learning English?
10. What else would you like to say about the signals?

Chinese version

1. 您在听实验中的信号时有何感觉？
2. 当您在听某种/某些信号时，您是否感到舒服/不舒服？请指明是哪个/哪些信号让您感到舒服/不舒服？
3. 您认为某种/某些信号可以让您听得更清楚/不清楚吗？请指明是哪种信号。
4. 是否有一种信号可以帮助您更好地理解句子的含义？请指明是哪种信号。
5. 是否有一种信号可以帮助您更好地理解句子的结构？请指明是哪种信号。
6. 是否有一种信号更容易理解？请指明是哪种信号。
7. 是否有一种信号让您花更少/更多的精力去理解句子？这是哪种信号？
8. 实验中的信号，您更喜欢听哪一种？
9. 用这些信号学习英语，您有什么看法？
10. 关于这些信号，您还想说些什么？

APPENDIX D

IOC analysis for the guiding questions in a semi-structured interview

Item	Expert A	Expert B	Expert C	Expert D	Expert E	Result of analysis
1	+1	+1	+1	+1	+1	√
2	+1	+1	+1	+1	+1	√
3	+1	+1	0	+1	+1	√
4	+1	+1	+1	+1	+1	√
5	+1	+1	+1	+1	+1	√
6	+1	+1	+1	+1	+1	√
7	+1	+1	+1	+1	+1	√
8	+1	+1	+1	+1	+1	√
9	0	0	+1	+1	0	√
10	+1	+1	+1	+1	+1	√
Total	9	9	9	10	9	√

Notes:

1. +1 = the item is congruent with the objective
2. -1 = the item is not congruent with the objective
3. 0 = uncertain about this item

Result of IOC:

(IOC = $\sum R/N$)

Item number: 10

R = 9 + 9 + 9 + 10 + 9 = 46 (Scores from experts)

N = 5 (Number of experts)

IOC = 46/5 = 9.2

Percentage: 9.2/10 × 100% = 92%

The table above shows that the result of IOC analysis is 9.2, and the percentage is 92% that is higher than 80%. Therefore, the items are suitable for adoption in the interview.

CURRICULUM VITAE

Cai Xirui received his Bachelor's degree in English from Xinyang Normal University, China in 2010. He obtained his Master's degree in Linguistics and Applied Linguistics in Foreign Languages from Yunnan Normal University, China in 2013. He is currently working as a lecturer at the Department of Foreign Languages, Kunming Medical University, China. He had substantial experience in teaching English language skills, translation, and academic writing courses.

In 2018, he started his Ph.D. study in the School of Foreign Languages, Institute of Social Technology, Suranaree University of Technology, Thailand. His research interests include Phonetics, Phonology, and Neurolinguistics.