To recognise a spoken word, one has to access the phonological knowledge of this word mentally represented at a lexical (whole–word) level and the knowledge of consonants and vowels encoded at a sublexical level. For Mandarin Chinese, the meaning of a word is decided by not only the combination of consonants and vowels but also lexical tones. In this book, Jinxing Yue investigates the cognitive and neural mechanisms underlying word recognition in native Mandarin speakers, with a particular focus on the role of lexical–level representations.

Chapter 1 reviews relevant studies and concepts. The two main issues addressed by the thesis are also introduced: the influence of lexical–level representation on the processing of sublexical features and the temporal and spatial features of the neural activities in the early phase of word recognition. Chapter 2 presents a study investigating the interaction between the lexical–level representation and the tonal representation at a sublexical level with auditory lexical decision tasks. Chapter 3 describes a study examining how lexical and sublexical representations influence form priming in monosyllabic tonal word–forms with tone contrasts in Mandarin Chinese. Chapter 4 presents an ERP study monitoring the rapid development of new cortical memory traces of a Mandarin pseudo–word. Chapter 5 reports the results of an ERP study exploring neural evidence of access to lexical–level representations in the N1 time window, which is temporally earlier than an MMN time window. In Chapter 6, a revised TRACE model, namely the TRACE–Tone model is described.

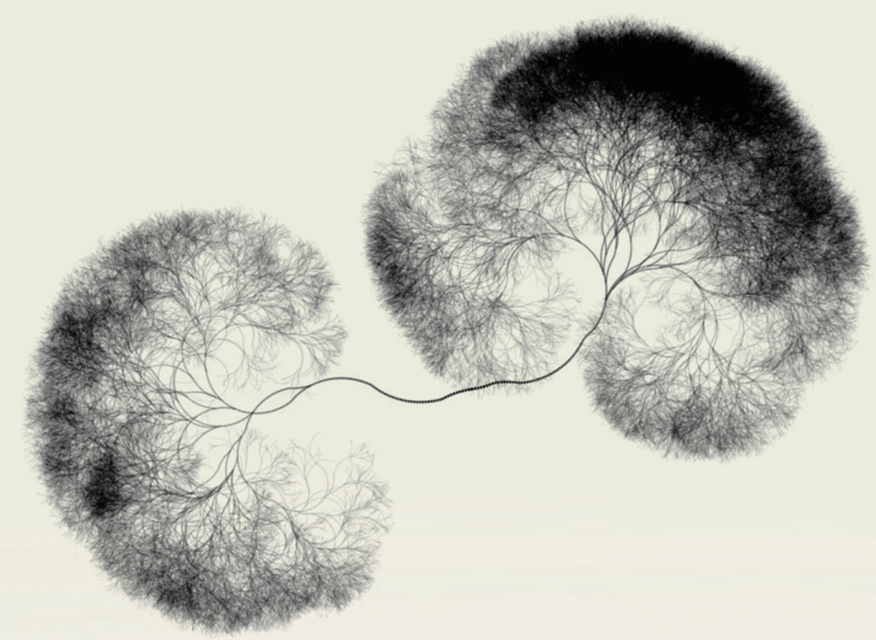
Tone-word Recognition in Mandarin Chinese • Jinxing Yue

143

Tone–word Recognition in Mandarin Chinese

Influences of Lexical–level Representations

Jinxing Yue



Tone-word recognition in Mandarin Chinese

Influences of lexical-level representations

by

Jinxing Yue

The research reported in this thesis has been carried out under the auspices of the Erasmus Mundus joint PhD programme International Doctorate for Experimental Approaches to Language and Brain (IDEALAB), of Universities of Groningen (NL), Newcastle upon Tyne (UK), Potsdam (DE), Trento (IT), and Macquarie University, Sydney (AU), and the generous Erasmus Mundus Joint Doctorate Fellowship issued by the European Commission (grant no. 2012-1713/001-001-EMII EMJD). Publication of the thesis was financially supported by the University of Groningen.

Groningen Dissertations in Linguistics 143

ISSN: 0928-0030

ISBN: 978-90-367-8434-4 (Printed version)

ISBN: 978-90-367-8433-7 (Digital version)

Cover design based on ‘Neuron Fractal’, an art work by amhillman

Downloaded from http://www.contextfreeart.org/gallery/view.php?id=1507

Printed by Fande Printing Service (泛德印务), Harbin, China

Tone-word recognition in Mandarin Chinese

Influences of lexical-level representations

PhD Thesis

to obtain the the joint degree of PhD at the University of Groningen, Newcastle University upon Tyne, the University of Potsdam, the University of Trento, and Macquarie University

on the authority of

the Rector Magnificus of the University of Groningen, Prof. E. Sterken, the Pro-Vice-Chancellor of the University of Newcastle upon Tyne, Prof. S. Cholerton, the President of the University of Potsdam, Prof. O. Günther, the Rector of the University of Trento, Prof. P. Collini, and the Deputy Vice Chancellor of Macquairie University, Prof. S. Pretorius and in accordance with the decision by the College of Deans of the University of Groningen.

This thesis will be defended in public on

Thursday 14 January 2016 at 12.45 hours

by

Jinxing Yue

Supervisor

Prof. Y.R.M. Bastiaanse

Prof. D. Howard

Co-supervisor

Dr. K. Alter

Assessment Committee

Prof. V.J.J.P. van Heuven

Prof. J.C.J. Hoeks

Prof. B.A.M. Maassen

Prof. Y. Shtyrov

Acknowledgements

First and foremost, I would like to thank Roelien Bastiaanse, Kai Alter, and David Howard for their supervision, help, and guidance throughout my joint PhD study in Groningen and Newcastle.

I would like to express special thanks to Roelien. She led me into studying neurolinguistics when I just began my master's courses in Groningen five years ago. During my IDEALAB study, she was not only my promotor, but also my daily supervisor in Groningen. It means that she spent a lot of time with me and gave me so much care. The weekly meetings with her greatly motivated and inspired me to progress; the resources and help she provided immensely facilitated my work. She read all my writings quickly and thoroughly and gave so many priceless comments. Without her, my PhD study won't be possible. Thank you Roelien for everything you have done for me!

My heartfelt gratitude also goes to Kai Alter. Kai was my daily supervisor in Newcastle. Thanks to him, I became aware of the magnificent view from Tyne Bridge. I have been more knowledgeable about the neuroscience of speech under his supervision. He also generously provided extra money to reward each participant of my first ERP experiment from his own budget. It was also him who gave helpful comments and insightful criticisms on all my drafts. I will never forget those days when I was working in his lab.

I want to express my great appreciation to David Howard. His contribution guaranteed that I could spend minimal energy on the administrative issues. I am very grateful to have such an experienced supervisor like him as every time I went in his office with confusions about my research, I could get answers when I came out.

My thanks go to my reading committee: John Hoeks, Vincent van Heuven, Ben Maassen, and Yury Shtyrov. I am greatly benefitted from their comments, corrections and revisions. This book also becomes stronger because of your contributions. My special thanks go to Ben and Yury for those inspiring discussions and encouraging personal communications during my PhD study.

Studying in the IDEALAB is the most valuable experience in my life. The IDEALAB is an international and warm family. I am deeply thankful to Ria de Bleser, Barbara Höhle, Gabriele Miceli, and Lyndsey Nickels for setting up instructive training courses and providing comfortable accommodations at Potsdam, Rovereto, and Sydney.

I am also extremely grateful to Alice Pomstra, Anja Papke, and Helena Trompelt for helping out during my IDEALAB study. I especially appreciate Alice's help during my graduation stage.

I am feeling indebted to all my colleagues in the IDEALAB programme. We 'reunited' every six months in a different country. I am so grateful for having you under the same umbrella, sharing our knowledge, cooking, and travelling around the world.

I wish to express gratitude to my colleagues and friends in Newcastle. I am so grateful to have all you wonderful people around me. Thank you all for sharing you ups and downs, assisting my experiment, hosting me to attend your journal clubs and group meetings, sharing your baking (the unforgettable ION bake-off) and tea in the 'brew room'.

I am indebted to my colleagues in Groningen. You created a friendly, interactive, and encouraging atmosphere for me, which ensured that I could be ‘drown’ in my project and still felt highly supported! My thanks also go to my paranimfen, Rimke Groenewold and Srdjan Popov. You helped me so much and still asked for ‘tasks’. My defence will be perfect with your accompany!

I thank Kelly Kallahan, Cecilia Devers, Xiaosong Rui, and Wenxing Zhang for proofreading my thesis. I thank Fei Wen, Jianhua Zhang, and Yuwei Bu for their assistance in data collection. I thank Yong Yang for helping with graph making.

In the end, my thanks go to my wife, my parents, and my parents in-low. Without your infinite love, I wouldn’t be able to spend enough time working on this book. This book is dedicated to my son who was born in the last year of PhD study. He is a wonder, making me reflect on my life in the past and look ahead.

Studies reported in this book were funded by Erasmus Mundus Joint Doctorate Fellowship issued by the European Commission under a grand number <grant no. 2012-1713/001-001-EMII EMJD>, Newcastle University, and a grant from Ministry of Education of P. R. China (12YJCZH262). The publication of this book is funded by the University of Groningen.

Contents

Acknowledgements .... **v**

Contents .... **ix**

Chapter 1 Introduction **19**

1.1. The roles of lexical-level representations in word recognition 20

1.2. Segmental and suprasegmental cues 22

1.3. Linguistic features of Mandarin Chinese 24

1.4. Cognitive and neural bases for tone-word recognition 26

1.4.1. Separate lexical-tone and segment perception 26

1.4.2. The role of lexical tones in word recognition 28

1.5. Models of tone-word recognition 31

1.6. The present thesis 33

Chapter 2 Interactions between Lexical Tone and Lexical-level Representations: Evidence from Lexical Decision Tasks **37**

2.1. Introduction 37

2.1.1. Spoken word recognition in Mandarin Chinese 39

2.1.2. Lexical tones in Mandarin Chinese 40
2.1.3. Lexical decisions in real words and pseudo-words 41
2.1.4. The present study 42
2.2. Method 43
2.2.1. Participants 44
2.2.2. Materials 44
2.2.3. Procedure and Apparatus 45
2.2.4. Data analysis 47
2.3. Results 48
2.3.1. Reaction times 48
2.3.2. Accuracy 50
2.4. Discussion 52
2.4.1. The influence of lexical-level representations on tone-word recognition reflected by RTs 52
2.4.2. Accuracy and perception of tonal features 54
2.4.3. Semantic and phonological contents at the lexical level 57
2.5. Conclusion 58

Chapter 3 Form Priming in Tonally Contrasted Word Forms With and Without Lexical-level Phonological Representations in Mandarin Chinese 59
3.1. Introduction 59
3.1.1. Form priming in spoken word recognition 60
3.1.2. Lexical access in Mandarin Chinese 62
3.1.3. The present study 65
3.2. Method 66
3.2.1. Participants 66
3.2.2. Materials 67
3.2.3. Procedure 68
3.2.4. Data analysis 69
3.3. Results 70
3.3.1. Reaction times 70
3.3.2. Correct percentages 74
3.4. Discussion 75
3.4.1. Form priming in real-word targets 76
3.4.2. Form priming in pseudo-word targets 78

3.4.3. Theoretical implications 79
3.5. Conclusion.. 81

Chapter 4 Representing Segment-tone Connections in the Human Cortex: Evidence from the Rapid Hebbian Learning of Novel Tone Words 83
4.1. Introduction 83
4.1.1. Representations of segments and lexical tones 85
4.1.2. Integrating novel word forms into the mental lexicon rapidly: hippocampus-based versus neocortex-based theory . 86
4.1.3. The electrophysiological indicator of long-term memory traces for words: the MMN 87
4.1.4. The present study 89
4.2. Method 91
4.2.1. Participants 91
4.2.2. Stimuli 91
4.2.3. Procedure... 95
4.2.4. EEG data acquisition 95
4.2.5. Processing of EEG data 96
4.3. Results 98
4.3.1. Behavioural results 98
4.3.2. ERP results 98
4.4. Discussion 106
4.4.1. Long-term memory traces for the novel word form reflected by the enhanced MMN 106
4.4.2. Neocortex-based rapid Hebbian learning 109
4.4.3. Effects in the N1 time window................................. 111
4.4.4. Limitations 113
4.5. Conclusion... 114

Chapter 5 Early Access to Lexical-level Representations of Tonal Word Forms: An Auditory Habituation Study 115
5.1. Introduction 115
5.1.1. The phonological representation of tonal word forms in Chinese 117
5.1.2. Habituation of AEPs... 120
5.1.3. The present study 123

5.2. Method 124
5.2.1. Participants 124
5.2.2. Stimuli 125
5.2.3. EEG data acquisition 128
5.2.4. EEG data processing and analysis 129
5.3. Results 131
5.3.1. Habituation of the N1 132
5.3.2. Habituation of the P2 141
5.4. Discussion 142
5.4.1. The habituation of N1 and P2 142
5.4.2. Lexicality modulation of the spoken-word-related N1 habituation 144
5.4.3. The temporal features of lexical access reflected by N1 habituation 145
5.4.4 The spatial feature of lexical access reflected by N1 habituation 146
5.4.5. Position-related lexical effects on N1 habituation...... 147
5.4.6. P2 habituation 148
5.4.7. Potential applications 149
5.5. Conclusion 150

Chapter 6 Discussion and Conclusion **.. 151**
6.1. General discussion 151
6.1.1. How do lexical-level representations interact with sublexical representations? 151
6.1.2 What is the neural mechanism behind access to lexical-level representations? 154
6.2. Theoretical considerations: TRACE-Tone model 155
6.2.1. The three-level structure 157
6.2.2. The operation of the TRACE-Tone model 159
6.3. Limitation and future directions 160
6.4. Conclusion 162

Summary 163

Samenvatting 167

Appendix **173**
A Appendix to Chapter 2: Experimental materials 174
B Appendix to Chapter 3: Experimental materials...................... 175

References **181**

Groningen dissertations in linguistics (grodil)............................ **203**

List of Figures

Figure 1.1 The f0 contours of the four tones.. 25

Figure 2.1 Schematic demonstrations of f0 contours. 46

Figure 2.1 Mean RTs of lexical decisions. 51

Figure 2.2 Mean accuracy rates of lexical decisions. 52

Figure 3.1 Average reaction times in the real-word targets. 72

Figure 3.2 Average reaction times in the pseudo-word targets. ..74

Figure 4.1 The design and normalisation of materials.. 94

Figure 4.2 ERPs recorded in at FCz.. 99

Figure 4.3 ERPs to the standards and deviants..........................100

Figure 4.4 MMNs deviant stimuli 101

Figure 4.5 Statistical analyses on the mean values of MMNs...103

Figure 4.6 Topographic maps of MMNs 104

Figure 5.1 A stimulation train of the habituation design. 127

Figure 5.2 The oscillograms and spectrograms 128
Figure 5.3 ERPs elicited by the four types of stimuli................137
Figure 5.4 Average habituation coefficient R data....................139
Figure 5.5 The statistical analyses on the N1 habituation 139
Figure 5.6 ERPs elicited by S1 and S4 of the stimuli 140
Figure 6.1 A diagram of the TRACE-Tone model....................156

List of Tables

Table 2.1 Mean RTs and accuracy rates in the LDTs....................49

Table 3.1 Results of conditions with real-word targets 71

Table 3.2 Results of conditions with pseudo-word targets..........73

Table 5.1 Spoken tonal word forms used in the experiment 126

Table 5.2 The peak latencies of N1 of the five presentations....133

Table 5.3 The peak latencies of P2 of the five presentations 134

Table 5.4 The N1 habituation coefficient Rs.............................135

Table 5.5 The P2 habituation coefficient Rs 136

CHAPTER 1

Introduction

In the long history of biological evolution, the faculty of language has become a unique trait of the human brain which distinguishes our species from an enormous amount of others (Hauser, Chomsky, & Fitch, 2002). Speech, the meaningful vocalisations used to convey language by humans (Fitch, 2000), is the primary tool of communication in most cultures. The physical form of speech is acoustic signals composed of various co-occurring frequencies (Fitch, Miller, & Tallal, 1997). Spoken words are recognised by analysing acoustically complex cues at remarkable speed. There has been wide consensus that spoken word recognition is realised by a specialised system (or processor) continuously mapping acoustic input onto multiple levels of representations (see Lieberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Samuel, 2001). These representations can be roughly divided into lexical and sublexical levels (Dahan & Magnuson, 2006). At the sublexical level, the recognition system encodes smaller units of spoken words, such as syllables, phonemes, and phonetic features; at the lexical level, the representations are lexicalised combinations of sublexical units. Apart from the phonological components, lexical representations can also include semantic components (Swingley & Aslin, 2000). The current thesis focuses on the phonological component of lexical representations, which are also referred to as 'whole-word phonological representations'.

1.1. The roles of lexical-level representations in word recognition

The means of access to various levels of representations during spoken word recognition has been a matter of debate, but multiple studies have shown that lexical-level representations can be accessed as a spoken word unfolds, and may even influence the perception of sublexical elements. Supportive evidence has been reported in studies with behavioural tasks, revealing that the perception of sublexical features is facilitated by feedback from lexical-level representations (Samuel, 1981; Connine & Clifton, 1987; Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Newman, Sawusch, & Luce, 1997; Samuel, 1996; Samuel, 2001; Magnuson, McMurry, Tanenhaus, & Aslin, 2003; Norris, McQueen, & Cutler, 2003; Mirman, McClelland, & Holt, 2005). For example, by using a phoneme identification task, Samuel (2001) asked native speakers of English to judge whether words ended with [s] (as in tremendous) or [š] (as in diminish). The results showed that participants reported more items with [s] endings when the identification task was preceded by an adaptation with words ending with [s], whereas participants reported more items with [š] endings when the adaptors ended with [š]. Their finding was straightforward evidence of top-down effects from lexical-level representations in processing via sublexical representations.

These data could be explained by the TRACE model, a speech perception model which allows both bottom-up and top-down processing via a three-layer structure of word representation (McClelland & Elman, 1986). The TRACE model proposed that the bottom layer was a detector of acoustic features; the middle layer stored the more abstract representations of phonemes; and the topmost layer was the lexical level of word representations. Interactions within a layer were competition via inhibitory connections, whereas between-layer connections were excitatory in nature and permitted facilitations.

In line with these psycholinguistic findings, studies recording neural activities with electroencephalography (EEG) and magneto-encephalography (MEG) have reported evidence that, under experimental contexts, the perception of a specific phoneme can lead to rapid access to lexical representations. For example, by using the mismatch negativity (MMN) as the neural indicator of access to lexical representations in the long-term memory, Pulvermüller et al. (2001) found divergent MMN responses between a real Finnish word stimulus and a pseudo-word stimulus around 150 ms after the recognition point, after which a word form can be recognised given the subsequent phoneme(s); the onsets of the phoneme corresponding with the underlined letter in *lakki* and pseudo-word **vakki* are the recognition points. In a more recent study, MacGregor, Pulvermüller, van Casteren, and Shtyrov (2012) used a passive listening task with multiple items derived from English, finding MEG evidence that lexical access could begin as early as 50-80 ms post the recognition point. More importantly, by carefully matching the acoustic features and setting control conditions, these studies revealed that the different electrical or magnetic responses were induced by lexicality contrasts instead of low-level acoustic and/or phonemic contrasts (for more evidence, see Shtyrov, Nikulin, & Pulvermüller, 2010; Shtyrov, 2011; Shtyrov, Kimppa, Pulvermüller, & Kujala, 2011).

Taking the above-mentioned studies together, the existing behavioural and neurophysiological data have provided two important implications for the role of lexical-level representations. First, word representations at a lexical level are dynamically involved in the analysis of acoustic input, which can even influence the perception of sublexical features in a top-down fashion. Second, access to lexical-level representations can occur readily even before attention sets in.

The materials used in most previous studies distinguished word meanings only based on segmental cues, namely consonants and vowels. For example, *pipe* is a meaningful English word but *pite* is not (Shtyrov et al., 2010). In contrast, many languages express distinctive meanings with the same segmental patterns when they are overlaid with different lexical tones, a suprasegmental cue. These languages are called tone languages and include Mandarin Chinese, Thai, and Shona; the words in these languages are called tone words. For a tone word, the meaning is determined not only by the phonemic sequence but also by the lexical tone. However, the role of lexical-level representations during recognition of words carrying lexical tones remains unclear. Thus, the present thesis follows the two indications of studies in non-tone languages to explore the role of lexical-level representations in the recognition of spoken tone words.

1.2. Segmental and suprasegmental cues

A lexical-level representation has been primarily regarded as a conceptualised combination of segmental cues in studies of European languages. These studies can shed limited light on words exploiting rich suprasegmental cues, another type of acoustic device used to distinguish word meanings between identical segment patterns. Suprasegmental cues in a broad sense refer to properties of speech such as timber, loudness, length, and pitch accent, features which cannot be derived from the phonemes (Nooteboom, 1997). Pitch accent is a feature typically represented by the fundamental frequency (f0) of a speech sound, which can be extended over a syllable, a word or even the length of a sentence segment (e.g., interrogative intonation).

For spoken words in most Indo-European languages, suprasegmental cues refer to a specific type of pitch accent: lexical stress, whose variation pattern in the same segment is quite limited. For example, the typical stress patterns of most English words like *abide*, *abnormal*, *situate*, and *operation* are fixed[1]. The poverty of suprasegmental variation may have led to an exclusion of stress patterns in the current psycholinguistic models of spoken word recognition in Indo-European languages (e.g., TRACE

---

[1] The stress patterns can vary in diverse contexts. However, these variations are not recognised as typical stress patterns, nor lead to changes in the lexical semantics.

model, McClelland & Elman, 1986; Neighbourhood Activation Model, Luce & Pisoni, 1998; Cohort Model, Marslen-Wilson, 1987), even though a number of studies have shown that lexical stress indeed plays a role in constraining word recognition in non-tone languages (e.g., van Heuven, 1985; Cutler & Van Donselaar, 2001; Donselaar, Koster, & Cutler, 2005).

The perception of segmental and suprasegmental cues has been found to be related to differential neural networks. The neural network involved in analysing segmental cues is more lateralised to the left hemisphere, whereas the network for processing pitch information in non-tonal languages has been primarily localised in the right hemisphere (e.g., right inferior frontal gyrus), activated together with involvement of the left-hemispheric homologue (Whalen & Liberman, 1987; Liberman & Whalen, 2000; Zatorre, Evans, Meyer, & Gjedde, 1992; Zatorre & Belin, 2001; Friederici & Alter, 2004; Zatorre & Gandour, 2008). Physically, segmental cues are rapidly changing cues with rich temporal information (e.g., voice onset time), whereas suprasegmental cues are slowly varying cues represented by spectral features (Zatorre & Gandour, 2008). One account for such hemispheric lateralisation is that the left and right hemispheres have biased sensitivities to analysing acoustic cues on different timescales (Boemio, Fromm, Braun, & Poeppel, 2005), and eventually develop specified neural representations to respond to more abstract phonological inputs such as phonemes and stress patterns (Zatorre & Gandour, 2008).

It is estimated that 60-70% of the languages in the world (Yip, 2006), including Mandarin, Cantonese, Thai, Vietnamese, and Shona, employ lexical tones to regulate word meanings together with segmental cues. Unlike most Indo-European languages, these tonal languages are comprised of word forms wherein the same segment can have completely different meanings based on contrastive tones. This role of lexical tone in determining word meanings is on a par with the role of phonemes in non-tone languages. Therefore, for tone languages, tones have been described as tonemes[2] in some psycholinguistic studies of Mandarin Chinese (Ye & Connine, 1999; Zhao, Guo, Zhou, & Shu, 2011).The way spoken words carrying lexical tones are recognised has become an intriguing issue. Not surprisingly, existing studies have mainly investigated the characteristics of lexical-tone perception (Ren, Yang, & Li, 2009; Wong, Parsons, Martinez, & Diehl, 2004; Hsieh, Gandour, Wong, & Hutchins, 2001; Gandour et al., 2000) and/or compared the similarity and disparity between the processing of segmental and tonal cues (Luo et al., 2006; Li et al., 2010; Zhao et al., 2011; Sereno & Lee, 2015). However, relatively little is known about the role of lexical-level representations in spoken tone-word recognition. The present thesis attempts to add to the understanding of how lexical-level representations operate in the recognition of monosyllabic words of Mandarin Chinese.

In the following part of this chapter, we will review some most important concepts of Mandarin Chinese and relevant studies on Mandarin word recognition. A few studies on Cantonese and Thai will also be mentioned as a supplement. The review section is followed by an elaboration of the aims and contents of this thesis.

1.3. Linguistic features of Mandarin Chinese

Mandarin Chinese, called *Putonghua* in Mainland China, is an official language in most Chinese speaking countries and regions. A syllable in Mandarin consists of an onset, a rime and a lexical tone. The onset, equivalent to *shengmu* in Chinese, refers to the initial position in the syllable and allows only one consonant but no consonant clusters (e.g., [sk]). The rime, *yunmu* in Chinese, can be a vowel or a diphthong followed by an optional coda which is restricted to one of two nasal consonants, namely [n] and [ŋ]. When the rime is a vowel, the vowel is the nucleus of the syllable. Diphthong rimes include two types: one type has a glide preceding a central vowel (e.g., /ia/; /ua/), whereas the other type has a

[2] Toneme is a linguistic concept, and was originally introduced by linguists around 1940s (Jones, 1944; Pike, 1947).

glide succeeding a central vowel (e.g., /ai/; /ei/). Mandarin has four lexical tones, denoted as Tone1, Tone2, Tone3, and Tone4. Lexical tones are physically determined by the fundamental frequency (f0) of a syllable. The shapes of the f0 contours of each tone are described as follows: Tone 1 is flat (−), Tone 2 is rising (/), Tone 3 is low dipping (V), and Tone 4 is high falling (\) (Duanmu, 2002).

The rime of each syllable must be overlaid with one lexical tone to distinguish the meanings of words having the same segment. A well-known example is the combination of /ma/ with the four tones, yielding four words with completely different meanings: /ma1/[3] (妈) 'mother', /ma2/ (麻) 'hemp', /ma3/(马) 'horse', and /ma4/(骂) 'to curse'. Figure 1.1 illustrates the f0 contours of the four words based on the segment /ma/ produced by a female speaker of Mandarin.

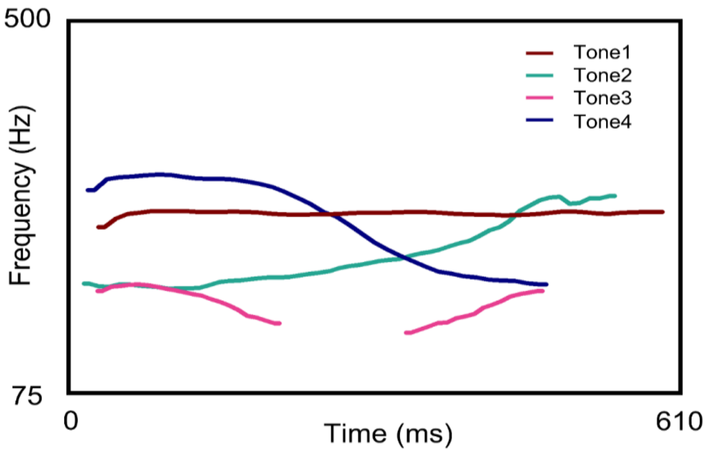


Figure 1.1 The f0 contours of the four tones. The demonstrated words are /ma1/ 'mother', /ma2/ 'hemp', /ma3/ 'horse', and /ma4/ 'to curse'.

[3] A spoken Chinese word form is transcribed in *Pinyin* script, an official Latinised Chinese transcription system, between / /. The tone number is written after the segments making up the syllable.

Syllables with lexical tones are the basic units of the Chinese lexicon. A syllable is a morpheme that is either a word by itself or is part of a polysyllabic word. For example, /chi1/ (吃) means ‘eat’ as an independent word. However, when it is combined with /xiao3/(小), forming the noun /xiao3chi1/(小吃), the meaning is ‘snack’ (literally translated as small eat). Monosyllabic words are the most frequently used word type in modern Chinese (Language Teaching Research Centre of Beijing Language Institute, 1986). Therefore, the current thesis targets monosyllabic spoken words for the initial probe into the role of lexical-level representations in spoken tone-word recognition.

1.4. Cognitive and neural bases for tone-word recognition

Lexical tone processing has been the topic of much debate by researchers of tone-word recognition. An overview of previous studies has led to two important indications. First, lexical tones and segmental cues can be processed separately via differential cognitive and neural mechanisms. Second, the processing of lexical tones and segmental cues is subject to top-down regulation of lexical-level representations.

1.4.1. Separate lexical-tone and segment perception

A body of literature has documented evidence that segmental cues (i.e., onset and rime) and lexical tone are recognised via separate cognitive and neural mechanisms, even though lexical tones are physically entangled with rimes. For example, Liang and van Heuven (2004) reported a case study of an aphasic speaker of Chinese with left-hemispheric lesions. They found selective impairment of lexical tone identification with spared production and perception of vowels. This case was strong evidence that the segmental and tonal aspects of Chinese words are processed separately via mechanisms lateralised to the left hemisphere. In an electrophysiological study, Luo et al. (2006) used Mismatch Negativity (MMN) to indicate the brain’s responses to consonants and lexical tones of Mandarin words in the pre-lexical processing stage. The MMN is an event-related brain potential (ERP) component, elicited by infrequent auditory events ('deviant' stimuli) presented intermittently among other, more frequently presented auditory events ('standard' stimuli).The MMN can reflect the neural activities involved in discriminating the acoustic difference between deviant and standard stimuli (Näätänen, 1995). Luo and colleagues found that MMNs in lexical-tone contrasts were more pronounced at the right-hemispheric recording electrodes relative to MMNs to consonant contrasts, which were stronger over the left-hemispheric recording sites. Moreover, the tone-related MMNs had earlier latencies than consonant-related MMNs. These findings suggested that lexical tones and segmental cues in Mandarin Chinese are processed differently in terms of their time course and the underlying neural anatomy.

In line with this right-hemispheric dominance in tone-related MMNs, one study with functional magnetic resonance imaging (fMRI) in an auditory immediate recognition task also revealed more prominent tone-related activation over the right hemisphere (Li et al., 2010). The researchers found that, although monitoring lexical-tone contrasts and segmental contrasts (i.e., onset / rime) in a word sequence led to activations in largely overlapped cortical regions in both hemispheres, stronger activation in the frontoparietal areas of the right hemisphere was identified in direct comparisons between tone-contrast monitoring and consonant-contrast monitoring, and between tone-contrast monitoring and rime-contrast monitoring. These findings indicated that lexical-tone perception is supported by cortical regions distributed over the left and right hemispheres, but it is the right-hemispheric neural response that distinguishes lexical-tone perception from segmental cue perception.

However, a few studies have shown that the right-hemispheric response to lexical tone is due to the perception of acoustic features rather than the activation of tonal representations with linguistic specifications. For example, the functional imaging data reported by Gandour et al. (2000) in a study with Positron Emission Tomography (PET), showed that in judging lexical tones of Thai words, the cortical activation in native speakers was lateralised to the left hemisphere, whereas the activation in both native Chinese (tone language) and English (non-tone language) speakers was dominant in the right hemisphere (for similar results, see Hsieh et al., 2001; Wong et al., 2004). These results suggest that lexical specification of tones in the left hemisphere hinges on linguistic experience, but the right hemisphere may be responsible for capturing acoustic cues which are slowly varying (Zatorre & Gandour, 2008).

This functional division of the two hemispheres in processing lexical tone is also supported by a PET study (Wong et al., 2004). The researchers found that the left insula was most active when native Chinese-speaking participants performed Chinese tone discrimination. However, when the stimuli were English words superimposed with Chinese tones, the activation of the right insula was similar in native English-speaking participants and native Chinese speakers. Moreover, native English speakers exhibited the same response pattern in the right hemisphere when listening to Chinese tone words. These findings suggest that the way a lexical tone is perceived is highly influenced by higher order auditory representations (Griffiths & Warren, 2002).

1.4.2. The role of lexical tones in word recognition

Whether lexical tones are processed in the same fashion as segmental cues is highly controversial. A number of studies with behavioural measurements have repeatedly shown that the processing of segmental cues has a primacy over the processing of lexical tone. Lexical tones are processed more slowly and less accurately than segmental cues (Repp & Lin, 1990; Experiments 2 and 3 in Cutler & Chen, 1997; Experiments 1 and 2 in Ye & Connine, 1999; Yip, 2001; Mattys, White, & Melhorn, 2005). For example, in a same-different judgment task with monosyllabic spoken words in Cantonese, Cutler and Chen (1997) found that both native Cantonese speakers and speakers without Cantonese knowledge (native Dutch speakers) were prone to more errors and longer response times when two successively presented words contrasted in tones relative to onset and rime contrasts. More recently, by using the ERP technique with high temporal resolution, several studies have found neural evidence that tonal information is processed later than segmental cues (Hu, Gao, Ma, & Yao, 2012 with Chinese idioms; Li, Wang, & Yang, 2014 with Chinese poems). For example, Li et al. (2014) asked participants to judge whether the last word in one line of a poem matched the rime and tone of the last word in the previous line as strictly required by the rhyming rules in ancient Chinese poetry. They found that compared to the congruent condition, rime violations, regardless of tone matching, elicited stronger negative-going ERPs time-locked to 300-500 ms post stimulus onset, but tone violations alone only triggered larger positive-going ERPs between 600-1000 ms.

Contradicting the above results, however, a few recent studies reported that during recognition of spoken tone-words, the time course of lexical tone processing did not deviate from that of segmental cue processing (Brown-Schmidt & Canseco-Gonzalez, 2004; Schirmer et al., 2005; Lee, 2007; Malins & Joanisse, 2010; Zhao et al., 2011; Malins & Joanisse, 2012). Zhao et al. (2011), for example, measured the ERPs time-locked to the onset of monosyllabic spoken words in Mandarin. These words were preceded by the presentation of a picture tagged with a corresponding written word (e.g., /bi2/ ‘nose’ and the picture of ‘nose’ tagged with the Chinese character 鼻 ‘nose’ in the congruent condition). By manipulating the phonological mismatches between the spoken word and the Chinese word represented by the picture, they found no differential N400 effects, a neural indicator of lexical access difficulty (Kutas & Hillyard, 1984; Kutas & Federmeier, 2011), amongst the tone, rime, and onset mismatch conditions in terms of amplitude and latency. Moreover, they identified earlier and stronger N400 responses in the whole-syllable mismatch condition (e.g., /bi2/ vs. /ge1/) relative to the other three partial mismatch conditions (e.g., tone mismatch /bi2/ vs. /bi3/, rime mismatch /bi2/ vs. /bo2/, and onset mismatch /bi2/ vs. /li2/). Their findings suggested that lexical tones can constrain tone-word recognition in a similar way as segmental cues. Thus, segmental cues do not have a processing primacy over tonal cues. More interestingly, this study indicated that in the access to lexical-level representations, a syllable is subject to holistic processing, in which the sublexical features are no longer recognised separately but integrated into one unit which can be mapped onto the lexical entries at the lexical level. These results implied top-down influence of lexical-level representations in the perception of sublexical features in tonal words.

The varying processing weights of lexical tones in different studies have indicated that lexical tones are processed differently under various contexts (Ye & Connine, 1999; Liu & Samuel, 2007). Studies reporting a segment processing primacy usually employed tasks requiring participants to focus on the phonological aspects of stimuli (e.g., same-different judgment task in Cutler & Chen, 1997; rhyme judgment task in Li et al., 2014). In contrast, those studies reporting comparable roles of segments and tones often adopted tasks requiring analyses at a semantic level (e.g., sentence congruence judgment task in Schirmer et al., 2005; cross-modal word matching task in Zhao et al., 2011). Focus on the phonological aspects may elicit processing via sublexical representations and therefore are more likely to reflect the bottom-up processing mechanism (Repp & Lin, 1990; Shuai & Gong, 2014). However, tasks requiring semantic-level analyses may give rise to the activation of lexical representations, leading to a prominent top-down influence on lexical-tone perception (e.g., Schirmer et al., 2005; Zhao et al., 2011).

Furthermore, it should be reiterated that the MMN responses to mismatched lexical tones occurred earlier than the MMNs to mismatched onset consonants as reported in Luo et al. (2006). This result is different from those studies that reported either earlier processing of segmental cues or similar speeds of the processing of segmental and tonal cues (e.g., Hu et al., 2012; Schirmer et al., 2005). The contradiction may be induced by differences in the experimental settings. First, in Luo et al. (2006), the participants did not need to perform any overt tasks that were related to the auditory stimuli, whereas in most previous studies, the participants had to perform some judgment either on the semantic or the phonological aspects. That is, in studies like the Luo et al. (2006), the brain may respond differently to a tone word compared to those studies with an overt task. Second, Luo and colleagues focused on the neural responses in an early time window centring around 200 ms post stimulus onset, whereas the studies with explicit tasks focused on later time windows between 400 ms and 600 ms after stimulus onset. It may be concluded that the disparities in those studies imply varying processing primacies of different aspects of tone words as a tone word unfolds.

1.5. Models of tone-word recognition

Attempts to model the recognition of monosyllabic Mandarin words have emerged in the last decade and a half. To date, the two well-documented models proposed in Ye & Connine (1999) and Zhao et al. (2011) are both based on the TRACE model (McClelland & Elman, 1986). Both models agreed that a spoken tone word was recognised via representations organised in a high-and-low structure. At the low level, representations were sublexical in nature. The bottom representational layer was a system identifying acoustic features of speech sounds (i.e., the rapidly changing temporal information and the slowly varying spectral information). Above the bottom layer was the representation of more abstract phonemes and tonemes. At the high representational level, lexical representations were composed of existing segment-tone patterns used as the forms of words.

Differing from Ye & Connine's (1999) model, Zhao et al. (2011) proposed that the lexical representation could be further divided into two layers: a morphemic layer storing the phonological patterns of segment-tone combinations, and a semantic layer encoding the conceptual components of words. Moreover, on a par with the original TRACE model, the two models describe excitatory interconnections between the lexical layer and the phonemic-tonemic layer, and mutually inhibitory connections within the same layer. The between-layer excitatory interconnections allowed bi-directional information flow, and thus made top-down influence of lexical-level representations on sublexical feature perception possible. Moreover, such structure also permitted direct access to lexical-level representations by an input of sublexical features.

Whilst the above models described the representational structure of tone words in Mandarin, two studies have also tapped into the neural mechanism of tone-word recognition. Luo et al. (2011) proposed a two-stage model, operating in a serial fashion. In the initial pre-attentive stage, the brain processed the speech signal of a spoken tone word as general acoustic cues; in the subsequent attentive stage, the processing occurred via a left-lateralised neural network based on linguistic experience. Luo et al.'s model provided an account for their data that in the pre-attentive processing stage, tone-related brain responses (MMN) were greater in the right hemisphere, but consonant-related responses were more pronounced in the left hemisphere. That is, the right hemisphere was sensitive to slowly varying acoustic cues (i.e., lexical tones), and the left hemisphere was responsible for capturing rapidly changing cues (i.e., segments) (Zatorre & Belin, 2001; Zatorre, Belin, & Penhune, 2002). However, this model did not predict any interaction between general acoustic processing and phonological processing.

In a more recent study, Shuai and Gong (2014) proposed a three-stage model of lexical-tone perception. A special focus of this model was how phonological knowledge affected the perception of acoustic cues at the acoustic-feature level. The three stages were marked by three auditory ERP responses, namely N1, P2, and N400. N1 and P2 are two auditory evoked brain potentials. The N1 is the first prominent negative deflection, usually centring around 100 ms post stimulus onset. Succeeding the N1, the P2 is a prominent positive-going component, peaking around 200 ms post stimulus onset. In the first stage, the processing of the initial phoneme begins about 100 ms after the onset of a tone word. Lexical tones were far from being recognised in such a short time window. Nonetheless, the limited input could be used to predict general syllable patterns, and lead to top-down influence which could be strengthened by contextual information. In the second stage (100-300 ms), the top-down influence was further enhanced with the accumulation of input cues, whilst the bottom-up processing of lexical tones and segmental cues was going on, as reflected by the P2. In the third stage (300-500 ms), top-down processing due to the access to semantic information became the main stream, along with some bottom-up processing, as indexed by the N400. Although this model elegantly accounts for the time course of tone-word processing in accordance with the time windows of the neurophysiological indices, the authors only considered two sublexical layers of representations, namely the acoustic analysis and the phonological representation, without any clear consideration of representations at the lexical level.

1.6. The present thesis

To advance the investigation of tone-word recognition, the present thesis aims to underpin the role of lexical-level representations by looking for answers to two general questions:

1) How do lexical-level representations interact with sublexical representations of lexical tones and segments?

2) What is the neural mechanism behind the activation of lexical-level representations?

The basic rationale is to investigate the way lexical-level representations affect tone-word recognition by comparing the responses to real words and phonotactically possible pseudo tone word forms. In the current thesis, all tonal pseudo-words are derived from legal segment patterns and tones, but the combinations of the segments and the lexical tones are meaningless in Mandarin. For example, the combination of /na/ and Tone1, */na1/ is a non-existent segment-tone pattern in Mandarin Chinese, even though /na/ is a legal segment pattern in the combinations with Tone2 (/na2/(拿) 'hold'), Tone3 (/na3/(哪) 'where') and Tone4 (/na4/(那) 'there'). We label this kind of pseudo-word a tone-manipulated pseudo-word. Compared with pseudo-words derived from non-existent combinations of onsets and rimes like */tua1/ (in Shuai & Gong, 2014), tone-manipulated pseudo-words can induce no segmental processing difficulties at the sublexical level, but still lack lexical-level representations. Therefore, such pseudo-words are an ideal tool for revealing top-down effects yielded by lexical-level representations.

Four studies are carried out to address the two general research questions, and will be presented in Chapter 2 - 5. The two studies reported in Chapters 2 and 3 are conducted by using auditory lexical decision tasks (LDTs) and a form priming paradigm to measure the behavioural responses in real and pseudo tonal word forms. In Chapter 2, with auditory LDTs, we examine whether the absence of lexical-level representations in pseudo-words leads to differential responses to pseudo-words in three tonal categories (Tone1, Tone2, and Tone4), and whether this response pattern is distinct from the one in real words of the three tones. Reaction time and accuracy data will be used to indicate the underlying processes (Goldinger, 1996). Two versions of the auditory LDT will be used to detect whether the semantic and phonological components of lexical-level representations influence top-down processing in the same way. Since the perception of the three lexical tones are different in nature (Wu & Shu, 2003; Lai & Zhang, 2008), we expect distinctive response patterns to pseudo-words in the three tonal categories rather than to real words because lexical decisions of real words should be sensitive to lexical factors instead of sublexical features. Moreover, if both semantic and phonological representations are stored at the same lexical level, we do not expect distinctive response patterns due to the different task instructions.

In Chapter 3, with a form priming paradigm, we systematically check the priming effects in lexical decisions of tonal word forms with ten types of prime-target pairs. In form-priming experiments, researchers usually manipulate the lexical decision performance in target words preceded by various types of prime words (Zwitserlood, 1996). The priming effects are characterised by facilitation and inhibition. Facilitatory priming is usually shown by faster responses, and inhibitory priming is reflected by slower reaction times, compared to proper baselines. The primary goal of this study is to reveal the priming effects in paired word forms sharing the same segment but with distinctive tones (word : word, pseudo-word : word, word : pseudo-word, pseudo-word : pseudo-word, in which the initial word form is a prime, followed by a target). These materials allow us to detect the interaction between lexical and sublexical access in tone-word recognition as reflected by facilitatory priming and inhibitory priming effects.

Although behavioural measurements such as lexical decisions give valuable insight into the organisational structure of the mental lexicon of tone words, such methods cannot faithfully reveal online processes because of the involvement of post-lexical processing and response planning during task completion (Balota & Chumbley, 1984). Therefore, in Chapters 4 and 5, we employ ERPs, a neurophysiological technique indexing post-synaptic activities in large bundles of neurons aligned in a parallel orientation, related with specific stimulus events (Coles & Rugg, 1995). This technique provides high temporal resolution at millisecond-level accuracy and allows researchers to investigate the hemispheric lateralisation of brain functions in response to auditory stimuli (Alho et al., 1998; Rinne et al., 1999; Shtyrov et al., 2010).

In Chapter 4, we investigate the time course of access to lexical-level representations in tone-word recognition by referring to the MMNs elicited in an oddball experiment. The speech-related MMN, usually identified between 150-250 ms post stimulus onset, has been used as an ERP indicator of auditory discrimination and automatic access to the long-term memory traces of spoken words (for reviews, see Pulvermüller & Shtyrov, 2006; Shtyrov & Pulvermüller, 2007). Moreover, the MMN over the left-hemispheric region has been linked with specified neural networks for language processing (Alho et al., 1998; Rinne et al., 1999; Näätänen et al., 1999; Koyama et al., 2000). Following a protocol of rapid word learning (Shtyrov et al., 2010), we compare the MMN responses elicited in real words and pseudo-words, repeatedly presented during perceptual training. If lexical-level representations can modulate MMN responses, the MMNs in the final phase of perceptual training should be enhanced relative to the MMNs in the early training period. Moreover, we expect the MMNs to the trained pseudo-word to resemble the MMNs of the real word. In addition, the expected learning effects should be over the left-hemispheric recording sites.

In Chapter 5, we explore whether lexical representations can be accessed in the pre-MMN time windows, namely the N1 and P2 time windows. The N1 and P2 are known to be highly sensitive to the physical features of stimuli, and thus, it is difficult to explore top-down processing in these time windows by referring to the two ERP components. To overcome this issue, we adopt a habituation paradigm which has been used to identify disparities between the N1 responses to auditory stimuli with high and low levels of representations, namely speech and non-speech sounds (Woods & Elmasian, 1986; Teismann et al., 2004). Based on these studies, we expect to observe differential auditory habituation patterns in real words and pseudo-words.

The following four chapters (Chapters 2 to 5) will report our studies in detail. The ‘Discussion and Conclusion’ in Chapter 6 will summarise and discuss the general findings of the thesis.

CHAPTER 2

Interactions between Lexical Tone and Lexical-level Representations: Evidence from Lexical Decision Tasks

2.1. Introduction

The human brain must process two types of acoustic cues to recognise a spoken word: segmental and suprasegmental cues. Segmental cues refer to information that can be used to identify the boundaries of phonemes and/or phonemic combinations as lexical forms unfold. Therefore, segmental cues are mainly related to phonemes. Suprasegmental cues in spoken words refer to acoustic features that can be extended over a syllable or even over several words. In linguistics, suprasegmental cues include stress patterns, intonation and lexical tones. In words of most Indo-European languages, the use of suprasegmental cues is restricted to stress patterns, and thus, the recognition of a spoken word form mainly relies on mapping the acoustic input onto entries of lexicalised combinations of segmental units (i.e., phonemes), also called segment patterns.

Current linguistic models have focused on how segmental features are represented and accessed in speech perception without taking suprasegmental cues into consideration (TRACE model in McClelland & Elman, 1986; Neighbourhood Activation Model in Luce & Pisoni, 1998; Cohort Model in Marslen-Wilson, 1987; Shortlist Model in Norris, 1994). These models agree that phonological knowledge is stored at two separate levels of mental representations, namely the lexical level and the sublexical level (Dahan & Magnuson, 2006). Representations at the sublexical level encode phonemic knowledge (e.g., consonants and vowels), whereas the lexical level consists of representational units in the form of whole words. For example, the combination of the sublexical features [k] [ʌ] [p] yields a lexical entry, cup, which is a representational unit at the lexical level.

Lexical-level representations and sublexical representations are interactively accessed during spoken word recognition. For example, a series of studies have shown that, given sufficient contexts, the input of a phoneme or a syllable can directly access the lexical-level representation leading to word responses (Pulvermüller et al., 2001; Shtyrov et al., 2010; Shtyrov et al., 2011; MacGregor et al., 2012). Furthermore, a considerable number of studies have shown that the existence of lexical-level representations in turn can affect the perception of phonemes (Samuel, 2001; Magnuson et al., 2003; Norris et al., 2003; Mirman et al., 2005). These studies strongly support the TRACE model (McClelland & Elman, 1986) which suggests that lexical and sublexical representations have interconnections allowing both bottom-up and top-down processing.

Unlike most Indo-European languages, approximately 60-70% of languages in the world, including Chinese and Bantu languages, employ systematic variation of suprasegmental cues to distinguish word meanings (Yip, 2006). Lexical tone is an example of such suprasegmental cues. It is not yet completely clear how spoken words are recognised via their mental representations in tonal languages. This study aimed to investigate this issue, with a special focus on how lexical-level representations interact with sublexical representations of lexical tones, by examining participants' lexical decisions of Mandarin word forms carrying different lexical tones.

2.1.1. Spoken word recognition in Mandarin Chinese

Mandarin Chinese is an ideal language for the purpose of the present study. Mandarin, referred to as *Putonghua* in mainland China, is the official language in most Chinese speaking countries and regions. A single syllable in Mandarin can be a word or a morpheme, and consists of an onset, a rime and a lexical tone. The onset only allows one consonant, no consonant clusters. The rime position hosts vowels and their nasalised forms with [n] and [ŋ]. Mandarin has four lexical tones, denoted as Tone1, Tone2, Tone3, and Tone4. Each syllable must be overlaid with one type of lexical tone in order to distinguish the different possible meanings of the same segment. For example, given the segment /ni/, /ni1/ (妮) means 'girl', /ni2/ (泥) means 'mud', /ni3/ (你) means 'you', and /ni4/ (腻) means 'greasy'.

Two models of monosyllabic Mandarin word recognition are based on a division of lexical and sublexical representations, on a par with models for European languages (Ye & Connine, 1999; Zhao et al., 2011). In Mandarin Chinese, the sublexical-level representations are phonemes (onset consonants and rime vowels) and tonemes (Tone1, Tone2, Tone3, and Tone4). Both of the models allow bi-directional processing between the lexical and sublexical levels of representation. That is, in word recognition, the perception of sublexical features can lead to access to lexical-level representations, which in turn can influence processing via representations at the sublexical level. The two models have one major contrast in the structure of the lexical-level representation. Ye and Connine (1999) proposed a word-level representation that includes all lexical information (see Malins, 2013 for a similar proposal), whereas Zhao et al. (2011) held that the lexical-level representation can be further divided into an upper layer of semantic representations, and a lower layer of phonological representations, storing segment-tone patterns (or morphemes). Furthermore, Zhao et al. (2011) argued that the two layers have excitatory interconnections as well.

2.1.2. Lexical tones in Mandarin Chinese

The four lexical tones in Mandarin Chinese have different recognition trajectories due to their peculiar characteristics. Lexical tones are physically determined by the fundamental frequency (f0) of speech sounds. The most important f0 parameters of Mandarin tones are pitch height and pitch contour (Howie, 1976). The contour shapes of the four tones are described as follows: Tone1 is flat (−), Tone2 is rising (/), Tone3 is dipping (\/), and Tone4 is falling (\) (Chao, 1968; Duanmu, 2002). Moreover, the typical onset and offset pitch heights of the four tones can be described in a 5-point scale: Tone1 is 55, Tone2 is 35, Tone3 is 24 and Tone4 is 51. Since 5 is the highest and 1 is the lowest on this scale, Tone1 and Tone4 have similar onset pitch heights which are higher than those of Tone2 and Tone3. Accordingly, Tone1 and Tone4 start with a high register, and Tone2 and Tone3 have low register onsets. Moreover, Tone1, Tone2 and Tone3 have similar shapes for the initial part of their pitch contours (see Figure 1.1 in Chapter 1 for an illustration).

The four tones are not isolated (i.e., correctly recognised) in the same way due to discrepancies in their pitch heights and contour shapes. For example, by using a gating paradigm, Wu and Shu (2003) observed the isolation trajectories in spoken words of four tones gated with a 40-ms increment based on the initial 80-ms proportion (i.e., Gate 1 presented 80-ms of a word, Gate 2 presented 120-ms of a word, Gate 3 presented 160-ms of a word, and so on). Participants were instructed to provide a word according to the acoustic information they heard. By analysing the tonal dimension of their answers in each gate, they found that Tone2 required the longest time to be isolated compared to Tone1, Tone3 and Tone4. Moreover, further analyses on tone confusion revealed that in early gates, Tone1 and Tone4 were mutually confusing. This effect was attributed to the similar onset pitch height of the two tones. Moreover, Tone2 and Tone3 words were more likely to be misidentified as Tone1 words in the initial gates. Considering that Tone1, Tone2, and Tone3 have similarly shaped of pitch contours, this kind of mistake likely reflects a default use of a high register tone with similar contour shapes when the available cue is limited for low register tones.

However, Wu and Shu (2003) did not match the occurrence probability of the segments. That is, some segment patterns occurred more often than the others. Consequently, the results can be confounded because the recurrence of the same segment might induce different processing strategies. To overcome this confounding factor, in another gating study, Lai and Zhang (2008) used eight word quadruplets, each of which consisted of words with the same segment carrying the four tones (i.e., /ma1/, /ma2/, /ma3/, and /ma4/). The researchers found that tones with high onset registers (Tone1 and Tone4) were recognised faster than those with low onset registers (Tone2 and Tone3). Furthermore, Tone1 was recognised faster than Tone4, but Tone2 and Tone3 had similar isolation points. The analyses revealed the same mistake patterns as in Wu and Shu (2003) that Tone1 and Tone4 were mutually confusing, and the onsets of Tone2 and Tone3 were more likely to be misidentified as Tone1, which has a higher onset pitch register but similar initial contour shapes as Tone2 and Tone3.

Although the pitch height and pitch contour have shown to affect lexical tone identification, it remains unclear whether the perception of pitch heights and pitch contours can be influenced by lexical-level representations. To examine this issue, we investigated lexical decision performance in Tone1, Tone2, and Tone4 real words and pseudo-words.

2.1.3. Lexical decisions in real words and pseudo-words

An auditory lexical decision task (LDT) entails timed classification of words and pseudo-words according to the lexical status of a spoken word form (Goldinger, 1996). Reaction time (RT) and accuracy are two dependent variables which have been used to infer the processes of spoken word perception. Higher RTs and lower accuracy rates are usually interpreted as indicating greater processing difficulty. The auditory LDT is believed to reflect lexical access because it is sensitive to word frequency (e.g., Slowiaczek & Pisoni, 1986), lexical competition (e.g., Gaskell & Dumay, 2003) and lexical status (e.g., Mimura, Verfaellie, & Millberg, 1997). The response latencies of real words are usually shorter than those of pseudo-words, and study participants are more likely to judge real words correctly relative to pseudo-words (Radeau et al., 1989).

Moreover, since pseudo-words do not have lexical entries in the mental lexicon, researchers can explore processing mechanisms at the lexical and sublexical levels of representations by comparing participants' behavioural and neural responses to real words and pseudo-words (Radeau et al., 1989; Gaskell & Dumay, 2003; Sumner & Samuel, 2007; Shtyrov, Osswald, & Pulvermüller, 2008; McGettigan et al., 2011). To generate pseudo-words, researchers usually substitute one or more phonemes of a real word, resulting in a phonotactically possible phoneme or phonemic sequence that is not a real word. This kind of pseudo-word has no lexical meaning and does not violate sublexical principles. Therefore, such segment-manipulated pseudo-words can reflect sublexical processing without triggering additional processing costs induced by phonotactically impossible combinations of phonemes (e.g., *[stk] is phonotactically impossible in English).

For the current study of Mandarin Chinese tone words, we created tonal pseudo-word forms by combining a legal segmental string (i.e., a syllable) with a tone in a novel way. For example, /se4/(色) is a lexical combination of /se/ and Tone4 that together means 'colour', whereas /se2/, consisting of /se/ + Tone2, is a pronounceable and phonotactically possible but meaningless combination in standard Mandarin Chinese. The most important characteristic of this kind of pseudo-word, named 'tone-manipulated pseudo-word', is that it does not involve any illegal segmental and suprasegmental elements, but the combination still lacks lexical representation at the lexical level.

2.1.4. The present study

The primary goal of the present study was to examine how lexical tones interact with lexical-level representations. For this purpose, we recorded the lexical decision performance in real words and pseudo-words carrying

Tone1, Tone2, and Tone4. As previous studies have revealed, the isolation time of the three tones can be ranked as Tone1 < Tone4 < Tone2 (Lai & Zhang, 2008). Moreover, Tone2 is likely to be misidentified as Tone1 in the initial phase of recognition because of their similar onset pitch contours, and Tone1 and Tone4 are mutually confusing due to their comparable onset pitch heights (Lai & Zhang, 2008; Wu & Shu, 2003). By comparing the lexical decision performance between Tone1, Tone2 and Tone4 in reals words, we expected that the lexical decisions on real words are sensitive to lexical factors rather than tonal features because of the top-down influence on tone perception. That is, when the lexical factor was controlled, we did not expect differences between lexical decisions in real words with Tone1, Tone2, and Tone4. In contrast, the lexical decision of pseudo-words should be influenced by the distinctive resolution speed of the three tones and the shared onset pitch features.

The secondary purpose of this study was to investigate whether a division of phonological and semantic representation levels is necessary for Mandarin word recognition (Zhao et al., 2011). For this purpose, we developed two tasks of lexical decision making. In one lexical decision task, participants were asked to decide whether an auditory word form was a meaningful word (LDT of semantics); in another version of the experiment, participants were instructed to judge whether a word form was an existent sound pattern (LDT of phonology). In order to complete the LDT of semantics, participants needed to access the semantic component of lexical representations, whereas in the LDT of phonology they only needed to rely on the phonological components of the lexical-level representations (*cf.* Poldrack et al., 1999). If the lexical segment-tone patterns were represented in a level below semantics, we would expect differential lexical decision patterns between the two tasks, as measured by reaction times (RT).

2.2. Method

This experiment was a 2×2×3 design with the between-subject factor of Task (LDT of semantics vs. phonology) and two within-subject factors, Lexicality (real word vs. pseudo-word) and Tone (Tone1 vs. Tone2 vs. Tone4). We did not include Tone 3 to avoid the reported mutual confusion between Tone2 and Tone3 based on their acoustic similarities (Wu & Shu, 2003; Lee, Tao, & Bond, 2008).

2.2.1. Participants

Eighty students from Harbin University of Commerce and Harbin Institute of Technology participated in the experiment. They were randomly divided into two equal groups. One group was instructed to perform auditory LDT of semantics (7 males, 33 females; *M* age = 19.8 years), whereas the other group was asked to perform auditory LDT of phonology (8 males, 32 females; *M* age = 22.3 years). They were all native speakers of Mandarin Chinese, as assessed by a questionnaire issued prior to the experiment. Of the 80 students, 75 were born and grew up in Heilongjiang Province and Jilin Province, known for their use of Standard Mandarin Chinese. Five participants were from the provinces of Tianjin City, Shanxi, Henan, and Hebei, and reported being monolingual speakers of Mandarin Chinese. The participants reported no hearing or language disorders.

2.2.2. Materials

The materials were monosyllabic Chinese spoken word forms, consisting of 27 real words (RW), 27 pseudo-words (PW) and 36 foils (27 real words and 9 pseudo-words). See Appendix A for a complete list of real and pseudo-words. Real words and pseudo-words were matched for high phonotactic probabilities (see Storkel, 2001), designated by the word frequency according to data from the Chinese Internet Word Frequency List of the Lancaster Corpus of Mandarin Chinese (McEnery & Xiao, 2004). The phonotactic probability of a pseudo-word is represented by the most frequently used word sharing the same segmental template but carrying different tones. Words appearing more than 100 times per million words were considered to have high phonotactic probability (*cf.* Zhang & Damian, 2009).

The real and pseudo-words consisted of equal numbers of stimuli with each tone (i.e., 9 each of Tone1, Tone2, and Tone3). The words of the three tonal categories were matched for word frequency (F (2, 24) = 1.26, p = .303); pseudo-words were matched for phonotactic probability (F (2, 24) = 0.425, p = .642). Real and pseudo-words were equally divided over three blocks, in each of which there was minimal overlap of onsets and rimes between the experimental stimuli and foils, except for one overlap of the initial onset /n/ for two pseudo-words in the third block.

All materials were produced by a female native Mandarin speaker in a sound proof studio with a U87 microphone, digitized via a Fireface 800 RME sound card, and recorded with Nuendo 6 (Steinburg Media Technology, Germany) at a sampling rate of 44.1 kHz. Each item was articulated 15 times. The initial and last four articulations were excluded from selection. Of the remaining items, one was selected based on two criteria: 1) the item was clearly articulated and easily recognised aurally, and 2) the f0 contour was clear and smooth based on visual inspection. All items were normalised for the same average intensity (65 dB) and duration (650 ms) using the acoustic software programme PRAAT (Boersma & Weenink, 2013). The f0 contours were not subject to normalisation in order to keep the items free from formant distortion. Figure 2.1 illustrates the f0 contours of all real and pseudo-words.

2.2.3. Procedure and Apparatus

Participants were tested individually in a sound attenuated room. The experimental stimuli presentation was programmed with DMDX V4.0.4.6 (Forster & Forster, 2003), and presented with an HP 540 laptop over a Cosonic CD-778MV headphone at 75 percent of the full volume of the laptop. Each participant was randomly assigned to either the LDT of semantics group or LDT of phonology group, and was asked to press ‘O’ for an existent word and ‘P’ for a non-existent word form on the laptop keyboard as quickly and accurately as they could. A practice session with nine trials was administered before the experimental session. A participant was not allowed to proceed to the first experimental block until the feedback showed no more than three erroneous judgments out of nine trials.

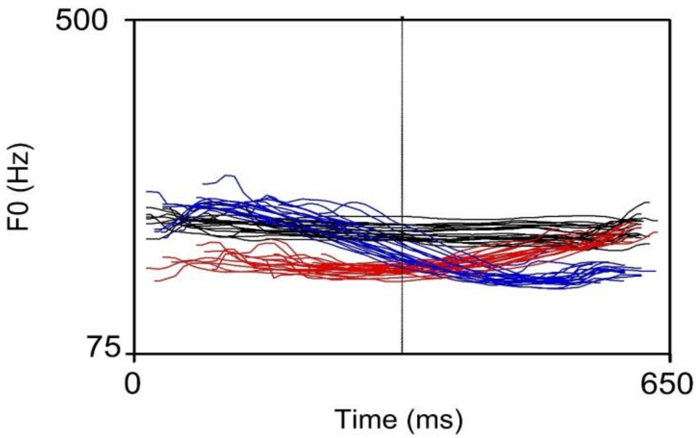


Figure 2.1 Schematic demonstrations of f0 contours of the 54 experimental items. Tone1 words are in black; Tone2 words are in red; Tone4 words are in blue. Note the similar onset pitch heights between Tone1 and Tone4 word forms, and the similar initial pitch contours between Tone1 and Tone2 word forms.

The experiment consisted of 30 trials in each block (9 real words and 9 pseudo-words). At the beginning of each trial, a fixation mark was presented for 1000 ms in the centre of the screen, followed by the aural presentation of a stimulus. After a 50 ms interval from the offset of the stimulus, a question mark was shown in the screen centre, prompting the participant for their judgment. In each block, the presentation of items was pseudo-randomised on an individual basis. The same experimental item type was not presented more than three times in a row. A 1-minute break was given between each block to avoid potential inference from the phonemic overlaps across blocks.

2.2.4. Data analysis

Data were pre-processed separately for the LDT of semantics and LDT of phonology responses. For the LDT of semantics, RTs of erroneous responses, and responses shorter than 50 ms (Luce, 1986) or longer than 3 *SD* (1527 ms) were omitted. The 314 omitted data points (195 erroneous, 21 shorter than 50 ms, 198 longer than 3 *SD*) accounted for 14% of the total observations. The rate was lower than the 15% level advised by Ratcliff (1993) for lexical decision studies. For the LDT of phonology, 183 erroneous data responses, 32 trials shorter than 50 ms, and 99 trials longer than 3 *SD* of 1445 ms were excluded, accounting for 14% of the total trials. The accuracy rates were obtained from the original data without data trimming.

Three-way repeated measures ANOVAs were first performed with the average RTs and accuracy rates by participant and by item separately. The by-item analysis was performed to rule out the possibility that any significant effects in by-subject analysis were caused by some specific items. If interactions between Task, Lexicality and Tone, and between Task and either Lexicality or Tone were identified, further analyses unpacked from the interactions were performed within the two tasks separately. If not, the data of the two tasks were combined for other analyses. Planned comparisons were conducted to identify which factors led to differential lexical decisions among word forms with the three tones. Lastly, we conducted multiple pairwise comparisons (paired *t*-tests) between every two-tone group (Tone1 vs. Tone2, Tone1 vs. Tone4, and Tone1 vs. Tone2) separately for real and pseudo-words with Bonferroni's correction to directly reveal the responses patterns. Greenhouse-Geisser correction was performed when the sphericity assumption in ANOVA was violated. The alpha level was set to $p < .05$. All p values were corrected when Bonferroni's correction or Greenhouse-Geisser correction were applied. The statistical analyses were conducted with SPSS V22 (IBM, Armonk, NY, USA).

2.3. Results

The average RTs and accuracy rates in real words and pseudo-words carrying three types of lexical tones in the two tasks can be seen in Table 2.1. Descriptive statistics showed that real words elicited longer RTs and higher accuracy rates than pseudo-words. Moreover, the LDT of semantics were generally longer than those of phonology.

2.3.1. Reaction times

Three-way ANOVAs identified Task as a marginally significant between-subject effect when the RTs were calculated by participant (F (1, 78) = 3.91, p = .052), and a significant main effect by item (F (1, 16) = 36.42, p < .001). The results suggest that the RTs in the LDT of phonology are shorter than those in the LDT of semantics. However, there is no interaction between Task and the other two within-subject factors, Lexicality and Tone (by subject F (2, 156) = 1.00, p = .370; by item F (2, 32) = .26, p = .775), nor interaction between Task and Tone (by subject F (2, 156) = 1.37, p = .257; by item F (2, 32) = 0.323, p = .726). These results show that the general patterns of RTs are the same in the two tasks. The only effect induced by the tasks is an overall slower response in the LDT of semantics than in the LDT of phonology.

Moreover, there was a main effect of Lexicality (by subject F (1, 78) = 39.49, p < .001; by item F (1, 16) = 35.85, p < .001). This result confirms that real words induce shorter RTs than pseudo-words. There was also a main effect of Tone (by subject F (2, 156) = 5.82, p = .004; by item F (2, 32) = 4.39, p = .021). However, the interaction between Lexicality and Tone was not significant (by subject F (2, 156) = 2.34, p = .100; by item F (2, 32) = 0.994, p = .381). This implies that the RT pattern in real words with Tone1, Tone2, and Tone4 does not differ from the pattern in pseudo-words with the three tones.

Table 2.1 Mean RTs and accuracy rates as functions of the lexicality of stimuli in the LDTs of semantics and phonology separately. (T1=Tone1, T2=Tone2, T3=Tone3, RW = real word, PW = pseudo-word)

	Reaction Time (ms)		Accuracy Rate	
	M	*SD*	*M*	*SD*
LDT of semantics				
T1RW	497.9	219.1	0.98	0.05
T2RW	486.8	208.4	0.96	0.06
T4RW	507.8	198.5	0.91	0.11
T1PW	569.5	251.7	0.81	0.19
T2PW	612.4	289.8	0.92	0.11
T4PW	646.3	263.2	0.91	0.09
LDT of phonology				
T1RW	425.6	160.6	0.96	0.06
T2RW	411.6	141.3	0.95	0.07
T4RW	431.2	142.9	0.92	0.10
T1PW	513.5	225.3	0.83	0.14
T2PW	498.7	224.0	0.95	0.09
T4PW	541.0	205.6	0.90	0.10

Nonetheless, Bonferroni corrected pairwise comparisons between the RTs in every two tone categories in real and pseudo-words still showed that Tone1 pseudo-words were recognised significantly more slowly than Tone4 pseudo-words by subject (t(79) = -3.18, p = .013), and with marginal significance by item (t(17) = -2.94, p = .054). The shorter RTs in Tone2 pseudo-words than Tone4 pseudo-words were a marginally significant effect in the comparison by subject (t(79) = -2.63, p = .06), and a significant effect in the comparison by item (t(17) = -3.35, p = .024). No significant effects were identified in other comparisons. The results are displayed in Figure 2.2.

2.3.2. Accuracy

Three-way repeated measures ANOVAs with the accuracy data showed that the between-subject factor Task was not a main effect (by subject F (1, 78) = 0.14, p = .706; by item F (1, 16) = 0.04, p = .844). This result indicates that accuracy rates of the two versions of LDT do not differ from each other. Moreover, there were no interactions between Task, Lexicality, and Tone (by subject F (2, 156) = 1.40, p = .250; by item F (2, 32) = 0.36, p = .70).

With respect to the within-subject factors, Lexicality was a main effect (by subject F (1, 78) = 32.71, p < .001; by item F (1, 16) = 7.93, p = .012). This result implies that real words have higher accuracy rates compared to pseudo-words. Moreover, a main effect of Tone and an interaction between Lexicality and Tone were identified (Tone: by subject F (2, 156) = 11.50, p < .001; by item F (2, 32) = 3.39, p = .046. Lexicality × Tone: by subject F (2, 156) = 29.77, p < .001; by item F (2, 32) = 7.65, p = .002). Planned comparisons revealed a main effect of Tone in both real words (by subject F (2, 156) = 11.37, p < .001; by item F (2, 32) = 3.77, p = .033) and pseudo-words (by subject F (2, 156) = 24.08, p < .001; by item F (2, 32) = 6.70, p = .004). These results suggest that accuracy rates do not vary with the task instructions, but the accuracy rates in real words with the three tones have different patterns from the accuracy rates in the three tonal categories of pseudo-words.

Multiple comparisons revealed significant differences between Tone1 real words and Tone4 real words (by subject $t(79) = 4.17$, $p < .001$; by item $t(17) = 3.24$, $p = .03$) and between Tone1 pseudo-words and Tone2 pseudo-words (by subject $t(79) = -6.31$, $p < .001$; by item $t(17) = -3.70$, $p = .012$). These results suggest that Tone1 words are more likely to be correctly judged than Tone4 words (see Figure 2.3). However, the LDTs in pseudo-words exhibit a different pattern that show Tone1 pseudo-words are prone to more erroneous judgment than Tone2 pseudo-words.

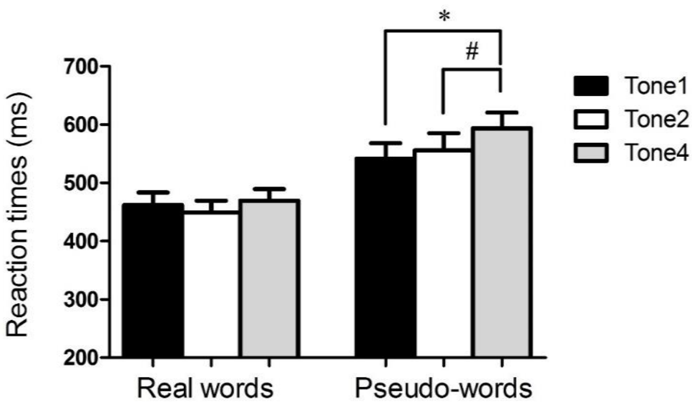


Figure 2.1 Mean RTs of lexical decisions on real words and pseudo-words with Tone1, Tone2, and Tone4. Error bars are standard error means. Note the RT of Tone4 pseudo-words is longer than the RTs of Tone1 and Tone2 pseudo-words. # $p = .06$, * $p < .05$ (significance marked according to analyses by subject).

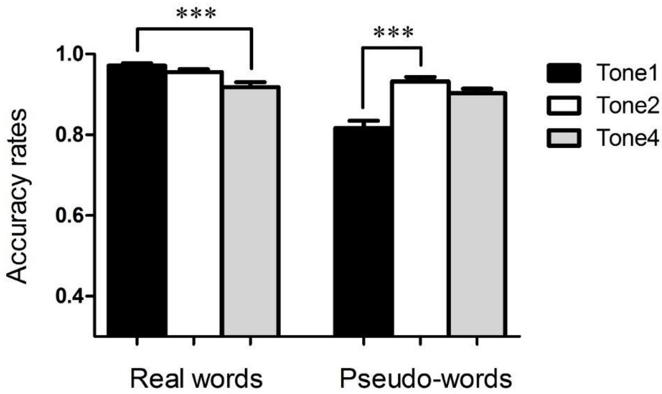


Figure 2.2 Mean accuracy rates of lexical decisions in real words and pseudo-words with Tone1, Tone2, and Tone4. Error bars are standard error means. *** $p < .001$ (significance marked according to analyses by subject).

2.4. Discussion

The present study used two versions of an auditory LDT to examine the interaction between lexical-level representations and lexical tone representations. We compared the RTs and accuracy rates of lexical decisions in three tonal categories of real words and in the same tonal categories of pseudo-words in Mandarin Chinese. Both the RT and accuracy data show that the pattern of decisions between real words with the three tones is distinct from the decision pattern in pseudo-words. Thus, these results suggest that lexical-level representations influence the role of lexical tones in Mandarin word recognition.

2.4.1. The influence of lexical-level representations on tone-word recognition reflected by RTs

The RT data show different response speeds for the decisions on pseudo-words with Tone1, Tone2, and Tone4, whereas the lexical decisions on real words do not exhibit differential RTs amongst the three tonal categories. These results imply that recognising real words is not affected by lexical-tone perception, whereas the speed of recognising a pseudo tone word may be related to the isolation of its lexical tone.

The lexical decision task has been used to reflect underlying processes in lexical access (Goldinger, 1996). Since the lexical frequencies between words of the three tonal categories are matched, comparable RTs are expected in Tone1, Tone2 and, Tone4 real words. This result would suggest that lexical tones in real words can be instantly integrated into segments and processed by accessing whole-word representations. Our interpretation is in agreement with the study by Zhao et al. (2011), which found that the N400 elicited in a lexical tone mismatch condition had similar latency and amplitude as the N400s in onset mismatch and rime mismatch conditions. The N400 is a neural indicator of difficulties in lexical access, usually observed around 400 ms post stimulus onset, as proposed by Kutas and Hillyard (1984). Their finding indicated that in lexical access, all segmental cues and tonal (i.e., suprasegmental) cues are utilised simultaneously.

In contrast with the similar RTs in reals words of the three tonal categories, Tone4 pseudo-words were judged more slowly than Tone1 and Tone2 pseudo-words in the present study. Due to the lack of lexical-level representations in pseudo-words, our results suggest that such temporal differences are a reflection of recognising pseudo-words via sublexical representations. Thus, we attribute the RT contrasts between Tone1 and Tone4 pseudo-words, and between Tone2 and Tone4 pseudo-words, to the differential time courses for isolating the three tones. The slower responses to Tone4 pseudo-words than to Tone1 pseudo-words are in line with one previous study showing that the isolation point of Tone1 is earlier than that of Tone4 (Lai & Zhang, 2008). However, contradictory to our data, existing studies have reported that Tone2 has a later isolation point than Tone1 and Tone4 (Wu & Shu, 2003; Lai & Zhang, 2008). In our study, there are no differences between the response latencies of Tone1 and Tone2 pseudo-words, and Tone2 pseudo-words show a tendency to be judged faster than Tone4 pseudo-words. One possible explanation is that it takes more time to distinguish Tone4 from Tone3. This is because Tone4 has a falling pitch contour (\), whereas Tone3 has a falling-dipping contour shape (\/). Therefore, the brain may recognise Tone4 only after confirmation that a falling tone is not accompanied by a dipping tone. These cognitive processes do not apply to the flat Tone1 (–) and rising Tone2 (/). However, since the current experiment does not involve Tone3 in its design, an alternative explanation is also needed.

We tentatively propose another account that may lie in the normalisation of stimulus duration. In the current study, the average duration of Tone1 pseudo-word recordings is 570 ms, Tone2 is 691 ms, and Tone4 is 654 ms. In order to normalise the duration into 650 ms, Tone1 pseudo-words needed to be stretched, while Tone2 pseudo-words needed to be compressed. Tone4 pseudo-words had to undergo very small temporal changes. Consequently, the isolation points might occur later for Tone1 and earlier for Tone2, while remaining more or less the same for Tone4. Therefore, we observe the tendency of slowest responses in Tone4 pseudo-words, and no difference in responses between Tone1 and Tone2 pseudo-words. However, further studies need to be conducted to provide greater evidence of this explanation.

2.4.2. Accuracy and perception of tonal features

Accuracy data show that the lexical decisions of Tone4 real words are more prone to errors than those of Tone1 real words, but Tone1 pseudo-words are more likely to be misjudged than Tone2 pseudo-words. Since Tone1 and Tone4 have similar onset pitch heights, and Tone1 and Tone2 share similar initial pitch contours, the accuracy data may reflect how lexical-level representations interact with the sublexical features of lexical tones.

It is striking that the accuracy data show quite a different picture from the one reflected by the RT patterns. Our data reveal that lexical tones not only lead to contrastive judgment between Tone1 and Tone2 pseudo-words, but also in real words carrying Tone1 and Tone4. Specifically, in lexical decisions on real words, Tone4 words are more likely than Tone1 words to be misjudged as pseudo-words; in decisions on pseudo-words, participants have greater difficulty in deciding the lexical status of Tone1 pseudo-words than Tone2 pseudo-words.

To explain these results, two questions must be answered. First, why are errors committed contrastively between two tones in real words (i.e., Tone1 vs. Tone4) and pseudo-words (i.e., Tone1 vs. Tone2)? Second, why are the error patterns different between real words and pseudo-words? We answer the two questions by considering the acoustic confusion between Tone1 and Tone2, and between Tone1 and Tone4.

Acoustically, Tone1 shares common points with both Tone4 and Tone2 (Guo, 1993). For example, as previously mentioned, Tone1 and Tone4 have the similar onset pitch heights, and the initial proportions of the pitch contours of Tone1 and Tone2 have similar flat shapes. However, the rising Tone2 and falling Tone4 are highly contrastive. In line with these linguistic facts, studies have found that when the tonal input is limited, Tone1 and Tone4 are mutually confusing, and Tone2 is more likely to be misidentified as Tone1 (Wu & Shu, 2003; Lai & Zhang, 2008). We argue that in our experiment, the processing of lexical tones may undergo the same tonal interference between Tone1 and Tone2, and between Tone1 and Tone4. Therefore, the errors become contrastive between Tone1 and Tone2 pseudo-words, and between Tone1 and Tone4 real words.

If our account holds true, how do we explain the distinctive error patterns between real words and pseudo-words? Our suggestion is that this lexicality selectivity in errors is induced by the different ways to perceive lexical tones in word forms with and without lexical-level representations. That is, when lexical-level information is available, there is interference between the perception of tone words with similar onset pitch heights (Tone1 vs. Tone4), implying that participants can perceive the onset pitch height but cannot use the pitch contour cue to differentiate the two tones. Contrastively, when there is no lexical-level representation, the lexical decision on pseudo-words is more error-prone in distinguishing between tones with similar pitch contours (Tone1 vs. Tone2), indicating sensitivity to pitch contour and hindrance to the perception of pitch height (Tone1 vs. Tone4).

It can be further inferred that real-word recognition may be sensitive to pitch height which is mediated by the top-down influence of lexical-level representations, whereas the better resolution of pitch contours in pseudo-word recognition suggests dependence on sublexical level processing, given a lack of top-down influence. Our inference is possible because top-down influence can facilitate lexical-tone perception in the very beginning stages of word recognition, whereas lexical-tone perception in pseudo-words does not have such facilitation, and naturally becomes more sensitive to slow varying cues like pitch contour.

One additional issue that must be noted is that interference between two contrastive tones has specific orientations. In our study, although Tone4 real words have lower accuracy rates, most Tone1 words are correctly judged. In contrast with this pattern, Tone1 pseudo-words are more likely to be misjudged, but decisions on Tone2 pseudo-words have high accuracy. This means that the proposed sublexical interference only affects the lexical decision of Tone4 real words and Tone1 pseudo-words. A parsimonious explanation is that Tone4 real words sound more like pseudo-words, and Tone1 pseudo-words sound more like real words. Furthermore, the lexical decision performance just so happens to be contrastive between Tone1 and Tone4 real words, and between Tone1 and Tone2 pseudo-words.

However, the factorial design of this study allows us to perform a rigorous examination of lexical performance, which reveals highly systemic variations contributed by an interaction between word forms' lexical status and the tones they carry. Therefore, we refute the coincidental view, and hold that the orientations of tonal interference are related to specific tonal perception modulated by lexical-level representations. Since Tone1 and Tone4 are mutually confusing in the initial phase of processing (Lai & Zhang, 2008; Wu & Shu, 2003), Tone1 and Tone4 may compete with each other in the early phase of processing for similar onset pitch heights. At the sublexical level, Tone1 can be isolated earlier than Tone4, making Tone4 word perception subject to more interference from the competition with Tone1. Such competition may induce lower quality of feedforward activation of the lexical-level representation, and eventually lead to less accurate lexical decisions in Tone4 words. For decisions in pseudo-words, due to the lack of top-down facilitation, lexical-tone perception occurs at a relatively slow pace via sublexical representations. Therefore, in the current study, Tone1 could be prone to competition not only with Tone2 (similar pitch contours) but also with Tone4 (similar onset pitch height), whereas Tone2 only needs to compete with Tone1. Consequently, the more complex competition interferes with lexical decision accuracy in Tone1 pseudo-words.

To summarise, the contrastive accuracy patterns between lexical decision performance in real words and pseudo-words carrying different tones imply divergent sensitivities to onset pitch height when lexical-level representation is available, and to pitch contour shape when lexical-tone processing is via sublexical representations. These results suggest that in the current experiment, accuracy and RT reflect different cognitive processes in tone-word perception. Future studies should be conducted using electrophysiological techniques to look for more online evidence for our explanations.

2.4.3. Semantic and phonological contents at the lexical level

The secondary purpose of the current study was to test whether there is a division between the semantic and phonological aspects of lexical-level representations (Zhao et al., 2011). By using the LDT of semantics and the LDT of phonology, our study achieved results that do not support this view. The task requiring participants to judge lexicality according to the semantic dimension (i.e., LDT of semantics) only led to generally prolonged RTs than in the task requiring participants to judge the existence of a sound pattern (i.e., LDT of phonology). This finding suggests that access to phonological representations is faster than that of semantic representations, as previous studies have shown (Friederici, 2002; Rodriguez-Fornells et al., 2002; Schmitt, Kutas, & Münte, 2002). Moreover, the overall faster responses in the LDT of phonology can be explained by both of the views that lexical phonology and semantics are at the same level (Ye & Connine, 1999; Malins, 2003) or by the view of separate layers (Zhao et al., 2011). Such evidence can be explained as both successful inhibition of semantic units within the same representational level or by top-down facilitation from the above semantic level.

Apart from the RT findings, there is no differentiation in accuracy between the two versions of LDT. This result suggests that the semantic and phonological representations at the lexical-level do not have substantial differences in terms of exerting top-down influence on the perception of sublexical features. Moreover, the two tasks did not influence the perception of word forms of the three tonal categories differently. This result indicates that task instructions that are biased on the word's semantics or phonology do not change the way lexical tones and lexical-level representations interact with each other. Therefore, our study suggests that the lexical-level representation can be a multi-dimensional structure hosting both semantic and phonological components. Moreover, it is not necessary to place the semantic representation and the phonological representation in two separate layers as suggested by Zhao et al. (2011).

2.5. Conclusion

In this study, we investigated how lexical tones and lexical-level representations affect lexical decision behaviours. Using two versions of an auditory LDT, we found evidence that lexical tones modulated lexical decisions differently in word forms with and without lexical-level representations. Moreover, different dependent variables in the LDT can indicate distinctive aspects of lexical-tone perception in spoken word recognition. In addition, the semantic and phonological components of the lexical-level representation modulate lexical-tone perception similarly.

CHAPTER 3

Form Priming in Tonally Contrasted Word Forms With and Without Lexical-level Phonological Representations in Mandarin Chinese

3.1. Introduction

One of the fundamental brain functions underlying human communication is the recognition of spoken word forms for further syntactic, semantic and pragmatic analyses. Successful word recognition relies on mapping continuous acoustic input onto mental representations of phonology with remarkable speed and accuracy (Vitevitch & Luce, 1998; Dahan & Magnuson, 2006). With respect to the structure of phonological representations, current models have reached a consensus that phonological representations are organised hierarchically with at least two levels: a low level representing sublexical features like phonemes and a high representational level for lexical (word) knowledge (e.g., TRACE model, McClelland & Elman, 1986; Shortlist model, Norris, 1994; Neighborhood Activation Model (NAM), Luce & Pisoni, 1998). Using priming paradigms, studies with behavioural measurements showed that during the recognition of a spoken word, a prime word sharing high form-similarity with a target word could induce competition at the lexical level (Slowiaczek & Pisoni, 1986; Slowiaczek & Hamburger, 1992; Gaskell & Dumay, 2003; Donselaar, Koster, & Cutler, 2005). However, less is known about the access to phonological representations at the lexical and sublexical levels in tonal languages. The current study used a form-priming paradigm to investigate how phonological representations are accessed when recognising monosyllabic tone words in Mandarin Chinese.

3.1.1. Form priming in spoken word recognition

Form priming is a well-established experimental paradigm which has been used to explore the structure of phonological representations and how those representations are accessed during word recognition (see Zwitserlood, 1996 for a review). In a form priming experiment, a target word form is usually presented after a prime word form. The rationale is that if a prime can activate the mental representation shared by a target, the prime should affect the recognition of the subsequent target. Priming effects can be quantified by reaction times measured in behavioural tasks such as lexical decision and shadowing, namely oral repetition (e.g., Radeau, et al., 1989; Hamburger & Slowiaczek, 1996; Bien, Bölte, & Zwitserlood, 2014).

Two typical priming effects have been identified in spoken word recognition: facilitation and inhibition. Facilitatory priming, as reflected by shorter reaction times relative to a baseline, has been associated with form-based overlaps between a prime and a target (Radeau, Morais, & Segui, 1995; Slowiaczek & Hamburger, 1992; Slowiaczek & Pisoni, 1986). Facilitation has been reliably observed in repetition priming when the prime and the target are identical word forms (Zwitserlood, 1996). Inhibitory priming, indexed by slower reactions compared to a baseline, has been attributed to the inhibition of a primed lexical competitor during the recognition of a phonologically similar target word (Slowiaczek & Pisoni, 1986). Lexical competitors are entries in the mental representation, having, for instance, overlapping phonemes from the word onset. Thus, given the input of the same portion of a speech signal, these entries compete with each other to be accessed before a word can be identified.

Form-priming effects induced by phonological representations at the lexical level have been investigated by comparing participants' responses to word forms primed by formally related real words or pseudo-words4. Studies in both lexical decision and shadowing showed that the high formal similarity between word primes and word targets (e.g., bland-black) did not lead to shorter reaction times than word targets sharing only one initial phoneme (e.g., burnt-black) or no phonemic overlaps (Slowiaczek & Pisoni, 1986; Radeau et al., 1989; Slowiaczek & Hamburger, 1992; Hamburger & Slowiaczek, 1996). The lack of facilitation has been described as lexical interference rather than inhibition because phonologically similar word primes do not always lead to prolonged lexical decision times in target words. This phenomenon has been attributed to a net effect of form-based facilitation and inhibition induced by lexical competition (Slowiaczek & Hamburger, 1992; Wagenmarkers, Zeelenburg, Steyvers, Shiffrin, & Raaijmakers, 2004; Lee, 2007). That is, although the phonological overlap can facilitate the processing of the form of target words, the cognitive system still needs to inhibit lexical competitors evoked by prime words. Consequently, lexical interference induces responses without evident facilitatory or inhibitory priming (Hamburger & Slowiaczek, 1996). In contrast, when the prime is a pseudo-word, the recognition of a target word with high phonological similarities (e.g., */blæt/-black) becomes faster (Slowiaczek & Pisoni, 1986; Radeau et al., 1989; Slowiaczek & Hamburger, 1992). Pseudo-words have been described as word forms without lexicalised phonological representations, so the facilitatory priming effect suggests that the re-activation of the same sublexical-level representation can boost the processing at the lexical level. Following this line of reasoning, several studies have successfully revealed that the lexicalisation of pseudo word

[4] Pseudo-words in studies with European languages are usually novel segment patterns that are meaningless but phonotactically possible.

forms after learning could induce word-like interference (Gaskell & Dumay, 2003; Qiao, Forster, & Witzel, 2009; Qiao & Forster, 2013).

3.1.2. Lexical access in Mandarin Chinese

While existing studies have mainly focused on recognising word forms with similar segments, it remains unclear that how phonological representations are accessed when lexical tones, a suprasegmental cue, play a critical role. Unlike most European languages, tone languages such as Chinese, Vietnamese, and Thai use systematic variations of prosody to differentiate meanings of the same segment. Mandarin Chinese is an ideal linguistic model to study this issue for the following reasons. First, Mandarin is an extreme example of tone languages employing four lexical tones with contrastive pitch contours. Second, the lexical tones in Mandarin are highly lexicalised because each syllable has one lexical tone. A well-known example is that the combinations of a segment /ma/ with the four tones lead to four different words, namely /ma1/ 'mother', /ma2/ 'hemp', /ma3/ 'horse', and /ma4/ 'curse'.

A number of studies have shown that lexical tones play a comparable role as the segmental cues (e.g., onset and rime) during word processing in Chinese languages (see Schirmer et al., 2005; Malins & Joanisse, 2010; Zhao et al., 2011 for data indicating an equal role; see Cutler & Chen, 1997; Mattys et al., 2005; Yip, 2001; Hu et al., 2012; Sereno & Lee, 2015; Wiener & Turnbull, 2015 for data suggesting a segmental processing primacy). Moreover, both behavioural and neurolinguistic evidence has suggested that the segmental and tonal features of Mandarin words are represented separately in the mental lexicon (Liang & van Heuven, 2004; Liu, et al., 2006; Tong, Francis, & Gandour, 2008). A few studies have also reported evidence that whole-word level representations, namely the lexicalised combinations of segmental and tonal representations, are crucial in the processing of spoken tone words (e.g., Ye & Connine, 1999; Zhao et al., 2011; Malins & Joanisse, 2012; Yue, Bastiaanse, & Alter, 2014). Taking the above-mentioned studies together, it can be concluded that the mental representation in Mandarin Chinese complies with a general hierarchical organisation: sublexical-level representations store knowledge of consonants, vowels, and lexical tones; meanwhile, lexical-level representations encode the whole-word phonological forms (i.e., the lexicalised combination of segments and tones).

How does a tone word activate the lexical- and sublexical-level representations of phonology during form priming? A few studies have touched upon this question with monosyllabic tonal word forms in Mandarin. Using a lexical decision task, Lee (2007) found that when a monosyllabic Mandarin word was primed by a real-word target with minimal tone contrasts (e.g., /lou3/-/lou2/) with an ISI of 250 ms, the response latency to the target word was not different from the baseline. This result was comparable to an earlier lexical decision study with minimal stress pairs. Like lexical tone, stress is another type of suprasegmental cue to distinguish segmental meanings. Cutler and Otake (1999, Experiment 3) used Japanese words to create prime-target pairs with minimal stress-pattern contrasts (e.g., haSHI ‘chopsticks’ and HAshi ‘bridge’, in which a stressed syllable is noted with capital letters). They found no difference between the reaction times in targets preceded by primes contrasting in the stress pattern and those in targets primed by phonologically unrelated words. The authors of these studies interpreted their results as an indication of immediate employment of suprasegmental cues to constrain word processing to rule out the primed lexical representations that would compete with the target word. Thus, the recognition of target words primed by segmentally overlapping words was similar to those primed with phonologically unrelated words. This lexical-constraint interpretation is almost equal to the claim that spoken words only contrasting in suprasegmental features are treated as totally different words in the mental lexicon because only in such a situation could the target words not possibly be influenced by the prime’s formal similarity.

Although the lexical-constraining idea could provide an account for priming with minimal tone pairs of words as in Lee (2007), it does not explain other priming results. For example, in a recent study of lexical decision with monosyllabic Mandarin words, Sereno and Lee (2015) found a fragile facilitation effect in minimal tone pairs5 (e.g., /ru3/-/ru4/) and consistent inhibition effects in minimal segment pairs (e.g., onset contrast /pu4/-/ru4/; rime contrast /re4/-/ru4/). If lexical-constraining efficiency is the mechanism to resolve the influence of primes on phonologically similar targets, Sereno and Lee's results imply that segmental cues were less effectively used than suprasegmental cues to rule out lexical competitors. However, such implication of a 'suprasegmental cue primacy' is not in favour of most experimental observations and is less practical for tone-word recognition, in that most existing studies found evidence of either segmental primacy or an equal role of the two types of cues (e.g.,Schirmer et al., 2005; Wiener & Turbull, 2015).

Alternatively, a more realistic interpretation might be that the seemingly non-primed responses are a consequence of lexical interference, as documented in early literature which found no facilitatory priming between segmentally similar words in English (e.g., Slowiaczek & Hamburger, 1992; Hamburger & Slowiaczek, 1996). That is, the lack of facilitation is a net effect of form-based facilitation and lexical competition-triggered inhibition. According to this view, the absence of priming effects in minimal tone/stress pairs (Lee, 2007; Cutler & Otaka, 1999) can be explained as follows: the overlapping segment may facilitate the processing of the sublexical features of the targets, but the inhibitory priming at the lexical level balances out the facilitation. This interpretation could somehow reconcile the facilitatory priming effects in minimal tone-pairs reported in Sereno and Lee (2015) and in an early Cantonese study by Yip (2001).

To disentangle the tone word priming issue, further experiments need to be implemented to overcome the limitations in previous studies. First, it was surprising that both Lee (2007) and Sereno and Lee (2015) did not report whether they had normalised their word form recordings. If no sound normalisation was performed, participants might have heard auditory stimuli with varying sound quality, duration, and intensity. This

[5] Minimal tone pairs are prime-target pairs contrasting only in tones.

limitation may have masked the true picture of form priming in tonal word forms. Second, although they claimed to use the same protocol as Lee (2007), Sereno and Lee reported response latencies that were approximately 300 ms shorter than those in Lee (2007). This quite striking difference might imply that Sereno and Lee (2014) did not measure reaction times in the same way Lee (2007) did. Another possible explanation is that the participants in both studies might not represent the same population, which could weaken the generality of the findings in the two studies. Third, despite the fact that Sereno and Lee matched the number of the four types of tones in the targets, they did not control for the number of homophonic words as Lee (2007) did. Lastly, since pseudo-words were only used as fillers, the best tool to detect sublexical processing had been overlooked (Slowiaczek & Pisoni, 1986, Radeau et al., 1989; Wagenmakers et al., 2004). That is, if interference is a net effect of form-based priming and lexical inhibition, pseudo-word primes should facilitate the processing of target words with minimal phonological contrasts (e.g., phonemes or tones).

3.1.3. The present study

Our study aimed to explore how phonological representations at lexical and sublexical levels are accessed in the form priming of tone words. The current study first sought to replicate Lee's (2007) study to confirm the priming patterns in identical word pairs and minimal tone pairs. Then, to approach the priming effect at the sublexical level, we took advantage of pseudo-words which share common segments with target words but carry different tones, referred to as tone-manipulated pseudo-words. These pseudo-words were phonotactically plausible but without meaning, created by combining a legal segment pattern with a lexical tone in Mandarin Chinese (e.g., */se2/ from /se/ + Tone2). Therefore, this kind of pseudo-word had only sublexical representations without lexical-level entries in the mental lexicon of Mandarin speakers. To study form priming effects at sublexical processing levels, we produced 'prime-target' pairs by pairing pseudo-word targets with segmentally overlapping real and pseudo-word primes.

For real-word targets, we expected no facilitation in the priming condition with tonally contrasted real words, in line with previous experimental findings (Lee, 2007). Moreover, if the lack of facilitation is a result of lexical interference (Slowiaczek & Hamburger, 1992), form-based facilitatory priming could be expected when target words are primed by pseudo-words with minimal tone contrasts due to the absence of entries at the lexical level.

It was difficult to predict the outcome in pseudo-word targets due to limited observation of tonal pseudo-words in form priming. Lexical decision has been assumed to reflect lexicon-based cognitive activities (Balota & Chumbley, 1984). However, due to the lack of lexical-level representations, pseudo-words can only access low-level phonological representations. That is, the priming effect on pseudo-words could only be induced via processing at the sublexical level. Since form-based facilitation has been considered as a sublexical effect, we tentatively predicted that the processing of pseudo-word targets can be facilitated by either pseudo-word or real-word primes with minimal tone contrasts.

3.2. Method

3.2.1. Participants

88 students and staff of Northeast Petroleum University at Daqing, China voluntarily participated in the current experiment (female: 50; mean age = 21, *SD* = 3.5). They were selected to take part in the current study according to a questionnaire issued before the experiment, in which they all reported Mandarin Chinese as their native language, no long-term exposure to other dialects of Chinese6, and no history of auditory diseases or speech disorders.

[6] Since the pseudo-words can be meaningful in other dialects of Chinese (e.g., */na1/ means 'you' in *Lanzhou* dialect and 'take' in *Xuzhou* dialect), we only included participants with Mandarin as their only native language.

3.2.2. Materials

We first selected 35 monosyllabic Mandarin Chinese real words and generated 35 pseudo-words as targets. For each target, we constructed five types of prime-target pairs. A real-word target was paired with: 1) an RSTR prime (an identical real word that shared the same segment and lexical tone with the real-word target, e.g., /lun4/-/lun4/); 2) an RSR prime (a real word that shared the same segment with the target, e.g., /lun2/-/lun4/); an RNOR prime (a real-word prime that did not overlap with the target in either the segmental or tonal proportion of the target, e.g., /pie3/-/lun4/); a PSR prime (a pseudo-word that only shared the same segmental structure with the real-word target, e.g., */lun3/-/lun4/); and a PNOR prime (a pseudo-word that had no phonological overlap with the target, */tai3/-/lun4/). The RNOR primes and the PNOR primes served as the baseline conditions to observe the priming effects in the two priming conditions with real-word primes (RSTR, RSR) and the priming condition with pseudo-word primes (PSR).

For pseudo-word targets, the five types of primes were: 1) a PSTP prime (a pseudo-word having the identical word form as the target, e.g., */zen2/-*/zen2/); 2) a PSP prime (a pseudo-word with the same segment but different tone than the target, e.g. */zen1/-*/zen2/); a PNOP prime (a pseudo-word that was unrelated to the target, e.g., /rui1/-*/zen2/); RSP prime (a real word that shared the segmental part with the target, /zen3/-*/zen2/); and an RNOP prime (a real word consisting of no common segment or lexical tone as the target pseudo-word, e.g., /xiu3/-*/zuan2/). The PNOP primes provided a baseline for the priming effects in the two pseudo-word priming conditions (PSTP, PSP), and the RNOP primes served as the baseline for the real word priming condition (RSP) to explore any facilitation or inhibition.

The experimental materials consisted of 333 word forms, 198 of which were used by Lee (2007). The word forms were equally distributed in five lists. 17 word forms were used twice in different lists. All lists had the same targets paired with different primes. The 70 targets were pseudo-randomly divided into 10 subsets (five for the real-word targets and five for the pseudo-word targets), so that each subset with seven items could take one type of prime in a list. Such materials not only permitted us to examine the lexical and sublexical modulations of tone-word priming, but it also balanced the probabilities of hearing real and pseudo-word forms. See Appendix B for a complete list of word and pseudo-word stimuli. Moreover, the reuse of word forms would not confound our results because no real or pseudo-words within a list appeared more than once except for the repetition priming conditions.

The real word stimuli have no homophones, judged according to the list of Chinese words in the Modern Chinese Frequency Dictionary (Language Teaching Research Centre for Beijing Language Institute, 1986). The average word frequencies of the five types of real-word primes (RSTR, RSR, RNOR, RSP, RNOP) were matched to avoid potential confoundings, supported by a one-way ANOVA revealing no significant difference between the average logarithm frequencies of the five prime types ($F(4, 190) = 0.076$, $p = .989$). The two words in each priming pair were lexically unrelated. That is, they could not form a disyllabic word.

The materials were produced by a female native Putonghua speaker in a sound proof cabin (Institute of Neuroscience, Newcastle University) with a high-quality Rode NT1-A microphone and E-MU 0404 recording system, digitised at a 16 bit, 44.1 kHz sampling rate on a Dell E5400 laptop via acoustic software programme PRAAT (Boersma & Weenink, 2013). Each word form was pronounced 15 times. Ignoring the first four and the last four recordings, one of the exemplars was carefully chosen. The selection criteria were the following. First, the chosen recording was clearly articulated. Second, the recording had an approximate duration around 550 ms. All experimental materials were normalised for the same duration of 550 ms and intensity of 75 dB SPL with PRAAT, and saved individually per word form.

3.2.3. Procedure

The participants were tested individually in a sound-attenuated room. They sat in front of a laptop (HP 540) with a distance of approximately 80 cm between their eyes and the screen. Auditory stimuli included 70 stimulus pairs, each of which consisted of a prime and a target, presented via a pair of Cosonic CD-778MV headphones at 75 percent of the full volume of the laptop. The presentation of stimuli was pseudo-randomised per participant to ensure that the same type of targets were not presented in more than three successive trials. Participants were randomly assigned to one list and instructed to decide whether the second word form (the target) in a stimulus pair was an existing word in Mandarin Chinese by pressing the button 'O' for a positive answer and the button 'P' for a negative answer as quickly and accurately as possible. They were also asked to pay close attention to the second word form (the target) in each stimulus pair and to ignore the first word form (the prime, *cf.* Shuai, Li, & Gong, 2012). A practice session with 10 prime-target pairs was administered before the experiment. During the practice session, feedback on reaction time and accuracy was presented on the screen immediately after a judgment was made. A participant was not allowed to proceed to the experimental session until the number of erroneous judgments was equal to or less than three.

A trial began with a set of fixation marks '*****' presented in the centre of the screen for 1000 ms. The prime was delivered 500 ms after the disappearance of the fixation marks, followed by the presentation of the target with an ISI of 250 ms between the prime and the target (Lee, 2007; Sereno & Lee, 2015). For each trial, participants were given a maximum 6 s to make a response. The next trial automatically started as soon as an answer was given. Response latency was measured from the onset of the target (Zwitserlood, 1996; Lee, 2007). Software programme DMDX V4.0.4.6 (Forster & Forster, 2003) was used to control the delivery of auditory stimuli and to record reaction times and accuracy. The entire protocol took about 10 minutes.

3.2.4. Data analysis

Reaction times and accuracy (correct percentage) were the dependent variables. Only correct judgments were included for statistical analyses.

Reaction times shorter than 400 ms and longer than 3000 ms were excluded (Lee, 2007; Shuai et al., 2012). Six participants out of 88 were excluded because they had less than four analysable responses out of seven trials in at least one type of prime-target pairs. In total, 9.2% of the data were rejected in the remaining 82 participants.

One-way repeated measures ANOVAs with the reaction times and correct percentages (accuracy) of real-word targets and pseudo-word targets were performed separately to examine whether Prime-type was a factor leading to systematically varying response patterns for the same targets. Reaction times and correct percentages were averaged both by subject and by item to ensure that the potential priming effects could be generalised rather than determined by some specific items used in our experiment. Greenhouse-Geisser correction was performed when the sperificity assumption was violated. If the Prime-type could be identified as a main effect, planned pair-wise comparisons defined a priori would be performed between a priming condition and the corresponding baseline condition (RSTR vs. RNOP, RSR vs. RNOR, and PSP vs. PNOR for real-word targets). The significance level was set to $p < .05$. All statistical analyses were conducted with SPSS V22 (IBM, Armonk, NY, USA).

3.3. Results

3.3.1. Reaction times

The experimental results of real-word targets paired with five types of primes are displayed in Table 3.1. ANOVAs revealed a reliable main effect of Prime-type (by subject F (4, 324) = 12.73, $p < .001$; by item F (4, 136) = 10.37, $p < .001$). Planned comparisons showed that the average reaction time of the targets preceded by the RSTR primes was 104 ms faster than that in the RNOR baseline condition (by subject t (81) = -4.80, p < .001; by item t (34) = -4.91, p < .001), and the mean reaction time of the targets primed by pseudo-words with segmental overlaps was 93 ms faster than the PNOR baseline condition (by subject t (81) = -5.12, p < .001; by item t (34) = -3.28, p = .002). However, the difference between RSR primes and RNOR primes was not significant (by subject t (81) = -0.48, p = .631; by item t (34) = -0.601, p = .552). These results suggest that tone word targets can be facilitated by word primes with identical forms and pseudo-word primes with minimal tone contrasts. In line with Lee (2007), there were no clear priming effects in word targets primed with tonally contrasted real words. These priming patterns (see Figure 3.1) are in line with our predictions based on previous studies and the idea of lexical interference.

Table 3.1 Results of conditions with real-word targets (RT = reaction time, CP = correct percentage).

	Example		Results	
	Prime	Target	RT (*SD*)	CP % (*SD*)
RSTR	/lun4/	/lun4/	1078 (266)	92.9 (11.7)
RSR	/lun2/	/lun4/	1173 (215)	89.4 (13.6)
RNOR	/pie3/	/lun4/	1182 (215)	87.1 (13.2)
PSR	*/lun3/	/lun4/	1092 (196)	93.4 (9.6)
PNOR	*/tai3/	/lun4/	1185 (230)	86.9 (12.9)

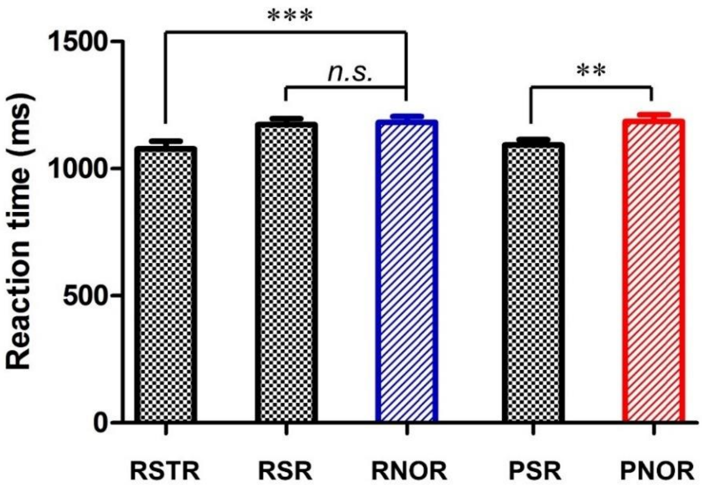


Figure 3.1 Average reaction times in the real-word targets preceded by the five conditions. The baseline condition is coloured in blue for the real-word priming condition; the red bar is the baseline for the pseudo-word-priming condition. *n.s.* p > .05, ** p < .01, *** p < .001.

For the pseudo-word targets, the experimental results are shown in Table 3.2. ANOVAs revealed a main effect of Prime-type (by subject F (4, 324) = 9.01, $p < .001$; by item F (4, 81) = 7.00, $p < .001$). Pair-wise comparisons showed that the recognition of tonal pseudo-word forms was facilitated for 59 ms by the identical pseudo-word primes (PSTP vs. PNOP: by subject t (81) = -2.40, p = .019; by item t (34) = -2.49, p = .018). However, the reaction latency became 88 ms longer when the pseudo-word targets were preceded by pseudo-word primes contrasting only in tones (PSP vs. PNOP: by subject t (81) = 4.35, $p < .001$; by item t (34) = 2.44, p = .020). Moreover, the effects induced by RSP primes (real-word primes with minimal tone contrasts) was unclear because no significant difference was identified between the RSP priming condition and the RNOP baseline condition (by subject t (81) = 0.46, p = .644; by item t (34)

= -0.05, p = .963). These results, as illustrated in Figure 3.2, suggest that the processing of pseudo-word targets is inhibited by the presence of pseudo-word primes contrasting in tones. No clear priming effects are produced by real-word primes sharing the same segments with pseudo-word targets. Both of these findings contradict our predictions that lexical decisions on pseudo-words can be facilitated by both real word and pseudo-word primes with minimal tone contrasts.

Table 3.2 Results of conditions with pseudo-word targets (RT = reaction time, CP = correct percentage).

	Example		Results	
	Prime	Target	RT (*SD*)	CP % (*SD*)
PSTP	*/dei2/	*/dei2/	1194 (300)	92 (10.8)
PSP	*/dei4/	*/dei2/	1341 (299)	89.2 (13)
PNOP	*/mu1/	*/dei2/	1253 (267)	91.3 (12)
RSP	/dei3/	*/dei2/	1314 (336)	93 (10.3)
RNOP	/leng4/	*/dei2/	1299 (327)	94.4 (10.7)

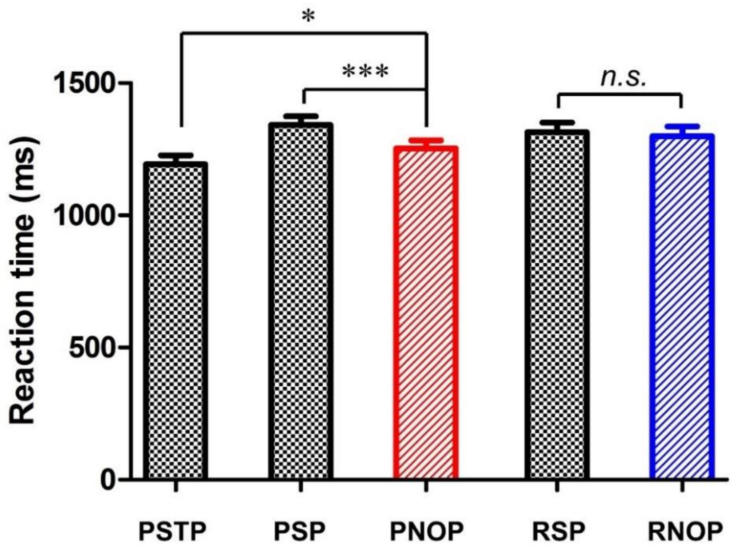


Figure 3.2 Average reaction times in the pseudo-word targets preceded by the five types of primes. The red bar is the baseline for the pseudo-word priming condition; the blue bar is the baseline for the real-word priming condition; n.s. $p > .05$, * $p < .05$, *** $p < .001$.

3.3.2. Correct percentages

The arcsine square root means of the percentage of correct items (Howard, Nickels, Coltheart, & Cole-Virtue, 2006; Ahrens, Cox, & Budhwar, 1990) were used for the same analysis procedures as those with reaction times. For real-word targets, ANOVAs identified a main effect of Prime-type (by subject F (4, 324) = 6.05, p < .001; by item F (4, 136) = 3.21, p = .015). Planned comparisons showed that RSTR primes and PSR primes led to higher correct percentages than identical primes (RSTR vs. RNOR: by subject t (81) = 3.68, p < .001; by item t (34) = 2.27, p = .030. PSR vs. RNOR: by subject t (81) = 3.76, p < .001; by item t (34) = 2.55, p = .015).

However, targets primed by real words contrasting in tones did not induce changes in accuracy (RSR vs. RNOR: by subject t (81) = 1.42, p = .160; by item t (34) = 0.719, p = .477). These results suggest that when real words are primed by identical real words or by pseudo-words with the same segments, the lexical decision on these real words is facilitated.

For pseudo-word targets, however, Prime-type was a significant main effect in the by-subject analysis (F (4, 324) = 3.73, p = .006) but not in the by-item analysis (F (4, 136) = 1.34, p = .260). Further pair-wise statistical analyses did not reveal significant differences in the planned comparisons with the correct percentage data averaged by participant (PSTP vs. PNOP: t (81) = 0.19, p = .849; PSP vs. PNOP: t (81) = -1.48, p = .144; RSP vs. RNOP: t (81) = -1.32, p = .192). These results suggest that the accuracy of lexical decisions in pseudo-words is not affected by different types of primes.

3.4. Discussion

The current study aimed to explore how high-level and low-level representations of phonology modulate word recognition in a form priming context. We compared the lexical decision performance in real word and pseudo-word targets preceded by 10 types of primes. The most important finding of our study is that the recognition of real words can be facilitated by pseudo-words with minimal tone contrasts, but there is no clear facilitation when primed by tonally contrasted real words. This finding suggests that the lack of facilitation in minimal tone pairs of real tonal words is a result of lexical interference. That is, real-word targets primed by real words contrasting in tones is affected by both sublexical facilitation and lexical inhibition due to the lexical competition between tonally contrasting word forms (Slowiaczek & Hamburger, 1992; Hamburger & Slowiaczek, 1996). Moreover, we also found robust inhibition in minimal tone pairs of pseudo-words, and no clear priming effects when pseudo-words are primed by tonally contrasted real words. These results suggest that inhibition is not a specific effect in lexical-level representations, neither is facilitation specifically at the sublexical level.

3.4.1. Form priming in real-word targets

Our results show that the processing of real words is speeded up by identical word primes, but becomes neither faster nor slower when primed by tonally contrasting words. The facilitation in word pairs with formally identical primes and targets is also reflected by the higher correct percentage relative to the baseline condition, whereas the targets in minimal tone-pairs of words do not elicit different error rates compared to the baseline condition. The facilitatory priming effect in identical word pairs is in line with the classic priming pattern in both tonal and non-tonal languages (Zwitserlood, 1996; Lee, 2007; Sereno & Lee, 2015). Together with the lack of facilitation in minimal tone-pairs of real words which are the same Lee (2007), these results attest that our data are comparable with those reported in previous form priming studies (e.g., Lee, 2007; Sereno & Lee, 2015).

Apart from replicating the two form priming effects, the current study revealed evidence of form-based facilitation produced by pseudo-word primes with minimal tone contrasts to the target words (e.g., */lun3/-/lun4/ 'argue'). To the best of our knowledge, this study is the first to investigate sublexical priming with pseudo-words in tone-word recognition. Our finding is similar to some early studies using English materials reporting faster responses to word targets preceded by pseudo-word primes with high phonological similarities (e.g., */blæt/-black, Slowiaczek & Pisoni, 1986; Radeau et al., 1989; Slowiaczek & Hamburger, 1992). The similarity implies that lexical tones are used as segmental cues (i.e., phonemes) during tone-word recognition. Therefore, in the same line of reasoning, we interpret the facilitatory priming effect produced by tonally contrasted pseudo-words in our experiment as evidence that the pre-activation of segment patterns at the sublexical representation level produces form-based priming for the recognition of tone-word targets.

Our findings are consistent with the lexical interference idea which has been used to explain the absence of facilitation in English words primed by real words with a large segmental overlap with their targets (Slowiaczek & Hamburger, 1992). In the current study, in order to judge the lexical status of target words by accessing the whole-word phonological representation, the cognitive system needs to inhibit the primed word competitors with minimal tone contrasts, even though the overlapping segments produce facilitation at the sublexical level. Consequently, the reaction times only reflect a net effect of the two opposing effects, without showing either overall facilitation or inhibition.

It should be noted that the lack of clear priming effects in minimal tone-pairs of real words can also be explained by the lexical-constraining idea. That is, there are no priming effects because lexical tones are efficient cues constraining lexical selection, so that prime words are treated as completely unrelated words than targets words, as in the baseline condition (Culter & Otake, 1999; Lee, 2007). This idea overlooks the apparent form overlaps in minimal tone-pairs, and predicts no priming effects in any minimal tone pairs. However, in the current study, pseudo-word primes facilitated lexical decisions of real-word targets with minimal tone contrasts. This result clearly shows the existence of form-based facilitation in overlapping segments, and thus endorses the lexical interference account. That is, the lack of facilitation does not mean that there is no facilitatory priming in the processing of target words due to formal overlaps, but the facilitation has been nullified by inhibition at the lexical level of phonological representation.

The lexical interference account also sheds light on some previous studies' findings that primes in minimal tone pairs of real words showed a tendency of facilitation in the processing of targets (e.g., Yip, 2001; Sereno & Lee, 2015). These data can be seen as a reflection of facilitatory priming induced by repeated activation of segmental representations at the sublexical level. However, it is less clear why the priming effect in those studies exhibited relatively stronger form-based facilitation than lexical inhibition. As we pointed out in the Introduction, the potential confounding made the results in different studies not fully comparable. Sereno and Lee (2015), for example, did not control for semantic relatedness between a prime and a target, and thus unexpected semantic priming effects in the baseline might mask the true form-priming effects. Therefore, this issue merits further studies.

3.4.2. Form priming in pseudo-word targets

The priming patterns in pseudo-word targets were largely unexpected. Our original prediction was parsimonious: as long as a prime shares some formal similarity with a pseudo-word target, the processing of this pseudo-word target should be speeded up. However, we only found facilitation in the repetition priming condition with two identical pseudo-words in a stimulus pair. In pseudo-word pairs with minimal tone contrasts, the priming effect is inhibitory rather than facilitatory. Moreover, the processing of pseudo-word targets preceded by real-word primes with minimal tone contrasts do not show clear priming effects relative to the priming baseline comprised of phonologically unrelated words and pseudo-words.

Despite limited knowledge about the priming effects in pseudo-words, explaining these unexpected priming patterns is not impossible. We propose tentative explanations by considering how phonological representations are accessed when a target word form does not have a lexical-level representation as follows. Since a pseudo-word can only access sublexical representations, the processing of pseudo-word targets primed by tonally contrasted pseudo-words must overcome stronger inhibition to access correct representations at the sublexical level. Specifically, the recognition system has to inhibit the primed segments and tones during the perception of a target pseudo-word, when the competition is induced by contrastive tones. As a result, lexical decisions of pseudo-words targets primed by pseudo-words with minimal tone contrasts slow down.

With respect to the absence of clear priming effects in pseudo-words contrasting in tones with real-word primes, we explain this effect as an outcome of two opposing effects on a par with the lexical interference account for the priming effect in minimal tone pairs of real words. Real-word primes access both the lexical and sublexical representations, but pseudo-words with minimal tone contrasts only access the sublexical representations, so the priming effects in pseudo-words can be contributed by two sources. The first source is the inhibitory priming produced by real-word primes at the sublexical level; the second source is the top-down facilitatory effect from the primed lexical-level representations on the processing of the segmental part of pseudo-words at the sublexical level. Consequently, the two effects counterbalance each other, leading to no prominent priming effects of either facilitation or inhibition.

3.4.3. Theoretical implications

Our study has two major theoretical implications for tone-word recognition. First, representations of segment patterns are at the sublexical level in the mental representation of Mandarin Chinese. Some previous studies have suggested that, similar to non-tone languages, lexical- and sublexical-level phonological representations in tone words of languages like Mandarin Chinese are also organised in a connectionist manner, which allows inhibitory connections within a level but excitatory links between two levels (McClelland & Elman, 1986; Slowiaczek & Hamburger, 1992; Zhao et al., 2011). Moreover, the between-level connections have been defined to be bi-directional, allowing bottom-up and top-down influences in parallel (see McClelland & Elman, 1986 for TRACE model based on English data; see Zhao et al., 2011, Malins, 2013 for recent revised TRACE models for Mandarin word recognition). Our study shows that, for Mandarin Chinese, a legal segment pattern is not a guarantee to access lexical-level representations. It is only a lexicalised combination of a segment pattern and a tone that can induce word responses. It can be further inferred that the sublexical level not only hosts representations of lexical tones, consonants (i.e., onsets) and vowels (i.e., rimes), but also encompasses segment patterns based on lower-order representations of segmental features (i.e., consonants and vowels).

Second, unlike the view that inhibition is induced by lexical competition and facilitation is form-based, the current study suggests that inhibitory priming is not a lexical-level specific phenomenon, nor is facilitatory priming only based on overlapping sublexical features. Although the traditional view can sufficiently explain the results in real-word targets, the unpredicted priming patterns in pseudo-word targets suggest that inhibition is not necessarily a specific lexical-level effect, and facilitation can also be induced by primed lexical-level representations. If facilitation and inhibition are level-specific priming effects, facilitatory priming should be the only effect in lexical decisions on pseudo-word targets preceded by real words or pseudo-words with minimal tone contrasts. However, our results show that the reaction times of pseudo-word targets in minimal tone pairs of pseudo-words are slower than the baseline, but real-word primes produce some facilitation in the recognition of pseudo-word targets with tone contrasts, leading to no clear priming effects. Therefore, updated theories are needed to explain the form priming in the recognition of tonal word forms.

We propose a general mechanism within a connectionist framework to explain the form priming in tone words. In a connectionist model, it is unnecessary to assume inhibition and facilitation to be level-specific, but is possible to associate the two priming effects with operations at lexical and sublexical representations via excitatory and inhibitory connections. Our basic assumptions are the following. First, in line with existing models, the phonological knowledge of spoken tone words are represented in a two-level structure: lexical and sublexical levels. The units within one level are linked by inhibitory connections, whereas the interconnections between the two levels are excitatory. Such representational organisation allows bi-directional information flow and instant feedforward and feedback communication across levels. Second, we assume that spoken real words can access both lexical and sublexical representations, whereas pseudo-words only have access to sublexical representations.

This mechanism of form priming can account for the priming patterns in the current study. More specifically, in minimal tone pairs consisting of pseudo-word primes and real-word targets, pseudo-word primes facilitate the segmental processing in real-word targets at the sublexical level, producing bottom-up facilitatory priming. The inhibition in tonally contrasting pseudo-word primes and targets is induced by competition at the sublexical level. The lack of evident priming effects in lexical decisions of real and pseudo-word targets preceded by tonally contrasting real-word primes is a net effect of the two opposite priming effects, namely facilitatory and inhibitory priming. For real-word targets, competition-induced inhibition is at the lexical level, and form-based facilitation is at the sublexical level. For pseudo-word targets, inhibition is at the sublexical level due to the competition between the primes and targets carrying identical segments with different tones, and facilitation is produced by the primed lexical-level representation via the top-down excitatory connections between lexical and sublexical representations. However, it merits future studies to examine whether our observations have task specificity or language specificity.

3.5. Conclusion

Our study investigated how lexical and sublexical representations modulate form priming patterns in words and pseudo-words derived from Mandarin Chinese during a lexical decision task. The main results were that pseudo-word primes could facilitate the processing of real-word targets with identical segments, but inhibit the recognition of pseudo-words with minimal tone pairs. Moreover, no prominent priming effects were revealed in real-word targets and pseudo-word targets primed by real words with minimal tone contrasts. These results suggest that the lack of priming effects may be an outcome of a net effect between facilitatory priming and inhibitory priming. Our finding indicates that inhibitory and facilitatory priming can be general mechanisms for Mandarin word recognition via access to phonological representations at the lexical and sublexical levels, but not specific for lexical competition and form-based sublexical processing respectively.

CHAPTER 4

Representing Segment-tone Connections in the Human Cortex: Evidence from the Rapid Hebbian Learning of Novel Tone Words[7]

4.1. Introduction

Spoken language is an important form of human communication. The capacity to learn new spoken words is a basic brain function for any individual during the life span. With exponential growth of information exchange, people encounter a large amount of novel words, such as neologism (e.g., *Google*), jargon (e.g., *app*), and terminology (e.g., *optogenetics*). In experimental contexts, word-like pseudo-words are used to represent novel words. A general practice for creating novel words is to derive novel segment patterns from existing words by replacing one

[7] This chapter is a slight adaption of Yue, Bastiaanse, and Alter (2014), published with the following title: *Cortical plasticity induced by rapid Hebbian learning of novel tonal word-forms: Evidence from mismatch negativity*.

phoneme (or syllable) with another phoneme (or syllable), following the phonological rules of the given language (e.g., the novel word **pite* versus the real English word *pipe* in Shtyrov et al, 2010; see Gaskell & Dumay, 2003 for more examples in English).

Unlike real words, novel words have neither semantic nor whole-word phonological representations in the mental lexicon. A whole-word phonological representation can be conceptualised as a network that consists of separate phonemic or syllabic representations at a low representational level. These low-level representations are interconnected as one unit at a high lexical level. This means that novel word input can activate only low-level representations of phonology such as phonemes or syllables due to the absence of lexicalised interconnections. Using behavioural and neural indicators that are sensitive to the representational distinction between real words and novel words, researchers can monitor the emergence of word-like responses to a novel word during training. This dynamics has been believed to reflect the plasticity of human brains triggered by integrating trained novel word forms into the long-term mental lexicon (for reviews, see Davis & Gaskell, 2009; Shtyrov, 2012).

The results of several recent experiments have drawn particular attention to a rapid learning mechanism in the human cortex (e.g., Shtyrov, 2011; Shtyrov et al., 2010). That is, the cortex can rapidly develop long-term memory traces for a novel word form after brief familiarisation. Furthermore, given that the learning contents were actually novel combinations of phonemes, the observed learning effects could be induced by the strengthened neuronal connections between those neural assemblies representing phonemes. This process has been explained by the Hebbian learning principle which proposes that if two neuronal assemblies, regardless of their loci in the brain, are co-activated repeatedly, the two assemblies will be associated, and develop into one cell assembly subserving the co-activation as a whole functional unit (Hebb, 1949; Pulvermüller, 1999).

Studying the rapid Hebbian learning mechanism with novel word forms can shed light on how native speakers of a language develop new lexical representations as an adaptation to the emerging new words in adulthood. However, it is not known whether novel word forms carrying lexical tones (a suprasegmental lexical cue) are also subjected to the rapid Hebbian learning mechanism in the neocortex. Our study aimed to answer this question with real and novel segment-tone patterns derived from Mandarin Chinese, a tonal language.

4.1.1. Representations of segments and lexical tones

Approximately 60-70% of languages in the world, including Chinese and Thai, use lexical tones to differentiate meanings (Yip, 2006). In Mandarin Chinese, addressed also as Putonghua in Mainland China, a monosyllabic word consists of a segmental part (i.e., onset and rime) and a suprasegmental part (i.e., lexical tone). The same segment can convey different meanings when it is combined with different tones. For example, the combinations of /ma/ with four tones (denoted with tone numbers 1, 2, 3 and 4) in Mandarin Chinese stand for completely different concepts: /ma1/ (妈) means ‘mom’; /ma2/ (麻) means ‘hemp’ or ‘numb’; /ma3/ (马) means ‘horse’; /ma4/ (骂) means ‘to curse’. A modified Trace model of Chinese word recognition suggested that the segment pattern and the lexical tone of a spoken word are represented separately in the low-level phonology, and in parallel, processed associatively as one lexical unit at the top level of word representation (Ye & Connine, 1999). This bi-directional dynamic processing structure is called ‘the Trace’ (McClelland & Ellman, 1986).

In Mandarin Chinese, when a segment pattern is not associated with a lexical tone in the mental lexicon, the combination of the two parts will yield a meaningless but phonologically plausible tonal pseudo-word. For example, /se4/(色) is a combination of /se/ and Tone 4, meaning ‘colour’, whereas /se2/, /se/ + Tone2, is a pronounceable but meaningless combination. Such novel segment-tone patterns cannot activate whole-word level representations, but are processed by the corresponding segmental and tonal representations separately, even though the segmental part could be encoded as one representational unit (Lee, 2007; Zhao et al., 2011), namely as one segment pattern. Therefore, a novel segment-tone pattern does not have a whole-word phonological representation only because of the lack of connections between the segmental and the tonal representations.

Using the representational difference between real and novel tonal word forms, the present study further addresses the rapid Hebbian learning mechanism in tonal languages with existing and novel monosyllabic segment-tone patterns derived from Mandarin Chinese.

4.1.2. Integrating novel word forms into the mental lexicon rapidly: hippocampus-based versus neocortex-based theory

A rich body of literature has reported that novel syllable sequences (i.e., novel segment patterns) can be integrated into the mental lexicon in adults and children only after overnight sleep, indexed by the emergence of lexicalised responses (Gaskell & Dumay, 2003; Davis, Di Betta, Macdonald, & Gaskell, 2009; Henderson, Weighall, & Gaskell, 2013). This means that the long-term lexical representation is formed slowly after consolidation in the neocortex. These data supported a Complementary Learning Systems (CLS) model of word learning (Davis & Gaskell, 2009; see McClelland, McNaughton, & O'Reilly, 1995 for the original CLS theory). This model proposed that in the initial phase of rapid learning, the hippocampus produces a specific memory of the learning target; in the subsequent slow learning phase, the specific memory of the target is consolidated and integrated into the generalised neocortical network via interactions between the hippocampus and the neocortex during overnight sleep (Davis & Gaskell, 2009). Accordingly, word-like responses to a novel word in the neocortex would not be possible without overnight consolidation in the slow learning phase.

Contrary to the hippocampus-based rapid learning, a small but growing number of studies have shown that lexicalised phonological representations for novel segment patterns can form rapidly in the neocortex after brief training (e.g., Shtyrov et al., 2010; Shtyrov, 2011; Sharon, Moscovitch, & Gilboa, 2011; Warren & Duff, 2014). Shtyrov and colleagues recorded electroencephalography (EEG) measurements at the scalp from native English speakers, finding enhanced electrical responses to a familiarised novel English word (**pite*) after a passive training for about 14 minutes (Shtyrov et al., 2010). The enhanced responses were believed to reflect the rapid formation of new cortical circuitries underlying long-term memory traces for the trained novel word. Furthermore, the source of the strengthened electrical responses was localised in the left perisylvian area which has been correlated with high level linguistic processing (Friederici, 2011). Taken together, these findings indicated a rapid Hebbian learning mechanism, which is neocortex based, for associating segmental representations and integrating novel word forms into the mental lexicon. This mechanism could lead to spontaneous long-term memory traces for novel words even without overnight consolidation. These data challenged the parsimonious CLS model of word learning which held that the hippocampus is responsible for rapid learning, and that neocortical plasticity is a result of slow learning.

However, the electrophysiological index (namely the novel word elicited potentials) used in these studies for quantifying cortical memory traces of spoken words can be affected by short-term neural plasticity. This is the case in auditory habituation (Thompson & Spencer, 1966) and/or sensitisation (Hughes & Rosenblith, 1957), induced by frequently repeated stimulation with the same novel word during training. These effects might have masked the true dynamics of developing new long-term lexical representations for novel words. To avoid this problem, we used the mismatch negativity (MMN) as the indicator for long-term memory traces of spoken lexical entries.

4.1.3. The electrophysiological indicator of long-term memory traces for words: the MMN

The MMN is an event-related potential (ERP) component elicited by infrequent auditory events (the ‘deviant’ stimuli) presented intermittently among another more frequently presented auditory events (the ‘standard’ stimuli). This paradigm is called an ‘oddball paradigm’. The classic method to obtain MMNs is to subtract the ERPs of the more frequent stimuli from the ERPs of the infrequent auditory events. MMNs are negatively going deflections, peaking between 100 ms to 250 ms past the time point when the infrequent event differs physically from the frequent event (Näätänen, 1995). The MMN, known since the mid 1970s (Näätänen, Gaillard, & Mäntysalo, 1978), has been linked to an automatic mechanism to discriminate any rare auditory changes, which is independent from attention. Moreover, it has also been shown that improved behavioural discrimination performance after training could be reflected by the enhanced amplitude of MMNs (Näätänen, Schröger, Karakas, Tervaniemi, & Paavilainen, 1993). With respect to using the MMN to explore the neural representation of spoken words, however, findings demonstrated that physical or acoustic distinctions between the frequent and the infrequent lexical stimuli do not necessarily determine the amplitude of language-related MMNs (Pulvermüller & Shtyrov, 2006).

The language-related MMN has been shown to be highly experience-dependent and thus regarded as a neurophysiological signature of long-term memory traces for lexical entries (Shtyrov & Pulvermüller, 2007). The latency of language-related MMNs is usually later than those of non-speech sounds, beginning approximately between 150 ms and 250 ms post stimulus onset, marking higher order lexical processing. Studies have shown that at the phoneme level, participants have larger MMN responses (MMN and MMNm, the magnetic equivalent of MMN measured with magnetoencephalography, MEG) to deviant phonemes in their native language relative to deviant foreign phonemes which are acoustically very similar to the native ones (Näätänen et al., 1997; Dehaene-Lambertz, 1997; Sharma & Dorman, 2000). In terms of word level perception, studies have revealed that the magnitude of MMNs to deviant stimuli is sensitive to lexicality. That is, MMNs to real words are larger than MMNs to pseudo-words (Pulvermüller et al., 2001; Gu et al., 2012). MMNs are also sensitive to word frequency, with high frequency words eliciting more prominent MMNs than low frequency words (Alexandrov, Boricheva, Pulvermüller, & Shtyrov, 2010; Shtyrov et al., 2011). In contrast, the lexical status of standard stimuli does not affect the size of the MMN significantly (Shtyrov & Pulvermüller, 2002). It can be concluded that when linguistic stimuli deviate at the level of phonemes or words, reduced MMNs to these stimuli indicate weak or non-existing word-specific neural assemblies underlying the long-term memory repertoire (Pulvermüller & Shtyrov, 2006; Shtyrov & Pulvermüller, 2007).

MMNs are typically distributed frontocentrally with maximum amplitude (Näätänen, 2001). Moreover, the MMN to speech deviants is also observable over the left hemisphere, sometimes even with a left-hemispheric dominance (e.g., Alho et al., 1998; Rinne et al., 1999; Näätänen et al., 1999; Koyama et al., 2000; Shtyrov, Kujala, Palva, Ilmoniemi, & Näätänen, 2000; Luo, et al., 2006). However, MMNs to non-speech deviants have shown a bilateral or a right-hemispheric dominant distribution (Paavilainen, Alho, Reinikainen, Sams, & Näätänen, 1991; Levänen, Ahonen, Hari, McEvoy, & Sams, 1996). The left-hemispheric distribution of the MMN has been attributed to the presence of lexical long-term memory traces housed in the left temporal and inferior frontal cortex (Pulvermüller & Shtyrov, 2006), which are highly specialised in the human brain for language processing (Catani, Jones, & ffytche, 2005). These spatial characteristics have enabled researchers to track neuronal changes during the learning of new linguistic patterns (e.g., Tremblay, Kraus, Carrell, & McGee, 1997). However, it should be mentioned that despite the dominant left-hemispheric pattern in the topography of the language-related MMN, both cerebral hemispheres have been found to be involved in the processing of the acoustic form of linguistic input (Näätänen, 2001; Friederici & Alter, 2004).

4.1.4. The present study

The MMN is a suitable indicator to explore whether a novel combination of a segment pattern and a lexical tone could trigger rapid plasticity in the human cortex, leading to long-term lexical representations. To elicit reliable MMNs, we created real and novel monosyllabic word forms based on Mandarin Chinese for a single-deviant oddball experiment. The single-deviant oddball design is known to generate word-specific MMN responses which can be used as a robust index of long-term memory traces (Shtyrov & Pulvermüller, 2007). Mandarin Chinese materials are ideal for this study for the following reasons. First, monosyllabic words are the most frequently used word type in Mandarin Chinese (Language Teaching Research Centre for Beijing Language Institute, 1986), so it is very likely that the expected MMN can reflect the processing at lexical level. Second, the simple onset-plus-rime structure of a syllable in Mandarin Chinese allows us to strictly match the acoustic features of novel and real word forms by manipulating only the onset part.

To monitor the MMN dynamics in novel and real tonal word forms when they are presented as the deviant 'oddball' in a stimulation sequence, a novel segment-tone pattern was presented as deviant stimuli for the experimental condition; for the control condition, a real word was delivered as deviant stimuli. On the basis of this experimental design, we expected that if rapid Hebbian learning can take place in the neocortex during exposure to novel tonal word forms, the lexicalisation of the neural connections between the segmental and tonal representations can be indexed by the MMN dynamics. Therefore, according to this neocortex-based rapid learning theory, we hypothesised that the MMN to the novel segment-tone pattern is weaker in the early exposure phase but becomes stronger in the late phase of exposure, whereas the MMN to the real word deviant does not change during the exposure (Hypothesis 1). Furthermore, we also hypothesised that changes in the MMN to the novel word form deviant between the early phase and the late phase are detectable at the midline and the left-hemispheric electrode sites (Hypothesis 2). This hypothesis was partially driven by the two EEG studies reporting enhanced electrical responses to trained novel segmental words in the midline scalp (Shtyrov et al., 2010; Shtyrov, 2011). Hypothesis 2 was also motivated by studies revealing that the neural bases of Chinese spoken word recognition have been localised in the left perisylvian areas, similar to findings with European languages (e.g., Hsieh et al., 2001; Xu, et al., 2006; Wong, Perrachione, & Parrish, 2007). Therefore, we also expected MMN changes over the left hemispheric region. However, we should point out that there is one hemodynamic study showing that part of the phonological knowledge of Mandarin Chinese can be nested in the right hemisphere (Li et al., 2010).

Conversely, according to the CLS theory, it is possible that rapid learning induces lexicalised plasticity only in the hippocampus, and long-term neocortical changes only occur in the slow learning phase after overnight consolidation. Following this line of thought, we would not expect the emergence of cortical connections between the segmental and the tonal representations for the trained novel tonal word form. Therefore, from the CLS view, we hypothesised that there is no difference in the MMN response to the novel tonal word form and the control real word during passive training (Hypothesis 3).

4.2. Method

4.2.1. Participants

Sixteen native Mandarin Chinese (i.e., Putonghua) speakers participated in the study (female: 8; average age = 25 years, SD = 2.58). In a pre-experimental questionnaire, all reported normal hearing, no history of neurological diseases, and no abuse of alcohol or drugs. All were right-handed, as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) with minor adaptations for Chinese culture. All participants gave their informed consent according to the Declaration of Helsinki under a procedure approved by the Ethical Committee of the Faculty of Medical Sciences, Newcastle University (UK). Each of them received £10 for participation.

4.2.2. Stimuli

Stimuli were created for two conditions: the novel word deviant condition (the experimental condition) and the real word deviant condition (the control condition). Each condition was delivered in one block using the oddball paradigm in which participants were exposed to two types of stimuli repeatedly: a standard stimulus with high recurrence probability and a deviant stimulus with low recurrence probability. The novel word deviant condition consisted of a real word /peng3/ as the standard stimulus and a tone-manipulated novel word form */teng3/ as the deviant stimulus, whereas the real word deviant condition was composed of a novel word */pang3/8 which was the standard stimulus and a real word /tang3/ as the deviant stimulus. This design, following Shtyrov et al. (2010), aimed to rule out any confounding effect produced by the acoustic differences and word–pseudo-word contrast.

Concerning the relatively simple syllable structures in Mandarin Chinese, the phonotactic probabilities of the two deviants were measured by the number of combinations of the two rimes (/eng/ and /ang/) with other consonants respectively. The two rimes were phonotactically matched because 20 consonants could combine with /eng/, and 21 consonants could combine with /ang/. Since MMNs for comparisons were elicited by a real word and by a pseudo-word, word frequencies were not checked.

A well-trained female native Putonghua speaker recorded 35 exemplars of /peng3/, */teng3/, and */pang3/ in a sound proof cabin (Auditory Group, Newcastle University) with a high quality Rode NT1-A microphone and E-MU 0404 recording system, digitised at a 16 bit, 44.1 thousand Hz sampling rate on a Dell E5400 laptop. One of the exemplars was carefully chosen by an experienced linguist as the base word for further manipulations and cross-splicing, regardless of the first and final five exemplars with the criterion as follows. First, the chosen word must be clearly articulated and the fundamental pitch contour can be neatly plotted by PRAAT (Boersma & Weenink, 2013). Second, to avoid potential distortion during normalisation, preferred candidates should have a duration close to 500 ms. A cross-splicing technique was used to cut and to replace sounds during the stimuli preparation. This technique guaranteed that the rimes within one condition and the onsets across

[8] The segment /pang/ can be combined with the other three tones forming meaningful words (/pang1/, 胖, swollen; /pang2/, 旁, side; /pang4/, 胖, fat). The segment /teng/ is a meaningful word when it is overlaid with Tone 1 and Tone 2 (/teng1/, 熥, to heat food with steam; /teng2/, 疼, to ache)

conditions were identical. Two novel words were recorded as base words to avoid any potential lexical impact on the articulation of the segment. For the novel word deviant condition, the /eng3/ in /peng3/ was cross-spliced with the counterpart in */teng3/. Then, for the real word deviant condition, */pang3/ was first split into /p/ and /ang3/, and deviant stimulus /tang3/ was subsequently produced by combining the /t/ in */teng3/ with the /ang3/ in */pang3/, and the background pseudo-word */pang3/ was derived by combining the /p/ in /peng3/ with the /ang3/ in */pang3/. The onsets /t/ and [p/, and the rimes /eng3/ and /ang3/ were normalised to the same duration (onset: 80 ms; rime: 420 ms), voice-onset time (80 ms) and average intensity (onset: 43 dB; rime: 46.5 dB). Considering that cross-splicing could potentially lead to unnatural formant transitions between the plosive onset and the rime (*cf.* Steinberg, Truckenbrodt, & Jacobsen, 2012), one more criterion for the base word selection was that optimal candidates for one condition (e.g., /peng3/ and */teng3/) should have very similar rime sections (e.g., the /eng3/ in /peng3/ and */teng3/) to minimise poor formant transitions.

The fundamental frequency of /eng3/ was overlaid onto /ang3/ to maximally avoid pitch differences across conditions (onset: 187 Hz, offset: 167 Hz; minimum: 123 Hz; maximum: 187 Hz). After the combination, the four experimental items were normalised to the same intensity (46 dB) with the same root-mean-square values of sound pressure (0.004 Pa). All procedures were performed in PRAAT (Boersma & Weenink, 2013). Figure 4.1 depicts the design of the stimuli, waveforms and spectrograms, and the cross-splicing procedure. Stimuli were presented in a passive oddball paradigm in two blocks. In one block, the standard stimulus was a real word (/peng3/) and the deviant stimulus was a novel word (*/teng3/), whilst in the other block, the standard stimulus was a novel word (*/pang3/) with a real word (/tang3/) as the deviant. The presentation order of the two blocks was counterbalanced for all participants. Beginning with 20 trials of standard stimuli, each block contained 1060 trials in total, of which 160 trials were the deviant stimuli (probability of 15%). Except for the first 20 trials, the delivery of deviant and standard stimuli was pseudo-randomised, meaning that at least two standard stimuli would be presented between the deliveries of two deviant stimuli. The inter-stimulus interval (ISI) was 450 ms. Control of auditory stimuli and trigger delivery were realised using Cogent 2000 developed by the Cogent 2000 team at University College London, which is a MATLAB based toolbox (The Mathworks, Natick, MA, USA).

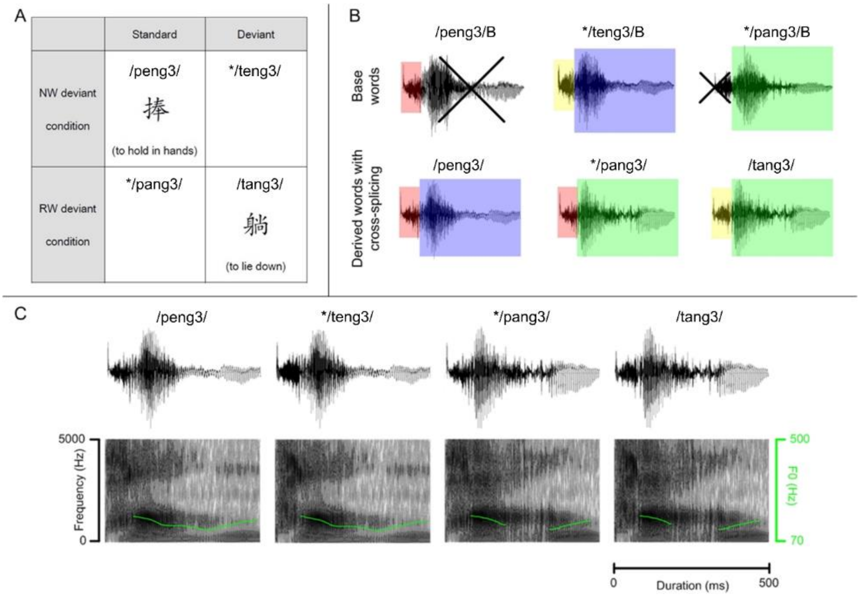


Figure 4.1 The design and normalisation of materials. (A) The orthogonal design of the oddball stimuli. NW = Novel word; RW = Real word. The scripts in the brackets designate the pronunciation of stimuli in Pinyin, below which are the corresponding Chinese characters for the RWs. The meaning of a RW is denoted in parentheses. (B) Cross-splicing procedure. The upper panel lists the three base words, /peng3/, */teng3/, and */pang3/, and their waveforms. The transparent colour backgrounds designate which part of the verbal sound has been cut. Segmentations to be replaced are overlaid with black crosses (×). The Pinyin of each base word is noted with a capital letter 'B'. The lower panel shows three words derived from cross-splicing. The background colour shows the origin of a segmentation in base words. (C) The sound waveforms, spectra, and fundamental frequency contours of the four stimuli after normalisation.

4.2.3. Procedure

The EEG experiment was conducted in an acoustically and electrically shielded booth with dim light. During the passive oddball experiment, participants were instructed to sit comfortably and watch a silent cartoon movie without subtitles at a distance of about 1.2 metres from the video monitor. They were informed that auditory stimuli would be delivered via the headphone, but that they should focus on the contents of the cartoon and ignore the auditory interference. The duration of one block was about 17 minutes and the entire EEG measurement lasted about 39 minutes, including a five-minute silent interval between two blocks. Participants were instructed to refrain from extraneous facial and body movement. A comprehension test was conducted immediately after the EEG experiment to examine whether the participant's attention was focused on the movie. In the test, each participant was required to judge the correctness of 10 statements about the contents of the cartoon by selecting either 'Right', 'Wrong' or 'Cannot remember'.

4.2.4. EEG data acquisition

The electroencephalography (EEG) of each participant was recorded using 32 Ag/AgCl electrodes mounted in an elastic cap at standard locations (the extended international 10-20 system) via SynAmps2 (Neuroscan, Sterling, VA, USA) at a 16 bit, 1000 Hz sampling rate. To record the artefacts caused by ocular movements, two electrodes were placed on the outer canthi of both eyes for horizontal movement detection, and three electrodes were situated on the right and left infraorbital ridges and on the glabella to monitor vertical activities. The ground electrode was AFz and the online reference electrode was Cz. The impedance of each electrode was kept below 5 kΩ.

4.2.5. Processing of EEG data

In the off-line data processing, EEG data were first bandpass (1-20 Hz) filtered with a finite impulse response digital filter (attenuation rate: 12 dB/octave). Epochs of 475 ms time windows were defined from 75 ms pre-stimulus onset to 400 ms post stimulus onset. Trials contaminated by artefacts whose gradient exceeded 100 μV were rejected. After the baseline was corrected to the averaged amplitude in -50 to 0 ms9(see Shtyrov et al., 2010; Gu et al., 2012 for using a 50-ms pre-stimulus baseline), ERPs evoked by standard and deviant stimuli were calculated by averaging the processed individual EEG data which were re-referenced to the average of all electrodes. In order to identify cortical plasticity correlated with the familiarisation of a novel word, we calculated the ERPs in the first and last 25% of trials to reflect the brain responses in the early and late exposure phases respectively. For the novel word deviant condition, an average of 33 out of 40 trials were obtained from each participant in the early exposure phase, and an average of 32 were obtained in the late exposure phase. An average of 32 and 33 trials in the early and late exposure phases respectively were acquired for the real word deviant condition. All procedures were performed using Scan V4.5 software (Compumedics Neuroscan, El Paso, TX, USA).

The time window for calculating MMNs was determined by paired sample *t*-test with the data collected at electrode site FCz using R V3.1 (R Foundation for Statistical Computing, Vienna, Austria). The comparisons were between the ERPs of all the standards and the ERPs of all the deviants for the real word deviant condition and the novel word deviant condition separately at each sampling point in the whole post-onset epoch (i.e., 0-400 ms, 1 point per millisecond, see Gu et al., 2012 for a similar

[9] The pre-stimulus baseline in -75–0 ms yielded unexpected three-way interactive effects related with the experiment design ($F(2, 30) = 4.75$, $p = 0.016$). Assuming that the effects may be caused by some late response, we shortened the baseline to -50–0 ms. The new baseline was not modulated by the experimental factors interactively ($F(2, 30) = 0.41$, $p = 0.617$), and guaranteed that the MMN in the subsequent time window was not affected by the baseline.

method). These comparisons can reveal typical word-elicited mismatch responses (i.e., MMNs) which are usually most prominent at frontocentral electrode sites (Shtyrov & Pulvermüller, 2002; Pulvermüller et al., 2001). Once the time window was defined, the MMNs were calculated by subtracting the ERPs of standard stimuli from those of deviant stimuli. To decrease the noise, we averaged the MMN voltage data in three regions of interest (ROIs) representing the left hemisphere (F3, FC3, C3, CP3), the midline scalp region (Fz, FCz, Cz, CPz), and the right hemisphere (F4, FC4, C4, CP4). This division of ROIs was defined a priori, which has been considered to index typically spoken-word evoked MMN components (e.g., Shtyrov, et al., 2010). The MMNs were submitted to repeated measures ANOVAs with three within-subject factors: Lexicality (real word versus novel word), Exposure-phase (early and late exposure phases), and Scalp-distribution (left hemispheric, right hemispheric and midline regions). When a three-way interaction was found, two-way repeated measures ANOVAs were conducted for Lexicality and Exposure-phase as the factors in the left, right and midline ROIs separately (*cf.* Luo et al., 2011). When a two-way interaction was identified, planned comparisons were performed between the early and late exposure phases for the two conditions separately (*cf.* Shtyrov, 2011). Huynh-Feldt correction was performed to control for sphericity violations. Since visual inspection also showed that mismatch responses in the pre-MMN epoch seemed to be a function of exposure phases (see Figure 4.4), the same statistical analyses were performed in one 50 ms time window defined post hoc, namely the N1 window in 60-110 ms10. Analyses were performed with SPSS V22 (IBM, Armonk, NY, USA).

[10] The N1 is an auditory evoked potential component. Since it is not the main focus of the current study, more information is introduced in the Discussion section.

4.3. Results

4.3.1. Behavioural results

In the cartoon comprehension test, either an erroneous judgment or a selection of 'Cannot remember' was counted as an error. The average error rate of the test was 0.1 (min = 0.0, max = 0.3, SD = 0.11). The low error rate showed that the participants were paying attention to the distracting visual task.

4.3.2. ERP results

When the ERPs were averaged for the entire exposure phase, the deviant stimuli appeared to elicit more negative responses as compared to the standard stimuli around 200 ms post onset (Figure 4.2). When the data were grouped for the early and late exposure phases separately, visual inspection revealed that in all left- and right-hemispheric and midline regions, MMNs elicited by the real word deviants and novel word deviants in the late exposure phase were consistently more negative relative to the standards around 200 ms post onset, whilst no evident negativity could be identified in the left-hemispheric region elicited by novel word deviants in the early exposure phase (Figure 4.3, Figure 4.4). In addition, the mismatch responses in the pre-MMN period seemed to be modulated by the exposure phase and the lexical status. Novel word deviants seemed to elicit more prominent negative deflections (i.e., N1) in the late phase of exposure, whereas real word deviants seemed to trigger this response pattern only in the early phase of exposure.

Definition of the MMN time window

The data showed that real word deviants elicited significantly ($p < 0.05$) more negative responses in 179–237 ms and 339–400 ms, and novel word deviants elicited significantly more negatively going ERP waves in 201–238 ms and 380–400 ms, as well as more positive responses between 268–307 ms. We considered the significant difference before 250 ms to be the MMN by referring to the MMN literature (Shtyrov & Pulvermüller, 2007).

Accordingly, a 40-ms window between 198–238 ms was selected for further MMN data analysis because it was the common time window for MMNs to both real and novel word deviants.

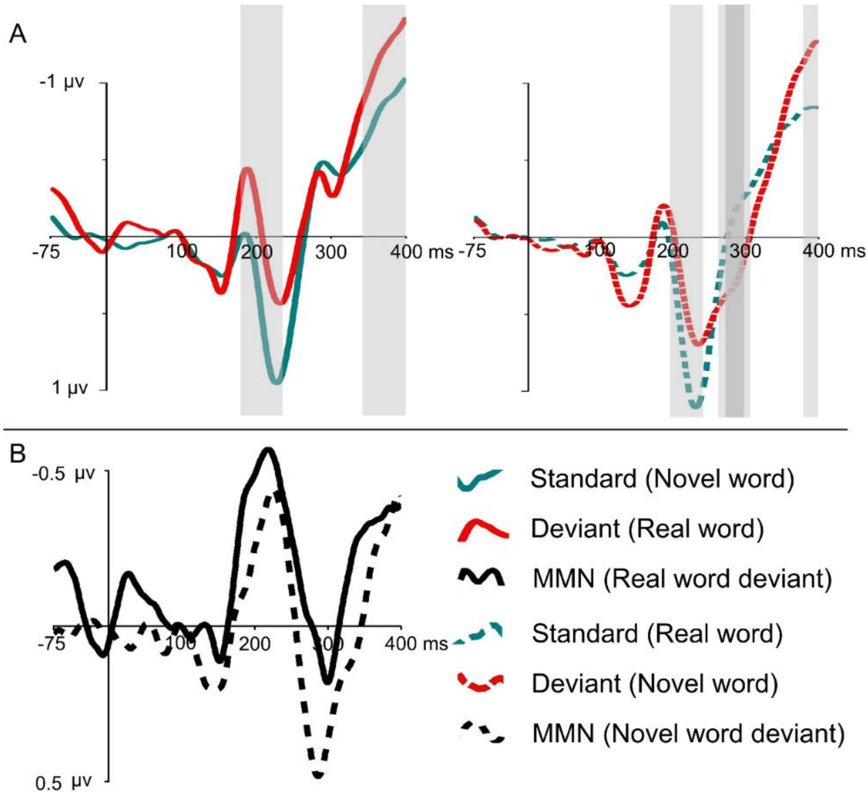


Figure 4.2 Grand-average ERPs recorded in the entire phase of exposure at FCz. (A) Electrical responses to rarely presented real word deviants and frequently presented novel word standards (left panel) and ERPs elicited by rarely presented novel word deviants and real word standards with a high recurrence probability (right panel). The grey bars designate the time window in which significant difference is identified in the *t*-tests (light grey: $p < 0.05$; dark grey: $p < 0.01$). (B) The subtracted MMNs to the real word (solid line) and novel word deviants (dash line).

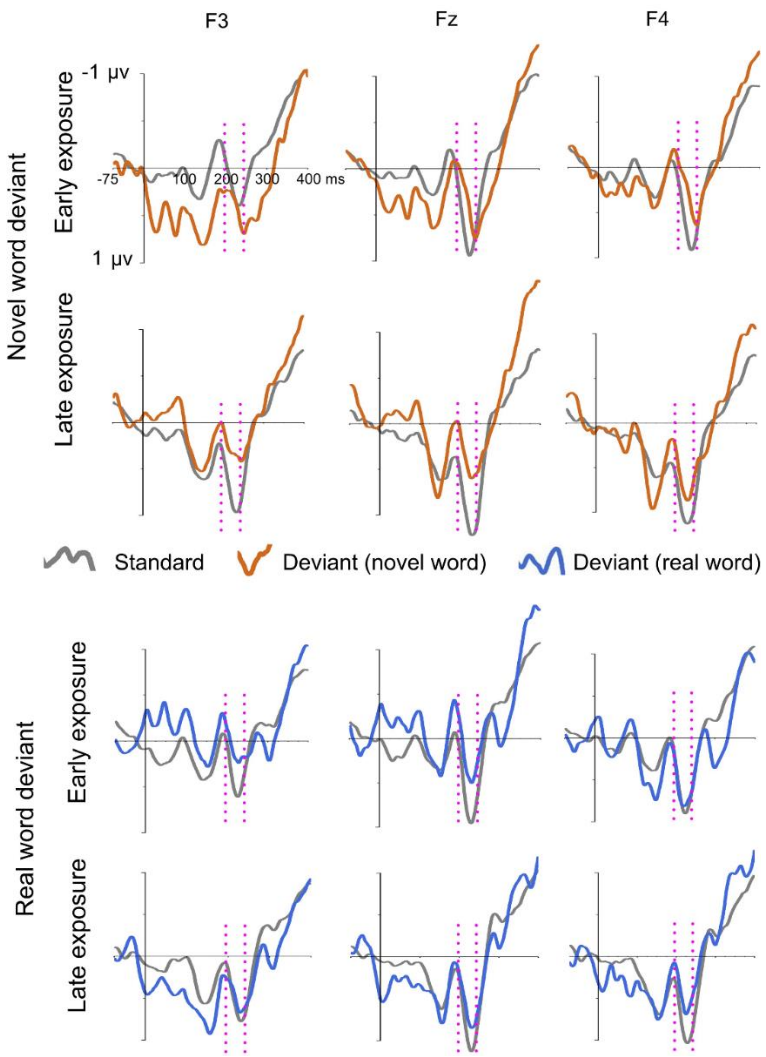


Figure 4.3 Grand-average ERP waveforms to the standards (in grey), the novel word deviants (in yellow) and real word deviants (in blue) measured at F3 (left column), Fz (middle column), and F4 (right column) representing the left, right, and the midline regions. The two pink dash lines delineate the time window of interest (198-238 ms).

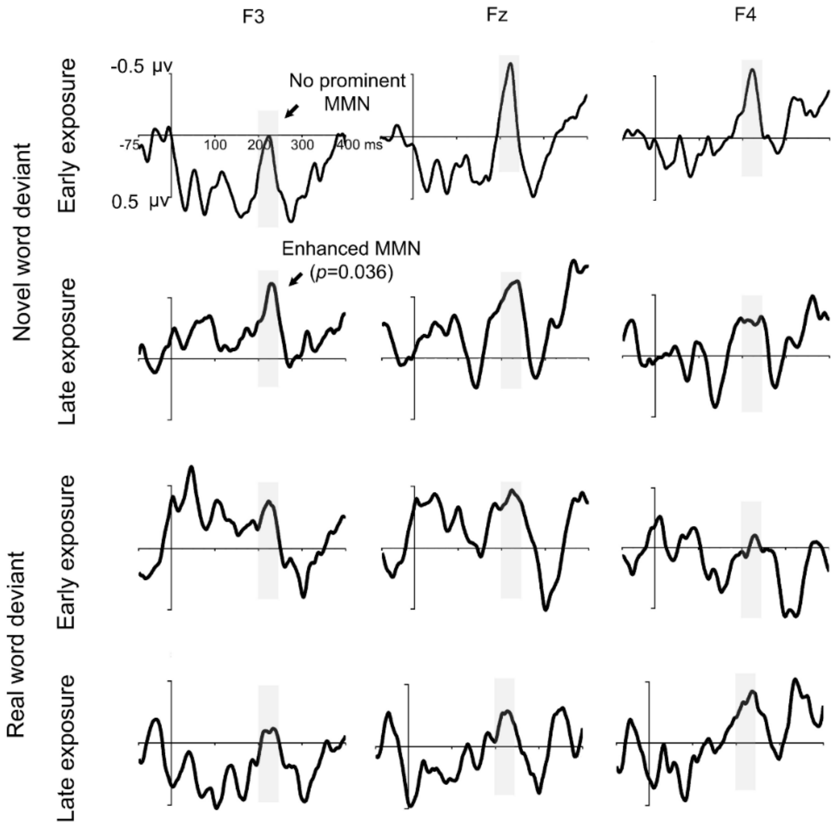


Figure 4.4 Grand-average MMN waveforms to the deviant stimuli measured at F3 (left column), Fz (middle column), and F4 (right column) for the left, right, and the midline regions. The upper two panels display the waveforms in the early and late exposure phases for the novel word learning condition, and the lower two panels show the two exposure phases for the real word learning condition. The grey areas denote the MMN window (198-238 ms). The arrows point at the two MMNs in the left scalp regions showing a significant MMN enhancement in the late exposure phase compared to the early exposure phase.

We assumed that this time window could capture the representative MMN responses to monosyllabic Mandarin Chinese word forms as an indicator of long-term memory traces for lexical input when the early and final phases of exposure were taken into account. It should be noted that this current study aimed to track the cortical dynamics of the rapid tonal word learning with the MMN, rather than to distinguish between the ERP patterns to real tonal words and novel tonal words. Thusly, we only used one representative time window for the MMN calculation, which was also helpful to increase the signal-to-noise ratio (Luck, 2005). Additionally, there was no significant difference in the pre-MMN period for the novel word deviant or real word deviant conditions.

MMN in 198-238 ms

A three-way ANOVA of averaged MMN amplitudes revealed a significant interaction among Lexicality × Exposure-phase × Scalp-distribution ($F(2, 30) = 3.85$, $p = .049$). This indicated that the three factors related to our hypotheses indeed mediate the MMN responses interactively. Separate two-way ANOVAs between Lexicality × Exposure-phase were conducted, identifying a significant interaction only in the left-hemispheric ROI, but not in the midline ROI or the right lateral ROI (left ROI: $F(1, 15) = 5.29$, $p = .036$; midline ROI: $F(1, 15) = 2.56$, $p = .130$; right ROI: $F(1, 15) = 1.36$, $p = .262$). Moreover, the interaction between Lexicality × Exposure-phase was not significant in the midline ROI ($F(1, 15) = 3.74$, $p = .07$) nor in the right lateral ROI ($F(1, 15) = .74$, $p = .40$). The left lateralised result first accepts our Hypothesis 1 which predicted an MMN enhancement in novel word deviants after passive exposure. It is also partially supportive of our Hypothesis 2 which predicted enhanced MMNs in the left hemisphere, but it rejects the other part of the Hypothesis 2 which expected learning effects in the midline ROI.

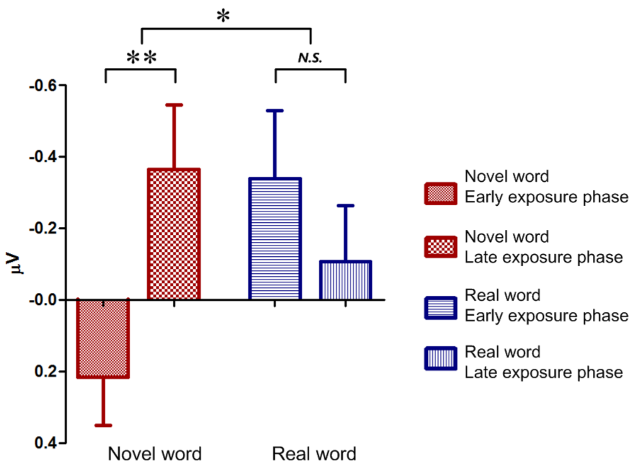


Figure 4.5 Statistical analyses on the mean values of MMN voltage to novel word and real word deviants in the early and late exposure phases at electrodes (F3, FC3, C3, CP3) in the left-hemispheric ROI. The error bars are standard error means. * p< .05, ** p < .01.

Two planned comparisons further confirmed that the Exposure-phase is a factor modulating the amplitude of the MMN to novel tonal word-form deviants rather than real tonal word deviants (novel word deviant: $F(1, 15) = 13.76$, $p = .002$; real word deviant: $F(1, 15) = .778$, $p = .392$). That is, the MMN to the novel segment-tone pattern in the early exposure phase was significantly less negative than the MMN in the late exposure phase over the left-hemispheric ROI (early phase: $M = 0.22$ μV, $SE = 0.13$; late phase: $M = -0.36$ μV, $SE = 0.18$); the averaged amplitude of MMNs to real tonal word deviants did not significantly change during the exposure (early phase: $M = -0.34$ μV, $SE = 0.19$; late phase: $M = -0.11$, $SE = 0.16$; see Figure 4.5). None of these results, however, support our CLS-model driven Hypothesis 3 which expected that no lexicalised memory traces for novel word forms can solidify before overnight consolidation. Figure 4.6 displays the dynamic changes of the topographic maps of MMNs in different exposure phases for both the novel word and real word deviant conditions.

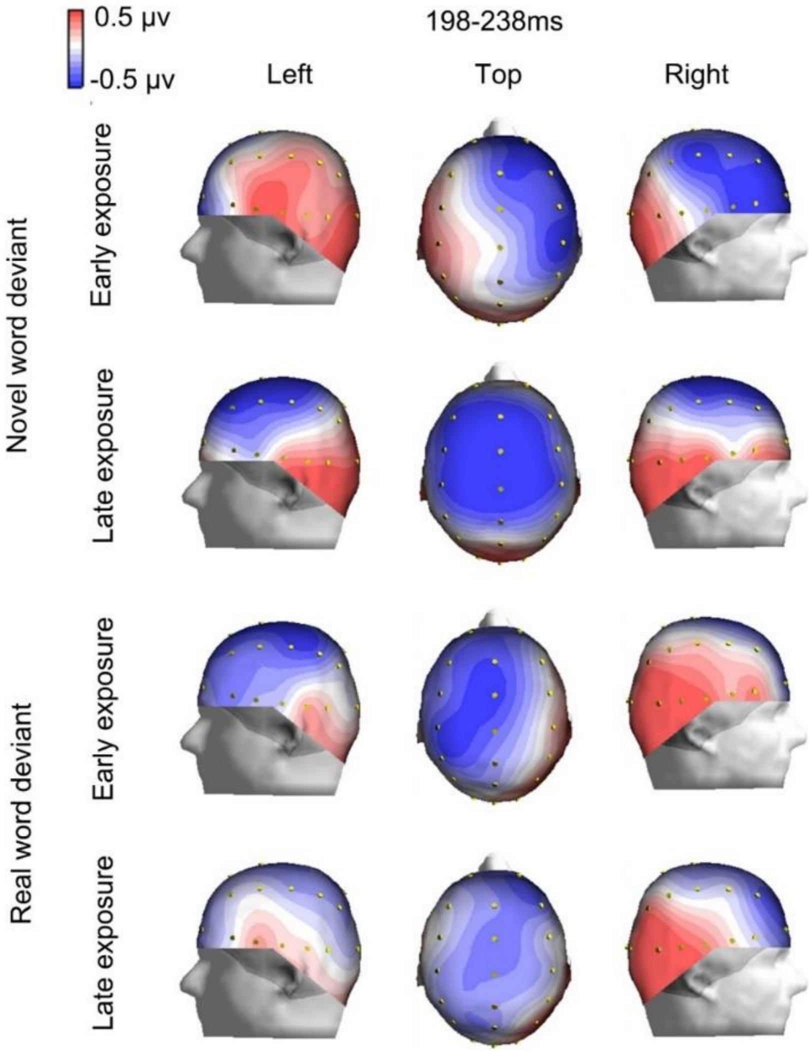


Figure 4.6 Topographic maps of MMNs averaged for the 198-238 ms time window, viewed from the left, top and right sides of a 3-D scalp (ASA 4.9, ANT, Enschede, The Netherlands). Note the enhanced MMN responses to novel word deviants over the left-hemispheric region in the late exposure phase.

N1 in 60-110 ms

Statistically significant effects were unexpectedly identified in this pre-MMN time window. A two-way interaction between Lexicality × Exposure-phase was found ($F(1, 15) = 7.62$, $p = .015$), as was a three-way interaction between Lexicality × Exposure-phase × Scalp-distribution ($F(2, 30) = 4.20$, $p = .041$). Two-way ANOVAs with Lexicality and Exposure-phase as two factors were further performed for each ROI separately (i.e., left, midline and right). An interaction between Lexicality × Exposure-phase was significant in the left ROI and the midline ROI (left ROI: $F(1, 15) = 8.21$, $p = .012$; midline ROI: $F(1, 15) = 8.15$, $p = .012$).

Planned comparisons were performed between the early exposure phase and the late exposure phase in the left and midline ROIs separately. The effect of Exposure-phase was significant for the novel word deviant condition both in the left and midline ROIs (left ROI: $F(1, 15) = 11.64$, $p = .004$; midline ROI: $F(1, 15) = 9.65$, $p = .007$) but not for the real word deviant condition (left ROI: $F(1, 15) = 2.55$, $p = .131$; midline ROI: $F(1, 15) = 1.50$, $p = .240$). The results, in part, support the visual inspection in the time window defined to represent N1 responses which showed that novel word deviants elicited stronger N1 effect in the late exposure phase as compared to the early exposure phase over the left ROI (early exposure phase: $M = 0.28$ μV, $SE = 0.13$; late exposure phase: $M = -0.34$ μV, $SE = 0.13$) and the midline ROI (early exposure phase: $M = 0.20$ μV, $SE = 0.17$; late exposure phase: $M = -0.41$ μV, $SE = 0.15$). However, the observed changes in the N1 time window for the real word condition are not supported statistically over the left ROI (early exposure phase: $M = -0.21$ μV, $SE = 0.16$; late exposure phase: $M = 0.20$ μV, $SE = 0.13$) nor the midline ROI (early exposure phase: $M = -0.26$ μV, $SE = 0.21$; late exposure phase: $M = 0.09$ μV, $SE = 0.17$). These results show that the training has strengthened the mismatch responses in the N1 time window in novel word-form deviants. This finding seems to imply some early top-down processing of the novel tonal deviant, which might be modulated by the exposure phase.

4.4. Discussion

Repeated exposure to novel lexical patterns can lead to rapid and lexicalised 'long-term' neural changes in the human cortex even outside the focus of attention (Shtyrov, 2012). Several recent experiments (e.g., Shtyrov et al., 2010; Shtyrov, 2011) also reported data indicating that the passive perceptual training in a short span of time can increase the brain's electrical responses to the trained novel words. This has been interpreted as an indication of newly formed long-term representations induced by the rapid Hebbian learning of novel spoken words. The present study obtained new electrophysiological evidence which is supportive of this neocortex-based rapid learning theory with novel segment-tone patterns.

4.4.1. Long-term memory traces for the novel word form reflected by the enhanced MMN

Using a passive oddball paradigm, we familiarised participants with a tone-manipulated novel word derived from Mandarin Chinese for approximately 17 minutes, and recorded their brain potential changes time-locked to the onset of standard and deviant stimuli. MMNs were calculated by subtracting the ERPs of the standard stimuli from the ERPs of the deviant stimuli in the 198–238 ms time window. This time window allowed us to capture the dynamics of MMN responses for the following reasons. First, this window is within the time range of the language-related MMN reported in previous studies (150–250 ms, Shtyrov & Pulvermüller, 2007). Second, even though this MMN time window is relatively later than those of many studies with non-tonal words (e.g., 150 ms in Pulvermüller et al., 2001; 140 ms in Alexandrov et al., 2010), it is similar to some prior findings that reported MMNs beginning about 200 ms after stimulus onset, using monosyllabic Chinese words with either onset contrast or tone contrast designs (e.g., onset contrast /bai4/ and /sai4/ in Luo et al., 2006; tone contrast /gai3/ and /gai4/ in Ren, Yang, & Li, 2009). Third, this kind of delayed MMN latency in studies with tonal words could be caused by the involvement of lexical-tone perception during spoken word recognition. In our study, the recognition of the deviant stimuli (i.e.,

*/teng3/ and /pang3/) required the perception of both the segment and Tone 3. Using a gating paradigm, Lai and Zhang (2008) reported that successful identification of the tone takes about 170 ms to 300 ms when the target monosyllabic Mandarin Chinese word carries an obstruent onset with a 520-ms stimulus duration. It might take less time to identify the tone of the deviants in the current oddball experiment by virtue of the repetition of the same stimulus, leading to lexical access of a deviant word form at around 200 ms post onset as is reflected by the MMN.

By comparing the MMN elicited by the novel word deviant and the real word deviant in both the early exposure phase (first 25% trials) and the late exposure phase (last 25% trials), we discovered significantly enhanced MMNs to the novel word deviant in the left-hemispheric region with the passage of training time for about 14 minutes (early phase 0.22 μV versus late phase -0.36 μV, see Figure 4.3, 4.4, 4.5), which was coherent with two previous EEG experiments indexed by oddball-evoked potentials, namely the ERPs to rarely presented deviant stimuli (Shtyrov et al., 2010; Shtyrov, 2011). The neurophysiological dynamics in our study can be considered a result of newly formed neuronal assembly representing the novel segment-tone pattern in the left hemisphere through the rapid Hebbian learning mechanism. The absence of such changes in the real word deviant condition can rule out the possibility that enhanced MMN responses to the passively familiarised novel word form are caused merely by a word-pseudo-word contrast or by the acoustic disparity between the standard and deviant stimuli, namely the onsets /p/ versus /t/.

It is also unlikely that the enhanced MMN to the novel word deviant only reflects that the brain is unable to discriminate between the two onset phonemes at the beginning of the familiarisation, but 'learns' to detect the acoustic mismatch in the late exposure phase (Näätänen et al., 1993). If this account were true, the MMN dynamics in the real word deviant condition would have exhibited the same MMN changing pattern as in the novel word deviant condition due to the same configuration of the onset consonants in the two conditions. Our data, however, show that there is no significant difference in the magnitude of the MMNs to the real word deviants in both the early and late exposure phases. Additionally, participants' attention cannot be a confounding factor for the current result, in that their attention had been successfully distracted by the cartoon as shown by the very low error rates in the comprehension task. Therefore, we believe that the MMNs obtained in the current experiment signify the long-term memory traces for tonal word forms.

Moreover, the emergence of left-hemispheric MMNs to novel word deviants modulated by the time course of training also resembles findings in a previous study reporting increased MMNs over the left hemisphere after days of training on discriminating voice-onset-time manipulated spoken word forms (Tremblay et al., 1997). This left hemispheric enhancement is in line with our Hypothesis 2 and implies that the trained novel segment-tone pattern has been integrated into the long-term mental lexicon, represented by a newly developed neuronal assembly. More specifically, through Hebbian learning rules, the repeated co-activation of the segment and the lexical tone of the target novel word */teng3/ may have strengthened the neuronal connections between the two types of phonological representations, merging the segmental and the lexical tone representations into one whole-word phonological representation.

The current results are only partially consistent with our Hypothesis 2 which predicted an enhancement of MMNs in the left-hemispheric and midline regions after familiarisation of the novel word. The reason could be that the novel word deviant in the present study was derived from a segment pattern, but prior studies that motivated us to generate our Hypothesis 2 used novel words derived from novel segment patterns (Shtyrov et al., 2010; Shtyrov, 2011). Conceptually, untrained novel segmental words (e.g., **pite*) can only be represented by lower-level phonemic knowledge, whereas novel tonal words have already had segment-pattern representations like some real tonal words. For example, the segment /teng/ in */teng3/ may have been lexicalised by real segment neighbours /teng1/, /teng2/, and /teng4/. Thus, the perception of segmental cues in a novel tonal word may activate the same segmental representations as those of real tonal words carrying different tones (e.g., */teng3/ and /teng2/). Consequently, the MMN in the midline region did not demonstrate significant differences between novel and real tonal words, even though the midline scalp is a reliable region for distinguishing between brain responses to novel segment patterns and real non-tonal words.

Interestingly, our data show that segment-pattern representations do not guarantee word-like brain responses to a novel tonal word form without training. A word-like MMN pattern was observed only after the passive perceptual training for about 14 minutes. This finding attests a finer structure of the whole-word phonological representation of tonal words in the human cortex by considering the association between segmental representations and suprasegmental features (e.g., lexical tone). That is to say, for Mandarin Chinese, the rapid cortical restructuring for novel word forms was unlikely to be induced by ‘learning’ any of the segmental or the tonal features, but the key to lexicalising a novel segment-tone pattern is to establish connections between the segmental representation and the tonal representation. This learning pattern is coherent with Hebbian learning. It means that the enhanced MMN over the left hemisphere can be related to the establishment of a connecting trace between the segment and the tone, forming a new neural assembly functionally representing the novel segment-tone pattern in the mental lexicon, supported by the rapid Hebbian learning mechanism.

4.4.2. Neocortex-based rapid Hebbian learning

The enhancement of MMNs to the novel word form in the late phase of passive exposure suggests that newly formed neural connections may be long-term per se as the typical language-related MMN has repeatedly been found to be modulated with language experience (Näätänen et al., 1997; Dehaene-Lambertz, 1997). Therefore, the neuronal plasticity reported in the current study is supportive of the recent theory claiming that the neocortex can develop long-term plasticity in the initial phase of novel word learning which is free from hippocampus-based consolidation (Shtyrov, 2012). Our finding challenges the CLS model which restricts initial, rapid word learning to the modulation of the hippocampus, and holds that long-term (i.e., lexicalised) changes in the neocortex occur only after an overnight neocortex-hippocampus interaction (Davis & Gaskell, 2009). Admittedly, due to the limitation of the method we used, it is not possible for us to directly measure the activity in the hippocampus. However, we still believe that the EEG data support our conclusions in two ways. First, the scalp-recorded EEG, which is sensitive to the activities on the cortical surface rather than deep inside the brain (i.e., hippocampus), is suitable for investigating neural changes in the cerebral cortex (Coles & Rugg, 1995). Therefore, the ERP signals we obtained mainly reflect the rapid plasticity in the neocortex, and are not contributed by the neural activities in the hippocampus.

Second, from the view of the CLS model, even if the hippocampus is activated during the experiment, it is not necessary to expect spontaneous changes in the neocortex due to the lack of overnight consolidation. Accordingly, the MMN responses during the entire exposure phase should be constant for the novel word deviant condition and the real word deviant condition as well. However, our data demonstrated an opposite pattern of MMN responses in the novel word deviant condition; namely, the amplitude of MMNs in the final phase of exposure became enhanced compared to the early exposure phase.

It must be noted that the neocortex-based rapid learning theory and the hippocampus-based rapid learning theory are not necessarily mutually exclusive, for a number of reasons. First, both of the theories agree that the rapid learning mechanism is subjected to Hebbian Learning principles (see Norman & O'Reilly, 2003 for an introduction of Hebbian learning in the CLS model). Second, given that the CLS theory is based on data showing that the hippocampus is specialised for memorising specific events while the neocortex is specialised for slowly extracting generalities from the environment based on accumulated experience, this theory could not predict the neocortical plasticity induced by rapid learning when the novel input has already had generalised cortical representations (e.g., novel word forms derived from new combinations of existing phonological units). This learning situation is exactly what the brain has to face during rapid novel word learning. Therefore, it is not possible to refute the existence of a neocortical rapid learning mechanism with the data in favour of the CLS theory. As such, neocortex-based rapid learning may be an independent module that can be added to the CLS model for spoken word learning, allowing direct plasticity in the neocortex for developing long-term memory traces for novel word forms, which is free from hippocampus-driven consolidation.

Neuropsychological data have also attested a hippocampus-independent rapid learning mechanism for novel words. For example, Sharon et al. (2011) found that associations of unknown animals and names could be successfully learned in severe amnesic patients, who had profound hippocampus damage but relatively spared left temporal lobe, through an implicit learning task. Taken together, we postulate that the hippocampus-based and the neocortex-based rapid learning mechanism may co-exist but work independently in the central nervous system. Furthermore, when the phonological features of a novel lexical pattern can be represented separately by the existing generalised cortical memory traces, the neocortex can rapidly develop new neural connections to encode whole-word level representations for the input, following Hebbian learning rules.

4.4.3. Effects in the N1 time window

Apart from the MMN dynamics, the present study also revealed pre-MMN effects in the N1 time window when the data were grouped into the early and late phases of exposure. The auditory N1 is an obligatory ERP component peaking negatively at about 100 ms after the stimulus onset with multiple neural sources in or near the primary auditory cortex. The N1 is functionally correlated with pre-attentive acoustic detection and encoding (Näätänen & Picton, 1987).

In the current study, the results partially support the visual observation that novel word deviants elicited more negative N1 measurements in the final phase of exposure relative to the early exposure phase. However, the N1 changing pattern to real word deviants is not supported by the statistics. Moreover, the distribution of the unexpected N1 effect was broader than the region demonstrating MMN changes (the

N1 effect: the left-hemispheric and midline ROIs; the MMN effect: the left-hemispheric ROI). Considering the discovery of rapid learning in this study, it seems reasonable to regard these results as an implication that the auditory system 'learns' the novel tonal word in the late exposure phase, and therefore, exhibits stronger N1 responses. If this interpretation is correct, these very early effects seem to be supportive evidence that the lexical status of an auditory input can be recognised at an ultra-rapid pace (MacGregor, et al., 2012).

However, this pre-MMN effect was moderated when we calculated the mismatch responses by subtracting the ERPs to the standard stimuli from the ERPs to the deviant stimuli in the entire exposure phase for the novel word deviant condition and the real word deviant condition separately. This finding subverts the preliminary learning account for the N1 dynamics in the novel word deviants because this effect is not robust enough to differentiate between brain responses to real words and pseudo-words in an oddball paradigm. Furthermore, these phase-sensitive effects may be induced by habituation of late auditory evoked potentials (e.g., N1 and P2) as a result of stimulus repetition (Woods & Elmasian, 1986; Rosburg, Zimmerer, & Huonker, 2010). That is, the N1 to the novel word deviants in the early exposure phase could be more habituated than the N1 responses in the late exposure phase (Shtyrov, 2011).

In contrast, the MMN has been shown to be a robust indicator of automatic lexical access. Therefore, the N1 effects cannot be functionally equal to the MMN, nor determinant of the subsequent MMN dynamic pattern, even though the effect in the N1 window might reflect some early top-down processing of monosyllabic Mandarin Chinese spoken word forms when the exposure phase is divided. Given that habituation of auditory evoked potentials has been thought to be a reflection of short-term post-synaptic plasticity (Wang & Knösche, 2013), the unexpected but interesting N1 changes could merit future research which might reveal how the strength of long-term neuronal connections influences the short-term neuronal plasticity within the neural network representing speech knowledge, as indexed by the MMN and the N1 respectively.

4.4.4. Limitations

Although the current study provides evidence that the rapid plasticity of the human cortex forms long-term memory traces for novel segment-tone patterns by strengthening the connection between the segment and the lexical tone representations over left hemisphere, our data cannot reveal fine-grained spatial information of the neural anatomy correlated with rapid tonal word learning. Moreover, research using neurophysiological indicators of MMNs or oddball-evoked potentials to monitor the cortical plasticity must employ acoustically similar real and novel words to construct an oddball paradigm. Consequently, it is difficult to decide whether the 'learning' effect is caused by randomly presenting the target novel word or by the recurrence of the real word as the background carrying similar phonemes, syllable patterns and/or lexical tones.

Finally, this study has low ecological validity, in that it only addressed the learning of phonological word forms without touching upon the learning of word meanings. Nevertheless, it is undoubtedly fundamental to acquire phonological word forms before the development of an intact mental lexicon with semantic features. Moreover, our findings have several implications for non-experimental word learning. First, it helps to explain how some dialects with tonal variations are quickly recognised and naturally learned without classroom teaching. Second, with respect to learning a tonal language in adulthood as a foreign language, it seems to be worthwhile to focus on training segments and lexical tones separately, which could facilitate the learning of new words derived from a trained segment and a lexical tone. Third, this learning paradigm may be applicable for language impaired populations to re-develop memory traces for tonal word forms with minimal active effort. Unfortunately, to the best of our knowledge, there is no documentation reporting studies on these issues. Therefore, it would be interesting to further the current research in these directions and to examine the role of the neocortex-based rapid learning mechanism in both phonological and semantic learning.

Even though our findings are limited to the learning of novel phonological patterns, the current study still reveals reliable evidence of neocortical rapid plasticity induced by the learning of novel tonal words, taking advantage of the MMN paradigm. First, this paradigm has been repeatedly shown to elicit automatic and word-specific electrical responses as a robust index of long-term memory traces of spoken word knowledge (Shtyrov & Pulvermüller, 2007). Second, using this paradigm, we can maximally match the physical features of the deviant and standard stimuli to minimise the possibility that the MMN responses are caused by the acoustic difference rather than the lexical access.

4.5. Conclusion

Using novel segment-tone patterns derived from Mandarin Chinese, we tested whether passive perceptual exposure to novel tonal word forms leads to rapid reconstruction of long-term neuronal connections. We found enhanced MMNs to the novel tonal word over the left-hemispheric electrodes after a 14-minute exposure. The results suggest that repeated exposure to a novel segment-tone pattern can trigger the rapid development of long-term memory traces of novel tonal word forms by strengthening neural connections between the segmental and tonal representations, following Hebbian learning rules. Our MMN findings are supportive of a recent neocortex-based rapid learning theory for spoken word learning, arguing that the initial phase of learning (i.e., rapid learning) could lead to long-term changes in the neocortex without overnight consolidation driven by hippocampal structures.

CHAPTER 5

Early Access to Lexical-level Representations of Tonal Word Forms: An Auditory Habituation Study

5.1. Introduction

Recognising a spoken word relies on rapid and accurate access to its lexical-level representation. The phonological part of lexical-level representations is composed of sound patterns corresponding with the spoken forms of words (Swingley & Aslin, 2000). Lexical-level representations of phonology can be based on low-level representations of phonemes and other sublexical phonological knowledge (Dahan & Magnuson, 2006). How these representations are accessed during spoken word recognition has attracted much attention in psycholinguistic research over the past decades.

Using neurophysiological recording techniques such as electroencephalography (EEG) and magnetoencephalography (MEG), a number of studies have revealed evidence that is supportive of rapid access to lexical-level representations. That is, the input of some sublexical features (e.g., phonemes and syllables) can rapidly activate neural responses related to lexical-level processing, even before attention sets in (Shtyrov & Pulvermüller, 2002; Pulvermüller et al., 2001; Alexandrov et al., 2010; Shtyrov et al., 2011; Gu et al., 2012). Many of these studies used mismatch negativity (MMN) to index activation of lexical-level representations. The MMN, measured with EEG and MEG, is a neural response to any *deviant* auditory stimulus with low recurrence probabilities in a sequence of more frequently presented *standard* stimuli (Näätänen et al., 1978; Näätänen, 1995). In a seminal study by Pulvermüller and colleagues (Pulvermüller et al., 2001), the researchers found that deviant Finnish words elicited more prominent MMNs than deviant pseudo-words in native speakers. The peaking latency of the MMN effect was about 150 ms after the time point (i.e., recognition point) when the deviant stimulus became physically different from the frequently presented standard stimulus (e.g., Finnish deviant *lakki* versus standard *lakko*; the recognition point is the onset of the sounds corresponding with the underlined letters). However, this effect was not identified in non-native Finnish speakers. These results indicate fast access to lexical-specific long-term memory traces represented in the cortical network.

Recently, an MEG study found the neurophysiological marker of an ultra-early divergence between the responses to passively delivered real words and pseudo-words approximately 50-80 ms after the onset of word-final consonants (MacGregor et al., 2012). In line with MMN studies, the researchers set the recognition point of a word form at the onset of the final phoneme after which the lexical status of a word pattern could be determined (e.g., *cut* and **cuck*). This result suggests that the human brain can rapidly access the neural circuitries representing word-level knowledge.

Despite the fact that these findings have contributed critical temporal information on the human brain's function for recognising auditory signals as lexical input in the perception system, it remains unclear how representations of words carrying lexical tones are accessed. This is of special interest to us because one recent study showed that a spoken monosyllabic Chinese word form could access its lexical-level representation instantly after the initial proportion of the spoken tonal word unfolds (Yue et al., 2014). Secondly, using passive experimental paradigms like an oddball paradigm, most studies presented acoustically similar words and pseudo-words alternatively in the same experimental block. In this case, the repeated occurrence of real and pseudo-word forms sharing a large proportion of their segments may lead to some rapid learning effect on the deviant evoked brain potentials (e.g., Shtyrov et al., 2010; Shtyrov, 2011; Yue et al., 2014). This impact may mask the distinctive response patterns between real word and pseudo-word recognition. In the current study, we further examined early access to neural representations of tonal word forms derived from Mandarin Chinese. To avoid any co-occurrence of acoustically similar real words and pseudo-words, we used a habituation design to measure the decrement of auditory evoked brain potentials (AEPs) with repeated word forms.

5.1.1. The phonological representation of tonal word forms in Chinese

In contrast to non-tonal Indo-European languages, approximately 60-70% of languages in the world use systemic variation of suprasegmental patterns over the same segment to express different meanings (Yip, 2006). Mandarin Chinese (also called *Putonghua* in mainland China), is a widely used language with four lexical tones. Physically, lexical tones are related to the fundamental frequency (F0) of a segment. In Mandarin Chinese, the typical F0 contours of the four tones are as follows: Tone 1 is high level flat (—), Tone 2 is high-rising (/), Tone 3 is low-dipping (∨), and Tone 4 is high-falling (\ ) (Duanmu, 2002). Monosyllabic morphemes (e.g., *zì* in Chinese) are the basic linguistic unit in Chinese. A large number of monosyllabic morphemes are words which have independent semantics and are the most frequently used word type in modern Chinese (Wang et al., 1986). For a monosyllabic word, the same segment pattern can convey different meanings when overlaid with contrastive tones. For example, /ma1/(妈) means mother, but /ma3/(马) means horse. Apart from the tones, Mandarin has simpler segmental structures than European languages. A syllable in Mandarin only allows single consonants to be the onset; for the rime, only single vowels, diphthongs or the nasalised form of the rime vowels with [n] and [ŋ] are permitted.

The linguistic characteristics of Chinese suggest different representational structures between tone and non-tone words. Although the weights of tonal and segmental cues in tone word recognition are still under debate (see Schirmer et al., 2005 and Zhao et al., 2011 for comparable roles of tones and segments; see Hu et al., 2012 and Wiener & Turnbull, 2015 for a primacy of segmental processing), many studies of Chinese languages have shown that segmental and tonal features are represented separately, based on behavioural and neural measures (Cutler & Chen, 1997; Lee, 2007; Luo et al., 2006; Li et al., 2009; Hu et al., 2012; Li et al., 2014). Therefore, it is necessary to consider lexical prosody when modeling the processing of tonal words. Several studies have attempted to construct models to explain how tonal word forms are mapped onto the phonological representations of monosyllabic Chinese words by adapting the TRACE model (Ye & Connine, 1999; Zhao et al., 2011; Malins, 2013). Despite some disparities, a consensus among these models is that phonological representations consist of two separate representational levels: a low-level representation of onsets (i.e., consonants), rimes (i.e., vowels), and tones, and a high-level representation of whole-word knowledge. The whole-word phonology at the higher representational level can be conceptualised as the lexicalised combinations of low-level representations. The division of high and low representational levels allows bi-directional top-down and bottom-up processing.

While the activation of a low-level phonological representation is incremental *per se* as a spoken word unfolds, it remains unclear when the lexical-level representation can be accessed in tonal word recognition. With the ERP technique that gives high temporal resolution, several studies found that the word-level processing of monosyllabic tone words in Chinese could take place in the N400 time window which is approximately between 300-500 ms after the onset of target auditory words (Schirmer et al., 2005; Zhao et al., 2011; Malins & Joanisse, 2012). The N400 is a negative-going ERP component, and has been functionally associated with semantic processing and/or access to long-term word memory (Kutas & Hillyard, 1984; Kutas & Federmeier, 2000; Kutas & Federmeier, 2011). These studies found larger amplitude of the N400 elicited by auditory Chinese words that mismatched a given context than the N400 for those words in a congruent context.

However, our recent study with an oddball paradigm showed that a rapidly learned tonal pseudo-word form could activate MMNs recorded over the left-hemispheric electrodes around 220 ms post stimulus onset (Yue et al., 2014). Furthermore, in a time window between 60 to 110 ms post stimulus onset, which is typical for the auditory evoked potential N1, we also identified an unexpected mismatch effect which differed between the ERP response to the 'unlearned' tonal pseudo-word form and the response to the same word form after perceptual training. These results were explained as evidence of very early access to lexical-level phonological representations of tone words in the time windows of N1 and MMN before the semantic access indexed by the late N400. However, when the MMN responses were calculated by using the data in the whole training session instead of being analysed in the early and late phases of perceptual training separately, the effect in the N1 time window vanished. This suggests that the oddball paradigm may not be a reliable method to investigate lexical-level processing in the pre-MMN time windows.

The present study further examined whether the existence of lexical-level phonological representations can induce differential brain responses to monosyllabic real and pseudo Mandarin words in the pre-MMN time windows. We targeted two AEPs, N1 and P2, which usually centre around 100 ms and 200 ms after the onset of an auditory stimulus. To avoid any confounding induced by low-level phonological processing difficulties, we created tone-manipulated pseudo-words by combining a legal segment pattern with a lexical tone to derive a novel segment-tone pattern which does not exist in Mandarin Chinese. For example, the combination of segment pattern /na/ and Tone 3, /na3/ (哪), means 'where', whereas the combination of /na/ and Tone 1, */na1/, is a meaningless pseudo-word form. Therefore, this type of pseudo-word has segmental and tonal representations at a sublexical-level without any lexical-level entry.

5.1.2. Habituation of AEPs

To investigate lexical processing in the pre-MMN time windows, we adopted an auditory short-term habituation design in the current study which showed sensitivity to N1 and P2 components. Habituation[11] has been considered an elementary form of non-associative learning, referring to a decrement in responsiveness with repeated stimulation which does not involve sensory adaptation or fatigue (Harris, 1943; Sharpless & Jasper, 1956; Thompson & Spencer, 1966; Rankin et al., 2009). Decreased responses can recover when the stimulation is withheld for a stimulus-free period (Pinsker, Kupfermann, Castellucci, & Kandel, 1970). Habituation of responses is believed to be a result of short-term neural plasticity in the respective neural network activated by the stimulation (see Kandel, 2001 for a review). The neural mechanism behind this phenomenon is that repeated stimulations, usually innocuous, reduce synaptic efficacy and connective strength between the neurons responding to input and the neurons receiving output, leading to decrement in the behavioural and electrophysiological responses (e.g., Castellucci, Pinsker, Kupfermann, & Kandel, 1970; Castellucci & Kandel, 1974).

In the auditory domain, habituation has been reliably reflected by the reduced amplitude of late-latency AEPs such as N1 and P2, or their MEG event-related field (ERF) homologues N1m and P2m[12], typically recorded around the vertex (e.g., Fruhstorfer, Soveri, & Järvilehto, 1970; Rust, 1977; Rosburg et al., 2010; Wang & Knösche, 2013). The auditory N1 is a

[11] Since the purpose of this study was to compare the response decrement patterns between real words and pseudo-words rather than to characterise the habituation phenomenon in word-related AEPs, we did not use a dishabituation design to check whether the response suppression observed in this study reached the working criteria for habituation (e.g., Rankin et al., 2009). Nevertheless, using a habituation design that has been repeatedly used in many previous studies, we believe the response decrement in the AEPs can reflect the auditory habituation with repeated stimulation of tonal word forms.

[12] The aim of this study was to examine the habituation of the neural responses within the N1 time window, so we used N1 to refer to both ERP N1 component and ERF N1m component. So did we with the P2.

negative-going ERP component, peaking about 100 ms post stimulus onset. Therefore, the N1 is also called N100 or N100m. The auditory P2, also called P200, which closely succeeds the N1, is a positive deflection, peaking about 200 ms post onset. The auditory N1 and P2 have multiple neural sources including the primary auditory cortex and prefrontal cortex (Wood & Wolpaw, 1982; Näätänen & Picton, 1987; Blenner & Yingling, 1994; Hyde, 1997; Crowley & Colrain, 2004). The auditory N1 has been correlated with both bottom-up sensory processing of the physical characteristics of stimuli (Keidel & Spreng, 1965; Jones, Longe, & Vaz Pato, 1998) and some top-down influences from cognitive functions (Curio, Neuloh, Numminen, Jousmäki, & Hari, 2000; Schafer & Marcus, 1973). There are no clear sources of the auditory P2. However, recent studies have shown that the P2 is correlated with phonological processing (Shuai & Gong, 2014; Huang, Yang, Zhang, & Guo, 2014).

Auditory habituation has consistently been reflected by the decrement in the AEP amplitudes to each repeated stimulus in a stimulation train as compared to the AEP amplitudes to the first presentation of the stimulus. The interval between each stimulus presentation (i.e., within-train interstimulus interval, ISI) is usually constant and shorter than 10 s (e.g., 1 s, 3 s, and 10 s used in Budd, Barry, Gordon, Rennie, & Michie, 1998; 0.6 s, 1.2 s, and 1.8 s in Rosburg et al., 2010). Moreover, shorter ISIs within stimulation trains can elicit larger decrements in the auditory responses than longer ISIs (e.g., Hari, Kaila, Katila, Tuomisto, & Varpula, 1982; Woods & Elmasian, 1986; Rosburg et al., 2010). This kind of habituation is also called 'short-term habituation[13] because decreased N1 responses can recover after 4 to 10 seconds of stimulation-free time (e.g., Rosburg, Haueisen, & Sauer, 2002; Rosburg et al., 2010; Wang & Knösche, 2013; Muenssinger et al., 2013). A typical decrement in short-term habituation is usually a negative

[13] The other type of habituation is auditory long-term habituation, which has also been observed by comparing N1s and P2s between experimental blocks or with an ISI longer than 10 s. However, since there is still no clear correlation between any underlying sensory processing and long-term habituation (e.g., Öhman & Lader, 1977; Woods & Elmasian, 1986), our study only focused on the short-term habituation.

exponential function of the number of stimulus presentations (Rankin et al., 2009). However, auditory habituation studies reported inconsistent patterns of response decrements even by using similar experimental designs. Some studies found classic, gradual decrease in the auditory responses with stimulus repetitions as concluded by the short-term habituation literature (e.g., Fruhstorfer et al., 1970; Woods & Elmasian, 1986; Rosburg et al., 2002). Other studies identified an unchanged decrement in the second presentation through the other presentations of a stimulus right after the first presentation, addressed as a straightforward decrement (e.g., Budd et al., 1998; Rosburg et al., 2006, 2010). Although some researchers have argued that the latter pattern is associated with neural refractoriness[14] rather than habituation (e.g., Ritter, Vaughan, & Costa, 1968; Budd et al., 1998; Rosburg et al., 2010), we still used the term 'habituation' to refer to the general response decrement because the differentiation between the two decreasing patterns in the auditory domain has not yet been established with reliable experimental methods.

A growing number of studies have shown that the short-term habituation pattern of the N1 to repeatedly presented speech sounds differs from the N1 habituation pattern to non-speech sounds. For example, one early study discovered that when the stimulus onset asynchronies (SOA) was 1000 ms between two repeated stimuli, N1 amplitudes to recurrent speech sounds showed greater habituation than those to non-speech tones (Woods & Elmasian, 1986). Specifically, the N1 amplitudes to repeated speech stimuli were about 49% of the amplitudes to the stimuli at the initial position of a train, whereas those to repeated tones were approximately 79% of the N1 to the first position stimuli. Using MEG, Teismann et al. (2004) found that repeated speech sounds elicited smaller attenuation of the auditory N1 responses in the left-hemispheric recording sites than in homologues over the right hemisphere, whereas the acoustically matched tone stimuli triggered an equivalent decrement in

[14] Neural refractoriness or refractory period refers to the brief period when a neuron is not able to respond to a successive stimulus immediately after the production of an action potential in response to the previous stimulus (Mildner, 2008).

both hemispheric regions. Traditionally, the left hemisphere has been considered to be specialised for language-related processing and the storage of phonological knowledge (e.g., Catani et al., 2005; Vigneau et al., 2006). In line with this traditional view, a robust literature has documented stronger neural responses to real words than pseudo-word forms in the left hemisphere (e.g., Shtyrov et al., 2008; Shtyrov et al., 2010; MacGregor et al., 2012; Yue et al., 2014). Since neural representations of speech knowledge has been regarded to be more abstract at a higher representational level than the representations of basic acoustic cues at a lower level, these studies suggested that the N1 habituation pattern can differ between acoustic stimuli encoded by different levels of neural representations.

5.1.3. The present study

Taking advantage of the sensitivity of AEP habituation to auditory processing in the early time windows, we used a habituation design to explore the time course of early access to lexical-level representations during tone-word recognition. To do this, we compared the habituation patterns of the auditory N1-P2 elicited by a real tone word (/ma1/) and a tonal pseudo-word form (*/na1/) with whole-head EEG recordings. If the short-term habituation of N1 can differentiate auditory stimuli represented by neural organisations at different levels (e.g., Woods & Elmasian, 1986; Teismann et al., 2004), we hypothesised that, due to the lack of neural representation of lexical-level phonology in the pseudo-word form, native Mandarin speakers would have a different N1 habituation pattern in the pseudo-word form than in the real word (Hypothesis 1), suggesting early access to the whole-word phonological representation (MacGregor et al., 2012; Yue et al., 2014). Moreover, if the neural representation of lexical-level phonology has a left-hemispheric dominance (e.g., Shtyrov et al., 2010; MacGregor et al., 2012), we hypothesised that the differential N1 habituation pattern between pseudo and real word forms is in left hemispheric recording sites (Hypothesis 2). To rule out the possibility that any observed effects were induced by the contrast between the onset consonants (i.e., /m/ vs. /n/), we also presented two real words with an identical onset contrast (i.e., /mi2/ vs. /ni2/) as control conditions. If the N1 short-term habituation pattern is modulated with lexical-level phonological representations, we hypothesised no difference between the N1 short-term habituation of the two control words (Hypothesis 3).

Even though it is unclear whether the habituation of P2 also differs between speech sounds and tone sounds, we still included the analysis of the P2 because several studies revealed that the P2 component was related to some aspects of phonological processing, such as voice-onset time (VOT) discrimination and perception (Dorman, 1974; Tremblay et al., 2001), identification of the existence of phonemic patterns (Liebenthal et al., 2010), phonological mismatch during spoken word recognition (Huang, et al., 2014), and the lexical priming effect of lexical tone recognition (Shuai & Gong, 2014). These findings, however, have only revealed the neural correlate between the P2 and the processing of sublexical features such as consonants, segment patterns and lexical tones. So far, there is no evidence that the P2 reflects any processing at the lexical level beyond phonological classification. Furthermore, there is also a lack of evidence that the short-term habituation of P2 is different between the responses to auditory stimuli having high and low levels of neural representations (i.e., speech vs. tones). Therefore, we tentatively hypothesised that the habituation patterns in the P2 window are not different between pseudo and real words (Hypothesis 4).

5.2. Method

5.2.1. Participants

Twenty native speakers of Mandarin Chinese were tested in this experiment, reporting normal hearing, no history of neurological diseases and no abuse of alcohol or drugs in a pre-experiment questionnaire (female: 12; *mean age* = 24 years; *SD* = 4). All had acquired Mandarin Chinese from early childhood without exposure to other dialects in which the pseudo word form /na1/ is meaningful (as it is in *Lanzhou* and *Xuzhou* dialects). All participants were right-handed, as assessed by the Edinburgh

Handedness Inventory (Oldfield, 1971) with minor adaptations for Chinese culture (e.g., replacing ‘knife’ with ‘chopsticks’). Written informed consent was given by participants prior to the experiment, according to the Declaration of Helsinki under a procedure approved by the Ethical Committee of the Faculty of Arts, University of Groningen (The Netherlands). Participants were paid as compensation for their participation.

5.2.2. Stimuli

To observe the habituation of AEPs, we utilised a short-term habituation design in which a stimulus is repeatedly presented at a short constant ISI in a stimulation train (Teismann et al., 2004; Rosburg et al., 2010; Wang & Knösche, 2013). By repeating a stimulation train for a certain number of times, reliable AEPs to the stimuli presented at each position in a train can be obtained by averaging the recorded responses across all stimulation trains. Habituation can be quantified by comparing the averaged amplitude of responses to the initial stimuli and the subsequently presented stimulations at each position of a train. We constructed four stimuli for the two experimental conditions and two control conditions. In one experimental condition, the stimulus was a tonal pseudo-word form */na1/; the other experimental condition consisted of a real tone word /ma1/ (mother, 妈). These two word forms were acoustically similar, allowing us to compare the habituation patterns between the AEP responses to stimuli with different lexical statuses. Moreover, two real tone words, /ni2/ (mud, 泥) and /mi2/ (mystery, 迷), were used to control for the habituation effect induced by the onset contrast (/n/ vs. /m/). See Table 5.1 for the tonal word form stimuli. The three words are all highly frequently used in modern Mandarin Chinese according to the List of Frequently Used Characters in Modern Chinese (Language Affairs Revolution Council of China, 1988).

Table 5.1 Spoken tonal word forms used in the experimental and control conditions

	Experimental conditions		Control conditions	
Pinyin	*/na1/	/ma1/	/ni2/	/mi2/
Chinese character		妈	泥	迷
Meaning		mother	Mud	mystery

During stimulus presentation, we presented one type of stimuli in each experimental block (Rosburg et al., 2010). One block consisted of 60 stimulation trains in each of which a stimulus was presented five times, namely S1, S2, S3, S4, and S5. The inter-train interval (ITI) was 5 s; the ISI was constantly 500 ms (see Figure 5.1). We delivered the four blocks in two orders: */na1/-/mi2/-/ma1/-/ni2/ and /mi2/-*/na1/-/ni2/-/ma1/. This arrangement was to avoid too many repetitions of the same lexical tone and rime vowel in two successive blocks. Additionally, in order to evade the potential rapid learning effect on the perception of a pseudo-word in a context where an acoustically similar real word was also presented (Shtyrov et al., 2010; Yue et al., 2014), the experimental block of */na1/ was always presented prior to the block of the real word /ma1/. Participants were randomly assigned to one presentation order. The two orders were counterbalanced between participants. There was a 1.5-minute stimulus-free period between the first and the second blocks, and between the third and the fourth blocks. A 2-minute break was set between the second and third blocks. Control of stimulus delivery and trigger sending were implemented by E-prime (Version 2.0, Psychology Software Tools Inc., Pittsburgh, PA, USA).

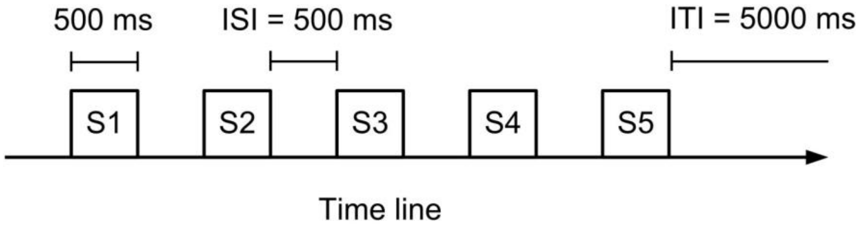


Figure 5.1 Illustration of a stimulation train of the habituation design. S1 to S5 refer to the first to the fifth presentation of stimuli.

To guarantee that the AEPs reflect natural-word-related responses, we used naturally articulated spoken word forms. A well-trained female native speaker of Mandarin Chinese produced 45 exemplars for each type of stimuli in a sound-proof recording studio with Adobe Audition (version 3.0) through a Dateq (BCS25) audiomixer, an Adirol (UA 25) USB audio capture device, and a Shure (SM27) microphone, digitised at a 16 bit, 44.1 kHz sampling rate. One of the exemplars for each condition was carefully selected according to the following criteria. First, the candidate word form should be neatly articulated with an approximate duration of 500 ms to improve the sound quality after normalisation. Second, preferred candidates with the same lexical tone (e.g., /ni2/ and /mi2/) should carry similar shapes of pitch contours.

The four chosen candidates were normalised for the same duration (500 ms). After this procedure, we replaced the fundamental frequency of candidate /ma1/ with the one of candidate */na1/ (F0 onset: 246 Hz; offset: 247 Hz; minimum: 237 Hz; maximum: 247 Hz); the F0 of candidate /mi2/ was overlaid by that of candidate /ni2/ (F0 onset: 197 Hz; offset: 280 Hz; minimum: 197 Hz; maximum: 280 Hz). Lastly, all four candidates were adjusted to the same average intensity (75 dB). To make sure that the control words served the experimental purpose well, we calculated the F2 values of the onset consonants of the four types of stimuli. The F2 is considered to be a key acoustic measure for distinguishing between /m/ and /n/ in Chinese (Li, Zhang, & Huang, 2009). The F2 values for the /m/ in normalised /ma1/ and /mi2/ were 1409 Hz and 1583 Hz, and those for the /n/ in */na1/ and /ni2/ were 2506 Hz and 2179 Hz, respectively. The data confirmed that the materials were qualified for revealing effects induced by the onset contrast. All these steps were performed with PRAAT (Boersma & Weenink, 2013). Figure 5.2 displays the oscillograms and spectrograms of the materials.

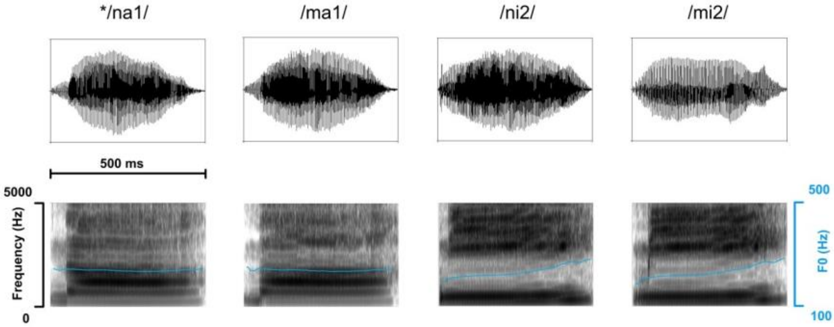


Figure 5.2 The oscillograms and spectrograms of the four word forms. Note the similarity between the envelopes of the two experimental stimuli (*/na1/, /ma1/) and those of the two control items (/ni2/, /mi2/), and the same pitch contours of the experimental items and the two control words.

5.2.3. EEG data acquisition

Participants were tested individually in a sound attenuated room with dim lighting, and seated comfortably in front of a video monitor at a distance of about 1.2 m. To control for the reported impact of selective attention of auditory stimuli on habituation (Öhman & Lader, 1972), participants were instructed to watch a silent cartoon movie and to try their best to remember the contents for a good performance in a post-experimental comprehension test. In the test, participants were asked to judge whether the 12 statements matched what they saw in the movie by choosing either 'Yes', 'No', or 'Cannot remember'. They were also asked to ignore any auditory interference (i.e., the stimuli) delivered binaurally via headphones at 55 dB SPL during the movie.

The EEG data were collected at a sampling rate of 500 Hz with the ANT acquisition system (Advanced Neuro Technology, ANT; Enschede, The Netherlands). 64-channel WaveGuard caps (the extended international 10-20 system) with shielded wires to decrease the influence of environmental electrical noise were used to record EEGs from the scalp, as well as from the two mastoids for offline data re-reference. To record artefacts caused by ocular movements and blinks, two electrodes were placed on the outer canthi of both eyes for horizontal movement detection, and another two electrodes were situated on the left infraorbital and supraorbital ridges to monitor vertical activities. The ground electrode was AFz. Average reference was used for the online EEG acquisition. The impedance of each electrode was kept below 5 kΩ. The EEG recording protocol took about 90 minutes to complete.

5.2.4. EEG data processing and analysis

In the offline data processing, the EEG recordings were first bandpass filtered at 1-30 Hz (24 dB/Oct). Then, epochs were segmented separately for the five presentations of each stimulus type for 600 ms, from -100 ms to 500 ms before and after the onset of stimulus. Trails contaminated with artefacts (> 100 µV voltage variation) were rejected. ERPs were obtained by grand averaging the remaining epochs (an average of 36 trials per presentation per condition) with the baseline corrected for the 100 ms pre-stimulus period.

The N1 and P2 peaks were detected at each electrode per participant. The N1 peak detection was confined to a time window between 80 and 170 ms, defined *a priori* to cover all latency ranges for the N1 components (Pang & Taylor, 2000; Mulert et al., 2005; Bell, Dentale, Buchner, & Mayr, 2010). The P2 peak was detected in a 90-ms time window between 170-260 ms (Hamm et al., 2012). The decrement in the N1 and P2 peak voltages induced by the habituation was indexed by the habituation coefficient R, calculated as $R = AEP_n / AEP_1$ ($n = 1, 2, 3, 4, 5$), in which AEP_n refers to the brain potential response to the nth presentation in a stimulation train (Teismann et al., 2004). The lower the value of an R, the larger the AEP response is habituated. This normalisation procedure was performed with the data collected at three electrodes, C3, Cz, and C4 in the vertex scalp region for each participant, to represent auditory processing in the left, midline and right hemispheric regions. The reason we analysed these three electrodes was that AEP habituation in the vertex region has been shown to be the most prominent (Fruhstorfer, 1970, 1971) and the electrical potential changes collected at these positions have been used to indicate typical neural activities for auditory function, especially when the lateralisation of auditory function is concerned (e.g., Prescott, Connolly, & Gruzelier, 1984; Shtyrov et al., 2010).

Three statistical analyses were performed to examine the AEP habituation effects. First, to confirm that our paradigm successfully elicited decreased auditory responses, we performed three-way repeated measures ANOVAs separately for the N1 and P2 amplitude data collected at the vertex electrode, Cz, where the auditory habituation in humans has been most intensively studied (e.g., Fruhstorfer, 1970, 1971; Rosburg et al., 2010). The three factors were Presentation (S1, S2, S3, S4, and S5), Condition (experiment vs. control), and Onset (/m/ vs. /n/). We expected a main effect of Presentation, which would suggest an overall difference between the AEPs (i.e., N1-P2) elicited by S1 and the other four repetitions in a stimulation train.

The second analysis tested whether and how the within-train short-term habituation effects were modulated by the lexical status of the stimuli. The analyses were carried out with the N1 and P2 habituation coefficient Rs separately. We first ran 2×2×3 repeated measures ANOVAs with three fixed-effect factors: Condition (experiment vs. control), Onset (/m/ vs. /n/), and Laterality (left, midline, and right hemispheric regions), by using the habituation coefficient R data through the 2nd, 3rd, 4th, and 5th presentations separately. If an interaction between the three factors can be identified, further ANOVAs are performed at the three electrodes separately to investigate which scalp region housed the different habituation effect between the real word and pseudo-word forms. If the time of Presentation shows an interaction between Condition and Onset, we investigate whether the pseudo-word forms and the real words elicited different hemispheric habituation patterns. This question was inspired by the finding that the brain's responses to speech sounds are less habituated in the left hemisphere than in the right hemisphere (Teismann et al., 2004). For this purpose, we performed four *post hoc* paired *t*-tests between the R data gleaned at C3 and C4 for the pseudo-word form, real word, and two control real word conditions separately.

If an interaction between Condition, Onset, and Laterality can be identified at one position of the four repetitions (i.e., S2, S3, S4, and S5), the last *post hoc* analysis is performed to investigate whether the paradigm used in this study triggers response decrease throughout a stimulation block due to fatigue (Muenssinger et al., 2013), and whether this kind of between-train effect interacts with the short-term habituation effect. To do this, we will first extract the averaged AEP amplitudes of the succeeding trials after artefact rejection in the first 30 trials and the last 30 trials in each type of stimuli. The responses to the first and last 30 trials represent the AEPs in the early and late experimental phases. A four-way repeated measures ANOVA is performed for a certain position in a train with four factors: Phase (early vs. late experimental phases), Condition, Onset, and Laterality. If there is a within-block decrement during the experiment, we would expect a main effect of Phase. However, studies using the current paradigm have not reported impacts from the between-train adaptation, so we did not expect an interaction between Phase and the other three factors. In all the ANOVAs, Greenhouse-Geisser correction was performed to control for sphericity violations. The significance level was set at $p < 0.05$.

5.3. Results

In the cartoon comprehension test, only an answer of 'Yes' was counted as correct. The average percentage of correct answers was 90% (*SD* = 10%). The high rate of correctness indicates that participants' attention was well distracted by the movie.

Figure 5.3 displays an overview of the epoched ERPs obtained at the vertex (Cz). Visual inspection shows decreased N1 and P2 responses around 110 ms and 200 ms post onset with stimulus repetition in all four conditions, indicating that the current auditory habitation paradigm fulfilled the purpose of eliciting short-term habituation in the auditory evoked potentials. This observation is coherent with the results of peak detection for the local maximum. The N1 and P2 peak latencies at the three representative electrode sites for the five presentations are shown in Tables 5.2 and 5.3 separately.

Descriptive statistics of the habituation coefficient R at the recording sites situated on the central scalp region showed that the N1 and P2 responses to repeated word forms were prone to decrements with stimulus repetition (see Tables 5.4 and 5.5). There seemed to be no evident differences between the habituated responses to various types of repeated stimuli, except the greater habituation of the N1 responses to S4 in the pseudo-word condition than in the other three real-word conditions (see the bold column in Table 5.4).

5.3.1. Habituation of the N1

Decrement in the N1 with repeated stimulation

As predicted, repeated measures ANOVAs with N1 peak amplitudes at Cz revealed a main effect of Presentation (F (4, 76) = 44.11, $p < .001$), which supports the amplitude difference between the AEPs to S1 and the subsequent four repetitions as shown in visual inspection. There was also an interaction of Condition × Onset (F (1, 19) = 6.56, $p = .016$), implying that the factors in the experimental design contribute to the modulation of the habituation pattern in spoken word forms.

Short-term habituation of the N1

An interaction between Condition × Onset × Laterality (F (2, 38) = 6.89, $p = .003$) was only found in the repeated measures ANOVA with the coefficient R data of the N1 peak amplitudes to S4, whereas interactions were not significant in the other presentation positions of the stimulus train (S2: F(2, 38) = 0.26, $p = .64$; S3: F(2, 38) = 0.47, $p = .58$; S5: F(2, 38) = 1.46, $p = .25$). Motivated by the three-way interaction, three ANOVAs were further conducted at the C3, Cz, and C4 electrode sites, revealing a significant two-way interaction between Condition and Onset only at C4 (F(1, 19) = 6.46, $p = .02$).

Table 5.2 The peak latencies of N1 (ms) of the five presentations at C3, Cz, and C4 for the four types of stimuli

	N1-S1		N1-S2		N1-S3		N1-S4		N1-S5	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
C3										
*/na1/	108	16	103	16	105	17	106	16	105	14
/ma1/	122	17	114	20	121	17	122	15	124	19
/ni2/	113	10	101	15	112	16	107	13	108	12
/mi2/	114	14	118	23	121	16	121	15	121	15
Cz										
*/na1/	108	18	102	16	103	14	103	16	102	13
/ma1/	121	19	114	20	119	18	118	16	121	20
/ni2/	112	9	103	13	111	17	104	15	106	11
/mi2/	114	14	120	22	115	17	123	17	120	18
C4										
*/na1/	107	16	108	16	107	13	108	17	106	18
/ma1/	122	19	113	19	120	19	113	17	121	20
/ni2/	112	9	107	14	107	15	103	16	107	12
/mi2/	113	14	117	20	118	18	122	21	119	17

Table 5.3 The peak latencies of P2 (ms) of the five presentations at C3, Cz, and C4 for the four types of stimuli

	P2-S1		P2-S2		P2-S3		P2-S4		P2-S5	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
C3										
*/na1/	198	14	198	17	198	17	192	17	198	20
/ma1/	207	15	210	22	206	18	209	25	205	15
/ni2/	195	19	194	18	197	21	193	15	191	15
/mi2/	208	19	209	21	207	20	212	25	207	17
Cz										
*/na1/	198	12	196	17	199	20	195	16	196	20
/ma1/	207	11	203	19	208	15	207	23	205	13
/ni2/	192	16	191	16	192	18	189	13	190	16
/mi2/	209	20	206	19	207	18	208	21	207	14
C4										
*/na1/	199	12	199	19	197	18	194	16	197	20
/ma1/	206	12	204	18	208	17	210	20	206	15
/ni2/	191	14	190	16	190	15	191	19	188	15
/mi2/	210	16	212	17	211	20	207	19	208	16

Table 5.4 The N1 habituation coefficient Rs of the four repetitions at C3, Cz, and C4 for the four types of stimuli

	N1-S2		N1-S3		N1-S4		N1-S5	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
C3								
*/na1/	0.52	0.28	0.56	0.26	**0.46**	0.39	0.60	0.33
/ma1/	0.62	0.36	0.68	0.30	**0.72**	0.38	0.64	0.43
/ni2/	0.63	0.43	0.75	0.54	**0.58**	0.31	0.59	0.59
/mi2/	0.48	0.23	0.60	0.30	**0.65**	0.24	0.50	0.29
Cz								
*/na1/	0.47	0.30	0.52	0.31	**0.36**	0.38	0.56	0.30
/ma1/	0.60	0.32	0.66	0.32	**0.65**	0.35	0.57	0.41
/ni2/	0.51	0.34	0.78	1.00	**0.54**	0.28	0.63	0.64
/mi2/	0.47	0.23	0.55	0.36	**0.55**	0.26	0.49	0.32
C4								
*/na1/	0.48	0.30	0.49	0.26	**0.32**	0.24	0.52	0.28
/ma1/	0.69	0.28	0.67	0.30	**0.69**	0.30	0.59	0.42
/ni2/	0.47	0.62	0.73	0.68	**0.67**	0.56	0.81	1.04
/mi2/	0.57	0.27	0.60	0.32	**0.61**	0.37	0.59	0.41

Table 5.5 The P2 habituation coefficient Rs of the four repetitions at C3, Cz, and C4 for the four types of stimuli

	P2-S2		P2-S3		P2-S4		P2-S5	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
C3								
*/na1/	0.54	0.18	0.56	0.27	0.57	0.22	0.52	0.26
/ma1/	0.61	0.20	0.59	0.35	0.66	0.30	0.65	0.37
/ni2/	0.63	0.31	0.50	0.22	0.47	0.20	0.52	0.32
/mi2/	0.66	0.78	0.61	0.46	0.49	0.26	0.70	0.41
Cz								
*/na1/	0.52	0.20	0.50	0.21	0.50	0.20	0.46	0.22
/ma1/	0.53	0.17	0.53	0.31	0.57	0.24	0.58	0.27
/ni2/	0.58	0.30	0.41	0.27	0.46	0.21	0.48	0.26
/mi2/	0.49	0.39	0.53	0.33	0.46	0.24	0.58	0.35
C4								
*/na1/	0.59	0.21	0.60	0.26	0.58	0.24	0.50	0.23
/ma1/	0.53	0.17	0.58	0.26	0.57	0.21	0.63	0.25
/ni2/	0.65	0.38	0.47	0.26	0.55	0.30	0.49	0.31
/mi2/	0.51	0.38	0.66	0.45	0.61	0.38	0.67	0.49

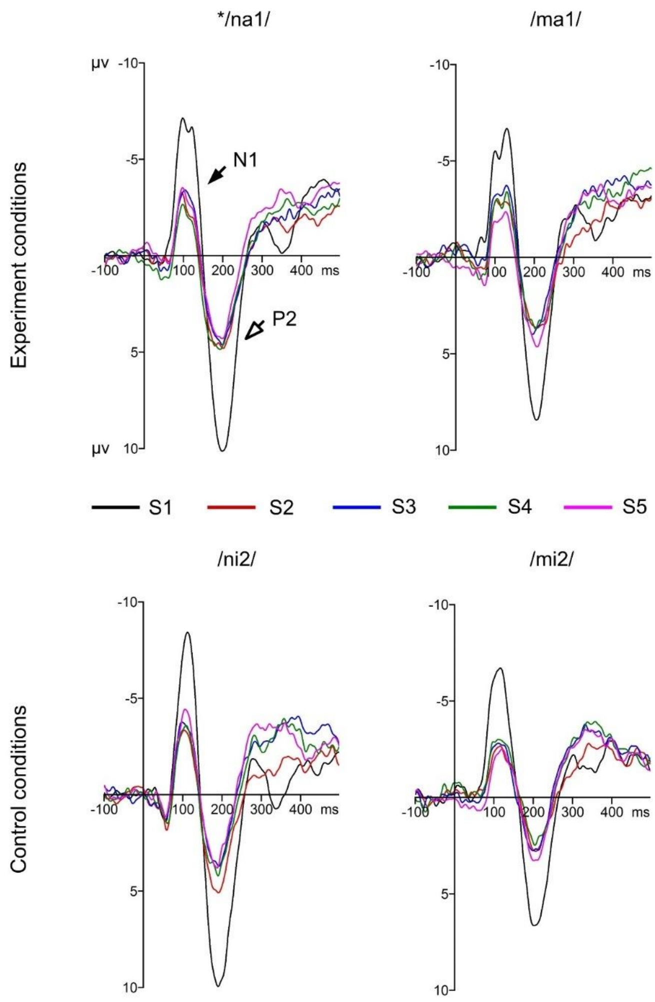


Figure 5.3 Grand averaged ERPs elicited by the four types of stimuli (*/na1/, /ma1/, /ni2/, /mi2/) with the five presentations collected at the vertex recording site (Cz). To illustrate the habituation effect, the ERP epochs to the five stimulations have been overlaid. Note the decreased N1 response (marked with a solid arrow) and P2 response (marked with a white arrow) in all four conditions compared to the ERPs to the first stimulation.

Two planned comparisons were conducted, showing that Onset was a main effect in the experimental conditions (i.e., */na1/ vs. /ma1/, $F(1, 19) = 15.20, p < .001$), but not in the control conditions (i.e., /ni2/ vs. /mi2/, $F(1, 19) = 0.34$, $p = .57$), as shown in Figures 5.4, 5.5, and 5.6. These results were consistent with the pattern identified by the descriptive statistics of the N1 habituation coefficient R. The averaged coefficient R at C4 for S4 of the pseudo-word was about 0.30 smaller than those of the experimental real word and the two control words (*/na1/: $M = 0.32$; /ma1/: $M = 0.69$; /ni2/: $M = 0.67$; /mi2/: $M = 0.61$). In addition, the latencies of N1 responses in S4 at C4 were around 110 ms post stimulus onset for the four types of stimuli, ranging from 103 to 122 ms post stimulus onset (*/na1/: $M = 107$; /ma1/: $M = 117$; /ni2/: $M = 103$; /mi2/: $M = 122$). These results are in favour of our Hypotheses 1 and 3 that the short-term habituation pattern can be used to distinguish between ultra-early lexical access in tonal real and pseudo-word forms. However, the greater N1 decrement at the right-hemispheric electrode is not supportive of our Hypothesis 2 which proposed lexical effect to be over the left-hemispheric region.

Given the difference in the N1 habituation patterns between the existent and non-existent tonal word patterns in S4, *post hoc* comparisons (paired sample *t*-tests) were performed between the coefficient R of the N1 in the 4th presentation gleaned at C3 and C4. As Figure 5.5 shows, the analyses revealed consistently weaker responsiveness to S4 of the pseudo-word form at the right electrode C4 than at the left C3 (t (19) = 2.97, $p = .008$). However, there was no difference between any C3-C4 pairs of the real tonal words (C3-C4 paired difference for */na1/: $M = 0.14$, $SD = 0.22$; /ma1/: $M = 0.03$, $SD = 0.30$; /ni2/: $M = -0.09$, $SD = 0.61$; /mi2/: $M = 0.04$, $SD = 0.30$). These results indicate that the short-term habituation of the N1 is modulated by the lexical status of a word form stimulus. They also suggest that the N1 decrement in the pseudo-word condition over the right hemispheric region is larger than in the left hemispheric scalp.

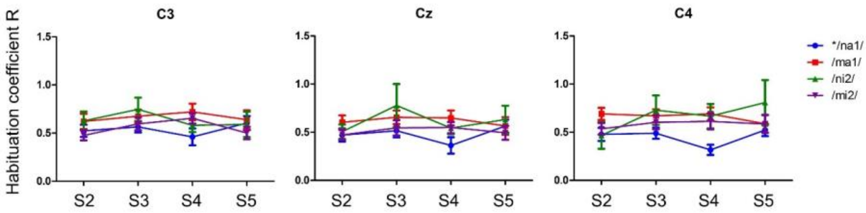


Figure 5.4 Average habituation coefficient R data at C3, Cz, and C4 in the four stimulus repetitions. The error bars are standard error means. Note the lower R value in S4 at C4 for the pseudo-word */na1/.

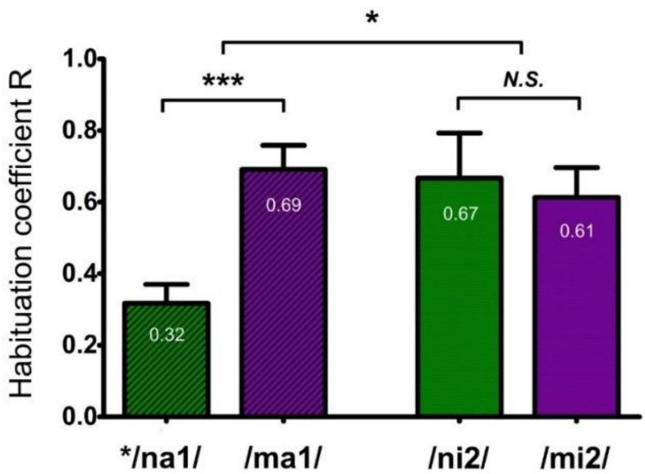


Figure 5.5 The statistical analyses on the N1 habituation coefficient R at the C4 electrode to the four types of stimuli with the 4th repetition of stimuli. The error bars are standard error means. **p* < .05, ****p* < .001, *n.s.* *p* > .05.

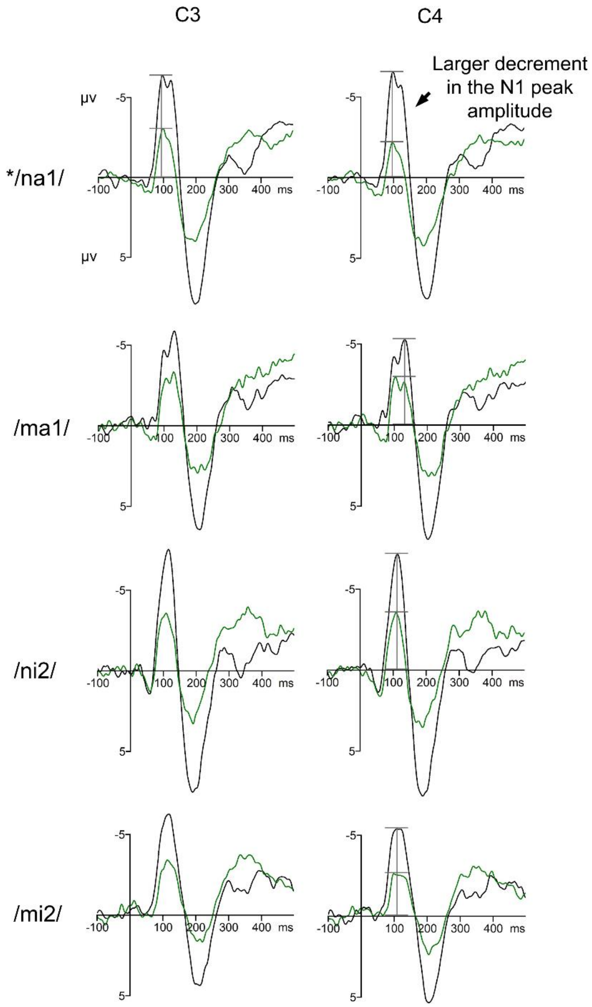


Figure 5.6 Grand averaged ERPs elicited by S1 and S4 of the four types of stimuli (i.e., */na1/, /ma1/, /ni2/, /mi2/) at the left (C3) and right (C4) central scalp regions. The AEP responses to S1 are shown with black lines, and those to S4 with green lines. The grey bars denote the difference between the N1 peak amplitudes with regard to the two times of stimulation. The upper bar shows the N1 peaks to the S1; the middle bar to S4; the lower bars are at baseline.

Between-train habituation of the N1

The last *post hoc* ANOVAs for the N1 were conducted to examine whether the short-term habituation effect at C4 was influenced by the potential decrease of the N1 occurring through the delivery of stimulation trains. The first four-way ANOVA revealed main effects of Phase (F (1, 19) = 32.84, $p < .001$) and Presentation (F (4, 76) = 35.12, $p < .001$) and a two-way interaction of Condition × Onset (F (1, 19) = 9.00, $p = .007$). These results suggest that the N1 peak amplitude becomes more attenuated in the late phase than in the early phase of one experimental block. However, the absence of an interaction of Phase × Condition × Onset implies that such amplitude attenuation through a stimulation block does not mediate the brain responses related to the factors of the experimental design.

5.3.2. Habituation of the P2

Similar to the results in the N1, the three-way ANOVA with P2 peak amplitudes at Cz identified a main effect of Presentation (F (4, 76) = 44.11, $p < .001$). Apart from this effect, the analysis revealed two more main effects of Condition (F (1, 19) = 25.99, $p < .001$) and Onset (F (1, 19) = 47.71, $p < .001$) and a two-way interaction of Condition × Onset (F(1, 19) = 5.61, $p = .029$). However, there was no significant interaction of Presentation × Condition × Onset (F(4, 76) = 0.28, $p = .83$). These effects confirm the visualised decrement in the P2 amplitudes with stimulus repetition, and also imply that the P2 component itself is sensitive to the phonological features of linguistic stimulation.

The four separate three-way ANOVAs with the P2 habituation coefficient R data through the 2nd and 5th presentations of repeated stimuli, however, did not reveal any interactions between the three factors (Condition × Onset × Laterality, S2: F(2, 36) = 0.28, $p = .83$; S3: F(2, 36) = 1.19, $p = .31$; S4: F(2, 36) = 3.07, $p = .09$; S5: F(2, 36) = 0.27, $p = .68$) [15]. This is consistent with our Hypothesis 3 that P2 habituation cannot

[15] One participant was excluded due to an extremely large difference between the P2 amplitude of the 4th and 1st presentation (the P2 to S4 is 139 times more than the 1st).

distinguish between the neural responses to pseudo and real tonal words. Furthermore, Laterality was identified as a main effect in S3 ($F(2, 36) = 4.02$, $p = .044$), S4 ($F(2, 36) = 5.13$, $p = .022$) and S5 ($F(2, 36) = 3.85$, $p = .031$). Onset reached significance as a main effect in S5 ($F(1, 18) = 4.59$, $p = .046$). These results show that the decrement of P2 amplitudes exhibits a uniform scalp distribution to repeated spoken word forms. However, the P2 habituation cannot mark the neural differences between the responses to real and pseudo-word forms.

5.4. Discussion

A growing number of studies have presented neurophysiological evidence that the human brain is able to access phonological knowledge at a lexical level very rapidly (Pulvermüller et al., 2001; MacGregor et al., 2012). By recording the whole-head EEG of native Mandarin Chinese speakers, the current study investigated whether word forms carrying lexical tones can rapidly access the lexical-level representation of phonology in the pre-attentive time window. Using an auditory habituation design, we observed the decrement of auditory AEP amplitudes in repeatedly presented spoken tonal word forms. We found that the short-term habituation of the auditory N1 elicited by the tonal pseudo-word forms was different from the habituation of the N1 responses to real words. To our knowledge, this study is the first reporting evidence that the lexical status of a tonal word form modulates the habituation of AEP component N1 between 103 and 122 ms post stimulus onset.

5.4.1. The habituation of N1 and P2

Using the current habituation design, we successfully observed the short-term habituation of spoken-word-related N1 and P2 components with repeatedly presented stimulation trains, each of which contained five identical auditory stimuli separated by a constant ISI of 500 ms. The size of the N1 habituation in our study is comparable with the results of one study of auditory habituation in speech sounds (Woods & Elmasian, 1986). In Woods and Elmasian's (1986) study, the N1 amplitudes to speech stimuli at vertex for the fourth presentation position (i.e., S4) were approximately 52% of the N1 responses to the initial stimulation (i.e., S1). Similarly, in this study, the N1 peak amplitudes of the three real-word stimuli /ma1/, /ni2/, and /mi2/ at Cz in S4 are approximately 65%, 54%, and 55% of the N1 peaks in S1.

Unlike the decreasing pattern described in classic habituation (Thompson & Spencer, 1966; Rankin et al., 2009), our results do not show a gradual decrement in either the N1 or P2 responses. That is, the N1 and P2 responses in this study have already reached an asymptotic level with the first repetition (i.e., S2), whereas the habituation of AEPs (or AEFs) in many studies has been found to complete by the 3rd presentation (e.g., Woods & Elmasian, 1986; Rosburg et al., 2002). This result seems to show that the auditory N1-P2 decrement is induced by neural refractoriness rather than habituation. However, we do not think that distinguishing between refractoriness and the classic habituation model is helpful in explaining our results. First, using similar experimental designs as the current study, previous studies have reported either gradual or straightforward declines in auditory evoked responses (e.g., gradual decrease in Woods & Elmasian, 1986; straightforward decrease in Budd et al., 1998). Therefore, the distinction between neural refractoriness and habituation in auditory responses is still far from being clear, when auditory responses are measured by far field potentials or the magnetic field from the scalp. Second, from the viewpoint of neurobiology, short-term habituation has been attributed to a decrease of releasable vesicles containing neurotransmitters during successive stimulation (Castellucci et al., 1970; Castellucci & Kandel, 1974; Sara, Mozhayeva, Liu, & Kavalali, 2002). Such a mechanism is very likely to be the physiological foundation shared by refractoriness and habituation. All in all, the results suggest that our habituation design allows us to capture the brain's electrical responsiveness to repeated word forms, reflecting how the human cortex becomes habituated to the input of existent and non-existent tonal word patterns.

5.4.2. Lexicality modulation of the spoken-word-related N1 habituation

We found that the N1 short-term habituation patterns differed between word form stimuli with and without lexical-level phonological representations. By using stimulation trains hosting five repeated stimuli, we discovered a reliably greater decrement of the N1 amplitudes to the tonal pseudo-word at the 4th stimulation position (i.e., S4) as reflected by a smaller habituation coefficient R, compared to the N1 responses to the acoustically similar real tone word at the C4 electrode. This result is in line with our Hypothesis 1 which predicted differences between the N1 short-term habituation of real and pseudo tonal word forms. The *post hoc* comparisons further revealed more persistent responsiveness in the left hemispheric electrode than the right side only in the pseudo-word condition. The pseudo-word in the current study does not involve any illegal segmental and suprasegmental cues, but still has no lexical-level representations. However, the input of tonal pseudo-words can still access neural representations of sublexical phonological elements such as consonants (i.e., onsets), vowels (i.e., rimes) and lexical tones. In contrast, the perception of real tone words activates high-order, lexicalised neural representations of whole-word phonology apart from those low-level phonological units (e.g., Zhao et al., 2011; Malins & Joanisse, 2012; Yue et al., 2014). These results suggest that the physiological disparity in the organisation of neural representations induces differential N1 habituation patterns between the real and pseudo tonal words. This finding is in line with previous studies reporting that the N1 habituation to speech sounds was different from the one to pure tones (Woods & Elmasian, 1986; Teismann et al., 2004).

Moreover, the lexical habituation effect in the N1 cannot be attributed to the onset difference between /n/ and /m/, because habituation in the control words /ni2/ and /mi2/ did not differ from each other. This result is consistent with our Hypothesis 2 which proposed no difference between the N1 habituation in the two control words (i.e., /ni2/ and /mi2/). This finding further confirms that the difference between the N1 habituation of the repeated pseudo and real tonal word forms in the experimental condition is induced by the existence of the lexical-level representation rather than by processing contrastive onsets at a low, phonemic level.

5.4.3. The temporal features of lexical access reflected by N1 habituation

The current study reveals temporal and spatial information of the neural signature related to access to the lexical-level representations of tone words. In terms of the temporal aspect, the lexical N1 habituation effect indicates that lexical access to whole-word phonological representations can occur in a very early time window which is about 110 ms after the onset of a spoken tonal word form. To be specific, the N1 peaking latencies reveal that the N1 habituation can reflect the lexical status of the word form input within a time range of 103 – 122 ms. This finding is consistent with the result of our recent ERP study reporting lexical mismatch responses in the 60 – 110 ms time window by using monosyllabic tonal word forms (Yue et al., 2014). Although the N1 time window has traditionally been shown to reflect sensory (i.e., bottom-up) processes of stimuli, our data together with a few other studies provide clear evidence of top-down processing observed in this very early time window especially in response to existent and non-existent lexical patterns. Our finding suggests that limited acoustic input of the initial portion of a tonal word form can directly access the respective whole-word phonological representation at the lexical representational level. However, with such a small amount of input, the lexical tones and rimes should be far from being recognisable (Shuai & Gong, 2014). For example, Wu and Shu (2003) reported that lexical tones could be successfully isolated approximately 160 – 200 ms after the word onset. Therefore, our results strongly suggest that lexical-level representations of phonology can be accessed before all sublexical features are recognised, at least in a highly repeated context.

Our finding is a complement to studies of rapid lexical access because most previous studies measured neural responses reflecting lexical-level processing from the onset of the last phoneme (e.g., Pulvermüller et al., 2001; Shtyrov et al., 2010; MacGregor et al., 2012). In such contexts, the brain has already received a large portion of the acoustic features of a word form stimulus before integrating the last phoneme. In contrast, we calculated electrical responses to word form stimuli from the onset point and still found lexicality modulated brain responses.

5.4.4 The spatial feature of lexical access reflected by N1 habituation

With respect to the spatial feature of lexical access, our data show a larger N1 decrement with the repetition of the tonal pseudo-word form at the right hemispheric electrode C4 relative to the N1 decrement at the contralateral C3 and to the reduction at C4 of real words. Since the only representational difference between the real and pseudo-words in our experiment is the existence of lexical-level representations, we attribute the lower responsiveness at C4 in the pseudo-word repetition to the lack of lexical-level phonological representations supported by a neural network lateralised to the right hemisphere.

However, this account opposes our Hypothesis 2, arguing that if the whole-word phonology is represented by a neural network lateralised to the left hemisphere, the lexicality modulation of the N1 decrement should be observed over the left hemispheric scalp region. Since the left-hemispheric dominance in response to lexical phonology has been well documented (e.g., Shtyrov et al., 2008; Shtyrov et al., 2010; MacGregor et al., 2012), including one study that reported lexical modulation of ERPs in the N1 time window (Yue et al., 2014), it was surprising for us to identify differential responses between the pseudo-word form and the real tonal words in the right-hemispheric recording site.

One plausible explanation is that the rapid activation of lexical-level phonology in the current study is based on the holistic word-form representation localised in the right hemisphere (see Vigneau et al., 2011 for a review). Some early neurophysiological studies have found that when spoken words are presented in a noisy context, the right hemisphere shows enhanced MMN responses relative to the left hemisphere in the pre-attentive period (Shtyrov et al., 1998; Shtyrov, Kujala, Ilmoniemi, & Näätänen, 1999). In line with these studies showing a role of the right hemisphere in speech perception, a few neurolinguistic studies have found that participants with left-hemispheric brain damage had relatively better preserved auditory word comprehension and formulaic language production, but failed to perform tasks requiring sub-lexical knowledge such as rhyme judgment (e.g., Sidtis, Volpe, Wilson, Rayport, & Gazzaniga, 1981; Gazzaniga, 2000; Sidtis, 2006). Based on these findings, it has been concluded that the right hemisphere hosts a limited lexicon consisting of holistic whole-word phonology, whereas the left hemisphere has fine-grained phonological representations (Vigneau et al., 2011).

Therefore, we consider our data as a reflection of lower responsiveness in the right hemisphere with the repetition of the pseudo-word due to the absence of any holistic word-phonological representation. That is, the N1 responses over the right hemisphere can be more active with real word repetition by virtue of the existence of holistic whole-word representations, which do not exist for pseudo-words. Furthermore, our data revealed that the N1 habituation in repeated pseudo-words was stronger at the left-hemispheric electrode relative to the right-hemispheric homologue, but not different from the habituation in real words. These results imply that in a short-term habituation design, sublexical neural representations hosted by the left hemisphere can be accessed similarly between real words and pseudo-words.

5.4.5. Position-related lexical effects on N1 habituation

The last issue that needs to be clarified in the N1 time window is that we found differential short-term habituation in the 4th presentation (i.e., S4) of the real and pseudo tonal word forms, but not in any other of the three repetitions (i.e., the 2nd, 3rd, and 5th presentations). This result is in line with two previous studies that reported divergent habituation patterns between pure tone and speech stimuli at the 4th position of a stimulation train (Woods & Elmasian, 1986; Teismann et al., 2004). However, we must point out that there is one more stimulation position in the train of our study. One of the previous studies used stimulation trains with only four positions (Teismann et al., 2004), and in the other study, 50% of stimulation trains carried a deviant stimulus in the 5th position (Woods & Elmasian, 1986).

Why did we not find the lexical effect in presentation positions other than S4? One possible reason may be that the N1 decrement through S2 to S5 is subject to some serial position effects in a stimulation train. First, one study showed that the dipole position of the N1m to repeated tones in the mediolateral cortex became stable after the 3rd presentation, even if the amplitudes of the N1m did not differ between the presentations (Rosburg, Haueisen, & Kreitschmann-Andermahr, 2004). This finding implied that, using a repetition design, brain activation after the 3rd presentation of a stimulus (e.g., the 4th and 5th) is different from the first three presentations. Second, some behavioural studies have shown that items presented in the first and the last positions of a sequence were more likely to be remembered than those in the middle (e.g., Mondor & Morin, 2004; Jutras, 2006). Therefore, it is possible that in the current study, a stimulus in the final position of a sequence (i.e., the 5th presentation) triggers extra cognitive processing other than just habituating access to the lexical-level representation. As a result, the 4th presentation becomes the most revealing position reflecting the N1 habituation modulated by the lexical status of a stimulus.

5.4.6. P2 habituation

Although the auditory P2 has often been studied together with the N1 as part of an N1-P2 Complex (see Tremblay et al., 2001), habituation of the P2 did not show the same differential patterns between responses to real and pseudo tonal words. This result is in line with our Prediction 4, which proposed no differences between the P2 decrement patterns between real words and pseudo tonal word forms. This result attests that the short-term habituation of P2 does not differentiate between auditory stimuli represented at different levels (e.g., pure tone versus speech sounds).

However, our data implies that the amplitude of P2 itself is sensitive to the phonological features of different onset consonants and rime vowels, and the short-term habituation of P2 in the 5th stimulation position can be modulated by the onset consonants. These results indicate that the P2 is a neural signature of phonemic classification via the representations at the sublexical level (e.g., Tremblay et al., 2001; Huang et al., 2014; Shuai & Gong, 2014).

5.4.7. Potential applications

There are various potential applications of the current experimental design for different experimental purposes. First, our findings may lead to the development of a novel neurophysiological method to test access to lexical-level representations of word knowledge. To date, the MMN has been used as a reliable indicator of the cortical representation of spoken word knowledge (Shtyrov & Pulvermüller, 2007; Pulvermüller & Shtyrov, 2006). However, considering the different time windows between the MMN and the N1 habituation effect and the contrastive scalp distributions of the two ERP effects (left-hemispheric MMN and right-hemispheric N1 habituation), the two measurements may reflect distinctive lexical access mechanisms. To be specific, the MMN may index access into word-level phonology which can be further divided into phonemic units. In contrast, the auditory habituation design may be especially useful to examine the holistic whole-word phonological representation in the right hemisphere which has been shown to play a critical role for language comprehension and production in patients with language impairment due to left-hemispheric damage (Vigneau et al., 2011).

Second, our experiment may lead to new investigations on auditory word perception in children experiencing developmental language disorders such as specific language impairment (SLI) and dyslexia, studies which could further delineate the neuropsychological impairments in these populations. Finally, the N1 habituation may be used as an indicator to probe the development of lexical memory traces in the cortex during word learning with a particular target of the neural representations of whole-words in the right hemisphere.

5.5. Conclusion

To conclude, our study presented evidence that the lack of lexical-level representations in tonal pseudo-word forms could result in differential N1 habituation patterns between real and pseudo-word forms derived from Mandarin Chinese in native listeners. The results suggest early access to lexical-level representations of tonal words given very limited acoustic input (between 103 and 122 ms after the word onset). These findings also imply that, differing from the traditional neural network of speech representation lateralised to the left hemispheric, the lexical-level phonological representation accessed in the early N1 time window in a habituation paradigm could have right-hemispheric dominance.

CHAPTER 6

Discussion and Conclusion

6.1. General discussion

The present thesis investigated the role of lexical representations (also called whole-word phonological representations) in tone-word recognition. Four studies were conducted to address the two general research questions by comparing the behavioural and neural responses to monosyllabic real tone-words and tone-manipulated pseudo-words. Each question will be discussed by taking the results of the four studies into account.

6.1.1. How do lexical-level representations interact with sublexical representations?

The perception of phonemes has been found to be influenced by lexical-level representations in Indo-European languages (e.g., Samuel, 2001; Magnuson et al., 2003; Norris et al., 2003; Mirman et al., 2005), and sometimes even triggers lexical responses directly (e.g., Pulvermüller et al., 2001; Shtyrov et al., 2010; MacGregor et al., 2012). Similarly, our studies found that in tone languages like Mandarin Chinese, representations at the lexical level dynamically interact with sublexical representations.

First, lexical tones are processed differently in word forms with and without lexical-level representations. The results of the experiment from Chapter 2 show that in lexical-decision tasks, when lexical frequency and phonotactic probability are controlled, lexical tones do not influence lexical decisions on real words, whereas for pseudo-words, the response latencies are different between word forms carrying Tone1 and Tone4, and between word forms carrying Tone2 and Tone4. Moreover, the lexical status of Tone4 real words is more likely to be misjudged than the lexical status of Tone1 real words. Furthermore, lexical decisions on Tone1 pseudo-words are more error-prone than decisions on Tone2 pseudo-words. Phonetically, Tone1 and Tone4 have similar onset pitch heights but contrastive pitch contours, whereas Tone1 and Tone2 have comparable shapes of the initial portion of their pitch contours but very different onset pitch heights. Therefore, these results indicate that real word recognition is sensitive to different kinds of tonal cues than pseudo-words containing the same lexical tones. Given that pseudo-words only have sublexical representations but no lexical-level representations like real words, our finding suggests dynamic top-down influence of lexical-level representations on lexical tone perception via the sublexical-level representations.

Second, lexical-level representations also have top-down influence on segmental cue perception. The results of the experiment presented in Chapter 3 demonstrate that when a prime and a target have the same segment but two distinctive tones, a pseudo-word prime facilitates (speeds up) the recognition of a real-word target, whereas a real-word prime induces lexical interference which balances out form-based facilitation. The lexical interference can be due to lexical competition between candidates carrying the same segment. These results suggest that when there are lexical-level representations, the perception of segment patterns can elicit lexical competitors, whereas when lexical-level representations are absent, the perception of segment patterns only triggers processing at the sublexical level. This finding supports the idea of the existence of excitatory interconnections between segmental and lexical-level representations (Ye & Connine, 1999; Zhao et al., 2011), which allow not only top-down facilitation but also bottom-up facilitatory priming.

The results of Chapter 4 and Chapter 5 present electrophysiological evidence that the perception of a very limited acoustic input can lead to access to lexical-level representations, at least in a highly repetitive experimental setting. In the oddball experiment (Chapter 4), participants were perceptually exposed to repeated tonal word forms, in which the deviant and standard stimuli contrasted only in the onset consonants (i.e., /t/ and /p/). In line with previous studies in English and Finnish (Pulvermüller et al., 2001; Shtyrov et al., 2010), real-word deviants elicit the neural signature of access to long-term memory traces of words. However, pseudo-word deviants only elicited word-like response patterns in the period after rapid learning occurred instead of when the pseudo-word was completely unknown.

Chapter 5 reports our study involving a habituation paradigm (Sharpless & Jasper, 1956; Rosburg et al., 2010). In our experiment, a spoken word form was presented five times in a stimulation train with a constant ISI. After a stimulus-free interval, the same stimulation train was repeated again. Auditory habituation is quantified by comparing the average ERP responses to the first presentation with responses to each subsequent repetition of the stimulus in all stimulation trains. Our data show differential habituation patterns of the auditory N1 response between a real word (/ma1/) and a pseudo-word (*/na1/). Although the two word forms have different onset consonants, the use of control conditions (two real words /mi2/ and /ni2/) rules out the possibility that the distinction in the N1 habituation patterns is caused by the onset contrast (i.e., /m/ vs. /n/). The N1 time window, around 100 ms post stimulus onset, has been considered to be too early to access lexical-level representations in recognising a tone word (Luo et al., 2006; Shuai & Gong, 2014). In contrast to this traditional view, our results confirm that the existence of whole-word phonological representations can facilitate lexical access even upon hearing the very initial portion of a tone word in an experimental context.

6.1.2 What is the neural mechanism behind access to lexical-level representations?

The temporal and spatial features of the neural mechanism underlying access to lexical-level representations are investigated by using the ERP technique. Previous studies with English and Finnish materials have shown that lexical-level representations can be automatically accessed between 50-150 ms after the recognition point of a word form (e.g., MacGregor et al., 2012; Pulvermüller et al., 2001). Moreover, speech processing has shown to be lateralised to the left hemisphere as revealed by ERPs and ERFs (Alho et al., 1998; Rinne et al., 1999; Näätänen et al., 1999; Koyama et al., 2000; Shtyrov et al., 2000; Luo et al., 2006; MacGregor et al., 2012). In Chapter 4, the MMN was used as a neural indicator of access to phonological representations at the lexical level. It is suggested that lexicality-related MMNs in monosyllabic tone words emerge around 220 ms (i.e., 198-238 ms) after the stimulus onset over the left-hemispheric recording sites. Unexpectedly, the mismatch responses also reveal neural activities related to the lexical status of stimuli in an earlier N1 time window, between 60-110 ms, over the left-hemispheric region. These results suggest that, in line with previous findings in European languages (e.g., MacGregor et al., 2012; Pulvermüller et al., 2001), lexical-level representations in tone languages can also be accessed at a remarkably rapid speed. Moreover, the lexical-level representations in Mandarin can be based on a neural network lateralised to the left hemisphere just as in languages that do not have lexical tones.

However, mismatch responses in an N1 time window vanish when the MMN responses are calculated by averaging the ERPs of the entire experimental session, rather than by analysing the early and late experimental sessions separately. This suggests that the cognitive processing in the N1 time window cannot be reliably detected by using an oddball paradigm. Therefore, in Chapter 5, we adopt a habituation paradigm to investigate the response decrement in the auditory-word evoked potentials N1 and P2. The results show a greater decrement of the peak amplitude of the N1 for the fourth presentation in the pseudo-word condition compared to the real word condition. By measuring the N1 peak latencies, it can be inferred that lexical access occurs around 110 ms (i.e., 103-122 ms) post stimulus onset. Our findings confirm that access to lexical-level representations in Mandarin is possible even in the N1 time window prior to MMN responses.

However, in contrast to the N1 effect with a left-hemispheric lateralisation as reported in Chapter 4, the differential N1 habituation patterns between the real words and pseudo-words are identified at a right-hemispheric recording site (C4). This finding suggests that the underlying neural network of lexical-level representations can also be lateralised to the right hemisphere. To reconcile the contrastive hemispheric patterns in the experiments of Chapter 4 and Chapter 5 in the same time window, we propose that these results reflect processing via two distinctive types of lexical-level representations. The right hemisphere has been shown to be involved in whole-word representation in a holistic manner (Vigneau et al., 2011; Sidtis, 2006). That is, the neural network representing a word form as one holistic and inseparable unit is lateralised to the right hemisphere. Contrastively, the lexical representation with a hierarchical architecture which can be subdivided into lower-levels of phonological units like phonemes is hosted by a neural network lateralised to the left hemisphere. On the basis of this assumption, our findings in Chapter 5 suggest that the holistic whole-word representations distributed over the right hemisphere can be accessed automatically and rapidly to differentiate between meaningful and non-existent word patterns.

6.2. Theoretical considerations: TRACE-Tone model

This thesis contributes to our understanding of how lexical-level representations in the cognitive system subserve tone-word recognition. Taking the new discoveries together, we propose a revised TRACE model for monosyllabic Mandarin word recognition, the *TRACE-Tone model* (see Figure 6.1). The TRACE-Tone model not only predicts the cognitive mechanisms of tone-word recognition, but also maps the corresponding brain functions onto hemispheric lateralisation.

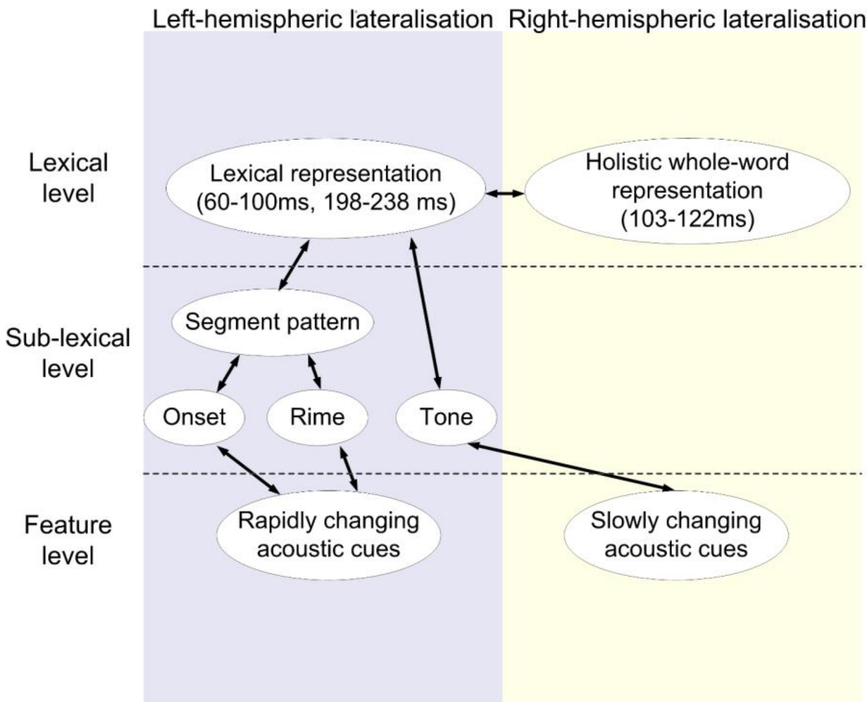


Figure 6.1 A diagram of the TRACE-Tone model of monosyllabic word recognition in Mandarin Chinese. The three representational levels are segregated with dashed lines. The arrows point out the possible directions of information flow via excitatory interconnections. The time windows of access to lexical representations and holistic whole-word representations are denoted. All modules below the lexical-level representations are sublexical modules. Modules subserved by neural networks lateralised to the left hemisphere have a violet background; modules supported by neural networks lateralised to the right hemisphere are marked with a yellow background.

The TRACE-Tone model has a three-level architecture which is in line with the model proposed in Ye and Connine (1999): a bottom level (the feature level) responsible for analysing input of acoustic cues; a middle level (the sublexical level) representing highly specified sublexical phonological units of onsets, rimes and lexical tones in Mandarin Chinese; and a top level (the lexical level) encoding lexicalised phonologies and concepts (i.e., semantics) and holistic whole-word representations. Representations of the same type of features are defined as a module. The interconnections between modules at different levels are excitatory, but the units within a module have inhibitory interconnections. The TRACE-Tone model has two processing routines at the sublexical and feature levels. A routine is defined as a set of modules and interconnections between various levels. One routine is specialised for segmental cue perception (i.e., onsets and rimes), whereas the other routine is for lexical tone perception. In the following two sections, we will first elaborate on the organisational structure of the TRACE-Tone model, and then introduce how the model operates during tone-word recognition.

6.2.1. The three-level structure

The bottom level consists of two modules. One module is for perceiving rapidly changing cues (the physical form of segmental cues) realised by a neural network lateralised to the left hemisphere. The other module responds to slowly varying cues (the physical form of suprasegmental cues) supported by a neural mechanism lateralised to the right hemisphere (Zatorre & Belin, 2001; Zatorre et al., 2002; Boemio et al., 2005; Luo et al., 2006).

Middle-level representations include phonological representations of onsets, rimes, and tones specialised for processing Mandarin tone words. All components at this level are stored in a neural network lateralised to the left hemisphere. Furthermore, the middle-level segmental domain is comprised of two separate layers: a lower layer of onsets and rimes modules, and an upper layer of segment patterns, specifically combinations of onsets and rimes. The upper layer of segmental representation is an important device because it takes charge of inspecting the fulfilment of phonotactic rules at word input. Although models based on the data in studies with Indo-European languages have treated this kind of segmental combination as a lexical-level representation, our studies show that legal combinations of segmental cues in Mandarin Chinese are not enough to elicit lexical responses (Chapters 2, 3, 4, and 5). Moreover, the existence of such combined segmental representations also explains how rapid learning of a novel segment-tone pattern can occur as reported in Chapter 4. In detail, the segment pattern and the tone of the pseudo-word */pang3/ already have sublexical representations in the mental lexicon. During perceptual training, the neural representations of the two parts are co-activated repetitively, and eventually develop a new neural representation specialised for the whole word following the Hebbian learning rules (Hebb, 1949; Pulvermüller, 1999).

The lexical-level representation entails two modules: the lexical representation and the holistic whole-word representation. The lexical representation is derived from the lexicalised combinations of segmental and tonal representations at lower representational levels; the holistic whole-word representation is composed of sound patterns that cannot be further divided into sublexical phonological units. As our study suggests, access to lexical representations is subserved by a network lateralised to the left hemisphere, whereas the network of holistic whole-word representations is lateralised to the right hemisphere. Moreover, in the current model, we do not explicitly separate the phonological component from other aspects of lexical-level representations like the semantic component (*cf.* Zhao et al., 2011). The reason for this is that participants' performance is similar in a classic lexical decision task and a lexical decision of word-sound patterns, except that the reaction times in the classic lexical decision task are generally longer than in the lexical decision task of a sound pattern (see Chapter 2). In addition, no difference in accuracy is found between the two tasks. Since all components of lexical-level representations are supposed to be exploited in a classic lexical decision paradigm, and lexical decisions of sound patterns only depend on lexical phonology, the results suggest no qualitative differences between the top-down influence of lexical phonology and other aspects of lexical representations on tone-word perception. Therefore, we propose a general module of lexical-level representations hosting all aspects of lexical knowledge, as in Ye and Connine (1999).

6.2.2. The operation of the TRACE-Tone model

The operation of the TRACE-Tone model is realised by between- and within-module interconnections. The interconnections between the three levels are defined to be excitatory, allowing bi-directional processing. Therefore, bottom-up and top-down processing can occur simultaneously in the whole range of the representational space. By contrast, in two previous models (Zhao, et al., 2011; Ye & Connine, 1999) the connections between the feature level and the sublexical level (i.e., phoneme-tone) only permit feedforward activation. In other words, acoustic input analysis can only induce activation upward to the phonological representations, and is exempted from any top-down influence. However, we assume that top-down influence can even modulate the acoustic cue analysis, in partial agreement with Shuai and Gong (2014). That is, the bi-directional excitatory interconnections in our model guarantee ultra-rapid registration of acoustic cues into the higher phonological representations and delivery of simultaneous feedback to the analysis of ongoing acoustic cues from the upper representational levels.

We also assume that units within each representational module are mutually inhibitory following the original TRACE model. This mechanism is required for rapidly mapping acoustic input onto correct representations by inhibiting potential competitors. For example, the results of the form priming study reported in Chapter 3 endorse inhibition in the sublexical modules. In that study, we did not observe the expected facilitatory priming effect in pseudo-word targets primed by pseudo-words with the same segments but contrastive tones (e.g., */zei1/-*/zei4/). In contrast, we found that the response latency of such pseudo-word targets is consistently slower than the baseline, exhibiting inhibitory priming effects. This result can be explained by the inhibitory connections within the sublexical modules.

We also take into consideration the time course of lexical-level activation for the TRACE-Tone model. After the onset of a spoken word, although the acoustic input is still limited, the perception system can rapidly access both the lexical representation and the holistic whole-word representation around 100 ms post stimulus onset (i.e., the N1 time window). Lexical representations can be further accessed around 220 ms (i.e., the MMN time window). The time course we propose here is a significant complement to previous studies in which the lexical-level representation were shown to be activated in time windows around 400 ms (N400, in Zhao et al., 2011) and 200 ms (P2, in Shuai & Gong, 2014) post stimulus onset. Moreover, since the N1 time window is too early to receive sufficient cues to recognise a word form (Shuai & Gong, 2014), these findings indicate that the lexical representation can be accessed long before the completion of sublexical feature perception, at least when the word form is repeatedly presented.

6.3. Limitation and future directions

Although the two studies with behavioural measurements (Chapters 2 and 3) and the two studies with neurophysiological methods (Chapters 4 and 5) showed the top-down influence of lexical-level representations, we could not compare the online neural responses to tonal word forms in active and passive word perception. In the lexical decision experiments, participants had to pay attention to auditory stimuli, whereas in the two ERP studies, participants were asked to concentrate on watching movies to avoid a voluntary shift of attention. Therefore, a comparison between the neural responses to tonal word forms in active and passive tasks could examine the role of attention in tone-word perception. For instance, the results of the experiments in Chapters 4 and 5 showed that access to lexical-level representations occurred as early as in the N1 time window: around 110 ms post stimulus onset. However, the N1 is known to be modulated by attention (Näätänen & Picton, 1987). Hence, more studies should be carried out to explore whether the lexicality-related N1 effects of our experiment are sufficiently robust to show up in attended experimental contexts such as lexical decision tasks.

Second, the four studies of this thesis only presented isolated word form stimuli without any preceding contexts. Moreover, in the two ERP experiments we presented only a few word forms repeatedly. This kind of experimental setting has two major methodological advantages. First, it can help obtain a higher signal-to-noise ratio of the electrophysiological signals. Second, the acoustic features of the experimental stimuli can be maximally controlled. However, reactions to such unnatural experiments may not reflect the same processes during real speech perception. Therefore, future studies of the top-down influence of lexical-level representations should be carried out by using more constraining contexts such as idioms (e.g., Hu et al., 2012) and ancient Chinese poems (e.g., Li et al., 2014).

Third, the two ERP studies as reported in Chapters 4 and 5 showed evidence of hemispheric lateralisation of the neural networks subserving lexical-level representations. However, our studies did not provide specific localisation of the neural activities. We found differential MMN and N1 response patterns between real words and pseudo words with lexical tones in both left-hemispheric regions and right-hemispheric regions under distinctive experimental settings. However, it is not possible to set up a direct link between the pertinent linguistic functions and the neural anatomy by referring to the scalp distribution (i.e., the topography) of the ERPs. To map cognitive processing onto cortical structures, techniques with high spatial resolution such as fMRT, PET, and Transcranial Magnetic Stimulation (TMS) should be used.

Last but not least, our studies add to the understanding of behavioural and neural responses to real words and pseudo-words only in Mandarin Chinese. Future research can be carried out by using other tonal languages such as Shona, Cantonese, or Thai. Moreover, it would be especially intriguing to use the few European languages with lexical tones, such as the west Limburgian dialects of Dutch (see Fournier, Gussenhoven, Jensen, & Hagoort, 2010 for the dialect of Roermond in The Netherlands; see Peters, 2008 for the dialect of Hasselt in Belgium) and some Scandinavian languages like Norwegian (e.g., Kempe, Bublitz, & Brooks, 2015; Wang, Behne, Jongman, & Sereno, 2004) and Swedish (Roll, Horne, & Lindgren, 2010; Felder, Jönsson-Steiner, Eulitz, & Lahiri, 2009).

6.4. Conclusion

The current thesis investigated the psycholinguistic interaction between lexical-level representations and sublexical level representations during tone-word recognition in Mandarin Chinese. Overall, our studies show that tonal cues and segmental cues are processed differently in word forms with and without lexical-level representations. Specifically, when lexical-level representations are available, the perception of sublexical features is subject to top-down influence. Moreover, feedback from lexical-level representations can lead to rapid lexical responses upon perception of limited acoustic cues. Furthermore, the electrophysiological data suggest that lexical-level representations consist of a left-hemispheric component and a right-hemispheric component. The left-hemispheric component can be further divided into sublexical-level representations, whereas the right-hemispheric component is stored in a holistic form. Based on our findings, we propose a revised TRACE model: that is, the TRACE-Tone model for monosyllabic word recognition in Mandarin Chinese. This model predicts the hemispheric lateralisation of the neural networks underlying each representational module, and the time courses of activation at the lexical level of representations in the pre-attentive processing stage.

Summary

This thesis investigated the role of lexical-level representations in spoken word recognition when lexical-tone perception is considered. Real words and pseudo-words derived from Mandarin Chinese were used to test healthy native speakers with behavioural and neural measurements. Four studies were conducted to address two general research questions: 1) How do lexical-level representations interact with sublexical representations of lexical tones and segments? 2) What is the neural mechanism behind the access to lexical-level representations?

Chapter 1 provides a survey of the existing literature on spoken word recognition, which revealed that the mental representation of word knowledge can be roughly divided into lexical and sublexical levels. Moreover, the two representational levels have excitatory interconnections allowing both top-down and bottom-up processing (McClellend & Elman, 1986). Studies in Mandarin Chinese have shown a similar representational structure: the sublexical-level representation which is composed of lexical tones and segmental cues (i.e., onsets and rimes), and the lexical-level representation which encodes lexicalised combinations of sublexical representations. Four models explaining the cognitive processes of Mandarin word recognition are reviewed and

compared (Ye & Connine, 1999; Zhao et al., 2011; Luo et al., 2006; Shuai & Gong, 2014).

In Chapter 2, we present a study investigating the interaction between the lexical-level representation and the tonal representation at a sublexical level. Two groups of native speakers of Mandarin Chinese were tested separately using two auditory lexical decision tasks. The two tasks were designed to examine whether the semantic and phonological aspects of lexical-level knowledge interact with tonal representations differently. The patterns of reaction times and accuracy rates in three tonal categories of real words were compared with the patterns in pseudo-words of the same tone categories. The reaction time data showed that Tone4 pseudo-words had a tendency to be judged more slowly than Tone1 and Tone2 pseudo-words, whereas no difference was identified between the real words of the three tones. Moreover, the accuracy data showed that Tone4 real words were more likely to be judged incorrectly than Tone1 real words, and Tone1 pseudo-words had lower lexical decision accuracy than Tone2 pseudo-words. These results suggest that the lexical-level representation has top-down influence on the perception of tone cues (e.g., isolation points, pitch levels, pitch contours) during real word recognition, leading to differential lexical decision patterns in real words and pseudowords. In addition, the patterns in auditory lexical decision of semantics were the same as those in lexical decision of phonology. These findings imply that it is not necessary to divide the semantic and phonological aspects of the lexical-level representation into two layers, as proposed by Zhao et al. (2011).

Chapter 3 describes a study examining how lexical and sublexical representations influence form priming in monosyllabic tonal word forms with tone contrasts in Mandarin Chinese. The behavioural responses of native Mandarin speakers to the same targets paired with ten types of primes were recorded in an auditory lexical decision task. The results showed that apart from the facilitation in repetition priming conditions (/lun4/-/lun4/; */lun3/-*/lun3/), pseudo-word primes facilitated the processing of real-word targets with minimal tone contrasts (*/lun3/-/lun4/), and slowed lexical decisions on pseudo-word targets in minimal tone pairs (*/zei1/-*/zei4/). Moreover, real-word primes did not elicit clear priming effects in either tonally contrasted word targets or pseudo-words (/lun4/-/lun2/; /lun4/-*/lun3/). These results suggest that the access to segmental-level representations can facilitate real word processing but inhibit pseudo-word processing. Therefore, the lack of clear priming effects in minimal tone pairs is the outcome of balanced inhibition and facilitation. A connectionist mechanism on how phonological representations of tone words are accessed in form priming was discussed.

Chapter 4 presents an ERP study monitoring the rapid development of new cortical memory traces of a Mandarin pseudo-word. Although several experiments reported rapid cortical plasticity induced by brief training of novel segmental patterns (Shtyrov et al., 2010; Shtyrov, 2011), few studies have devoted attention to the neural dynamics during the rapid learning of novel tonal word forms in tonal languages, such as Chinese. In this study, native speakers of Mandarin Chinese were exposed to acoustically matched real and novel segment-tone patterns. By recording their Mismatch Negativity (MMN) responses (an ERP indicator of long-term memory traces for spoken words), we found enhanced responses to the novel word forms (i.e., pseudowords) in the N1 time window and the MMN time window over the left-hemispheric region in the late exposure phase relative to the early exposure phase. In contrast, no significant changes were identified in MMN responses to the real words during familiarisation. The results suggest a rapid Hebbian learning mechanism in the neocortex for developing long-term memory traces of a novel segment-tone pattern by establishing new associations between segmental and tonal representations. This indicates that lexical-level representations can be rapidly accessed only in real words and learned pseudo-words, whereas unfamiliar pseudo-words can only be processed via sublexical representations.

Chapter 5 reports the results of an ERP study exploring neural evidence of access to lexical-level representations in the N1 time window, which is temporally earlier than an MMN time window. To investigate this, this study used an auditory habituation design to examine whether the whole-word phonology of tonal word forms can be accessed in the pre-attentive processing stage. We measured the auditory N1 and P2 elicited by a Mandarin Chinese monosyllabic tonal word and an acoustically comparable pseudo-word form. A stimulus was presented five times in each stimulation train. The results showed that the pseudo-word was consistently subject to a larger N1 decrement than the real word at the right-hemispheric recording site (C4) for the fourth presentation position of a stimulus train. Moreover, for the same stimulus position, the N1 responses to the repeated pseudo-word form were more persistent at the left-hemispheric recording site (C3) than at the right homologue (C4). Two words with the same contrastive onsets were used to control for the effects induced by different onset consonants. These results suggest that whole-word phonological representations of tonal word forms can be rapidly accessed in the N1 time window (around 110 ms post stimulus onset). Additionally, the accessed whole-word representation as reflected by auditory habituation may be supported by a neural network more lateralised to the right hemisphere.

In Chapter 6, the results of the four studies are discussed, followed by a description of the revised TRACE model, namely the TRACE-Tone model which is based on the findings of the present thesis. Limitations of the experiments conducted and directions for future studies are presented. The last part of this chapter addresses the answers to the two general research questions raised in Chapter 1. First, the results suggest that the lexical-level representation has a top-down influence on both lexical tone and segment perception. In addition, the access to lexical-level representations occurs very rapidly even when the acoustic input is limited. Lexical-level representations are accessed at least in the N1 time window (around 60-110 ms post stimulus onset) and the MMN time window (around 220 ms post stimulus onset). Moreover, the lexical-level representation is composed of two separate modules: one based on phonological knowledge represented by a neural network with a left-hemispheric specialisation, and the other based on holistic whole-word phonological representations that is subserved by a neural network lateralised to the right hemisphere.

Samenvatting

Dit proefschrift beschrijft de rol van lexicale-niveaurepresentaties in gesproken woordherkenning waarbij lexicale-toonperceptie van belang is. Echte woorden en pseudeowoorden afgeleid van het Mandarijn Chinees werden gebruikt om gezonde moedertaalsprekers te testen middels gedragsmetingen en neurale metingen. Er werden vier studies uitgevoerd om de volgende twee centrale onderzoeksvragen te beantwoorden: 1) Wat is de wisselwerking tussen lexicale-niveaurepresentaties en sublexicale representaties van lexicale tonen en segmenten? 2) Wat is het neurale mechanisme achter de toegang tot lexicale-niveaurepresentaties?

In de Introductie (Hoofdstuk 1) bleek uit een literatuurreview over gesproken-woordherkenning dat de mentale representatie van woordkennis grofweg verdeeld kan worden in een lexicaal en een sublexicaal niveau. Verder kwam naar voren dat de twee representatieniveaus activerende verbindingen hadden, waardoor zowel top-down als bottum-up verwerking mogelijk was (McClellend & Elman, 1986). Studies in het Mandarijn Chinees hebben een vergelijkbare representatiestructuur laten zien: de sublexicale-niveaurepresentatie bestaande uit lexicale tonen en segmentale cues (oftewel aanzetten en ritmes) en de lexicale-niveaurepresentatie die lexicale combinaties van sublexicale representaties codeert. Vier modellen die de cognitieve processen van woordherkenning in Mandarijn uiteenzetten werden beschreven en vergeleken (Ye & Connine, 1999; Zhao et al., 2011; Luo et al., 2006; Shuai & Gong, 2014).

In Hoofdstuk 2 werd een onderzoek uitgevoerd naar de interactie tussen lexicale-niveaurepresentatie en tonale representatie op sublexicaal niveau. Twee groepen moedertaalsprekers van Mandarijn Chinees werden afzonderlijk getest met behulp van twee auditieve lexicale-decisietaken. De twee taken waren ontwikkeld om te onderzoeken of de semantische en fonologische aspecten van lexicale-niveaukennis op verschillende manieren interacteren met tonale representaties. De reactietijden en accuraatheid in drie tonale categorieën van echte woorden werden vergeleken met de patronen van pseudowoorden van dezelfde tooncategorieën. Uit de reactietijden dat Toon4 pseudowoorden over het algemeen trager werden beoordeeld dan Toon1 en Toon2 pseudowoorden, terwijl er geen verschil werd vastgesteld tussen de echte woorden van de drie tonen. Verder bleek uit de accuraatheidsdata bleek dat Toon4 echte woorden eerder foutief werden beoordeeld dan Toon1 echte woorden, en dat Toon1 pseudowoorden een lagere lexicale-decisieaccuraatheid hadden dan Toon2 pseudowoorden. Deze resultaten suggereren dat de lexicale-niveaurepresentatie top-down invloed heeft op toonperceptiecues (bijvoorbeeld isolatiepunten, pitchniveaus en pitchcontouren) gedurende herkenning van echte woorden, wat leidt tot verschillende lexicale-decisiepatronen in echte woorden en pseudowoorden. Bovendien waren de patronen in auditieve lexicale decisie van semantiek gelijk aan die van lexicale decisie van fonologie. Deze bevindingen impliceren dat het niet nodig is om de semantische en fonologische aspecten van lexicale-niveaurepresentaties te verdelen in twee lagen zoals voorgesteld door Zhao et al. (2011).

In Hoofdstuk 3 wordt het onderzoek beschreven waarin we nagingen op welke wijze lexicale en sublexicale representaties van invloed zijn op vorm-priming in monosyllabische tonale woordvormen met tooncontrasten in Mandarijn Chinees. De gedragsreacties van moedertaalsprekers van Mandarijn op dezelfde doelwoorden, gecombineerd met tien typen primes werden opgenomen in een auditieve lexicale-decisietaak. Uit de resultaten bleek dat, afgezien van het faciliteren in herhalings-priming-condities (/lun4/-/lun4/; */lun3/-*/lun3/), pseudowoord-primes de verwerking van echte woorden met minimale tooncontrasten (*/lun3/-/lun4/) faciliteerden en lexicale decisies op pseudowoorddoelen in minimale toonparen (*/zei1/-*/zei4/) vertraagden. Daarnaast lokten echte-woord-primes geen duidelijke priming-effecten uit in tooncontrasterende woorddoelen of pseudowoorden (/lun4/-/lun2/; /lun4/-*/lun3/). Deze resultaten suggereren dat de toegang tot segment-niveaurepresentaties echte-woordverwerking kan faciliteren, maar pseudowoordverwerking inhibeert. Daarom kan het ontbreken van duidelijke primingeffecten in minimale toonparen het resultaat zijn van een evenwichtige combinatie van inhibitie en facilitering. Ten slotte werd een connectionistisch mechanisme besproken over hoe fonologische representaties van toonwoorden worden bereikt in vorm-priming.

Hoofdstuk 4 presenteerde een ERP-onderzoek dat de snelle ontwikkeling van nieuwe corticale geheugensporen van een pseudowoord in Mandarijn monitort. Hoewel diverse experimenten snelle corticale plasticiteit, veroorzaakt door korte training van nieuwe segmentale patronen, rapporteerden (Shtyrov et al., 2010; Shtyrov, 2011), hebben weinig studies aandacht besteed aan de neurale dynamiek gedurende het snelle leren van nieuwe tonale woordvormen in toontalen, zoals het Chinees. In dit onderzoek werden moedertaalsprekers van Mandarijn Chinees blootgesteld aan akoestisch gematchte echte en nieuwe-segmenttoonpatronen. Door het opnemen van hun Mismatch Negativity (MMN)-reacties (een ERP-indicator van lange-termijngeheugensporen voor gesproken woorden), vonden we in de linkerhersenhelft in het N1-tijdsvenster en het MMN-tijdsvenster voor nieuwe woordvormen (oftewel pseudowoorden) in de late-blootstellingsfase verbeterde reacties ten opzichte van de vroege-blootstellingsfase. Daarentegen werden geen significante veranderingen geïdentificeerd in MMN-reacties op echte woorden tijdens gewenning. Deze resultaten duiden op een snel Hebbian-leermechanisme in de menselijke neocortex om langetermijngeheugensporen voor een nieuwe segmenttoonpatroon te ontwikkelen door nieuwe associaties tussen segmentale en tonale representaties te bewerkstelligen. Onze bevinding duidt erop dat lexicale-niveaurepresentatie uitsluitend in echte woorden en geleerde pseudowoorden snel bereikt kan worden, terwijl onbekende pseudowoorden slechts via sublexicale representaties verwerkt kunnen worden.

Hoofdstuk 5 rapporteert een ERP-onderzoek naar het neurale bewijs van toegang tot de lexicale-niveaurepresentatie in het N1 tijdsvenster dat temporeel eerder is dan een MMN tijdsvenster. Om dit probleem te onderzoeken, werd in dit onderzoek een auditief gewenningsontwerp gebruikt om na te gaan of de hele-woordfonologie van tonale woordvormen bereikt kan woorden in het pre-aandachtige verwerkingsstadium. We hebben de auditieve N1 en P2 gemeten die werden uitgelokt door een Mandarijn Chinees monosyllabisch tonaal woord en een akoestisch vergelijkbare pseudowoordvorm. Een stimulus werd herhaaldelijk vijf keer aangeboden in elke stimulatietrein. Door de reacties op de constratieve onsetmedeklinkers te controleren, toonde ons onderzoek aan dat het pseudowoord consistent onderhevig was aan grotere N1-afname dan de echte woorden op de opnameplek in de rechterhemisfeer (C4) voor de vierde presentatiepositie van een stimulustrein. Bovendien waren de N1 reacties op de herhaalde pseudowoordvorm persistenter in de opnameplek in de linkerhemisfeer (C3) dan in de rechter homoloog (C4). Twee woorden met dezelfde contrastieve onsets werden gebruikt om te controleren voor de effecten die werden veroorzaakt door de verschillende onset-consonanten. Deze resultaten suggereren dat hele-woord fonologische representaties van tonale woordvormen snel bereikt kunnen worden in het N1 tijdsvenster (circa 110 ms post stimulus onset). Bovendien kan het zo zijn dat de hele-woordrepresentatie zoals weerspiegeld door auditieve gewenning wordt bevorderd door een neuraal netwerk dat meer gelateraliseerd is naar de rechterhemisfeer.

In hoofdstuk 6 werden de resultaten van de vier onderzoeken besproken, gevolgd door een beschrijving van een herzien TRACE model, te weten het TRACE-Tone model, dat is gebaseerd op de bevindingen van dit proefschrift. De beperkingen van de onderzoeken werden besproken, alsmede richtlijnen voor toekomstige studies. Het hoofdstuk eindigde met een conclusie. Dit proefschrift heeft geresulteerd in een reeks bevindingen die helpen bij de beantwoording van de twee algemene onderzoeksvragen die zijn gepresenteerd in Hoofdstuk 1. Ten eerste wijzen de resultaten erop dat de lexicale-niveaurepresentatie een top-down invloed had op zowel lexicale-toon en segmentperceptie. Daarnaast bleek dat het bereiken van lexicale-niveaurepresentaties zeer snel kon plaatsvinden, zelfs wanneer de akoestische input beperkt was. Lexicale-niveaurepresentaties konden op zijn minst worden bereikt in het N1 tijdsvenster (circa 60-110 ms post stimulus onset) en het MMN tijdsvenster (circa 220 ms post stimulus onset). Verder kon de lexicale-niveaurepresentatie bestaan uit twee aparte modules. Eén module kon gebaseerd zijn op fonologische kennis gerepresenteerd door een neuraal netwerk met een linkerhemisferische specialisatie en de andere op holistische hele-woord fonologische representaties die ondersteund zouden kunnen worden door een neuraal netwerk dat rechts gelateraliseerd is.

Appendix

A Appendix to Chapter 2: Experimental materials

Experimental materials in *Pinyin* script

	Block 1	Block 2	Block 3
Pseudo-words	gei1	nin1	ken1
	mang1	ze1	nong1
	ning1	mai1	na1
	ce2	bei2	te2
	bin2	ken2	shun2
	xiu2	zhui2	jiang2
	ran4	fou4	fo4
	za4	reng4	mou4
	qun4	chun4	zu4
Real words	dao1 (刀, knife)	jing1 (精, accurate)	gao1 (高, high)
	tuo1 (脱, take off)	gan1 (干, dry)	san1 (三, three)
	chong1 (冲, dash)	tie1 (贴, paste)	qie1 (切, cut)
	shen2 (神, divine)	cong2 (从, from)	pi2 (皮, skin)
	peng2 (朋, friend)	xi2 (习, learn)	liu2 (流, flow)
	lai2 (来, come)	hao2 (豪, luxury)	cheng2 (成, success)
	zhu4 (住, reside)	da4 (大, big)	dui4 (对, correct)
	jie4 (借, borrow)	qu4 (去, go)	bing4 (病, sick)
	yue4 (月, moon)	lang4 (浪, tide)	hai4 (害, harm)

B Appendix to Chapter 3: Experimental materials

B.1 Stimulus used for the five priming conditions with real-word targets (RWT)

RWT	RSTR	RSR	RNOR	PSR	PNOR
heng1	heng1	heng2	wa2	heng3	wo2
(哼, humph)	(哼, humph)	(横, transverse)	(娃, baby)		
nie1	nie1	nie4	chong1	nie3	fou4
(捏, pinch)	(捏, pinch)	(镍, nickel)	(冲, dash)		
sui1	sui1	sui2	mian3	sui3	keng4
(虽, although)	(虽, although)	(随, follow)	(免, exempt)		
niang2	niang2	niang4	shou1	niang3	huai3
(娘, mother)	(娘, mother)	(酿, brew)	(收, receive)		
shuai3	shuai3	shuai1	cong1	shuai2	die3
(甩, throw)	(甩, throw)	(衰, decline)	(葱, onion)		
guang3	guang3	guang4	ri4	guang2	zhuo4
(广, broad)	(广, broad)	(逛, stroll)	(日, day)		
lun4	lun4	lun2	pie3	lun3	tai3
(论, argue)	(论, argue)	(轮, wheel)	(撇, cast aside)		
zeng1	zeng1	zeng4	cao3	zeng2	ruo2
(增, grow)	(增, grow)	(赠, give as a gift)	(草, grass)		
reng1	reng1	reng2	wo3	reng4	yue3
(扔, toss)	(扔, toss)	(仍, still)	(我, I)		
rao2	rao2	rao4	lang4	rao1	rong1
(饶, pardon)	(饶, pardon)	(绕, surround)	(浪, tide)		
shui3	shui3	shui2	an3	shui1	dai2
(水, water)	(水, water)	(谁, who)	(俺, I)		
luan3	luan3	luan4	mi1	luan1	nin4
(卵, egg)	(卵, egg)	(乱, massive)	(眯, narrow one's eyes)		
niao3	niao3	niao4	seng1	niao1	tuan3
(鸟, bird)	(鸟, bird)	(尿, urine)	(僧, monk)		

mie4	mie4	mie1	chuang2	mie3	kuang3
(灭, extinguish)	(灭, extinguish)	(咩, bleat)	(床, bed)		
zang4	zang4	zang1	hu3	zang2	te2
(葬, interment)	(葬, interment)	(脏, dirty)	(虎, tiger)		
tie1	tie1	tie3	ba2	tie2	run3
(贴, paste)	(贴, paste)	(铁, iron)	(拔, pull out)		
mai2	mai2	mai3	suo1	mai1	lue1
(埋, bury)	(埋, bury)	(买, buy)	(缩, shrink)		
leng3	leng3	leng2	sao1	leng1	mie2
(冷, cold)	(冷, cold)	(楞, edge)	(骚, coquettish)		
er3	er3	er2	pang4	er1	gun1
(耳, ear)	(耳, ear)	(儿, child)	(胖, fat)		
guai3	guai3	guai1	rou4	guai2	nong3
(拐, turn)	(拐, turn)	(乖, behaved)	(肉, flesh)		
qia4	qia4	qia3	nong4	qia2	nuo1
(恰, proper)	(恰, proper)	(卡, stuck)	(弄, get)		
che1	che1	che4	qian3	che2	en3
(车, car)	(车, car)	(撤, retreat)	(浅, shallow)		
chun1	chun1	chun3	gua3	chun4	jue3
(春, spring)	(春, spring)	(蠢, stupid)	(寡, few)		
min2	min2	min3	kua1	min1	zei1
(民, people)	(民, people)	(敏, sensitive)	(夸, exaggerate)		
ru3	ru3	ru4	bai1	ru1	nuan1
(乳, milk)	(乳, milk)	(入, entre)	(掰, breakoff)		
kuan3	kuan3	kuan1	le4	kuan2	rou1
(款, money)	(款, money)	(宽, wide)	(乐, laugh)		
wai4	wai4	wai1	jiong3	wai2	ken1
(外, out)	(外, out)	(歪, slanting)	(窘, embarrassed)		
yong4	yong4	yong1	lia3	yong2	nv1
(用, use)	(用, use)	(拥, cuddle)	(俩, two)		
tun1	tun1	tun2	sha3	tun3	za4
(吞, swallow)	(吞, swallow)	(屯, village)	(傻, foolish)		

pen1 (喷, spray)	pen1 (喷, spray)	pen2 (盆, basin)	shua3 (耍, play with)	pen3	cang4
kang2 (抗, resist)	kang2 (抗, resist)	kang1 (糠, bran)	mou3 (某, some)	kang3	zui1
ran3 (染, dye)	ran3 (染, dye)	ran2 (燃, burn)	tou1 (偷, steal)	ran4	dei4
kan3 (砍, chop)	kan3 (砍, chop)	kan4 (看, look)	tui4 (退, recede)	kan2	sou4
nu4 (怒, rage)	nu4 (怒, rage)	nu3 (弩, crossbow)	tiao3 (挑, hang)	nu1	lia2
shai4 (晒, bask)	shai4 (晒, bask)	shai1 (筛, sift)	hong3 (哄, coax)	shai2	miu3

B.2 Stimulus used for the five priming conditions with pseudo-word targets (PWT)

PWT	PSTP	RSP	RNOP	PSP	PNOP
gun1	gun1	gun4 (棍, stick)	mei4 (妹, sister)	gun2	ze3
nuo1	nuo1	nuo2 (挪, move)	beng2 (蹦, jump)	nuo3	ben2
gei2	gei2	gei3 (给, give)	suan1 (酸, soar)	gei1	nve1
zhuai2	zhuai2	zhuai4 (拽, pluck)	men1 (闷, stuffy)	zhuai1	nang4
nen3	nen3	nen4 (嫩, tender)	po2 (婆, old woman)	nen1	kuai1
te3	te3	te4 (特, special)	pin2 (聘, employ)	te2	ren1
diu4	diu4	diu1 (丢, loss)	tai1 (胎,tyre)	diu3	hang3
jiong1	jiong1	jiong3 (窘, embarrassed)	duan3 (短, short)	jiong4	chai3
re1	re1	re4 (热, hot)	zong4 (纵, vertical)	re2	nan3
ri2	ri2	ri4 (日, day)	huai4 (坏, bad)	ri3	cen3
zhun2	zhun2	zhun3 (准, allow)	lao1 (捞, drag for)	zhun4	mie3
nü4	nü4	nü3 (女, woman)	ye2 (爷, grandfather)	nü1	se1
pie4	pie4	pie3 (撇, cast aside)	gou3 (狗, dog)	pie2	qun3
fo4	fo4	fo2 (佛, Buddha)	gai3 (改, change)	fo1	ken2
lai1	lai1	lai2 (来, come)	dong3 (懂, understand)	lai3	nve2

ruan1	ruan1	ruan3 (软, soft)	bo3 (跛, crippled)	ruan2	pei3
shuan2	shuan2	shuan1 (栓, bar)	jue4 (倔, stubborn)	shuan3	he3
zuan2	zuan2	zuan1 (钻, drill)	pei4 (配, match)	zuan3	que3
nong3	nong3	nong4 (弄, get)	que1 (缺, lack)	nong1	qun4
zhui3	zhui3	zhui1 (追, chase)	la1 (拉, pull)	zhui2	nei2
fou4	fou4	fou3 (否, negate)	qing3 (抢, rob)	fou1	ri1
mou1	mou1	mou2 (某, some)	xiang2 (降, surrender)	mou4	ban2
run1	run1	run4 (润, moist)	fei3 (匪, bandit)	run2	pou4
zen2	zen2	zen3 (怎, how)	xiu3 (宿, night)	zen1	rui1
chui3	chui3	chui1 (吹, blow)	rao4 (绕, surround)	chui4	qiong4
reng3	reng3	reng1 (扔, throw)	dao1 (刀, knife)	reng4	xiong4
zun3	zun3	zun1 (尊, respect)	gao4 (告, tell)	zun2	tou3
lia4	lia4	lia3 (俩, two)	kou1 (扣, buckle)	lia1	kong2
nin1	nin1	nin2 (您, you)	cang2 (藏, hide)	nin3	dai2
dei2	dei2	dei3 (得, have to)	leng4 (愣, stupefied)	dei4	mu1
zhua2	zhua2	zhua1 (抓, grasp)	ti3 (体, body)	zhua4	pen3
ming3	ming3	ming4 (明, light)	gu1 (姑, ante)	ming1	yue2

sai3	sai3	sai1 (塞, stuck)	geng4 (更, more)	sai2	rong1
ca4	ca4	ca1 (擦, wipe)	ying3 (影, shadow)	ca2	chen3
neng4	neng4	neng2 (能, can)	bei3 (北, north)	neng3	pou3

References

Ahrens, W. H., Cox, D. J., & Budhwar, G. (1990). Use of the arcsine and square root transformations for subjectively determined percentage data. *Weed Science, 38*, 452–458.

Alexandrov, A. A., Boricheva, D. O., Pulvermüller, F., & Shtyrov, Y. (2011). Strength of word-specific neural memory traces assessed electrophysiologically. *PLoS ONE, 6*, e22999.

Alho, K., Connolly, J. F., Cheour, M., Lehtokoski, A., Huotilainen, M., Virtanen, J., … Ilmoniemi, R. J. (1998). Hemispheric lateralization in preattentive processing of speech sounds. *Neuroscience Letters*, *258*, 9–12.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology. Human Perception and Performance*, *10*, 340–357.

Bell, R., Dentale, S., Buchner, A., & Mayr, S. (2010). ERP correlates of the irrelevant sound effect. *Psychophysiology, 47*, 1182–1191.

Bien, H., Bölte, J., & Zwitserlood, P. (2014). Do syllables play a role in German speech perception? Behavioral and electrophysiological data from primed lexical decision. *Frontiers in Psychology, 5*, 1544.

Blenner, J. L., & Yingling, C. D. (1994). Effects of prefrontal cortex lesions on visual evoked potential augmenting/reducing. *International Journal of Neuroscience, 78*, 145–156.

Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*, 389–395.

Boersma, P., & Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.39). Institute of Phonetic Sciences of the University of Amsterdam. <http://www.praat.org>.

Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research, 33*, 103–135.

Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology, 31*, 51–68.

Castellucci, V. F., & Kandel, E. R. (1974). A quantal analysis of the synaptic depression underlying habituation of the gill-withdrawal reflex in Aplysia. *Proceedings of the National Academy of Sciences of the United States of America, 71*, 5004–5008.

Castellucci,V., Pinsker, H., Kupfermann, I., Kandel, E. R. (1970). Neuronal mechanisms of Habituation and dishabituation of the gill-withdrawal reflex in Aplysia. *Science, 167*, 1745–1748.

Catani, M., Jones, D. K., & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology, 57*, 8–16.

Chao, Y. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Coles, M. G. H. & Rugg, M. D. (1995). Event-related Brain Potentials: An Introduction. In M.D. Rugg, & M.G.H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition* (pp. 1–26). Oxford, UK: Oxford University Press.

Connine, C. M., & Clifton, C. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology. Human Perception and Performance*, *13*, 291–299.

Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*, 463–480.

Crowley, K. E., & Colrain, I. M. (2004). A review of the evidence for P2 being an independent component process: Age, sleep and modality. *Clinical Neurophysiology, 115*, 732–744.

Curio, G., Neuloh, G., Numminen, J., Jousmäki, V., & Hari, R. (2000). Speaking modifies voice-evoked activity in the human auditory cortex. *Human Brain Mapping, 9*, 183–191.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59*, 165–179.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America, 105*, 1877–1888.

Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*, 171-195.

Dahan, D., & Magnuson, J. S. (2006). Spoken-word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 249–283). Amsterdam, The Netherlands: Academic Press.

Davis, M. H., Di Betta, A. M., Macdonald, M. J. E., & Gaskell, M. G. (2009). Learning and consolidation of novel spoken words. *Journal of Cognitive Neuroscience, 21*, 803–20.

Davis, M. H., & Gaskell, M. G. (2009). A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical*

Transactions of the Royal Society B, Biological Sciences, 364, 3773–3800.

Dehaene-Lambertz, G. (1997). Electrophysiological correlates of categorical phoneme perception in adults. *Neuroreport, 8*, 919–924.

Dorman, M. F. (1974). Auditory evoked potential correlates of speech sound discrimination. *Perception, 15*, 215–220.

Duanmu, S. (2002). *The Phonology of standard Chinese*. Oxford, England: Oxford University Press.

Felder, V., Jönsson-Steiner, E., & Eulitz, C. (2009). Asymmetric processing of lexical tonal contrast in Swedish. *Attention, Perception & Psychophysics*, *71*(8), 1890–1899.

Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, *4*, 258–267.

Fitch, R. H., Miller, S., & Tallal, P. (1997). Neurobiology of speech perception. *Annual Review of Neuroscience, 20*, 331–353.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behaviour Research Methods, Instruments, & Computers, 35*, 116–124.

Fournier, R., Gussenhoven, C., Jensen, O., & Hagoort, P. (2010). Lateralization of tonal and intonational pitch processing: An MEG study. *Brain Research*, *1328*, 79–88.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews, 91*, 1357–1392.

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: The dynamic dual pathway model. *Brain and Language, 89*, 267–276.

Fruhstorfer, H. (1971). Habituation and dishabituation of the human vertex response. *Electroencephalography and Clinical Neurophysiology, 30*, 306–312.

Fruhstorfer, H., Soveri, P., & Järvilehto, T. (1970). Short-term habituation of the auditory evoked response in man. *Electroencephalography and Clinical Neurophysiology, 28*, 153–161.

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*, 207–222.

Gaskell, M. G., & Dumay, N. (2003). Lexical competition and the acquisition of novel words. *Cognition, 89*, 105–132.

Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: Does the corpus callosum enable the human condition? *Brain, 123*, 1293–1326.

Goldinger, S. D. (1996). Auditory lexical decision. *Language and Cognitive Processes, 11*, 37–41.

Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*, 348–353.

Gu, F., Li, J., Wang, X., Hou, Q., Huang, Y., & Chen, L. (2012). Memory traces for tonal language words revealed by auditory event-related potentials. *Psychophysiology, 49*, 1353–1360.

Guo， J (郭锦桴).（1993）. *汉语声调语调阐要与探索* (The expatiation and exploration of tones in Chinese). Beijing, China: Beijing Language and Culture University Press

Hamburger, M., & Slowiaczek, L. M. (1996). Phonological priming reflects lexical competition. *Psychonomic Bulletin & Review, 3*, 520–525.

Hamm, J. P., Ethridge, L. E., Shapiro, J. R., Stevens, M. C., Boutros, N. N., Summerfelt, A. T. ... Clementz, B. A. (2012). Spatiotemporal and frequency domain analysis of auditory paired stimuli processing in schizophrenia and bipolar disorder with psychosis. *Psychophysiology, 49*, 522–530.

Hari, R., Kaila, K., Katila, T., Tuomisto, T., & Varpula, T. (1982) Interstimulus interval dependence of the auditory vertex response and its magnetic counterpart: Implications for their neural generation. *Electroencephalography and Clinical Neurophysiology, 54*, 561–569.

Harris, I.D. (1943). Habituatory response decrement in the intact organism. *Psychological Bulletin, 40*, 385–422.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hebb, D. O. (1949). *The organization of behaviour: A neuropsychological theory*. New York: John Wiley.

Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood: Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology, 116*, 572–592.

Howard, D., Nickels, L., Coltheart, M., & Cole-Virtue, J. (2006). Cumulative semantic inhibition in picture naming: Experimental and computational studies. *Cognition, 100*, 464–482.

Hsieh, L., Gandour, J., Wong, D., & Hutchins, G. D. (2001). Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience. *Brain and Language, 76*, 227–252.

Hu, J., Gao, S., Ma, W., & Yao, D. (2012). Dissociation of tone and vowel processing in Mandarin idioms. *Psychophysiology, 49*, 1179–1190.

Huang, X., Yang, J.-C., Zhang, Q., & Guo, C. (2014). The time course of spoken word recognition in Mandarin Chinese: A unimodal ERP study. *Neuropsychologia, 63*, 165–174.

Hughes, J. R., & Rosenblith, W. A. (1957). Electrophysiological evidence for auditory sensitization. *The Journal of the Acoustical Society of America, 29*, 275–280.

Hyde, M. (1997). The N1 response and its applications. *Audiology and Neuro-otology, 2*, 281–307.

Jones, D. (1944). Chronemes and tonemes. *Acta Linguistica, 4*, 1-10.

Jones, S. J., Longe, O., & Vaz Pato, M. (1998). Auditory evoked potentials to abrupt pitch and timbre change of complex tones: Electrophysiological evidence of "streaming"? *Electroencephalography and Clinical Neurophysiology, 108*, 131–142.

Jutras, B. (2006). Serial position effects for acoustic stimuli among children with and without hearing loss. *American Journal of Audiology, 15*, 57–65.

Kandel, E. R. (2001). The molecular biology of memory storage: A dialogue between gene and synapses. *Science, 294*, 1030–1038.

Keidel, W. D., & Spreng, M. (1965). Neurophysiological evidence for the Stevens power function in man. *The Journal of the Acoustical Society of America, 38*, 191–195.

Kempe, V., Bublitz, D., Brooks, P. J. (2015). Musical ability and non-native speech-sound processing are linked through sensitivity to pitch and spectral information. *British Journal of Psychology, 106*, 349–366.

Koyama, S., Gunji, A., Yabe, H., Oiwa, S., Akahane-Yamada, R., Kakigi, R., & Näätänen, R. (2000). Hemispheric lateralization in an analysis of speech sounds. *Cognitive Brain Research*, *10*, 119–124.

Kutas, M., & Federmeier, K.D. (2000). Electrophysiology reveals semantic memory use in 900 language comprehension. *Trends in Cognitive Science, 4*, 463–470.

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*, 621–641.

Kutas, M., & Hillyard, S.A., (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Lai, Y., & Zhang, J. (2008). Mandarin lexical tone recognition: The gating paradigm. *Kansas Working Papers in Linguistics, 30*, 183–194.

Language Affairs Revolution Council of China. (1988). *List of frequently used characters in modern Chinese*. <http://www.china-language.gov.cn/wenziguifan/shanghi/013c.htm>

Language Teaching Research Centre for Beijing Language Institute. (1986). *现代汉语频率词典* (Modern Chinese Word Frequency Dictionary). Beijing, China: Beijing Language Institute Press.

Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech, 50*, 101–123.

Lee, C. Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics, 36*, 537–563.

Levänen, S., Ahonen, A., Hari, R., McEvoy, L., & Sams, M. (1996). Deviant auditory stimuli activate human left and right auditory cortex differently. *Cerebral Cortex, 6*, 288–296.

Li, W., Wang, L., & Yang, Y. (2014). Chinese tone and vowel processing exhibits distinctive temporal characteristics: An electrophysiological perspective from classical Chinese poem processing. *PLoS ONE, 9*, e85683.

Li, N. (李宁), Zhang, X.-D.(张晓丹), & Huang, Z.-M.(黄昭明) (2009). 汉语鼻辅音共振峰的比较研究(A comparative study on the formants of nasal consonants in Chinese). *中国听力语言康复科学杂志 (Chinese Scientific Journal of Hearing and Speech Rehabilitation), 5*, 36–38.

Li, X., Gandour, J. T., Talavage, T., Wong, D., Hoffa, A., Lowe, M., … Dzemidzic, M. (2010). Hemispheric asymmetries in phonological processing of tones versus segmental units. *Neuroreport, 21*, 690–694.

Liang, J., & van Heuven, V. J. (2004). Evidence for separate tonal and segmental tiers in the lexical specification of words: A case study of a brain-damaged Chinese speaker. *Brain and Language, 91*, 282–293.

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

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4, 187–196.

Liebenthal, E., Desai, R., Ellingson, M. M., Ramachandran, B., Desai, A., & Binder, J. R. (2010). Specialization along the left superior temporal sulcus for auditory categorization. *Cerebral Cortex, 20*, 2958–2970.

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., & Chen, C. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *NeuroImage, 29*, 515–523.

Liu, S., & Samuel, A. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes, 22*, 566–594.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39*, 155–158.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.

Luo, H., Ni, J.-T., Li, Z.-H., Li, X.-O., Zhang, D.-R., Zeng, F.-G., & Chen, L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 19558–19563.

MacGregor, L.J., Pulvermüller, F., van Casteren, M., & Shtyrov, Y. (2012). Ultra-rapid access to words in the brain. *Nature Communication, 3*, 711.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003). Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science, 27*, 801–805.

Malins, J. (2013). Taking Tone into Account: Cognitive Neuroscientific Investigations of Mandarin Chinese Spoken Word Processing. Unpublished doctoral dissertation. The University of Western Ontario.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 62*, 407–420.

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia, 50*, 2032–2043.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*, 71–102.

Marslen-Wilson, W. D., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review*, *101*, 653–675.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477.

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

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning-systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102*, 419–457.

McEnery, A. M., & Xiao, R. Z. (2004). The Lancaster Corpus of Mandarin Chinese: A corpus for monolingual and contrastive language study. In M. T. Lino, M. F. Xavier, F. Ferreira, R. Costa, R. Silva (Eds), *Proceedings of the Fourth International Conference on Language*

Resources and Evaluation (LREC) 2004 (pp. 1175–1178). Lisbon, Portugal: Centro Cultural de Belem.

McGettigan, C., Warren, J.E., Eisner, F., Marshall, C.R. Shanmugalingam, P., Scott, S.K. (2011). Neural Correlates of Sublexical Processing in Phonological Working Memory. *Journal of Cognitive Neuroscience, 23*, 961–977.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1363–1389.

Mildner, V. (2008). The Cognitive Neuroscience of Human Communication. New York: Lawrence Erlbaum Associates.

Mimura, M., Verfaellie, M., & Milberg, W. P. (1997). Repetition priming in an auditory lexical decision task: effects of lexical status. *Memory & Cognition, 25*, 819–825.

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory and Language*, *52*, 424–443.

Mondor, T. A., & Morin, S. R. (2004). Primacy, recency, and suffix effects in auditory short-term memory for pure tones: Evidence from a probe recognition paradigm. *Canadian Journal of Experimental Psychology, 58*, 206–219.

Muenssinger, J., Stingl, K. T., Matuz, T., Binder, G., Ehehalt, S., & Preissl, H. (2013). Auditory habituation to simple tones: Reduced evidence for habituation in children compared to adults. *Frontiers in Human Neuroscience, 7*, 377.

Mulert, C., Jäger, L., Propp, S., Karch, S., Störmann, S., Pogarell, O., … Hegerl, U. (2005). Sound level dependence of the primary auditory cortex: Simultaneous measurement with 61-channel EEG and fMRI. *NeuroImage, 28*, 49–58.

Näätänen, R. (1995). The mismatch negativity: A powerful tool for cognitive neuroscience. *Ear and Hearing, 16*, 6–18.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*, 1–21.

Näätänen, R., & Picton, T. (1987) The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*, 375–425.

Näätänen, R., Gaillard, A.W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica, 42*, 313–329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Houtilainen, M., Iivonen, A., ... Luuk, A. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature, 385*, 432–434.

Näätänen, R., Schröger, E., Karakas, S., Tervaniemi, M., & Paavilainen, P. (1993). Development of a memory trace for a complex sound in the human brain. *Neuroreport, 4*, 503–506.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology. Human Perception and Performance*, *23*, 873–889.

Nooteboom, S. (1997). The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), *The Handbook of the phonetic sciences* (pp. 640–673). Cambridge: Blackwell Publishers.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 205–238.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologia, 9*, 97–113.

Öhman, A., & Lader, M. (1972). Selective attention and “habituation” of the auditory averaged evoked response in humans. *Physiology and Behaviour, 8*, 79–85.

Paavilainen, P., Alho, K., Reinikainen, K., Sams, M., & Näätänen, R. (1991). Right hemisphere dominance of different mismatch negativities. *Electroencephalography and Clinical Neurophysiology, 78*, 466–479.

Pang, E. W., & Taylor, M. J. (2000). Tracking the development of the N1 from age 3 to adulthood: An examination of speech and non-speech stimuli. *Clinical Neurophysiology, 111*, 388–397.

Peters, J. (2008). Tone and intonation in the dialect of Hasselt. *Linguistics, 46,* 983–1018.

Pike, K. L. (1947). *Phonemics: A technique for reducing languages to writing*. Ann Arbor, MI: University of Michigan Press.

Pinsker, H., Kupfermann, I., Castellucci, V., & Kandel, E. (1970). Habituation and Dishabituation of the GM-Withdrawal Reflex in Aplysia. *Science, 167*, 1740–1742.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *NeuroImage, 10*, 15–35.

Prescott, J., Connolly, J. F., & Gruzelier, J. H. (1984). The augmenting/reducing phenomenon in the auditory evoked potential. *Biological Psychology, 19*, 31–44.

Pulvermüller, F. (1999).Words in the brain's language. *Behavioural and Brain Sciences, 22*, 253–336.

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity as a tool for studying higher cognitive processes. *Progress in Neurobiology, 79*, 49–71.

Pulvermüller, F., Kujala, T., Shtyrov, Y., Simola, J., Tiitinen, H., Alku, P. ... Näätänen, R. (2001). Memory traces for words as revealed by the mismatch negativity. *NeuroImage, 14*, 607–616.

Qiao, X., & Forster, K. I. (2013). Novel word lexicalization and the prime lexicality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*, 1064.

Qiao, X., Forster, K., & Witzel, N. (2009). Is banara really a word? *Cognition, 113*, 254–257.

Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition, 17*, 525–535.

Radeau, M., Morais, J., & Segui, J. (1995). Phonological priming between monosyllabic spoken words. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1297.

Rankin, C. H., Abrams, T., Barry, R. J., Bhatnagar, S., Clayton, D. F., Colombo, J., … Thompson, R. F. (2009). Habituation revisited: An updated and revised description of the behavioural characteristics of habituation. *Neurobiology of Learning and Memory, 92*, 135–138.

Ratcliff, R., Hockley, W., & McKoon, G. (1985). Components of activation: repetition and priming effects in lexical decision and recognition. *Journal of Experimental Psychology: General, 114*, 435–450.

Repp, B. H., & Lin, H.-B. (1990). Integration of segmental and tonal information. *Journal of Phonetics*, *18*, 481–495.

Ren, G. Q., Yang, Y., & Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience*, *162*, 87–95.

Rinne, T., Alho, K., Alku, P., Holi, M., Sinkkonen, J., Virtanen, J., … Näätänen, R. (1999). Analysis of speech sounds is left-hemisphere predominant at 100-150 ms after sound onset. *Neuroreport*, *10*, 1113–1117.

Ritter, W., Vaughan, H. G., & Costa, L. D. (1968). Orienting and habituation to auditory stimuli: a study of short term changes in average evoked responses. *Electroencephalography and Clinical Neurophysiology, 25*, 550–556.

Rodriguez-Fornells, A., Schmitt, B. M., Kutas, M., & Münte, T. F. (2002). Electrophysiological estimates of the time course of semantic and phonological encoding during listening and naming. *Neuropsychologia, 40*, 778–787.

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology-ERPs of Swedish word processing. *Brain Research*, *1330*, 114–123.

Rosburg, T., Haueisen, J., & Kreitschmann-Andermahr, I. (2004). The dipole location shift within the auditory evoked neuromagnetic field components N100m and mismatch negativity (MMNm). *Clinical Neurophysiology, 115*, 906–913.

Rosburg, T., Haueisen, J., & Sauer, H. (2002). Habituation of the auditory evoked field component N100m and its dependence on stimulus duration. *Clinical Neurophysiology, 113*, 421–428.

Rosburg, T., Trautner, P., Boutros, N. N., Korzyukov, O. a., Schaller, C., Elger, C. E., & Kurthen, M. (2006). Habituation of auditory evoked potentials in intracranial and extracranial recordings. *Psychophysiology, 43*, 137–144.

Rosburg, T., Zimmerer, K., & Huonker, R. (2010). Short-term habituation of auditory evoked potential and neuromagnetic field components in dependence of the interstimulus interval. *Experimental Brain Research, 205*, 559–570.

Rust, J. (1977). Habituation and the orienting response in the auditory cortical evoked potential. *Psychophysiology, 14*, 123–126.

Samuel, A. G. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology. General*, *110*, 474–494.

Samuel, A. G. (1996). Does lexical information influence the perceptual restoration of phonemes? *Journal of Experimental Psychology: General, 125*, 28–51.

Samuel, A. G. (2001). Knowing a word affects the fundamental perception of the sounds within it. *Psychological Science*, 12, 348–351.

Sara, Y., Mozhayeva, M.G., Liu, X., Kavalali, E.T. (2002). Fast vesicle recycling supports neurotransmission during sustained stimulation at hippocampal synapses. *The Journal of Neuroscience, 22*, 1608–1617.

Schirmer, A., Tang, S.-L., Penney, T. B., Gunter, T. C., & Chen, H.-C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience, 17*, 1–12.

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech, 58*, 131–151.

Sharma, A., & Dorman, M. (2000). Neurophysiologic correlates of cross-language phonetic perception. *The Journal of the Acoustical Society of America, 107*, 2697–2703.

Sharon, T., Moscovitch, M., & Gilboa, A. (2010). Rapid neocortical acquisition of long-term arbitrary associations independent of the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 1146–1151.

Sharpless, S., & Jasper, H. (1956). Habituation of the arousal reaction. *Brain, 79*, 655–680.

Shtyrov, Y. (2011). Fast mapping of novel word forms traced neurophysiologically. *Frontiers in Psychology, 2*, 340.

Shtyrov, Y. (2012). Neural bases of rapid word learning. *The Neuroscientist, 18*, 312–319.

Shtyrov, Y., Kimppa, L., Pulvermüller, F., & Kujala, T. (2011). Event-related potentials reflecting the frequency of unattended spoken words: A neuronal index of connection strength in lexical memory circuits? *NeuroImage, 55*, 658–668.

Shtyrov, Y., Kujala, T., Ahveninen, J., Tervaniemi, M., Alku, P., Ilmoniemi, R. J., & Näätänen, R. (1998). Background acoustic noise and the hemispheric lateralization of speech processing in the human brain: Magnetic mismatch negativity study. *Neuroscience Letters, 251*, 141–144.

Shtyrov, Y., Kujala, T., Ilmoniemi, R. J., & Näätänen, R. (1999). Noise affects speech-signal processing differently in the cerebral hemispheres. *Neuroreport, 10*, 2189–2192.

Shtyrov, Y., Kujala, T., Palva, S., Ilmoniemi, R.J., & Näätänen, R. (2000). Discrimination of speech and of complex nonspeech sounds of different temporal structure in the left and right cerebral hemispheres. *NeuroImage, 12*, 657–663.

Shtyrov, Y., Nikulin, V. V., & Pulvermüller, F. (2010). Rapid cortical plasticity underlying novel word learning. *The Journal of Neuroscience, 30*, 16864–16867.

Shtyrov, Y., Osswald, K., & Pulvermüller, F. (2008). Memory traces for spoken words in the brain as revealed by the hemodynamic correlate of the mismatch negativity. *Cerebral Cortex, 18*, 29–37.

Shtyrov, Y., & Pulvermüller, F. (2002). Neurophysiological evidence of memory traces for words in the human brain. *Neuroreport, 13*, 521–525.

Shtyrov, Y., & Pulvermüller, F. (2007) Language in the passive auditory oddball: Motivations, benefits and prospects. *Journal of Psychophysiology, 21*, 176–186.

Shuai, L., & Gong, T. (2014). Temporal relation between top-down and bottom-up processing in lexical tone perception. *Frontiers in Behavioural Neuroscience, 8*, 97.

Shuai, L., Li, B., & Gong, T. (2012). Priming Effects of Tones and Segments in Lexical Processing in Mandarin. *In 6th International Conference on Speech Prosody*.

Sidtis, D.V. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology, 20*, 411–426.

Sidtis, J. J., Volpe, B. T., Wilson, D. H., Rayport, M., & Gazzaniga, M. S. (1981). Variability in right hemisphere language function after callosal section: Evidence for a continuum of generative capacity. *The Journal of Neuroscience, 1*, 323–331.

Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1239.

Slowiaczek, L. M., & Pisoni, D. B. (1986). Effects of phonological similarity on priming in auditory lexical decision. *Memory & Cognition, 14*, 230–237.

Steinberg, J., Truckenbrodt, H., & Jacobsen, T. (2011). The role of stimulus cross-splicing in an event-related potentials study: Misleading formant transitions hinder automatic phonological processing. *The Journal of Acoustic Society of America, 131*, 3120–3140.

Storkel, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44*, 1321–1337.

Sumner, M., & Samuel, A.G. (2007). Lexical inhibition and sublexical facilitation are surprisingly long lasting. *Journal of Experimental Psychology: Learning, Memory and Cognition, 33*, 769–790.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166.

Teismann, I. K., Sörös, P., Manemann, E., Ross, B., Pantev, C., & Knecht, S. (2004). Responsiveness to repeated speech stimuli persists in left but not right auditory cortex. *Neuroreport, 15*, 1267–1270.

Thompson, R. F., & Spencer, W. A. (1966). Habituation: A model phenomenon for the study of neuronal substrates of behaviour. *Psychology Review, 73*, 16–43.

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, 23, 689–708.

Tremblay, K., Kraus, N., Carrell, T., & McGee, T. (1997). Central auditory system plasticity: Generalization to novel stimuli following listening training. *The Journal of Acoustic Society of America, 102*, 3762–3773.

Tremblay, K., Kraus, N., McGee, T., Ponton, C., & Otis, B. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear and Hearing, 22*, 79–90.

Van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology Section A, 58*, 251–273.

Van Heuven, V. J. (1985). Perception of stress pattern and word recognition: Recognition of Dutch words with incorrect stress position. *Journal of the Acoustical Society of America*, 78, S21.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Duffau, H., Crivello, F., Houdé, O., … Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage, 30*, 1414–1432.

Vigneau, M., Beaucousin, V., Hervé, P. Y., Jobard, G., Petit, L., Crivello, F., … Tzourio-Mazoyer, N. (2011). What is right-hemisphere contribution to phonological, lexico-semantic, and sentence processing? Insights from a meta-analysis. *NeuroImage, 54*, 577–593.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*, 325–329.

Wagenmakers, E. J. M., Zeelenberg, R., Steyvers, M., Shiffrin, R., & Raaijmakers, J. G. (2004). Nonword repetition in lexical decision: Support for two opposing processes. *The Quarterly Journal of Experimental Psychology Section A, 57*, 1191–1210.

Wang, Y., Behne, D. M., Jongman, A., Sereno, J. A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Applied Psycholinguistics, 25*, 449–466.

Wang, X., Gu, F., He, K., Chen, L., & Chen, L. (2012). Preattentive extraction of abstract auditory rules in speech sound stream: A

Mismatch Negativity study using lexical tones. *PLoS ONE, 7*, e30027.

Wang, P., & Knösche, T. R. (2013). A realistic neural mass model of the cortex with laminar-specific connections and synaptic plasticity: Evaluation with auditory habituation. PloS ONE, 8, e77876.

Wang, X. D., Wang, M., & Chen, L. (2013). Hemispheric lateralization for early auditory processing of lexical tones: Dependence on pitch level and pitch contour. *Neuropsychologia, 51*, 2238–2244.

Warren, D. E., & Duff, M. C. (2014). Not so fast: Hippocampal amnesia slows word learning despite successful fast mapping. *Hippocampus, 24*, 920–933.

Whalen, D. H., & Liberman, A. M. (1987). Speech perception takes precedence over nonspeech perception. *Science*, *237*, 169–171.

Wiener, S., & Turnbull, R. (2015). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. Language and Speech. doi: 10.1177/0023830915578000

Wong, P. C. M., Parsons, L. M., Martinez, M., & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: The effect of linguistic contexts. *The Journal of Neuroscience*, *24*, 9153–9160.

Wong, P. C. M., Perrachione, T. K., & Parrish, T. B. (2007). Neural characteristics of successful and less successful speech and word learning in adults. *Human Brain Mapping, 28*, 995–1006.

Woods, D. L., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology, 65*, 447–459.

Wood, C. C., & Wolpaw, J. R. (1982). Scalp distribution of human auditory evoked potentials. II. Evidence for multiple sources and involvement of auditory cortex. *Electroencephalography and Clinical Neurophysiology, 54*, 25–38.

Wu, N. (武宁宁), & Shu, H. (舒华) (2003). Gating 技术与汉语听觉词加工 (The gating paradigm and spoken word recognition of Chinese). *心理学报*(Acta Psychologica), 35 (5): 582–590.

Xu, Z (徐志刚). (1992). *诗词韵律* (Prosody in poems). Jinan, China: Jinan Press.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y. ... Lowe, M. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping, 27*, 173–183.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14*, 609–630.

Yip, M. (2006). Tone: Phonology. In Brown, K. (Ed.), *Encyclopaedia of Language & Linguistics* (2nd ed., pp. 761–764). Oxford, UK: Elsevier.

Yip, M. C. (2001). Phonological priming in Cantonese spoken-word processing. *Psychologia, 44*, 223–229.

Yue, J., Bastiaanse, R., & Alter, K. (2014). Cortical plasticity induced by rapid Hebbian learning of novel tonal word-forms: Evidence from mismatch negativity. *Brain and Language, 139*, 10–22.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11*, 946–953.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*, 37–46.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*, 846–849.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical*

Transactions of the Royal Society of London. Series B, Biological Sciences, *363*, 1087–1104.

Zhang, Q., & Damian, M. F. (2009). The time course of segment and tone encoding in Chinese spoken production: An event-related potential study. *Neuroscience, 163*, 252–65.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia, 49*, 1761–1770.

Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes, 10*, 545–600.

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes, 11*, 589–596.

Groningen dissertations in linguistics (grodil)

1. Henriëtte de Swart (1991). *Adverbs of Quantification: A Generalized Quantifier Approach*.
2. Eric Hoekstra (1991). *Licensing Conditions on Phrase Structure*.
3. Dicky Gilbers (1992). *Phonological Networks. A Theory of Segment Representation*.
4. Helen de Hoop (1992). *Case Configuration and Noun Phrase Interpretation*.
5. Gosse Bouma (1993). *Nonmonotonicity and Categorial Unification Grammar*.
6. Peter I. Blok (1993). *The Interpretation of Focus*.
7. Roelien Bastiaanse (1993). *Studies in Aphasia*.
8. Bert Bos (1993). *Rapid User Interface Development with the Script Language Gist*.
9. Wim Kosmeijer (1993). *Barriers and Licensing*.
10. Jan-Wouter Zwart (1993). *Dutch Syntax: A Minimalist Approach*.
11. Mark Kas (1993). *Essays on Boolean Functions and Negative Polarity*.

12. Ton van der Wouden (1994). *Negative Contexts*.

13. Joop Houtman (1994). *Coordination and Constituency: A Study in Categorial Grammar*.

14. Petra Hendriks (1995). *Comparatives and Categorial Grammar*.

15. Maarten de Wind (1995). *Inversion in French*.

16. Jelly Julia de Jong (1996). *The Case of Bound Pronouns in Peripheral Romance*.

17. Sjoukje van der Wal (1996). *Negative Polarity Items and Negation: Tandem Acquisition*.

18. Anastasia Giannakidou (1997). *The Landscape of Polarity Items*.

19. Karen Lattewitz (1997). *Adjacency in Dutch and German*.

20. Edith Kaan (1997). *Processing Subject-Object Ambiguities in Dutch*.

21. Henny Klein (1997). *Adverbs of Degree in Dutch*.

22. Leonie Bosveld-de Smet (1998). *On Mass and Plural Quantification: The case of French 'des'/'du'-NPs*.

23. Rita Landeweerd (1998). *Discourse semantics of perspective and temporal structure*.

24. Mettina Veenstra (1998). *Formalizing the Minimalist Program*.

25. Roel Jonkers (1998). *Comprehension and Production of Verbs in aphasic Speakers*.

26. Erik F. Tjong Kim Sang (1998). *Machine Learning of Phonotactics*.

27. Paulien Rijkhoek (1998). *On Degree Phrases and Result Clauses*.

28. Jan de Jong (1999). *Specific Language Impairment in Dutch: Inflectional Morphology and Argument Structure.*

29. H. Wee (1999). *Definite Focus.*

30. Eun-Hee Lee (2000). *Dynamic and Stative Information in Temporal Reasoning: Korean tense and aspect in discourse.*

31. Ivilin P. Stoianov (2001). *Connectionist Lexical Processing.*

32. Klarien van der Linde (2001). *Sonority substitutions.*

33. Monique Lamers (2001). *Sentence processing: using syntactic, semantic, and thematic information*.

34. Shalom Zuckerman (2001). *The Acquisition of "Optional" Movement*.

35. Rob Koeling (2001). *Dialogue-Based Disambiguation: Using Dialogue Status to Improve Speech Understanding*.

36. Esther Ruigendijk (2002). *Case assignment in Agrammatism: a cross-linguistic study.*

37. Tony Mullen (2002). *An Investigation into Compositional Features and Feature Merging for Maximum Entropy-Based Parse Selection.*

38. Nanette Bienfait (2002). *Grammatica-onderwijs aan allochtone jongeren.*

39. Dirk-Bart den Ouden (2002). *Phonology in Aphasia: Syllables and segments in level-specific deficits.*

40. Rienk Withaar (2002). *The Role of the Phonological Loop in Sentence Comprehension.*

41. Kim Sauter (2002). *Transfer and Access to Universal Grammar in Adult Second Language Acquisition.*

42. Laura Sabourin (2003). *Grammatical Gender and Second Language Processing: An ERP Study.*

43. Hein van Schie (2003). *Visual Semantics.*

44. Lilia Schürcks-Grozeva (2003). *Binding and Bulgarian.*

45. Stasinos Konstantopoulos (2003). *Using ILP to Learn Local Linguistic* Structures.

46. Wilbert Heeringa (2004). *Measuring Dialect Pronunciation Differences using Levenshtein Distance*.

47. Wouter Jansen (2004). *Laryngeal Contrast and Phonetic Voicing: ALaboratory Phonology*.

48. Judith Rispens (2004). *Syntactic and phonological processing indevelopmentaldyslexia*.

49. Danielle Bougaïré (2004). *L'approche communicative des campagnes de sensibilisation en santé publique au Burkina Faso: Les cas de la planification familiale, du sida et de l'excision*.

50. Tanja Gaustad (2004). *Linguistic Knowledge and Word Sense Disambiguation.*

51. Susanne Schoof (2004). *An HPSG Account of Nonfinite Verbal Complements in Latin*.

52. M. Begoña Villada Moirón (2005). *Data-driven identification of fixed expressions and their modifiability.*

53. Robbert Prins (2005). *Finite-State Pre-Processing for Natural Language Analysis.*

54. Leonoor van der Beek (2005) *Topics in Corpus-Based Dutch Syntax*

55. Keiko Yoshioka (2005). *Linguistic and gestural introduction and tracking of referents in L1 and L2 discourse.*

56. Sible Andringa (2005). *Form-focused instruction and the development of second language proficiency*.

57. Joanneke Prenger (2005). *Taal telt! Een onderzoek naar de rol van taalvaardigheid en tekstbegrip in het realistisch wiskundeonderwijs.*

58. Neslihan Kansu-Yetkiner (2006). *Blood, Shame and Fear: Self-Presentation Strategies of Turkish Women's Talk about their Health and Sexuality*.

59. Mónika Z. Zempléni (2006). *Functional imaging of the hemispheric contribution to language processing.*

60. Maartje Schreuder (2006). *Prosodic Processes in Language and Music.*

61. Hidetoshi Shiraishi (2006). *Topics in Nivkh Phonology.*

62. Tamás Biró (2006). *Finding the Right Words: Implementing Optimality Theory with Simulated Annealing.*

63. Dieuwke de Goede (2006). *Verbs in Spoken Sentence Processing: Unraveling the Activation Pattern of the Matrix Verb.*

64. Eleonora Rossi (2007). *Clitic production in Italian agrammatism.*

65. Holger Hopp (2007). *Ultimate Attainment at the Interfaces in Second Language Acquisition: Grammar and Processing.*

66. Gerlof Bouma (2008). *Starting a Sentence in Dutch: A corpus study of subject- and object-fronting.*

67. Julia Klitsch (2008). *Open your eyes and listen carefully. Auditory and audiovisual speech perception and the McGurk effect in Dutch speakers with and without aphasia.*

68. Janneke ter Beek (2008). *Restructuring and Infinitival Complements in Dutch.*

69. Jori Mur (2008). *Off-line Answer Extraction for Question Answering.*

70. Lonneke van der Plas (2008). *Automatic Lexico-Semantic Acquisition for Question Answering.*

71. Arjen Versloot (2008). *Mechanisms of Language Change: Vowel reduction in 15th century West Frisian.*

72. Ismail Fahmi (2009). *Automatic term and Relation Extraction for Medical Question Answering System.*

73. Tuba Yarbay Duman (2009). *Turkish Agrammatic Aphasia: Word Order, Time Reference and Case.*

74. Maria Trofimova (2009). *Case Assignment by Prepositions in Russian Aphasia.*

75. Rasmus Steinkrauss (2009). *Frequency and Function in WH Question Acquisition. A Usage-Based Case Study of German L1 Acquisition.*

76. Marjolein Deunk (2009). *Discourse Practices in Preschool. Young Children's Participation in Everyday Classroom Activities.*

77. Sake Jager (2009). *Towards ICT-Integrated Language Learning: Developing an Implementation Framework in terms of Pedagogy, Technology and Environment.*

78. Francisco Dellatorre Borges (2010). *Parse Selection with Support Vector Machines.*

79. Geoffrey Andogah (2010). *Geographically Constrained Information Retrieval.*

80. Jacqueline van Kruiningen (2010). *Onderwijsontwerp als conversatie. Probleemoplossing in interprofessioneel overleg.*

81. Robert G. Shackleton (2010). *Quantitative Assessment of English-American Speech Relationships*.

82. Tim Van de Cruys (2010). *Mining for Meaning: The Extraction of Lexico-semantic Knowledge from Text.*

83. Therese Leinonen (2010). *An Acoustic Analysis of Vowel Pronunciation in Swedish Dialects.*

84. Erik-Jan Smits (2010). *Acquiring Quantification. How Children Use Semantics and Pragmatics to Constrain Meaning*.

85. Tal Caspi (2010). *A Dynamic Perspective on Second Language Development.*

86. Teodora Mehotcheva (2010). *After the fiesta is over. Foreign language attrition of Spanish in Dutch and German Erasmus Student.*

87. Xiaoyan Xu (2010). *English language attrition and retention in Chinese and Dutch university students.*

88. Jelena Prokić (2010). *Families and Resemblances*.

89. Radek Šimík (2011). *Modal existential wh-constructions*.

90. Katrien Colman (2011). *Behavioral and neuroimaging studies on language processing in Dutch speakers with Parkinson's disease.*

91. Siti Mina Tamah (2011). *A Study on Student Interaction in the Implementation of the Jigsaw Technique in Language Teaching.*

92. Aletta Kwant (2011).*Geraakt door prentenboeken. Effecten van het gebruik van prentenboeken op de sociaal-emotionele ontwikkeling van kleuters.*

93. Marlies Kluck (2011). *Sentence amalgamation.*

94. Anja Schüppert (2011). *Origin of asymmetry: Mutual intelligibility of spoken Danish and Swedish.*

95. Peter Nabende (2011).*Applying Dynamic Bayesian Networks in Transliteration Detection and Generation.*

96. Barbara Plank (2011). *Domain Adaptation for Parsing.*

97. Cagri Coltekin (2011).*Catching Words in a Stream of Speech: Computational simulations of segmenting transcribed child-directed speech.*

98. Dörte Hessler (2011).*Audiovisual Processing in Aphasic and Non-Brain-Damaged Listeners: The Whole is More than the Sum of its Parts.*

99. Herman Heringa (2012). *Appositional constructions.*

100. Diana Dimitrova (2012). *Neural Correlates of Prosody and Information Structure.*

101. Harwintha Anjarningsih (2012).*Time Reference in Standard Indonesian Agrammatic Aphasia*.

102. Myrte Gosen (2012). *Tracing learning in interaction. An analysis of shared reading of picture books at kindergarten.*

103. Martijn Wieling (2012). *A Quantitative Approach to Social and Geographical Dialect Variation*.

104. Gisi Cannizzaro (2012). *Early word order and animacy*.

105. Kostadin Cholakov (2012). *Lexical Acquisition for Computational Grammars.A Unified Model.*

106. Karin Beijering (2012). *Expressions of epistemic modality in Mainland Scandinavian. A study into the lexicalization-grammaticalization-pragmaticalization interface.*

107. Veerle Baaijen (2012). *The development of understanding through writing.*

108. Jacolien van Rij (2012).*Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults.*

109. Ankelien Schippers (2012). *Variation and change in Germanic long-distance dependencies.*

110. Hanneke Loerts (2012).*Uncommon gender: Eyes and brains, native and second language learners, & grammatical gender.*

111. Marjoleine Sloos (2013). *Frequency and phonological grammar: An integrated approach. Evidence from German, Indonesian, and Japanese*.

112. Aysa Arylova. (2013) *Possession in the Russian clause. Towards dynamicity in syntax.*

113. Daniël de Kok (2013). *Reversible Stochastic Attribute-Value Grammars.*

114. Gideon Kotzé (2013). *Complementary approaches to tree alignment: Combining statistical and rule-based methods*.

115. Fridah Katushemererwe (2013). *Computational Morphology and Bantu Language Learning: an Implementation for Runyakitara*.

116. Ryan C. Taylor (2013). *Tracking Referents: Markedness, World Knowledge and Pronoun Resolution.*

117. Hana Smiskova-Gustafsson (2013). *Chunks in L2 Development: A Usage-based Perspective.*

118. Milada Walková (2013). *The aspectual function of particles in phrasal verbs.*

119. Tom O. Abuom (2013). *Verb and Word Order Deficits in Swahili-English bilingual agrammatic speakers.*

120. Gülsen Yılmaz (2013). *Bilingual Language Development among the First Generation Turkish Immigrants in the Netherlands.*

121. Trevor Benjamin (2013). *Signaling Trouble: On the linguistic design of other-initiation of repair in English conversation.*

122. Nguyen Hong Thi Phuong (2013). *A Dynamic Usage-based Approach to Second Language Teaching.*

123. Harm Brouwer (2014). *The Electrophysiology of Language Comprehension: A Neurocomputational Model.*

124. Kendall Decker (2014). *Orthography Development for Creole Languages.*

125. Laura S. Bos (2015). *The Brain, Verbs, and the Past: Neurolinguistic Studies on Time Reference.*

126. Rimke Groenewold (2015). *Direct and indirect speech in aphasia: Studies of spoken discourse production and comprehension.*

127. Huiping Chan (2015). *A Dynamic Approach to the Development of Lexicon and Syntax in a Second Language.*

128. James Griffiths (2015). *On appositives*.

129. Pavel Rudnev (2015). *Dependency and discourse-configurationality: A study of Avar.*

130. Kirsten Kolstrup (2015). *Opportunities to speak. A qualitative study of a second language in use.*

131. Güliz Güneş (2015). *Deriving Prosodic structures.*

132. Cornelia Lahmann (2015). *Beyond barriers. Complexity, accuracy, and fluency in long-term L2 speakers' speech.*

133. Sri Wachyunni (2015). *Scaffolding and Cooperative Learning: Effects on Reading Comprehension and Vocabulary Knowledge in English as a Foreign Language.*

134. Albert Walsweer (2015). *Ruimte voor leren. Een etnogafisch onderzoek naar het verloop van een interventie gericht op versterking van*

het taalgebruik in een knowledge building environment op kleine Friese basisscholen.

135. Aleyda Lizeth Linares Calix (2015). *Raising Metacognitive Genre Awareness in L2 Academic Readers and Writers.*

136. Fathima Mufeeda Irshad (2015). *Second Language Development through the Lens of a Dynamic Usage-Based Approach.*

137. Oscar Strik (2015). *Modelling analogical change. A history of Swedish and Frisian verb inflection.*

138. He Sun (2015). *Predictors and stages of very young child EFL learners' English development in China.*

139 Marieke Haan (2015). *Mode Matters*. *Effects of survey modes on participation and answering behavior*.

140. Nienke Houtzager (2015). *Bilingual advantages in middle-aged and elderly populations.*

141. Noortje Joost Venhuizen (2015). *Projection in Discourse: A data-driven formal semantic analysis.*

142. Valerio Basile (2015). *From Logic to Language: Natural Language Generation from Logical Forms*.

143. Jinxing Yue (2016). *Tone-word Recognition in Mandarin Chinese: Influneces of Lexical-level Representations.*

GRODIL

Center for Language and Cognition Groningen (CLCG)

P.O. Box 716

9700 AS Groningen

The Netherlands