



Universiteit
Leiden
The Netherlands

Tone and intonation processing: from ambiguous acoustic signal to linguistic representation

Liu, M.

Citation

Liu, M. (2018, November 1). *Tone and intonation processing: from ambiguous acoustic signal to linguistic representation*. LOT dissertation series. LOT, Utrecht. Retrieved from <https://hdl.handle.net/1887/66615>

Version: Not Applicable (or Unknown)

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/66615>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/66615> holds various files of this Leiden University dissertation.

Author: Liu, M.

Title: Tone and intonation processing: from ambiguous acoustic signal to linguistic representation

Issue Date: 2018-11-01

Tone and intonation processing:

**From ambiguous acoustic signal to
linguistic representation**

Published by

LOT

Trans 10

3512 JK Utrecht

The Netherlands

phone: +31 30 253 6111

e-mail: lot@uu.nl

<http://www.lotschool.nl>

Cover illustration: Surface signals and beyond, by Feifei Wang.

ISBN: 978-94-6093-299-1

NUR 616

Copyright © 2018: Min Liu. All rights reserved.

**Tone and intonation processing:
From ambiguous acoustic signal to
linguistic representation**

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op donderdag 1 november 2018
klokke 11.15 uur

door

MIN LIU

geboren te Shaanxi, China
in 1988

Promotor: Prof.dr. Niels O. Schiller

Co-promotor: Dr. Yiya Chen

Promotiecommissie: Prof.dr. Rint Sybesma (secretaris)
Prof.dr. Carlos Gussenhoven (Radboud
University)
Prof.dr. Claartje Levelt
Prof.dr. Herbert Schriefers (Radboud
University)

Table of Contents

Acknowledgements	IX
Chapter 1 General introduction	1
1.1 Neural correlates of tone and intonation processing in Standard Chinese	5
1.2 Context effects on tone and intonation processing in Standard Chinese	8
1.3 Tonal mapping between Standard Chinese and Xi'an Mandarin.....	10
1.4 Cross-dialect phonological similarity in segment and tone on bi-dialectal spoken word recognition	12
Chapter 2 Online processing of tone and intonation in Standard Chinese: Evidence from ERPs	15
2.1 Introduction.....	17
2.2 Method	23
2.2.1 Participants.....	23
2.2.2 Materials	23
2.2.3 Recording and stimuli preparation.....	25
2.2.4 Task.....	28
2.2.5 Procedure.....	29
2.2.6 EEG data recording	29
2.2.7 Behavioral data analysis	30
2.2.8 EEG data analysis.....	30
2.3 Results.....	32
2.3.1 Behavioral results.....	32
2.3.2 ERP results	34
2.4 General discussion.....	37
2.5 Conclusion	42
Chapter 3 Context matters for tone and intonation processing in Standard Chinese	43

3.1	Introduction.....	45
3.2	Experiment 1	49
3.2.1	Method	49
3.2.2	Results.....	53
3.3	Experiment 2.....	56
3.3.1	Method	56
3.3.2	Results.....	58
3.4	Experiment 1 vs. Experiment 2.....	61
3.4.1	Response accuracy.....	61
3.4.2	Reaction time.....	62
3.5	General discussion.....	63
3.6	Conclusion	67
Chapter 4 Tonal mapping of Xi'an Mandarin and Standard Chinese		69
4.1	Introduction.....	71
4.2	Experiment 1	76
4.2.1	Method	76
4.2.2	Results.....	79
4.3	Experiment 2.....	82
4.3.1	Method	83
4.3.2	Results.....	88
4.4	General discussion.....	91
4.5	Conclusion	96
Chapter 5 Effects of cross-dialect phonological similarity in segment and tone on bi-dialectal auditory word recognition: Evidence from Xi'an Mandarin and Standard Chinese.....		97
5.1	Introduction.....	99
5.2	The present study	102
5.3	Method	105
5.3.1	Participants.....	105
5.3.2	Stimuli.....	107
5.3.3	Stimuli recording.....	110
5.3.4	Procedure.....	111

5.3.5 Data analysis	112
5.4 Results.....	113
5.4.1 Response accuracy.....	113
5.4.2 Reaction time.....	114
5.5 General discussion.....	119
5.6 Conclusion.....	125
Chapter 6 General discussion.....	127
Appendix A.....	137
Appendix B.....	139
Appendix C.....	141
Appendix D.....	143
References.....	163
Summary.....	183
Samenvatting.....	189
摘要.....	195
Curriculum vitae.....	201

Acknowledgements

I could not have finished my PhD dissertation without the help and guidance of many people. My deepest gratitude goes to my supervisors Yiya Chen and Niels Schiller. I thank Yiya for shaping the mindset of me as an entry-level researcher, for sharing her expertise in research and wisdom in life with me, and for giving constructive comments and suggestions on my work and career. I am grateful to Niels for his continuous encouragement over the years. I thank him for giving me the freedom to pursue my research interests and for providing prompt feedback at the last stage of my dissertation revisions. I have learnt a lot from both my supervisors. It has been a great privilege working with them.

I owe special thanks to Prof. Qingfang Zhang and Dr. Xuhai Chen. I appreciate that they gave me access to lab facilities when I conducted experiments in China. Qun Yang, Tingting Zheng, thank you for offering technical support and helping me recruit participants for the experiments. To all my participants in Leiden and in China, thank you for expressing interest in my experiments and showing up at the lab on time. You made my life much easier.

As I look back over all these years at LUCL, my heart is filled with such gratitude for the many amazing friends and colleagues I have met here. Yifei, thank you for sticking with me through thick and thin along the journey. I cherish the time we spent together attending conferences, traveling, and sharing happiness and sorrows. Qian, thank you for enlightening me on the issue of statistics and cheering me up when I was down. Man, thank you for sharing your knowledge of psycholinguistics with me, and I really enjoyed the lovely dinners you hosted. Ting, it was so much fun discussing research with you and wandering around cities. Han, you inspired me to relish and experience life. Thank you. Bobby, I owe you special thanks for helping me set up my experiments. I am extremely grateful that you helped me overcome frustration when things did not work out. Arum, Nazar, Nurenzia, Amanda, my dear office mates, I feel incredibly lucky to work side by side with you. Your loving and caring nature has made my days much more productive. I am indebted to

Cesko, my paranymph. Thank you for translating my Nederlandse samenvatting and helping me out with the final preparations. I would also like to thank Kate Bellamy and Eric Shek for proofreading the dissertation. Thanks also go to the anonymous reviewers of my submitted papers, and to the doctoral committee of this dissertation: Prof. Rint Sybesma, Prof. Carlos Gussenhoven, Prof. Herbert Schriefers and Prof. Claartje Levelt. Some of the other colleagues inside and outside LUCL I would like to thank are Aliza Glabergen-Plas, Bastien Boutonnet, Bei Peng, Eleanor Dutton, Elisabeth Mauder, Hanna Fricke, Hang Cheng, Lei Huang, Kalinka Timmer, Lesya Ganushchak, Leticia Pablos, Libo Geng, Lisette Jager, Jiang Wu, Jos Pacilly, Junru Wu, Menghui Shi, Mengru Han, Qi Wang, Qing Yang, Olga Kepinska, Saskia Lensink, Stella Gryllia, Xin Li, Yang Yang and Yunus Sulistyono.

It has been a really challenging but deeply rewarding five years. One person that made all this happen in the first place is my MA supervisor, Prof. Jinsong Zhang. For that, I owe him special thanks. I thank him for arousing my interest in research several years ago and I thank him, in particular, for encouraging me to pursue my PhD abroad.

I have enjoyed my life in Leiden greatly. I am especially thankful to two individuals who made significant contributions to it, James and Cynthia. Thank you for helping me settle in when I first came, and thank you for treating me like family all the time.

A huge thanks to my wonderful friends: Feifei, Shuju, Yanhua, Shanshan, Xiaolong, Cuicui, Fanhong, Xiaorong and Xiaotang. I feel so blessed to have you always there for me. Thank you.

Last but not least, no words can describe how thankful I am to my family: my mum and dad, my elder sister and my younger brother. Thank you for your unconditional love and endless belief in me. You have given me the strength to keep going.

This dissertation was funded by a PhD scholarship from the Chinese Scholarship Council. I would also like to acknowledge the partial financial support from the Leiden University Centre for Linguistics and the Leids Universiteits Fonds.

Chapter 1

General introduction

Spoken language processing is a task that humans continuously perform from birth. In this process, different aspects of linguistic information are involved, such as lexical, semantic, syntactic and prosodic information (Friederici, 2002; Isel, Alter, & Friederici, 2005). As “the organizational structure of speech” (Beckman, 1996), prosody is a determinant of the form of spoken language (Cutler, 1997). Spoken language processing therefore cannot be successful without a proper understanding of the prosodic information conveyed in spoken utterances.

Prosodic information is not always explicitly represented with a clear-cut interpretation. Spoken language is by its very nature a stream of speech signals, which are acoustically realized in terms of, for example, fundamental frequency (F0), duration, and intensity (Wagner & Watson, 2010). Speech signals are inherently ambiguous (Mirman, 2008). Not uncommonly, the same form of a speech signal can represent different prosodic information and therefore cause ambiguity. For example, a high, level phrase-final pitch contour that does not occur sentence-finally in English can indicate either an intermediate phrase boundary or an intonation phrase boundary (Speer & Blodgett, 2006). The question that arises is how ambiguous acoustic signals representing different prosodic information affect spoken language processing.

The most prominent prosodic feature of tonal languages such as Standard Chinese is lexical tone. F0 has been identified as the primary acoustic correlate of tones in Standard Chinese (Howie, 1976; Yip 2002), with T1 having a high-level contour (55¹), T2 a mid-rising contour (35), T3 a low-dipping (214), and T4 a high-falling contour (51). Tones distinguish lexical meanings at a lexical level. The same segment *ma* means *mother*, *bemp*, *horse* and *scold*, respectively, when it is combined with the four lexical tones.

In Standard Chinese, F0 is not only used to distinguish lexical meanings at the lexical level, it is also recruited to signal post-lexical information such as intonation types at the sentential level (Shen, 1985; Wu, 1982; Xu & Wang,

¹ Tone values are transcribed using a 5-point scale notation system according to Chao (1968); each tone is described by the initial and the end point of the pitch level.

2001). Question intonation in Standard Chinese is generally realized as an upward trend of the F0 contour while statement intonation is realized as a downward trend (Ho, 1977; Gårding, 1987; Liu & Xu, 2005). Previous production studies have demonstrated that the upward trend of F0 in question intonation is more pronounced at the end of sentences than at the beginning (Kratochvil, 1998; Liu & Xu, 2005; Xu, 2005; Peng et al., 2005), although some studies also reported an overall F0 rising of sentences in questions compared to statements (Ho, 1977; Shen, 1989).

Consequently, the dual functions of F0 lead to the interaction of tone at the lexical level and intonation at the sentential level in Standard Chinese. This raises the question of how tone and intonation are processed when the surface pitch contour cues both linguistic functions (i.e., tone and intonation). Existing studies have shown that the dual functions of F0 in Standard Chinese cause ambiguity in speech signals and result in pitch processing difficulties at the behavioral level (Yuan, 2011; Xu & Mok, 2012a, 2012b). However, what are the underlying neural mechanisms leading to the eventual behavioral decisions of tone and intonation processing? How do native listeners resolve the pitch processing difficulties? These issues are less well-understood and further research on tone and intonation processing is needed. Chapters 2 and 3 of this dissertation therefore set out to address these issues.

Speech ambiguity can also arise when the same or similar pitch contour(s) cues the same linguistic representation (e.g., tone), but different categories of that representation in two linguistic systems (e.g., different tonal categories in two tonal systems) of the same speaker. Indeed, most Standard Chinese speakers also speak another local Chinese dialect (Li & Lee, 2008; Wiener & Ito, 2014). Some of the dialects differ from Standard Chinese in both segmental and tonal information, whereas others, such as dialects within the Mandarin family, overlap largely in segmental information and tonal features with Standard Chinese. Across these Mandarin dialects and Standard Chinese, the same or similar pitch contour(s) is often used to characterize two different tonal categories, which in turn result in different lexical meanings. For example, the high-level pitch contour with a pitch value of 55 corresponds to a different tonal category in Xi'an Mandarin than in Standard Chinese (Luo & Wang, 1981;

Ren, 2012; Zhang & Shi, 2009). In the former, the syllable *ma55* means *scold*. In the latter, it means *mother*. For bi-dialectal speakers of such Mandarin dialects and Standard Chinese, the question arising is whether the same or similar pitch contours from the two tonal systems are taken as representations of the same tone in pitch processing. Furthermore, what role does tone play in the activation and processing of bi-dialectal lexical representation? Would ambiguous acoustic signals due to cross-dialect phonological similarity in segment and tone affect bi-dialectal listeners' lexical access during spoken word recognition? If yes, would the bi-dialectal listeners benefit or suffer from cross-dialect phonological similarity? Moreover, are the effects of cross-dialect phonological similarity on bi-dialectal auditory word recognition similar or different from the effects of cross-language phonological similarity on bilingual auditory word recognition? Chapters 4 and 5 attempt to answer these questions.

Currently, relatively little empirical research has been conducted to investigate the tonal features of other Mandarin dialects except for Standard Chinese. Even less empirical research concerns the phonological similarities or differences between the tonal system of a Mandarin dialect and that of Standard Chinese. Of all the Mandarin dialects, Xi'an Mandarin offers a very interesting test case to investigate cross-dialect phonological similarity effects. This is because each Xi'an Mandarin tone has a corresponding tone in Standard Chinese with which it shares similar tonal contour and pitch value, resulting in a seeming one-to-one correspondence pattern in tones between the two dialects (Li, 2001; Zhang, 2009). Using Xi'an Mandarin as a test case, Chapters 4 and 5 investigate how a tonal system of a closely related dialect of Standard Chinese (i.e., Xi'an Mandarin) affects tone processing (Chapter 4) and lexical access (Chapter 5) of bi-dialectal tonal language speakers.

Overall, this dissertation investigates how ambiguous acoustic signals representing different prosodic information affect spoken language processing. Specifically, it investigates how pitch is processed within a linguistic system (i.e., Standard Chinese) and across two linguistic systems (i.e., Standard Chinese and Xi'an Mandarin) when the same pitch contour cues different linguistic functions (i.e., tone and intonation) or different categories of the same linguistic function (i.e., tone). Chapter 2 taps into the neural correlates of tone and

intonation processing in Standard Chinese. Chapter 3 examines the role of semantic context in resolving pitch processing difficulties in tone and intonation processing in Standard Chinese. Chapter 4 empirically compares the tonal systems of Xi'an Mandarin and Standard Chinese in tone production and perception. Based on the tonal similarity results between Standard Chinese and Xi'an Mandarin, Chapter 5 further investigates if and how cross-dialect phonological similarity in segment and tone affects bi-dialectal lexical access during spoken word recognition. The rest of this chapter will introduce the background to these research questions and provide a brief overview of each chapter.

1.1 Neural correlates of tone and intonation processing in Standard Chinese

Both tone and intonation in Standard Chinese adopt F0 as their primary acoustic correlate (Ho, 1977; Shen, 1985; Wu, 1982; Xu & Wang, 2001; Xu, 2004). The dual functions of F0 lead to the interaction of final lexical tone and sentence intonation. When a statement ends with a falling tone (T4) or a question ends with a rising tone (T2), the F0 encodings of the final lexical tone and sentence intonation are in congruency. However, when a statement ends with a rising tone (T2) or a question ends with a falling tone (T4), the F0 encodings of the final lexical tone and sentence intonation are in conflict. This raises the question of how tone and intonation are processed in Standard Chinese when their F0 encodings are in conflict or in congruency.

There have only been a handful of studies on the effect of intonation on tone perception and vice versa. Connell, Hogan, and Rozsypal (1983) ran a tone perception experiment in Standard Chinese and found that intonation-induced F0 has little effect on tone perception. Tone identity is maintained in question intonation. With regard to the effect of tone on intonation perception, Yuan (2011) found that in Standard Chinese, questions ending with T4 (falling tone) were easier to identify than questions ending with T2 (rising tone).

The general consensus, therefore, is that it is difficult to disentangle tonal and intonational information from the surface pitch contour when the F0 encodings of tone and intonation are in congruency. The dual functions of F0

can cause intonation processing difficulty at the behavioral level in Standard Chinese. However, the underlying neural mechanisms leading to the eventual behavioral decisions are not yet clear. To shed light on this issue, Chapter 2 in this dissertation taps into the neural correlates of tone and intonation processing in Standard Chinese using the event-related potential (ERP) technique.

The ERP technique is a non-invasive technique which can be used to reveal the neural responses in the brain to ongoing specific events (Luck, 2005). It has been used to investigate online pitch processing, mostly tone processing, due to its high temporal resolution. Very few studies have examined the online processing of both tone and intonation. Ren, Yang, and Li (2009) and Ren, Tang, Li, and Sui (2013) probed native listeners' brain activities underlying the processing of tone and intonation in Standard Chinese at the pre-attentive stage with a one-syllable sentence. They found a mismatch negativity (MMN) effect for the question-statement contrast when the intonation is combined with T4, but not when the intonation is combined with T2. As the MMN is linked to higher order perceptual processes underlying stimulus discrimination (Pulvermüller & Shtyrov, 2006), these two studies suggest that at the pre-attentive stage, native listeners can tease apart question intonation from statement intonation when the intonation is combined with T4, but not when the intonation is combined with T2, just as what Yuan (2011) reported with behavioral perceptual judgment data.

Given that intonation is a feature at the sentential level that typically spans over several lexical items, it is necessary to investigate how tone and intonation are processed online when the length of an utterance is extended from one syllable to several syllables. Moreover, at the pre-attentive stage, there is no correlated measure of the behavioral effects given the deviant acoustic stimulus in the stream, making it impossible to distinguish between automatic neural responses arising from acoustic variability and responses related to "attention capture" (Fritz, Elhilali, David, & Shamma, 2007). The investigation of online processing of tone and intonation therefore needs to be extended from the pre-attentive to the attentive stage. Additionally, semantic context has been shown to affect the processing of tone (Ye & Connine, 1999). Little is known,

however, about how semantic context affects the processing of both tone and intonation when there is interaction between them. A study of tone and intonation processing in a neutral semantic context can serve as a baseline comparison for further research. Thus, Chapter 2 investigated the online processing mechanisms of tone and intonation in Standard Chinese over a broader sentence domain at the attentive stage under neutral semantic context.

In this chapter, native Standard Chinese listeners were presented with semantically neutral Standard Chinese sentences, which contrast in final tones (rising T2 or falling T4) and intonations (Question or Statement). During this process, their electrophysiological responses were recorded. One second after the offset of the sentences, listeners were asked to perform either a tone identification task or an intonation identification task. The behavioral results showed that while the identification of tone was not hindered by intonation, the identification of intonation was greatly impeded due to the interference of tone. In the T4 conditions, question intonation was rather difficult to identify correctly, whereas the identification of statement intonation presented almost no difficulty at all. In the T2 conditions, question intonation was still difficult to identify, while identification of statement intonation also tended to be problematic. Regarding the ERP results, a clear P300 effect was observed for the question-statement contrast in sentences ending with T4, but no ERP effect was found for the question-statement contrast in sentences ending with T2. These results in Chapter 2 provide ERP evidence for the interaction of tone and intonation in Standard Chinese, confirming the findings from behavioral metalinguistic data that native Standard Chinese listeners can distinguish between question and statement intonation when the intonation is associated with a final T4, but fail to do so when the intonation is associated with a final T2. The study reported in Chapter 2 extends our understanding of online processing of tone and intonation 1) from the pre-attentive stage to the attentive stage; and 2) within a larger domain (i.e. multi-word utterances) than a single word utterance.

1.2 Context effects on tone and intonation processing in Standard Chinese

The interaction of tone and intonation leads to intonation processing difficulty in Standard Chinese. Yuan (2011) found that in natural sentences, questions ending with T4 (falling tone) were easier to identify than questions ending with T2 (rising tone). A similar asymmetrical pattern of perception was also reported in Xu and Mok (2012a). However, in a follow-up study using low-pass filtered speech (Xu & Mok, 2012b), the pattern was reversed; Standard Chinese listeners were found to be better at identifying questions ending with T2 than questions ending with T4. The reversed patterns are very thought-provoking: what could be the reasons for them? Intuitively, these reversed perception patterns might result from many factors, such as prosodic features and lexical intelligibility, among which a potentially very important factor is sentence context.

Sentence context has been shown to facilitate the processing of tone. Ye and Connine (1999) investigated tone processing in Standard Chinese with the target syllables occurring in sentence-final position in a semantically highly constraining context (i.e., idiomatic context) or a semantically neutral context. They found that the semantically highly constraining context considerably facilitated the processing of tone.

Sentence context can also play a role in disentangling tonal information from intonational information when tone and intonation interact. In Cantonese, another Chinese variety with lexical tone, tone and intonation interact and cause perceptual processing difficulty of low tone in questions. When embedding the low tone words sentence-finally in a semantically neutral context versus a semantically strong biasing context (i.e., a disyllabic word context), Kung, Chwilla, and Schriefers (2014) found that the latter led to much better lexical-identification performance for words with a low tone at the end of questions.

In contrast to tone processing difficulty in questions in Cantonese, the interaction of tone and intonation leads to intonation processing difficulty in Standard Chinese (Xu & Mok, 2012a, 2012b; Yuan, 2011). This contrast calls for further research on the potential typology of the interaction between tone

and intonation in tonal languages. Moreover, while we know that context facilitates tone processing in Standard Chinese, the specific role of context, in particular its role in intonation processing and in disentangling intonation from tone processing, remains unclear.

Chapter 3 therefore investigated how tone and intonation are processed in Standard Chinese, as a function of semantic context when F0 encodings of the final lexical tone and sentence intonation are in conflict or in congruency. Two experiments were conducted to address this issue. Experiment 1 examined tone and intonation processing in a semantically neutral context, while Experiment 2 examined tone and intonation processing in a semantically constraining context. In each sentence context, tone and intonation identification experiments were performed using the same design with the same group of native speakers of Standard Chinese, allowing for a direct systematic comparison of tone versus intonation identification. The resulting measurements included the commonly-reported response accuracy, as well as an additional measurement, reaction time.

Results showed that the overall performance of tone identification was better than that of intonation. Tone identification was seldom affected by intonation information irrespective of semantic contexts. Intonation identification, particularly question intonation, however, was susceptible to the final lexical tone identity and was greatly affected by the semantic context. Specifically, in the semantically neutral context, questions were difficult to identify regardless of the lexical tone identity. In the semantically constraining context, questions ending with a falling tone were much better identified than questions ending with a rising tone, suggesting that top-down information provided by the semantically constraining context can play a facilitating role for listeners to disentangle intonational information from tonal information, especially in sentences with a lexical falling tone in the final position. Therefore, to resolve pitch processing difficulties more efficiently in tone and intonation processing in Standard Chinese, not only acoustic cues, but also semantic context has to be taken into consideration.

1.3 Tonal mapping between Standard Chinese and Xi'an Mandarin

The dual or multiple linguistic functions of pitch in a single linguistic system such as Standard Chinese and the corresponding pitch processing costs have received widespread attention among researchers. Less attention has been paid, however, to the fact that the same or similar pitch contours may cue the same function of pitch (e.g., tone) but different categories of that function in two tonal systems of the same speaker. As mentioned earlier, this is not rare for many Standard Chinese speakers, as most speakers of Standard Chinese also speak a local dialect (Li & Lee, 2008; Wiener & Ito, 2014), which may share tonal similarities with Standard Chinese.

It is of both practical and theoretical value to systematically investigate the tonal similarity or difference between different Chinese dialects and Standard Chinese. Such investigations can be the prerequisite a) to developing dialect-oriented speech synthesis and speech recognition technology (Czap & Zhao, 2017), b) to guiding language pedagogy in teaching Standard Chinese to dialectal speakers (Lam, 2005; Wong & Xiao, 2010), and c) to addressing issues such as whether the phonological information of one's two or more dialects are stored separately or integrally (Wu, 2015), or how cross-dialect phonological similarity or difference affects lexical access in the minds of bi-dialectal tonal language speakers.

The target language for this set of studies is Xi'an Mandarin, which overlaps largely in segmental features with Standard Chinese. The tones of Xi'an Mandarin seem to have one-to-one correspondence with those of Standard Chinese (Li, 2001; Zhang, 2009), providing a very informative test case for tonal similarities. As in Standard Chinese, there are four tonal categories in Xi'an Mandarin, and they are referred to as T1, T2, T3 and T4. Here, the terms T1-T4 are adopted to suggest that words which share the same tonal categories across the two dialects are etymologically-related translation equivalents in most cases. On the 5-point scale notation system (Chao, 1930; 1968), the pitch value of the Standard Chinese tones has been established as 55 (T1), 35 (T2), 214 (T3) and 51 (T4). However, studies have shown variance as to the specific pitch value of each Xi'an Mandarin tone. The majority of existing studies on Xi'an

Mandarin tones have been based on impressionistic observations (e.g., Bai, 1954; Luo & Wang, 1981; Peking University, 2003; Sun, 2007; Wang, 1996; Yuan, 1989). The very few acoustic studies on Xi'an Mandarin tones either sampled from a very limited number of speakers (e.g., two in Ma (2005); one in Ren (2012)) or lacked control of the lexical properties of the stimuli used (e.g., Zhang & Shi, 2009). It is not known to what degree these results can represent the typical tonal patterns of Xi'an Mandarin. Nevertheless, the basic tonal contour shape of Xi'an Mandarin tones tends to be largely consistent across studies, and each Xi'an Mandarin tone seems to have a corresponding tone in Standard Chinese with which it shares similar tonal contour and pitch value. It appears that there is a systematic mapping between the two tonal systems (Li, 2001; Zhang, 2009). However, better-designed empirical research is needed to establish the acoustic similarities or differences between the two tonal systems. Additionally, it is unclear whether the same or similar pitch contours across the two dialects are taken as representations of the same tone in pitch processing.

Chapter 4 therefore empirically compared the tonal systems of Xi'an Mandarin and Standard Chinese by means of a tone production and a tone perception experiment. Tonal categories with similar contours from Xi'an Mandarin and Standard Chinese were paired. Both tone production and perception experiments were carried out on highly proficient bi-dialectal speakers of Xi'an Mandarin and Standard Chinese, to reveal whether each pair of tones is acoustically and perceptually similar or different. Acoustic results showed that the F₀ difference ranged from no F₀ difference (level contour tone pair) through F₀ curvature difference (rising contour tone pair) to F₀ height difference (falling contour tone pair) and F₀ contour difference (low contour tone pair). Except for the falling contour tone pair, all the other tone pairs also exhibited differences in tone duration. The varying acoustic differences in different tone pairs, together with the phonological rule, resulted in varying degrees of tonal similarity in tone perception, but tones with similar contours between the two dialects were basically perceived to be the same. The two experiments together showed that there are indeed systematic mappings of tones between Xi'an Mandarin and Standard Chinese.

1.4 Cross-dialect phonological similarity in segment and tone on bi-dialectal spoken word recognition

The systematic mapping of tones between Standard Chinese and Xi'an Mandarin, as shown in Chapter 4, together with the large overlap of segmental features between the two dialects, makes cross-dialect homophones prevalent in the two languages. Cross-dialect minimal tone pairs (i.e., syllables sharing the segmental structure but not tonal contour) are also common in Standard Chinese and Xi'an Mandarin. How cross-dialect phonological similarity in segment and tone affects bi-dialectal spoken word recognition is the focus of interest in Chapter 5.

Little research has been conducted on bi-dialectal word recognition due to phonological similarity in tonal languages. However, there has been a considerable amount on bilingual word recognition due to phonological similarity in non-tonal Indo-European languages. An extreme case of phonological similarity is homophony. Bilingual word recognition studies have consistently shown that bilingual speakers find it harder to process interlingual homophones than non-homophonous control words. Moreover, the effect is robust across experimental tasks and modalities, be it a lexical decision task (Dijkstra, Grainger, & Van Heuven, 1999; Doctor & Klein, 1992; Lagrou, Hartsuiker, & Duyck, 2011; Nas, 1983), a gating task (Grosjean, 1988), or a word form priming task (Schulpen, Dijkstra, Schriefers, & Hasper, 2003), and be stimuli presented in the visual (Dijkstra et al., 1999; Doctor & Klein, 1992) or the auditory modality (Lagrou et al., 2011). These studies suggest parallel activation of homophone candidates from both languages and an interference effect of cross-language phonological similarity on word recognition. For tonal languages, phonological similarity between languages can be due to overlap in segment and/or tone. The question that arises is whether in bi-dialectal lexical processing, homophones co-activate and interfere, as in the bilingual situation. Furthermore, for tonal language speakers, what role does tone play in the activation and processing of bi-dialectal lexical representations during spoken word recognition?

Studies on the role of tone in lexical processing have mostly been conducted on Chinese varieties (mostly Standard Chinese and Cantonese) in a

monolingual context. The general consensus is that tonal information is used in recognition (Ching, 1985; Fox & Unkefer, 1985). However, contradictory results have been obtained as to whether tonal information constrains lexical activation. Using an auditory-auditory priming paradigm, Lee (2007) found a facilitatory priming effect when primes and targets overlapped in both segment and tone. Segment-only overlap (minimal tone pair) or tone-only overlap did not produce any priming effect, comparable to the baseline condition where primes and targets overlapped in neither segment nor tone. Sereno and Lee (2015), however, raised the concern that Lee (2007) did not control for the tonal similarity of the prime-target pairs. They conducted a follow-up study with balanced tonal distribution in the prime-target pairs and replicated the identity priming effect in Lee (2007) for the segment and tone overlap condition. In addition, they found a segment-only overlap facilitation effect, though smaller than the identity priming effect. Tone-only overlap, on the other hand, produced significant inhibition.

Given the conflicting results, more research is clearly needed to establish the role of lexical tone in auditory word recognition in Standard Chinese. It is also important to note that most speakers of Standard Chinese are bi-dialectal speakers. Existing studies have not controlled for participants' dialect background, which could be a potential cause of the different roles of tone and segment found in the literature. This study therefore set out to directly tap into their role(s) in bi-dialectal speakers' lexical processing. Specifically, Chapter 5 investigated the effect of cross-dialect phonological similarity in segment and tone on auditory word recognition in a bi-dialectal context (i.e., Standard Chinese and Xi'an Mandarin) using the auditory-auditory priming paradigm.

Balanced bi-dialectal speakers of Xi'an Mandarin and Standard Chinese took part in an auditory-auditory priming experiment with a generalized lexical decision task. The primes were monosyllabic homophones from either Xi'an Mandarin or Standard Chinese while the targets were disyllabic Xi'an Mandarin or Standard Chinese words. Primes and the first syllable of the target words overlapped in both segment and tone within a dialect (identical) or across two dialects (interdialectal homophones), or they overlapped in segment only within a dialect or across two dialects. In addition, a control condition was included

where primes and targets shared neither tone nor segment. Results showed that Standard Chinese primes did not yield significant priming effects for within- or cross-dialect segment-only overlap targets. Standard Chinese primes did not produce significant priming effects for within-dialect identical targets either. However, they did yield significant inhibitory priming effects for cross-dialect homophone targets. This overall pattern was reversed for Xi'an Mandarin primes because these primes were not treated differently from their interdialectal homophonous primes in the current mixed dialect setting. These results suggest that cross-dialect phonological similarity in segment alone does not affect lexical access in bi-dialectal auditory word recognition while cross-dialect phonological similarity in both segment and tone poses a threat to the recognition system of bi-dialectal listeners. It was concluded that tonal information plays a significant role in constraining word activation in bi-dialectal auditory word recognition. The results extend our understanding of the role of segment and tone in auditory word recognition in tonal languages from the monolingual context to the bi-dialectal context, and reveal a non-selective processing mechanism in bi-dialectal lexical access during auditory word recognition, as in bilingual lexical access.

Chapter 2

Online processing of tone and intonation in Standard Chinese: Evidence from ERPs²

² A version of this chapter is published as: Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia*, 91, 307-317.

Abstract

Event-related potentials (ERPs) were used to investigate the online processing of tone and intonation in Standard Chinese at the attentive stage. We examined the behavioral and electrophysiological responses of native Standard Chinese listeners to Standard Chinese sentences, which contrast in final tones (rising Tone2 or falling Tone4) and intonations (Question or Statement). A clear P300 effect was observed for the question-statement contrast in sentences ending with Tone4, but no ERP effect was found for the question-statement contrast in sentences ending with Tone2. Our results provide ERP evidence for the interaction of tone and intonation in Standard Chinese, confirming the findings from behavioral metalinguistic data that native Standard Chinese listeners can distinguish between question intonation and statement intonation when the intonation is associated with a final Tone4, but fail to do so when the intonation is associated with a final Tone2. Our study extends the understanding of online processing of tone and intonation 1) from the pre-attentive stage to the attentive stage and 2) within a larger domain (i.e., multi-word utterances) than a single word utterance.

Keywords: Tone2, Tone4, intonation, Standard Chinese, attentive processing, P300

2.1 Introduction

In spoken language processing, different aspects of linguistic information are involved, such as lexical, semantic, syntactic and prosodic information (Friederici, 2002; Isel, Alter, & Friederici, 2005). Among these aspects, prosodic information, especially pitch information, has been shown to be indispensable for spoken language processing in tonal languages such as Standard Chinese (e.g., Li, Chen, & Yang, 2011). Tone and intonation have been considered the two most significant prosodic features of Standard Chinese speech (Tseng & Su, 2014). At the lexical level, F0 is employed to differentiate the four lexical tones (Tone1 - high-level, Tone2 - mid-rising, Tone3 - low-dipping and Tone4 - high-falling), which contrast lexical meanings (Cutler & Chen, 1997; Yip, 2002). At the sentential level, F0 is also used to convey post-lexical information, for example, intonation types (e.g., question intonation, statement intonation) (Ladd, 2008). Although other acoustic correlates (such as duration, intensity and phonation) have also been shown to contribute to cue tonal and intonational contrasts (Garellek, Keating, Esposito, & Kreiman, 2013; Hu, 1987; Shi, 1980; Xu, 2009; Yu & Lam, 2014), F0 has been identified as the primary acoustic correlate of both tone and intonation in Standard Chinese (Ho, 1977; Shen, 1985; Wu, 1982; Xu & Wang, 2001; Xu, 2004). It may therefore not be surprising that tone and intonation interact with each other both in production and perception.

The interaction of tone and intonation in Standard Chinese has aroused great interests among researchers, and several models or theories have been put forward based mainly on acoustic data since the first studies on the topic by Chao (Chao, 1929, 1933). Of these acoustic studies, a general belief is that question intonation has higher F0 than statement intonation (Cao, 2004; Gårding, 1987; Shen, 1989; Wu, 1996). However, there is controversy about the temporal scope of such higher F0 in question intonation. Two alternative views have been established. One holds that there is an overall F0 rising of sentences in questions compared to statements (Ho, 1977; Shen, 1989; Yuan, 2011). The other claims that the F0 difference between questions and statements is more pronounced towards the end of the sentences (Kratochvil, 1998; Liu & Xu, 2005; Xu, 2005; Peng et al., 2005).

Different from the above acoustic studies, Liang and Van Heuven (2007) conducted intonation perception experiments with a seven-syllable sentence containing merely high-level tone syllables. They manipulated both the overall pitch level of the sentence and the pitch level of the final tone. Results showed that manipulating the final rise has a much stronger effect on the perception of intonation type than manipulation of the overall pitch level, indicating that the F0 of the final tone is more important than that of the whole sentence for intonation perception.

Not unique to Standard Chinese, the final rise has been shown to be a language-universal perceptual cue for question intonation (Gussenhoven & Chen, 2000). In a made-up language, Gussenhoven and Chen (2000) tested the perceptual cues for question intonation across three different language groups. All listeners tended to take the higher peak, the later peak and the higher end rise as cues for question intonation perception. In Cantonese, another representative language other than Standard Chinese within the Sinitic language family, Ma, Ciocca, and Whitehill (2011) also found that the perception of questions and statements relies primarily on the F0 characteristics of the final syllables.

Apart from studies on the temporal domain of perceptual cues of intonation, there has also been research, though regrettably little, on the effect of intonation on tone perception and vice versa. Connell, Hogan, and Rozsypal (1983) ran a tone perception experiment in Standard Chinese and found that intonation-induced F0 has little effect on tone perception. Tone identity is maintained in question intonation. With regard to the effect of tone on intonation perception, Yuan (2011) found that in Standard Chinese, questions ending with Tone4 (falling tone) were easier to identify than questions ending with Tone2 (rising tone). Three mechanisms were proposed for question intonation: an overall higher phrase curve, higher strengths of sentence-final tones and a tone-dependent mechanism. The tone-dependent mechanism conflicts with the strength mechanism on the final Tone2, possibly accounting for the difficulty of question identification in sentences ending with Tone2. In sentences ending with Tone4, the tone-dependent mechanism flattens the falling slope of the

final falling, making question intonation perceptually more salient for falling tone (Yuan, 2006).

Unlike in Standard Chinese, the intonation-induced F0 affects tone perception in Cantonese. Low tones (21, 23, 22) (tone values in 5-point scale notation, each tone is described by the initial and the end point of the pitch level) were misperceived as the mid-rising tone (25) at the final positions of questions (Fok-Chan, 1974; Kung, Chwilla, & Schriefers, 2014; Ma et al., 2011). This is probably because with a rising tail superimposed on all tone contours by question intonation (Ma, Ciocca, & Whitehill, 2006), the F0 contour of the low tones in questions resembles that of a mid-rising tone in questions. As for the effect of tone on intonation perception, native listeners were least accurate of all the six tones in Cantonese in distinguishing statements and questions for sentences ending with Tone 25 (Ma et al., 2011), suggesting that listeners confused the rising contour of Tone 25 with the final rise of question intonation.

Taken together, potential conflicts exist between tone and intonation in Standard Chinese and Cantonese, causing processing difficulties at the behavioral level. However, the underlying neural mechanisms leading to the eventual behavioral decisions are not yet clear. To shed light on this issue, research is needed to investigate the online processing of tone and intonation.

In recent years, a number of neurophysiological studies in regard to pitch processing have emerged, mainly with lesion, dichotic listening and functional neuroimaging techniques (Gandour et al., 1992; Klein, Zatorre, Milner, & Zhao, 2001; Van Lancker & Fromkin, 1973; Wang, Sereno, Jongman, & Hirsch, 2003). However, due to the low temporal resolution of these techniques, event-related potentials (ERPs), a high temporal resolution measure was introduced to pitch processing, offering more precise temporal information of online processing.

The majority of ERP studies relevant to pitch processing focus on the neural mechanisms of tone processing at the pre-attentive stage, where participants are directed to watch a silent movie or read a book and to ignore the auditory input (Fritz, Elhilali, David, & Shamma, 2007). In these studies, the ERP component of interest is the Mismatch-Negativity (MMN), an indicator of acoustic change detection (Näätänen, 2001; Pulvermüller & Shtyrov, 2006). Only two studies

examined the online processing of both tone and intonation in Standard Chinese, to our knowledge. Ren, Yang, and Li (2009) constructed an oddball sequence. A word with lexical Tone4 (i.e., /gai4/³), uttered with statement intonation was presented as the standard stimulus, and /gai4/ with question intonation was presented as the deviant stimulus to native Standard Chinese listeners. Their results showed a clear MMN effect when subtracting the waveform of the standard from that of the deviant. In another study, Ren, Yang, Li, and Sui (2013) adopted a three-stimuli oddball paradigm. The standard stimulus was /lai2/ with statement intonation. The deviant stimuli included an intonation deviant (/lai2/ with question intonation) and a lexical tone deviant (/lai4/ with statement intonation). Results showed an MMN for the tone deviant but not for the intonation deviant. As the MMN is linked to higher order perceptual processes underlying stimulus discrimination (Pulvermüller & Shtyrov, 2006), the above two studies suggest that at the pre-attentive stage, native listeners can tease apart question intonation from statement intonation when the intonation is combined with Tone4, but they are not able to tease apart the two types of intonation when the intonation is combined with Tone2, just as what Yuan (2011) has reported with behavioral perceptual judgment data. This correspondence of the online MMN results with the offline behavioral results validates the initial ERP evidence of the interaction of tone and intonation in Standard Chinese.

In addition to Standard Chinese, ERP evidence of online interplay of tone and intonation is also revealed in Cantonese (Kung et al., 2014). In this study, Cantonese participants were asked to perform a lexical-identification task, i.e., choosing the right word they heard from six Cantonese words on the screen in the form of Chinese characters, and the six words were tonal sextuplets of the critical word. ERP analyses revealed a P600 effect for low tone in questions relative to low tone in statements. The P600 effect was explained as an indicator of reanalysis, in the presence of a strong conflict of two competing

³ The number following the letters in Standard Chinese Pinyin represents Standard Chinese tone. “1” is Tone1 (high-level tone); “2” is Tone2 (mid-rising tone); “3” is Tone3 (low-dipping tone), and “4” is Tone4 (high-falling tone).

representations activated in questions ending with low tones. The two representations are a lexical representation with a low tone on the one hand and a lexical representation with a high rising tone on the other. Special attention should be paid to the fact that Kung et al. (2014) found a P600 effect in the semantically neutral sentence context. In their subsequent study, when introducing a highly constraining semantic context to the target words, the P600 disappeared, suggesting that semantic context plays a role in resolving the online conflict between intonation and tone.

Several remaining issues may be noticed given the above ERP studies on the processing of tone and intonation. First, the MMN studies of Standard Chinese restricted their attention to the interaction of tone and intonation in a one-syllable-sentence domain. Given that intonation is a feature at the sentential level that typically spans over several lexical items, it would be not only interesting but also necessary to investigate how tone and intonation information are processed when the length of an utterance is extended from one syllable to several syllables. Specifically, the question that arises here is whether native Standard Chinese listeners can disentangle intonation information from tone information over a broader sentence domain. Second, the extant ERP studies (Ren et al., 2009, 2013) investigated tone and intonation processing at the pre-attentive stage where the attention of the participants was directed to elsewhere. In this kind of design, there is no way to measure the behavioral effects of the deviant acoustic stimulus in the stream, making it impossible to distinguish between automatic neural responses arising from acoustic variability and responses related to “attention capture” (Fritz et al., 2007). Therefore, the present study aimed to extend the online processing of tone and intonation from the pre-attentive stage to the attentive processing stage. Moreover, we are interested in whether the processing of tone and intonation at the attentive stage differs from that at the pre-attentive stage. Third, the ERP study on Cantonese (Kung et al., 2014) extended the online processing of tone and intonation to a broader sentence domain. However, this study employed a lexical identification task rather than a pitch identification task *per se*. In this way, a potential concern is that the interaction of tone and intonation is not directly examined. Also, as stated earlier, intonation distorts

tone identity in Cantonese but not in Standard Chinese. There arises the question of whether the mechanisms underlying tone and intonation processing in Standard Chinese are different from that in Cantonese. Fourth, semantic context affects the processing of tone and intonation in Cantonese. It has also been proven that in Standard Chinese a constraining semantic context facilitates the processing of tone (Ye & Connine, 1999) and intonation (Liu, Chen, & Schiller, 2016a). In this study, we therefore took semantic context as a control variable and set it to be neutral so that it can serve as a baseline comparison for further research. In short, the present study was designed to investigate the online processing of tone and intonation in Standard Chinese over a broader sentence domain at the attentive stage under neutral semantic context.

The ERP component that is of our particular interest in the present study is the P300 (the P3b in particular). The P300 is a positive-going deflection peaking at around 300 ms in a time window of about 250 to 500 ms, or even to 900 ms (Patel & Azzam, 2005). It is thought to be elicited in the process of decision making (Hillyard, Hink, Schwent, & Picton, 1973; Nieuwenhuis, Aston-Jones, & Cohen, 2005; Rohrbaugh, Donchin, & Eriksen, 1974; Smith, Donchin, Cohen, & Starr, 1970; Verleger, Jaśkowski, & Wascher, 2005), reflective of processes involved in stimulus evaluation or categorization (Azizian, Freitas, Watson, & Squires, 2006; Frenck-Mestre et al., 2005; Johnson & Donchin, 1980; Kutas, McCarthy, & Donchin, 1977).

In the present study, we examined participants' behavioral and electrophysiological responses to Standard Chinese sentences contrasting in final tones (Tone2 or Tone4) and intonations (Question or Statement). We employed two pitch identification tasks. Participants were asked to categorize the final tone or intonation of the stimuli. A two-alternative forced choice (2AFC) task was adopted in this study. 2AFC is considered a highly simplified decision making condition, in which a choice must be made between two responses based on limited information about which is correct (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006). It best captures the essence of decision-making, and therefore, fits for the purpose of eliciting P300. To decrease fast guesses in the task, a 0.5/0.5 probability of each category was implemented, as in Pfefferbaum, Ford, Johnson, Wenegrat, and Kopell (1983). The same 2AFC

task was performed on all the stimuli, not just on one specific category to avoid selective tuning, which has been proved to be unnecessary and insufficient for P300 enhancement (Hillyard et al., 1973; Rohrbaugh et al., 1974).

We hypothesized that under neutral semantic context, at the attentive stage, native Standard Chinese listeners should be able to disentangle question intonation from statement intonation when the intonation concurs with a final Tone4. Behaviorally, this should be reflected in high identification accuracy. Electrophysiologically, we expect a P300 effect for questions ending with Tone4 relative to statements ending with Tone4. In the case of Tone2, due to the difficulty in teasing apart intonation information from tone information for participants, the behavioral performance is expected to show a lower accuracy. No clear P300 is expected between questions ending with Tone2 and statements ending with Tone2.

2.2 Method

2.2.1 Participants

Twenty right-handed native speakers of Standard Chinese from Northern China were paid to participate in the experiment. They were undergraduate or graduate students at Renmin University. Five of the participants were excluded from the analysis because of excessive artifacts in their EEG data. Age of the remaining 15 participants (7 male, 8 female) ranged from 20 to 28 ($M \pm SD$: 23.8 ± 2.8). None of them had received any formal musical training or had reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

2.2.2 Materials

Forty monosyllabic minimal word pairs varying by tone (Tone2 or Tone4) with otherwise identical segments were selected. Each minimal Tone2_Tone4 word pair contains words of comparable word frequency, homophone density, and syntactic word category. To avoid any word frequency effect, only frequent words with more than 4,500 occurrences in a corpus of 193 million words were used (Da, 2004). Following Chen, Vaid, and Wu (2009), homophone density

was defined as the number of homophone mates of a word, i.e., words that contain exactly the same phonetic segments and lexical tones. Tone2 words have similar homophone densities as their Tone4 equivalents. The forty word pairs comprise mainly pairs of nouns (32 pairs), but pairs of verbs (6 pairs) and adjectives (2 pairs) were also included to guarantee sufficient number of stimuli.

All the critical words were embedded in the final position of a five-syllable carrier sentence, i.e., *ta1 gang1gang1 shuo1 X* (English: She just said X), produced with either a statement or a question intonation. Only high-level tones (Tone1) were contained in the carrier sentence. This is to avoid downstep effect and to minimize the contribution of tone to the observed F0 movement (Shih, 2000). The carrier sentence was semantically meaningful but offered neutral semantic context to the target stimuli. By using the semantically neutral carrier, intonation information was successfully elicited. On the other hand, potential confound of semantic context with sentence prosody was excluded (Kung et al., 2014).

In total, 160 target sentences (40 Syllables \times 2 Tones \times 2 Intonations) were designed (see Table 1 for an example). Moreover, another 240 filler sentences were constructed. They resembled the target sentences in carrier sentence structure but differed from them in critical syllables in terms of either segmental composition or lexical tone (e.g., Tone1/Tone3). Pooling target sentences and filler sentences resulted in 400 sentences, which were uttered for the perception experiment.

Table 1. *An example of the experimental design.*

Condition		Example				
Tone	Intonation					
Tone2	Statement	Characters	她	刚刚	说	X (财)。
		Pinyin	ta1	gang1gang1	shuo1	cai2
		IPA	[^h A1]	[kaŋ1 kaŋ1]	[ʃuo1]	[ts^hai2]
		English	She	just	said	money.
Tone2	Question	Characters	她	刚刚	说	X (财)?
		Pinyin	ta1	gang1gang1	shuo1	cai2
		IPA	[^h A1]	[kaŋ1 kaŋ1]	[ʃuo1]	[ts^hai2]
		English	She	just	said	money?
Tone4	Statement	Characters	她	刚刚	说	X (菜)。
		Pinyin	ta1	gang1gang1	shuo1	cai4
		IPA	[^h A1]	[kaŋ1 kaŋ1]	[ʃuo1]	[ts^hai4]
		English	She	just	said	vegetable.
Tone4	Question	Characters	她	刚刚	说	X (菜)?
		Pinyin	ta1	gang1gang1	shuo1	cai4
		IPA	[^h A1]	[kaŋ1 kaŋ1]	[ʃuo1]	[ts^hai4]
		English	She	just	said	vegetable?

Note. The critical syllables are in bold.

2.2.3 Recording and stimuli preparation

One female native speaker of Standard Chinese, who was born and raised in Beijing, recorded the sentences. The recordings took place in a soundproof recording booth at the Phonetics Lab of Leiden University. Sentences were randomly presented to the speaker and recorded at 16-bit resolution and a sampling rate of 44.1 kHz. To eliminate paralinguistic information, the speaker was instructed to avoid any exaggerated emotional prosody during the recording.

This female speaker's recordings were chosen for the experiment for the clarity and consistency of the articulation. More importantly, the acoustic results of her recordings (see Figures 1 and 2) showed comparable F0 realization of tone and intonation to a prior study (Yuan, 2006) and were

therefore taken as the prototypical patterns for the perception study. In the subsequent perception experiment, the amplitude of all the sentences was normalized in Praat (Boersma & Weenink, 2015).

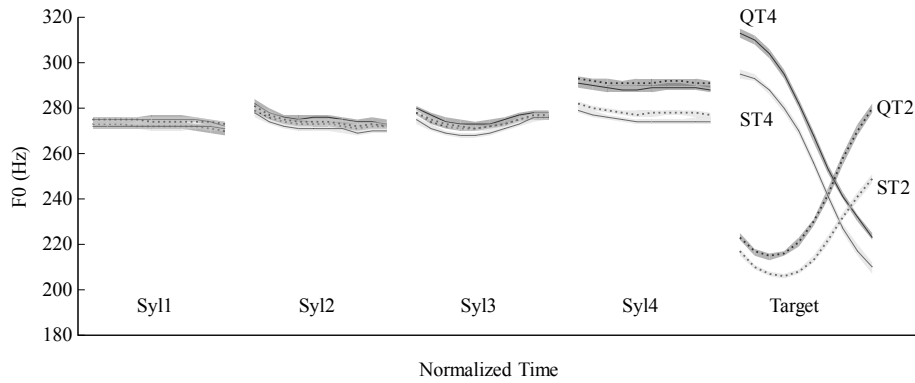


Figure 1. *F0 contours of the four experimental conditions. Each experimental condition is a combination of the levels of the factors Tone (Tone2, Tone4) and Intonation (Question, Statement), for example, QT2 refers to questions ending with Tone2. Syl1 to Syl4 are the carrier syllables, whereas Syl5 is the critical syllable. Dark solid lines indicate the mean F0 contours of QT4 (dark grey areas for ± 1 SD of mean), and light solid lines indicate the mean F0 contours of ST4 (light grey areas for ± 1 SD of mean). The corresponding dark dotted lines and light dotted lines indicate the mean F0 contours of QT2 and ST2, respectively.*

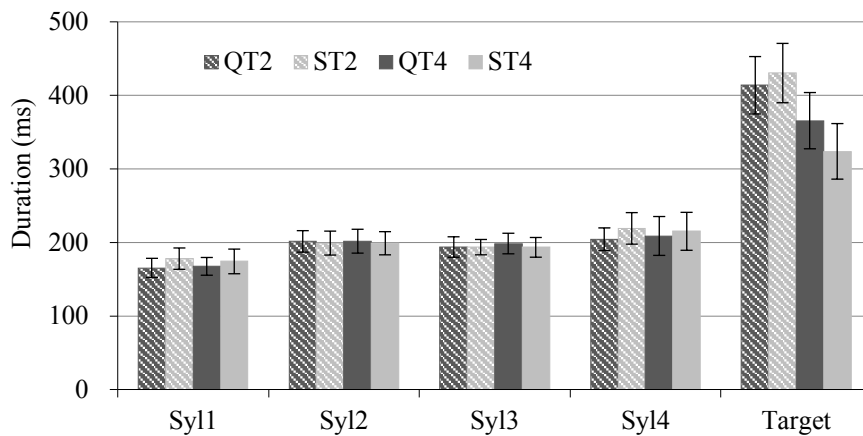


Figure 2. *Duration means ± 1 SD for each syllable of the four experimental conditions.*

The acoustic results (see Table 2) of the speech files showed that over our target stimuli pairs (which are five syllables long), the F0 and duration of the first three syllables revealed no differences between the two types of intonation for sentences ending with both Tone2 and Tone4 (all p s > .05). However, intonation affected the fourth syllable with a significantly raised F0 for the level tone over a question in comparison with that over a statement in both the Tone2 ($t(39) = 7.22, p < .001$, Cohen's $d = 1.64$) and the Tone4 ($t(39) = 7.17, p < .001$, Cohen's $d = 1.64$) conditions. When it comes to the critical syllable (i.e., the fifth syllable of the utterance), question with a final Tone2 had a significantly wider F0 range than its statement counterpart ($t(39) = 9.87, p < .001$, Cohen's $d = 1.90$). The minimum F0 showed a slight increase ($t(39) = 6.37, p < .001$, Cohen's $d = 1.27$), while a sharp rise of the maximum F0 was observed in the question condition relative to the statement condition ($t(39) = 11.47, p < .001$, Cohen's $d = 2.30$), suggesting a trend of sharper final rising. For questions with a final Tone4, there was an overall higher F0 contour than statements with a final Tone4. Though the F0 range was comparable between the two conditions ($t(39) = 1.24, p = .22$, Cohen's $d = 0.30$), the initial F0 ($t(39) = 8.65, p < .001$, Cohen's $d = 1.40$) and the final F0 ($t(39) = 4.48, p < .001$, Cohen's $d = 0.80$) of the high-falling tone were significantly higher in the question intonation (both p s < .001). Consistent with previous findings (Cao, 2004; Wu, 1996), our data showed that the pitch contour of Tone2 and Tone4 is maintained in both question and statement intonations, but pitch height differs between the two intonations across final tone identities. In addition to F0, duration also played a role in making a distinction between question intonation and statement intonation. Consistent with Yuan (2006), final Tone4 syllables in question intonation had significantly longer duration than those in statement intonation ($t(39) = 6.73, p < .001$, Cohen's $d = 1.11$). However, final Tone2 syllables tended to be slightly shorter in question intonation than in statement intonation ($t(39) = -3.10, p < .01$, Cohen's $d = -0.41$), whereas similar duration was found between the two intonation types for final Tone2 syllables in Yuan (2006). Considering the temporal scope of the interaction of tone and intonation, our data lend no support to the global or to the strictly local theory of question intonation, but are sympathetic towards the final-rising

theory given that F0 did not start to increase significantly until the pre-final syllable.

Table 2. *Acoustic properties of the experimental materials (SDs in parentheses).*

Parameter	Syllable	Tone2, Statement	Tone2, Question	Tone4, Statement	Tone4, Question
Duration (ms)	Syl1	178 (15)	165 (13)	174 (17)	167 (12)
	Syl2	199 (16)	201 (15)	199 (16)	202 (16)
	Syl3	194 (10)	194 (14)	193 (13)	199 (14)
	Syl4	219 (21)	204 (15)	215 (26)	209 (26)
	Syl5	430 (40)	414 (39)	324 (38)	366 (38)
Mean F0 (Hz)	Syl1	274 (11)	272 (10)	272 (9)	275 (8)
	Syl2	274 (9)	274 (8)	272 (7)	276 (8)
	Syl3	274 (8)	274 (7)	272 (8)	276 (8)
	Syl4	278 (7)	292 (9)	275 (8)	289 (9)
Max F0 (Hz)	Syl5	250 (12)	280 (14)	296 (10)	313 (13)
Min F0 (Hz)	Syl5	205 (6)	214 (8)	210 (16)	223 (11)
F0 range (Hz)	Syl5	45 (10)	66 (12)	86 (20)	90 (11)

2.2.4 Task

Participants were asked to perform two pitch identification tasks. The tone identification task consisted of half of the trials (1 sentence in a trial, 400 sentences in total), whereas the intonation identification task consisted of the other half. The specific task was randomly allocated from trial to trial. Task types were indicated by tone and intonation marks in Standard Chinese Pinyin system. When “ˊ ˋ” marks (“ˊ” stands for Tone2; “ˋ” stands for Tone4) appeared on the screen, participants were asked to identify the final tone of the sentence they heard. When “。 ?” marks appeared (“。” stands for statement intonation; “?” stands for question intonation), they were asked to identify whether the previously presented sentence bears a statement intonation or a question intonation. To complete the tasks, participants were requested to press the corresponding button within a two-second time limit. The tone and intonation marks used here are acquired at a very early age by native Standard

Chinese speakers. No participants had reported difficulty in understanding the tasks.

2.2.5 Procedure

Participants were tested individually in a soundproof booth. Stimuli were randomly presented using E-Prime 2.0 software through loudspeakers at a comfortable listening level of a 75 dB sound pressure level at source. Instructions were given to participants visually on screen and orally by the experimenter in Standard Chinese before the experiment.

The whole experiment included one practice block and four experiment blocks. The practice block contained 12 trials. Each experiment block encompassed 100 trials. Between each block, there was a 3-minute break. An experiment trial started with a 100 ms warning beep, followed by a 300 ms pause. After that an auditory sentence was presented while a red fixation cross appeared on the screen. Participants were instructed to gaze on the fixation cross and not to blink or move during the presentation of the sentence. In the meantime, participants were instructed to pay special attention to the final tone and the intonation of the sentence. One second after the offset of the stimuli, they were asked to perform either a tone identification task or an intonation identification task as accurately as possible within a two-second time limit. By doing so, the ERP effects of interest can be prevented from being confounded by motor-related processes (Kung et al., 2014; Salisbury, Rutherford, Shenton, & McCarley, 2001). The Inter Stimulus Interval (ISI) was 500 ms.

2.2.6 EEG data recording

EEG was recorded from 64 Ag/AgCl electrodes mounted in an elastic cap (Neuroscan system) with a sampling rate of 500 Hz. The right mastoid served as the reference electrode and AFz as the ground. Electrooculograms (EOGs) were monitored vertically and horizontally. Vertical EOGs were recorded by electrodes placed above and below the left eye, while horizontal EOGs were recorded by electrodes placed at the outer canthi of the left and right eye. The impedance of all electrodes was kept below 5K Ω .

2.2.7 Behavioral data analysis

Given that the behavioral responses were performed one second after the presentation of the stimuli, these delayed reaction time measurements were not further analyzed. Only Identification Rate (IR) was analyzed. IR was defined as the percentage of correct identification of tone in the tone identification task, and as the percentage of correct identification of intonation in the intonation identification task.

Statistical analyses were carried out with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.1.2 (R Core Team, 2015). Binomial logistic regression models were constructed for the dependent variable Response (Correct or Incorrect) with Task, Tone, Intonation and their interactions as fixed factors, and Subjects and Items as random factors. The fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. To capture the binary nature of the dependent variable, a logistic link function was applied. The estimate (β), ξ and p -values are reported.

2.2.8 EEG data analysis

The EEG data were analyzed with Brain Version Analyzer (Version 2.0). A 0.05-20 Hz band-pass filter was applied offline to the original EEG data. ERP epochs were defined in a 1,200 ms interval from -400 ms to 800 ms time-locked to the onset of the critical word. The baseline was calculated from -400 ms to -200 ms. In our acoustic data, F0 differences among the experimental conditions have been observed from the pre-final syllable. We therefore defined the time interval before the pre-final syllable as the baseline. Epochs with excessive eye movements and blinks were discarded. The criteria for artifact rejection were a maximal sudden voltage change of $25 \mu\text{V}$ in 100 ms, a maximal amplitude difference of $100 \mu\text{V}$ in a time window of 200 ms and a low amplitude activity within a range of $0.5 \mu\text{V}$ in a time window of 100 ms.

Prior to averaging, trials with incorrect behavioral responses were excluded. However, ERP data were collapsed across task types during averaging because on the one hand, the tone identification task evoked very few response errors,

and on the other hand, we did not observe differences in the ERP waveforms between the tone identification task and the intonation identification task under each experimental condition. To gain more statistical power, we aggregated all the correctly identified artifact-free trials from the tone identification task and the intonation identification task in the ERP analyses. As a result, a total of 30% of the data points were rejected. We found a clear peak in only one of the experimental conditions. Thus, in the following, we will exclusively report mean amplitudes.

A set of 27 electrodes was used for analyses, including 3 midline electrodes (Fz, Cz and Pz) and 24 lateral electrodes (F3/4, F1/2, FC3/4, FC1/2, C3/4, C1/2, CP3/4, CP1/2, P3/4, P1/2, PO5/6, PO3/4). The lateral electrodes were divided into six areas comprising four electrodes each (see Figure 3). These six areas were Left Frontal (F3, F1, FC3, FC1), Right Frontal (F2, F4, FC2, FC4), Left Central (C3, C1, CP3, CP1), Right Central (C2, C4, CP2, CP4), Left Posterior (P3, P1, PO5, PO3) and Right Posterior (P2, P4, PO4, PO6). For each area, the mean amplitude of the four electrodes was calculated and used in the following analyses. Due to the different numbers of the midline electrodes and the lateral electrodes, we decided to run repeated measures ANOVAs on the midline electrodes and the lateral electrodes separately.

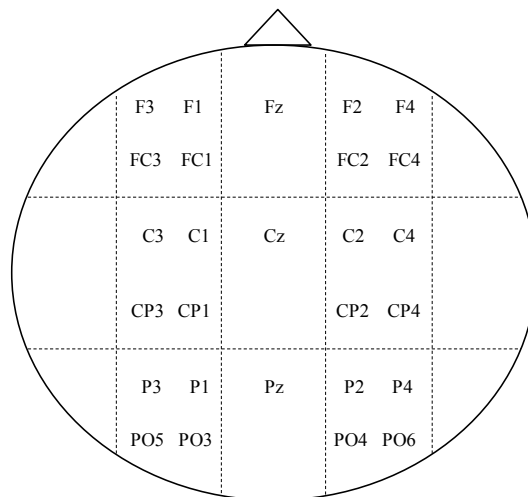


Figure 3. *Electrode areas used in the analyses. For the lateral electrodes, the amplitude of the four electrodes within each area was averaged.*

To establish the exact onset and range of the ERP effects, we ran repeated measures ANOVAs for 16 successive 50 ms time windows starting from the onset of the critical word up to 800 ms (following Schirmer, Tang, Penney, Gunter, & Chen, 2005). For the midline electrodes, within-subject variables included Tone (Tone2, Tone4), Intonation (Question, Statement) and Region (Frontal, Central, Posterior). For the lateral electrodes, within-subject variables included Tone (Tone2, Tone4), Intonation (Question, Statement), Region (Frontal, Central, Posterior) and Hemisphere (Left, Right). Statistical significance was computed using the Greenhouse-Geisser correction when the assumption of sphericity was violated. Corrected p -values are reported.

2.3 Results

2.3.1 Behavioral results

Figure 4 presents the identification rate of the four experimental conditions under different tasks (see also Table A1 in Appendix A for details). Tone stands for the tone identification task, and Intonation stands for the intonation identification task.

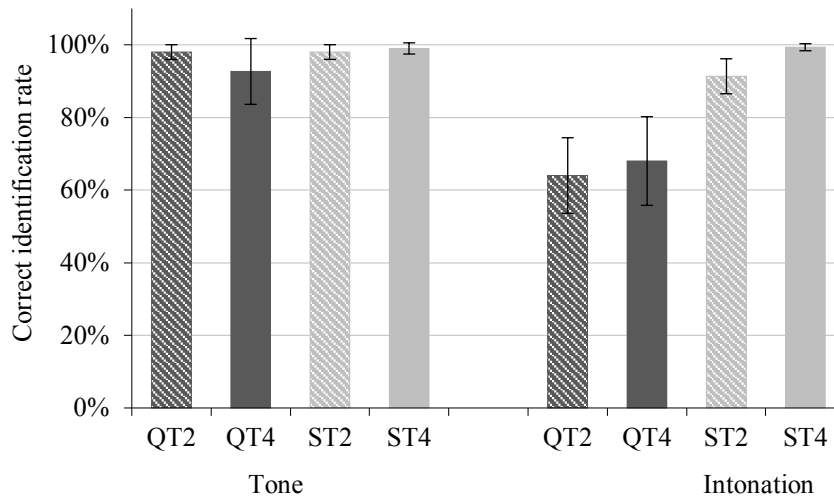


Figure 4. Identification rate for each experimental condition under different tasks. Tone indicates the tone identification task; Intonation indicates the intonation identification task.

Results showed a significant main effect of Task ($\beta = 0.69$, $\chi = 2.61$, $p < .01$) and Intonation ($\beta = 1.99$, $\chi = 5.45$, $p < .01$) on the odds of correct responses over incorrect responses. Two-way interactions, i.e., Task \times Tone, Task \times Intonation, and Tone \times Intonation, also reached significance (all $ps < .01$). There was no three-way interaction of Task, Tone and Intonation ($p > .1$). Separate models for subset data of Tone and Intonation revealed that the effects of Task were manifested in that the tone identification task had much higher identification rate than the intonation identification task in questions ending with Tone2 and Tone4, and also in statements ending with Tone2 (all $ps < .05$). Due to the near-ceiling level of identification performances in both tasks, no task difference was observed for statements ending with Tone4 ($p > .05$). Apparently, the tone identification task was much easier than the intonation identification task for the participants.

Separate models were also constructed for subset data of different tasks. For the tone identification task, a significant interaction of Tone \times Intonation ($\beta = 2.17$, $\chi = 2.50$, $p < .05$) was found. Identification rate of Tone4 was lower than that of Tone2 in questions ($\beta = -1.54$, $\chi = -2.95$, $p < .01$), but no difference was found between the two in statements ($\beta = 0.71$, $\chi = 1.00$, $p > .05$). No intonation effect was found in either Tone2 or Tone4 sentence pairs (both $ps > .05$). Overall, tone identification almost reached ceiling level across conditions. This suggests that the identity of tone was not hindered by intonation information.

With respect to the intonation identification task, a significant main effect of Intonation ($\beta = 2.06$, $\chi = 5.14$, $p < .01$) and an interaction of Tone \times Intonation were discovered ($\beta = 2.61$, $\chi = 2.95$, $p < .01$). Question intonation had a much lower identification rate than statement intonation regardless of the final tone identities (both $ps < .01$). Despite the fact that no difference was found for IR of question intonation between questions ending with Tone2 and questions ending with Tone4 ($\beta = 0.20$, $\chi = 0.51$, $p > .05$), a higher identification rate of statement intonation was indeed discovered for statements ending with Tone4 relative to statements ending with Tone2 ($\beta = 2.68$, $\chi = 3.61$, $p < .01$). This suggests that participants had more difficulties perceiving question intonation than statement intonation.

2.3.2 ERP results

Figure 5 shows the grand average ERP waveforms time-locked to the onset of the critical syllables for 9 electrodes. Figure 6 presents the topographic maps obtained in all the 64 electrodes. Since the focus of interest of this paper is tone and intonation effects, only tone-related and intonation-related effects and the corresponding interactions are discussed below.

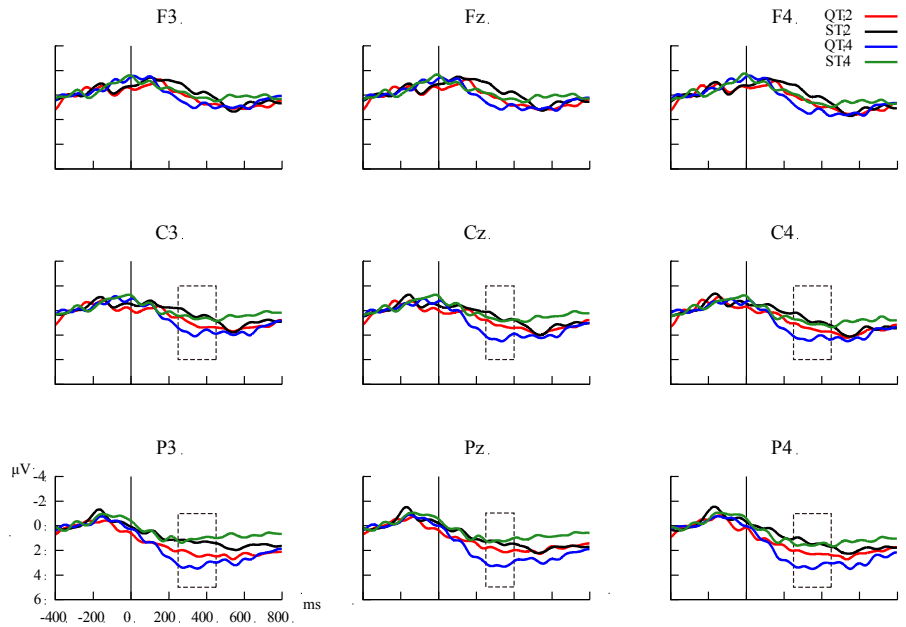


Figure 5. Grand average waveforms time-locked to the onset of the critical syllables with a baseline from -400 ms to -200 ms for nine representative electrodes. Negativity is plotted upwards. The boxes with dash lines mark the P300 time-window for the questions ending with Tone4 condition.

A summary of the time-course analyses for the midline electrodes and the lateral electrodes are presented in Table A2 and Table A3 (see Appendix A), respectively. Regions of Interest (ROIs) were identified as the time period when effects were consistently significant in two or more consecutive 50 ms time windows. Visual inspection of the waveforms also served as a complementary tool for the identification of ROIs. Consequently, we chose a ROI of 250-400

ms for the midline electrodes. For the lateral electrodes, a larger time window of 250-450 ms was identified as the ROI.

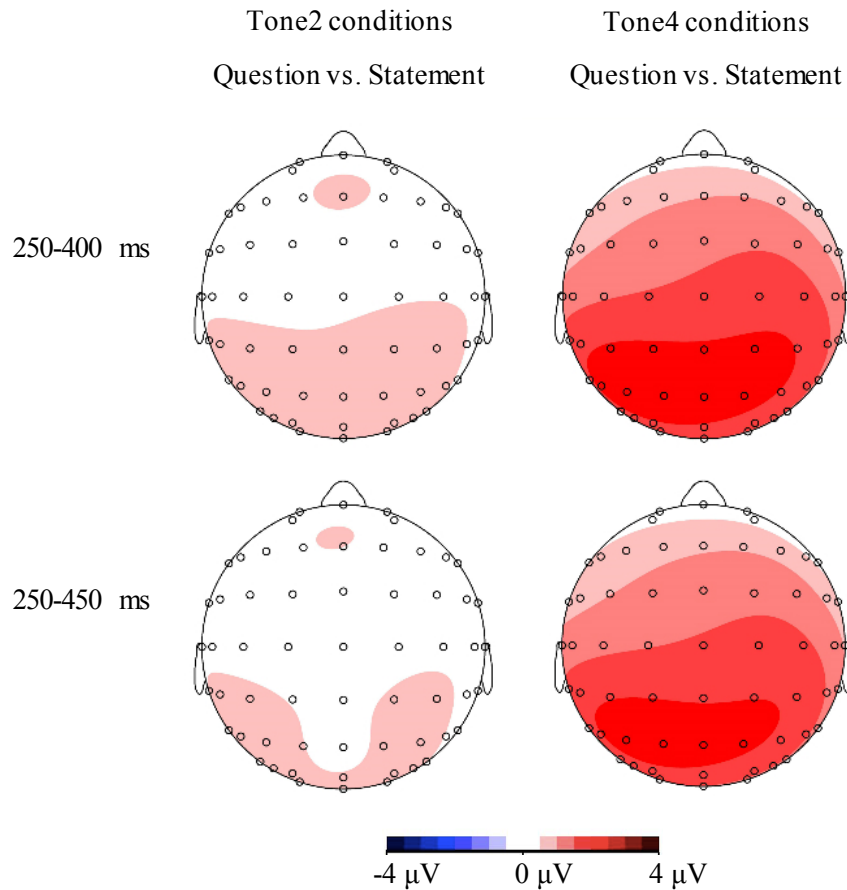


Figure 6. Topographic maps obtained from all 64 electrodes. The maps were calculated by subtracting the waveforms in statements from those in questions for the Tone2 conditions (the left column) and the Tone4 conditions (the right column), respectively. The upper row shows the topographic maps in a time window of 250-400 ms, where the midline electrodes show P300 effect. The bottom row shows the topographic maps in a larger time window of 250-450 ms, where the lateral electrodes show P300 effect.

The overall ANOVA for the mean amplitude of the midline electrodes in the time window of 250-400 ms revealed a main effect of Intonation ($F(1, 14)$

= 8.89, $p < .05$, $\eta_p^2 = .39$), and a three-way interaction of Tone \times Intonation \times Region ($F(1.56, 21.79) = 4.55$, $p < .05$, $\eta_p^2 = .25$). Follow-up ANOVAs were then performed for each level of Tone. Comparisons between QT2 and ST2 revealed neither a main effect of Intonation nor an interaction of Intonation \times Region (both p s $> .05$). However, the analysis comparing QT4 and ST4 yielded a significant main effect of Intonation ($F(1, 14) = 6.81$, $p < .05$, $\eta_p^2 = .33$) and a significant interaction of Intonation \times Region ($F(1.85, 25.83) = 9.05$, $p < .01$, $\eta_p^2 = .39$). Separate ANOVAs for each level of Region revealed a significant main effect of Intonation at the central ($F(1, 14) = 6.55$, $p < .05$, $\eta_p^2 = .32$) and posterior sites ($F(1, 14) = 13.49$, $p < .01$, $\eta_p^2 = .49$), with a larger positivity for QT4 than for ST4. No effect of Intonation was found at the frontal sites ($p > .05$).

As for the lateral electrodes, the overall ANOVA for the mean amplitude in the time window of 250-450 ms revealed a main effect of Intonation ($F(1, 14) = 10.18$, $p < .01$, $\eta_p^2 = .42$), a three-way interaction of Tone \times Intonation \times Region ($F(1.44, 20.08) = 4.87$, $p < .05$, $\eta_p^2 = .26$), and also a three-way interaction of Intonation \times Region \times Hemisphere ($F(1.56, 21.88) = 3.84$, $p < .05$, $\eta_p^2 = .22$). Follow-up ANOVAs for each level of Tone yielded no effects between QT2 and ST2 (all p s $> .05$), but a significant main effect of Intonation ($F(1, 14) = 4.78$, $p < .05$, $\eta_p^2 = .25$), a significant two-way interaction of Intonation \times Region ($F(1.95, 27.27) = 12.08$, $p < .01$, $\eta_p^2 = .46$) and a significant three-way interaction of Intonation \times Region \times Hemisphere ($F(1.60, 22.35) = 8.09$, $p < .05$, $\eta_p^2 = .37$) between QT4 and ST4. Subsequent separate ANOVAs for each level of Region between QT4 and ST4 showed a significant main effect of Intonation at the central ($F(1, 14) = 4.66$, $p < .05$, $\eta_p^2 = .25$) and posterior sites ($F(1, 14) = 12.31$, $p < .01$, $\eta_p^2 = .47$), with QT4 eliciting more positivity than ST4 at these regions. Despite the statistical insignificance ($p > .05$), it is worth emphasizing that the amplitude difference between QT4 and ST4 was more prominent at the posterior sites than at the central sites. At the frontal sites, however, no effect of Intonation was found ($p > .05$).

In sum, the results of ANOVAs did not reveal any effect for QT2 versus ST2. In contrast, an ERP effect was observed for QT4 versus ST4. The ERP effect took place in different time windows for the midline electrodes and the

lateral electrodes, with a central-posterior distribution from 250-400 ms for the midline electrodes and a central-posterior distribution from 250-450 ms for the lateral electrodes. Through visual inspection of the waveforms, we identified a positive-going waveform peaking at about 300 ms after the onset of the critical word in the QT4 condition versus the ST4 condition in both the midline electrodes and the lateral electrodes. Taking together the polarity and the topographical distribution of the effect, we conclude that a P300 effect was found for QT4 versus ST4, whereas no effect was present for QT2 versus ST2.

2.4 General discussion

The present study investigated the online processing of tone and intonation in Standard Chinese at the attentive stage. We examined the behavioral and electrophysiological responses of native Standard Chinese listeners to Standard Chinese sentences, which contrast in final tones (Tone2 or Tone4) and intonations (Question or Statement). The context of these sentences was manipulated to be semantically neutral. Our behavioral results showed that while the identification of tone was not hindered by intonation, the identification of intonation was greatly impeded by tone. In the Tone4 conditions, question intonation was rather difficult to be correctly identified, whereas identification of statement intonation almost showed no difficulty at all. In the Tone2 conditions, question intonation was still difficult to identify, while identification of statement intonation also tended to be problematic. Regarding ERP results, we found a clear P300 effect for questions ending with Tone4 relative to statements ending with Tone4. No ERP difference was found between questions ending with Tone2 and statements ending with Tone2.

According to previous studies, P300 reflects neurophysiological mechanisms of decision-making and categorical processing (Azizian et al., 2006; Courchesne E., Hillyard, & Courchesne R., 1977; Kotchoubey & Lang, 2001; Kutas et al., 1977). When the categorization becomes more difficult, P300 amplitudes become smaller (Polich, 2007; Verleger, Gasser, & Möcks, 1985). The P300 requires attention. Previous studies also suggested that P300 amplitude is larger when participants devoted more effort to a task (Johnson, 1984, 1986; Isreal, Chesney, Wickens, & Donchin, 1980). Intuitively, one may expect that an

increase in task difficulty leads to investment of more effort and should thus elicit large P3 amplitude (Kok, 2001). However, the P300 amplitude decreases when tasks become perceptually or cognitively more demanding (Luck, 2005). Therefore, our ERP results above suggest that at the attentive processing stage, the question-statement contrast in Tone4 conditions is easier to categorize, whereas categorization of the question-statement contrast in the Tone2 conditions is much more demanding for native Standard Chinese listeners. These results are highly consistent with the MMN studies examining the online processing of tone and intonation in Standard Chinese at the pre-attentive stage. In those two studies (Ren et al., 2009, 2013), listeners are able to perceive the difference between question and statement intonation when the final tone is Tone4 (reflected in an MMN effect), but they cannot make a distinction between question and statement intonation when the final tone is Tone2 (reflected in no MMN). The MMN studies used one-syllable sentences, while our study extended the length of the utterances from one syllable to five syllables. Results in our study seem to confirm that the online processing patterns of tone and intonation in Standard Chinese are maintained from the pre-attentive stage to the attentive stage over a longer utterance.

Though consensus has been reached that question intonation can be distinguished online from statement intonation when the sentences end with Tone4, one may question why the P300 effect is elicited in questions ending with Tone4 compared to statements ending with Tone4, rather than vice versa. In the above MMN studies (Ren et al., 2009, 2013) and many other P300 studies, a target and a non-target are preset in the design. Very often participants just respond to the target stimuli. The corresponding ERP effect is therefore detected in the target category relative to the non-target category. In our study, no target category is set in advance, and neither is it the case that participants responded only to one category. With equal probability of the two intonation types for Tone4, we also did not expect any bias due to selective tuning (Hillyard et al., 1973; Rohrbaugh et al., 1974). All in all, the two conditions hold equal possibility of eliciting P300 effect in principle. So what makes question ending with Tone4 outperform its statement counterpart? Evidence from the resource framework has shown that P300 amplitude is

sensitive to the allocation of resources. P300 amplitude is larger when participants devoted more effort to a task (Johnson, 1984, 1986; Isreal et al., 1980). The advantage of questions endings with Tone4 over statements ending with Tone4 in the present study, therefore, seems to suggest that participants devote more processing effort to the former compared to the latter. As mentioned earlier, P300 is typically elicited in the target condition relative to the non-target condition. Azizian et al. (2006) proposed that equally probable stimuli that were easily evaluated as non-target required less mental work for discrimination and produced no P300-like components. From this line of reasoning, with a final Tone4, question intonation seems to be evaluated as possessing target-like properties, whereas statement intonation is evaluated as possessing non-target-like properties. Interestingly, this assumption happens to coincide with the view that statement intonation is a default intonation type and question intonation is a marked intonation type (Peters & Pfitzinger, 2008; Yuan, 2011). It appears that as a default intonation type, statement intonation occupies less mental attention or effort to be identified, leading to attenuated or even diminished P300 amplitude. In contrast, question intonation requires extra mental attention in order to be identified, resulting in increased P300 amplitude.

Interestingly, a comparison of our ERP results with the behavioral results revealed a discrepancy. In sentences ending with Tone4, question and statement contrast can be perceived electrophysiologically (reflected in P300). Behaviorally, listeners still had difficulty identifying question intonation (reflected in an identification rate of 68%, which was only marginally significantly ($p < .05$) above chance level, i.e., 50%) from statement intonation. This discrepancy could possibly be ascribed to the different processing that the neural responses and the behavioral responses reflect. In our study, the recorded EEG was time-locked to the onset of the critical syllable, whereas the behavioral responses were collected one second after the presentation of the stimuli. With such setup, one would not expect an isomorphic mapping between the neural responses and the behavioral responses, as P300 reflects cognitive processing restricted to a limited set of scalp electrodes within a limited temporal window, whereas the behavioral responses reflect whole brain processing.

Despite the discrepancy between the ERP results and the behavioral results for Tone4 conditions, the behavioral results in the present study are not uninterpretable. Our results showed that tone identity was seldom affected by intonation, while intonation identification was greatly affected by tone. Specifically, question intonation had a much lower identification rate than statement intonation regardless of final tone identities. For statement intonation identification, statements ending with Tone4 showed a significantly higher identification rate (99.3%) than statements ending with Tone2 (91.3%). These results are in line with previous studies (Xu & Mok, 2012a; Yuan, 2011). However, our results concerning question intonation identification were different from what was reported in Yuan (2011). Yuan (2011) found that questions ending with Tone4 were easier to identify than questions ending with Tone2. In our study, no statistically significant difference was found between questions ending with Tone4 (68%) and questions ending with Tone2 (64%). A closer comparison between these studies led us to infer that neutral semantic context might pose great difficulty to question intonation identification in questions ending with Tone4. In another study (Liu et al., 2016a), we examined context effects on question intonation identification in the questions ending with Tone2 condition and the questions ending with Tone4 condition in two behavioral experiments. One experiment embedded the target syllables in a neutral semantic sentence context; the other embedded the target syllables in a highly constraining semantic sentence context. What we found is that in the neutral semantic context, questions ending with Tone4 were not easier than questions ending with Tone2 for question intonation identification. However, in the highly constraining semantic context, question intonation was much better identified in questions ending with Tone4 than in those ending with Tone2, as was in Xu and Mok (2012a) and Yuan (2011). Even more interesting is that in low-pass filtered speech context, Xu and Mok (2014) found that questions ending with Tone2 had a higher accuracy rate than questions ending with Tone4. From the hierarchy of question intonation identification in Tone2 and Tone4 conditions in the above studies, it seems that context plays a role in question intonation identification. The stronger the linguistic context is (highly constraining semantic context > neutral semantic context > low-pass filtered

context), the better the identification of question intonation in questions ending with Tone4.

The opposing pattern was observed for questions ending with Tone2, with better identification of question intonation for weaker linguistic context. We infer that with less semantic information, frequency code (Ohala, 1983), high or rising pitch to mark questions, and low or falling to mark statements are more likely to be applied to intonation identification, resulting in relatively better identification of questions ending with Tone2. However, under no circumstance could listeners disentangle question intonation from Tone2 easily.

Semantic contexts affect question intonation perception. Speech contexts, however, impact question intonation production. Acoustic analyses in Yuan (2006) revealed that question intonation was realized as higher F0 at the end and steeper F0 slope of the final Tone2 than statement intonation in sentences ending with Tone2, and as higher F0 at the end of the final Tone4 than statement intonation in sentences ending with Tone4. Our acoustic results are in agreement with Yuan's results for Tone2, but not for Tone4. We discovered not only a higher F0 at the end in questions than in statements as in Yuan (2006), but also a distinctively higher F0 at the initial contour of the final Tone4 for questions than statements in our data. These different acoustic realizations between Yuan (2006) and our study might result from different coarticulation patterns by the preceding tone contexts. In Yuan's speech materials, the target tone was preceded by a low tone. Tonal coarticulation causes F0 lowering at the initial part of the falling contour of Tone4. Question intonation thus has to be realized as rising in F0 at the end. In comparison, the target tone was preceded by high-level tones in our speech materials. Tonal coarticulation led to a rising in the initial F0 and made the high feature of Tone4 more prominent. Meanwhile, F0 at the end of final Tone4 maintained its rising trend in question intonation.

Finally, it is interesting to compare the present study with the one conducted in Cantonese (Kung et al., 2014). Both studies investigated the online processing of tone and intonation under neutral semantic context. Using similar designs, the present study discovered a P300 effect for questions ending with Tone4 relative to statements ending with Tone4, while Kung et al. (2014)

observed a P600 effect for low tone in questions relative to low tone in statements. The P300 effect in Standard Chinese reflected the ease with which question and statement intonation can be distinguished in sentences with a final Tone4. However, the P600 effect in Cantonese revealed the strong conflicts and processing difficulties when intonation-induced F0 changes lead to the activation of two competing lexical representations. The two ERP components revealed different realizations of interaction of tone and intonation in Standard Chinese and Cantonese. In Standard Chinese, tone identity is maintained with the presence of intonation. Intonation identification is, however, greatly susceptible to the final tone identity. Question intonation is easier to be distinguished from statement intonation if the sentences bear a final Tone4, whereas the difference between intonation types is harder to perceive if the sentences bear a final Tone2. In Cantonese, tone identity is heavily distorted by intonation. The F0 contour of the low tones obtain a rising tail in questions, making it resemble the F0 contour of the mid-rising tone and therefore, causing processing difficulties in lexical identification.

2.5 Conclusion

To conclude, the present study provides online evidence that listeners can distinguish between question intonation and statement intonation when the intonation is associated with a final Tone4, but fail to do so when the intonation is associated with a final Tone2 at the attentive stage of processing. This study extended the sentence scope from one syllable to several syllables, expanded ERP evidence from the pre-attentive stage of processing to the attentive stage of processing, and revealed different realizations of interaction of tone and intonation in Standard Chinese and Cantonese.

Chapter 3

**Context matters for tone and intonation
processing in Standard Chinese**

Abstract

In tonal languages such as Standard Chinese, both lexical tone and sentence intonation are primarily signaled by F0. Their F0 encodings are sometimes in conflict and sometimes in congruency. The present study investigated how tone and intonation, with F0 encodings in conflict or in congruency, are processed and how semantic context may affect their processing. To this end, tone and intonation identification experiments were conducted in both semantically neutral and constraining contexts. Results showed that the overall performance of tone identification was better than that of intonation. Specifically, tone identification was seldom affected by intonation information irrespective of semantic contexts. Intonation identification, particularly question intonation, however, was susceptible to the final lexical tone identity and was greatly affected by the semantic context. Specifically, in the semantically neutral context, questions were difficult to identify (as evident in the lower response accuracy and longer reaction time) regardless of the lexical tone identity. In the semantically constraining context, both intonations took significantly less time to be identified than in the semantically neutral context, and questions ending with a falling tone were much better identified than questions ending with a rising tone. These results suggest that top-down information provided by the semantically constraining context can play a facilitating role for listeners to disentangle intonational information from tonal information, especially in sentences with the lexical falling tone in the final position.

Keywords: tone; intonation; Standard Chinese; neutral context; constraining context

3.1 Introduction

Different languages may have different ways of marking questions. One common way of marking questions in various languages is with the use of syntactic means, including changing word order (see, e.g., Dewaele, 1999 for French; Durrell, 2011 for German; Quirk, Greenbaum, & Leech, 1972 for English), employing *wh*-question words (see, e.g., Dornisch, 1998 for Polish; Koutsoudas, 1968 for English; Rojina, 2004 for Russian), or adding interrogative particles (see, e.g., Chao, 1968 for Standard Chinese; Kuong, 2008 for Cantonese; Tsuchihashi, 1983 for Japanese). Another way frequently adopted across languages to signal questions is via prosodic means, known as intonation. In fact, intonation may be the only means to distinguish questions from statements in syntactically-unmarked yes-no questions (Ultan, 1978; Vaissière, 2008). In such cases, to express question-statement contrasts, a prominent feature of intonation is its modulation of F0 at the sentential level. However, F0 is not only recruited to convey post-lexical intonation information, it is also used to distinguish lexical meanings in many tonal languages such as Standard Chinese.

Tone and intonation are commonly considered as two significant prosodic features of Standard Chinese speech. Although other acoustic correlates (such as duration and intensity) have also been shown to contribute to cue tonal and intonational contrasts (Whalen & Xu, 1992; Xu, 2009; Yuan, 2006), F0 has been identified as the primary acoustic correlate of both tone and intonation (Shen, 1985; Wu, 1982; Xu & Wang, 2001). At the lexical level, the four lexical tones in Standard Chinese are characterized by distinctive F0 contours, with T1 having a high-level contour (55), T2 a mid-rising contour (35), T3 a low-dipping (214), and T4 a high-falling contour (51). At the sentential level, question intonation in Standard Chinese is generally realized as an upward trend of the F0 contour while statement intonation is realized as a downward trend (Ho, 1977; Gårding, 1987; Liu & Xu, 2005). Previous production studies have demonstrated that the upward trend of F0 in question intonation is more pronounced at the end of sentences than at the beginning (Kratochvil, 1998; Liu & Xu, 2005; Xu, 2005; Peng et al., 2005), although some studies also reported an overall F0 rising of sentences in questions compared to statements

(Ho, 1977; Shen, 1989). A perception study by Liang and Van Heuven (2007) found that manipulating the final rise has a much stronger effect on the perception of intonation type than manipulation of the overall pitch level, indicating that the F0 of the final tone is more important than that of the whole sentence for intonation perception. Thus, when a statement ends with a falling tone (T4) or a question ends with a rising tone (T2), the F0 encodings of the final lexical tone and sentence intonation are in congruency. However, when a statement ends with a rising tone (T2) or a question ends with a falling tone (T4), the F0 encodings of the final lexical tone and sentence intonation are in conflict. This raises the question of how tone and intonation are processed in Standard Chinese when their F0 encodings are in conflict or in congruency.

Few studies have tested the effect of intonation on tone perception and vice versa. Connell, Hogan, and Rozsypal (1983) ran a tone perception experiment in Standard Chinese and found that intonation-induced F0 had little effect on tone perception and that tone identity was maintained in question intonation. With regard to the effect of tone on intonation perception, Yuan (2011) found that in Standard Chinese, questions ending with T4 were easier to identify than questions ending with T2. This is interesting considering that in the former, the F0 encodings of question intonation and the final T4 were in conflict, whereas in the latter, the F0 encodings of question intonation and the final T2 were in congruency. In other words, an asymmetrical intonation perception pattern was observed for different F0 encodings of question intonation and final lexical tone. A similar asymmetrical pattern of perception was also reported in Xu and Mok (2012a). However, in a follow-up study using low-pass filtered speech (Xu & Mok, 2012b), the pattern was reversed; Standard Chinese listeners were found to be better at identifying questions ending with T2 than questions ending with T4. These reversed perception patterns might result from many factors, such as prosodic features and lexical intelligibility, among which a potentially very important factor is sentence context.

Sentence context has been shown to play a non-negligible role in language comprehension. It helps comprehenders resolve the identity of lexically ambiguous words (i.e., words with multiple meanings; see Simpson, 1994 for a review). In adverse listening conditions, sentence context helps listeners

compensate for noisy or degraded speech input (Patro & Mendel, 2016; Sheldon, Pichora-Fuller, & Schneider, 2008). Moreover, sentence context has been consistently reported to facilitate language processing, reflected in, for example, reduced processing time or attenuated neural activity of a word (N400) in a highly constraining context versus a weakly constraining context (e.g., Ehrlich & Rayner, 1981; Kutas & Hillyard, 1984). The contribution of sentence context to language processing may be attributed to the role of prediction in language processing. Over the last decades, there has been increasing evidence which suggests that the human brain constantly generates predictions to facilitate the processing of incoming information (for reviews, see, e.g., Federmeier, 2007; Kuperberg & Jaeger, 2016; Kutas, DeLong, & Smith, 2011). Such context-dependent predictive processing has been reported to be present at multiple levels of linguistic representation, such as semantic (Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Van Petten, Coulson, Rubin, Plante, & Parks, 1999), syntactic (Van Berkum, Brown, Zwitterlood, Kooijman, & Hagoort, 2005; Wicha, Bates, Moreno, & Kutas, 2003), phonological (DeLong, Urbach, & Kutas, 2005), and prosodic (Cole, Mo, & Hasegawa-Johnson, 2010; Buxó-Lugo & Watson, 2016; Bishop, 2012) information.

Of particular relevance to the present paper are studies that reported the effect of context on tone and intonation processing. In Standard Chinese, Ye and Connine (1999) investigated tone processing with the target syllables occurring in sentence-final position in a semantically highly constraining context (i.e., idiomatic context) or a semantically neutral context. They found that the semantically highly constraining context considerably facilitated the processing of tone. In Cantonese, another Chinese variety with lexical tone, a rising tail is superimposed on all tone contours by question intonation (Ma, Ciocca, & Whitehill, 2006), making the F0 contour of the low tones (21, 23, 22) in questions resemble that of the mid-rising tone (25). The interaction between tone and intonation in acoustic realization, in turn, causes perceptual processing difficulty of tone in questions. Low tones are easily misperceived as the mid-rising tone in the final positions of questions (Ma, Ciocca, & Whitehill, 2011). When embedding the low and mid-rising tone words sentence-finally in a semantically neutral context versus a semantically strong biasing context (i.e., a

disyllabic word context), Kung, Chwilla, and Schriefers (2014) found that the latter led to much better lexical-identification performance for words with a low tone at the end of questions. This led them to conclude that semantic context plays a major role in disentangling tonal information from intonational information.

It is important to note that, in contrast to the tone processing difficulty in questions in Cantonese, the interaction of tone and intonation leads to intonation processing difficulty in Standard Chinese (Xu & Mok, 2012a, 2012b; Yuan, 2011). This contrast invites further research on the potential typology of the interaction between tone and intonation in tonal languages. Moreover, while we know that context facilitates tone processing in Standard Chinese, the specific role of context, in particular its role in intonation processing and in disentangling intonation from tone processing, remains unclear. The present study was therefore designed to investigate how top-down information provided by semantic contexts affects tone and intonation processing in Standard Chinese when F0 encodings of the final lexical tone and sentence intonation are in conflict or in congruency.

Two experiments were conducted to address this issue. Experiment 1 aimed to examine tone and intonation processing in a semantically neutral context, while Experiment 2 examined tone and intonation processing in a semantically constraining context. In each sentence context, tone and intonation identification experiments were performed using the same design with the same group of participants, allowing for a direct systematic comparison of tone versus intonation identification. The resulting measurements included the commonly-reported response accuracy, as well as an additional measurement, reaction time, which measures the amount of time participants need to identify tone/intonation. Reaction time serves as a good indicator of the degree of difficulty of a perceptual decision: the more difficult a perceptual decision is, the longer the reaction time (Schneider, Dogil, & Möbius, 2011). We expect that reaction time would reveal the pitch processing difficulties entailed in listeners' judgement of tonal and intonational features. Together, they would provide strong evidence for the role of semantic context in resolving pitch processing difficulties in Standard Chinese.

3.2 Experiment 1

3.2.1 Method

3.2.1.1 Materials

Forty monosyllabic word pairs with minimal tonal contrast (T2 vs. T4) and otherwise identical segments were selected. Each minimal T2_T4 word pair contained words of comparable word frequency, homophone density and syntactic word category. To avoid any word frequency effect, only frequent words with more than 4,500 occurrences in a corpus of 193 million words were used (Da, 2004). The average frequencies were 91,631 for T2 and 112,144 for T4 words ($t(39) = -0.56, p = .58$). Following Ziegler, Tan, Perry, and Montant (2000) and Chen, Vaid, and Wu (2009), homophone density was defined as the number of homophone mates of a word, i.e., words that contain exactly the same phonetic segments and lexical tones. We ensured that T2 words had similar homophone densities to their T4 equivalents (15 vs. 15, $p = 1$). The forty word pairs comprised mainly pairs of nouns (32), but pairs of verbs (6) and adjectives (2) were also included to guarantee sufficient number of stimuli.

All the critical words occurred in the final position of a five-syllable carrier sentence, i.e., *ta1 gang1 gang1 shuo1 X* (English translation: “She just said X”), produced with either a statement (S) or a question (Q) intonation. Note that only high-level tones were contained in the carrier sentence. This is to avoid the down-step effect and to minimize the contribution of tone to the observed F0 movement (Shih, 2000). The carrier sentence was semantically meaningful but offered neutral semantic information to the target stimuli and will thus be referred to as the semantically neutral context hereafter. In total, 160 target sentences (40 Syllables \times 2 Tones \times 2 Intonations) were constructed (see Table 1 for an example). In addition, 240 filler sentences were included for the perception experiment. The filler sentences possess the same carrier but different critical syllables in terms of either segmental composition or lexical tone (e.g., T1/T3).

Table 1. *An example of the experimental design in Experiment 1.*

Condition		Example				
Tone	Intonation					
Tone2	Statement	Characters	她	刚刚	说	X(财) 。
		Pinyin	ta1	gang1gang1	shuo1	cai2
		English	She	just	said	money.
Tone2	Question	Characters	她	刚刚	说	X(财)?
		Pinyin	ta1	gang1gang1	shuo1	cai2
		English	She	just	said	money?
Tone4	Statement	Characters	她	刚刚	说	X(菜) 。
		Pinyin	ta1	gang1gang1	shuo1	cai4
		English	She	just	said	vegetable.
Tone4	Question	Characters	她	刚刚	说	X(菜)?
		Pinyin	ta1	gang1gang1	shuo1	cai4
		English	She	just	said	vegetable?

Note. The critical syllables are in bold.

3.2.1.2 Recording and stimuli preparation

One female native speaker of Standard Chinese, who was born and raised in Beijing, recorded the sentences in a soundproof recording booth at the Phonetics Laboratory of Leiden University. Sentences were randomly presented to the speaker using an HTML JavaScript and recorded with a Sennheiser MKH416T microphone at 16-bit resolution and a sampling rate of 44.1 kHz using Adobe Audition 2.0. To eliminate paralinguistic information, the speaker was instructed to avoid any exaggerated emotional prosody during the recording.

This female speaker's recordings were clear and consistent. The acoustic results of the recordings (see Chapter 2 and also Liu, Chen, & Schiller (2016b) for details) showed comparable F0 realization of tone and intonation to a prior study (Yuan, 2006) and were taken as the representative patterns of question and statement intonation in Standard Chinese, and therefore suitable for the perception study. In the subsequent perception experiment, the RMS amplitude

of all the sentences was normalized to 75dB SPL in Praat (Boersma & Weenink, 2015).

To further verify the validity of the intonation patterns perceptually, a panel of five phonetically trained researchers was asked to evaluate the typicality of the intonation of the sentences on a five-point scale (1 = very typical statement; 5 = very typical question). While all tokens produced by our speaker were included in the perception experiment, only tokens identified as typical of their corresponding intonation category (i.e., score ≤ 1.5 for statements and score ≥ 3.5 for questions) by at least three out of the five researchers were selected for the data analysis reported below. The average typicality rating score for the final set of selected stimuli in the data analysis was 1.1 for statements and 4.5 for questions, resulting in an exclusion of 13.1% of the data points.

3.2.1.3 *Participants*

Eighteen native speakers of Standard Chinese (10 females, 8 males) from Northern China were paid to participate in the experiment. They were undergraduate or graduate students at Beijing Language and Culture University, between 19 and 27 years old ($M \pm SD$: 23.6 ± 2.3). None of them had received any formal musical training or reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

3.2.1.4 *Procedure*

Participants were tested individually in a sound-attenuated room. Four-hundred sentence trials (including 160 targets and 240 fillers) were randomly presented to the participants using E-Prime 2.0 software through headphones (AKG K242HD) at a comfortable listening level. Instructions were given both visually on screen and orally by the experimenter in Standard Chinese before the experiment.

The experiment included a practice block and four experimental blocks. The practice block contained 12 trials. Each experimental block contained 100 trials. Between two blocks there was a 3-minute break. An experimental trial started with a 100 ms warning beep, followed by a 300 ms pause. After that an auditory

sentence was presented while a visual task interface appeared on the screen. Participants were requested to carry out either a tone identification task or an intonation identification task as quickly and accurately as possible. For each test block, half of the trials contained the tone identification task while the other half contained the intonation identification task; the task varied randomly from trial to trial. Task types were indicated by tone and intonation marks in Standard Chinese Pinyin system, the official romanization system for Standard Chinese, which all participants knew very well. For example, when “ ´ ` ” marks (“ ´ ” stands for T2; “ ` ” stands for T4) appeared on the screen, participants were asked to identify whether the final tone of the sentence was T2 or T4. When the “。 ? ” marks appeared on the screen (“。 ” stands for statement intonation; “? ” stands for question intonation), they were asked to identify whether the sentence bore statement or question intonation. Listeners were given up to 2 seconds after the offset of the sentence to respond. No participants reported difficulty in understanding the tasks. The inter-stimulus interval was 500 ms.

3.2.1.5 Data analysis

Previous studies on intonation perception have typically only reported response accuracy (Xu & Mok, 2012a, 2012b; Yuan, 2011). In this study, in addition to response accuracy, reaction time was included as a dependent variable. Response accuracy here was defined as the percentage of correct identification of tone in the tone identification task, and as the percentage of correct identification of intonation in the intonation identification task. Reaction time was defined as the response time relative to the onset of the last syllable for correct responses. To normalize the distribution, raw reaction times were transformed using the natural logarithm.

Statistical analyses were carried out with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.1.2 (R Core Team, 2015). Analysis of response accuracy was performed using binomial logistic regression models, and analysis of reaction time was performed using linear mixed-effects regression models. The models included Task, Tone, Intonation, and their interactions as fixed factors, and Subjects and Items as random factors. The

fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. For models of reaction time, trials with absolute standardized deviations exceeding 2.5 from the mean were considered as outliers and removed from further analysis. We also considered trial-by-trial dependency in model constructions. However, it did not significantly improve the model fit, and was therefore excluded in the final model.

3.2.2 Results

3.2.2.1 Response accuracy

To test whether tone and intonation are processed differently, we first examined the effect of Task. Results (see Figure 1 and also Table B1 in Appendix B) showed a significant main effect of Task ($\chi^2(1) = 47.19, p < .001$) and a significant interaction of Task \times Intonation ($\chi^2(1) = 12.50, p < .001$).

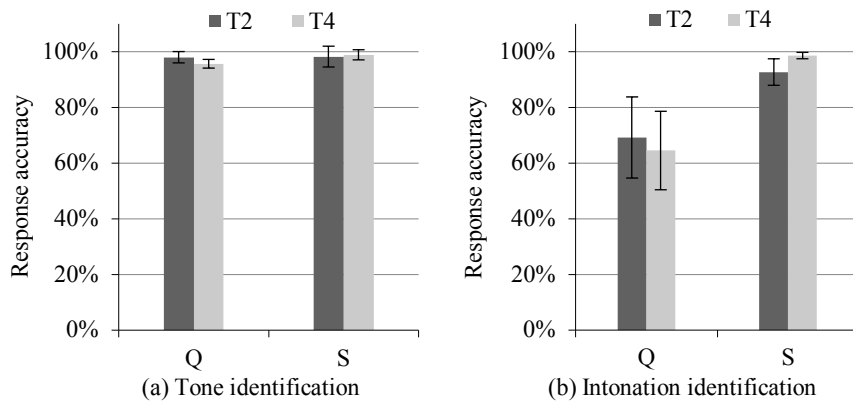


Figure 1. Response accuracy as a function of final lexical tone and sentence intonation for the tone identification task (a) and the intonation identification task (b) in the semantically neutral context. The error bars represent the 95% confidence interval of the means across participants.

Separate models for subset data of different intonation types revealed a notable asymmetry between question and statement intonation. Specifically, in

question sentences, the response accuracy of tone was much higher than that of intonation ($\beta = 3.22$, $\zeta = 9.48$, $p < .001$). This, however, was not observed in statement sentences, where near ceiling-level identification was found in both tasks.

Separate models were also constructed for subset data of different tasks. For the tone identification task, results showed a main effect of Intonation ($\chi^2(1) = 4.14$, $p = .04$); the response accuracy of tone was slightly lower in questions than in statements. No other effect was found. Given that tone identification was almost at ceiling level across the experimental conditions, the very few incorrect responses were likely motor-related errors as a result of the speed requirement of the task. Overall, it seems that the identity of lexical tone was not hindered by the intonation information. With respect to the intonation identification task, we found a significant main effect of Intonation ($\chi^2(1) = 76.55$, $p < .001$) as well as a significant interaction of Tone \times Intonation ($\chi^2(1) = 12.19$, $p < .001$). Question intonation tended to be more difficult to identify than statement intonation regardless of the final lexical tone identities. Separate models for subset data of different types of intonation showed that statement intonation was more accurately identified in statement sentences ending with T4 than in those ending with T2 ($\beta = 1.55$, $\zeta = 2.08$, $p = .04$), whereas question intonation was equally difficult to identify in question sentences ending with T2 and T4 ($\beta = -0.42$, $\zeta = -0.80$, $p = .42$).

3.2.2.2 Reaction time

Twenty-five trials (1.1%) were identified as outliers and removed from further analysis. Analysis of the remaining data points revealed a significant main effect of Task ($\chi^2(1) = 69.79$, $p < .001$); final tone was identified faster than sentence intonation (see Figure 2 and also Table B1 in Appendix B). Other factors such as Tone ($\chi^2(1) = 20.47$, $p < .001$), Intonation ($\chi^2(1) = 16.76$, $p < .001$) and the interaction of Tone \times Intonation ($\chi^2(1) = 25.83$, $p < .001$) were also significant.

Separate models for subset data of different intonations showed that across task types, there was a main effect of Tone for sentences with statement intonation ($\chi^2(1) = 49.82$, $p < .001$), but not for those with question intonation ($\chi^2(1) = 0.16$, $p = .68$). That is, irrespective of task types, a much shorter

reaction time was observed for statements ending with T4 relative to statements ending with T2 ($\beta = -0.13$, $t = -8.36$, $p < .001$), whereas no reaction time difference was found between questions ending with T2 versus T4 ($\beta = 0.01$, $t = 0.41$, $p = .68$).

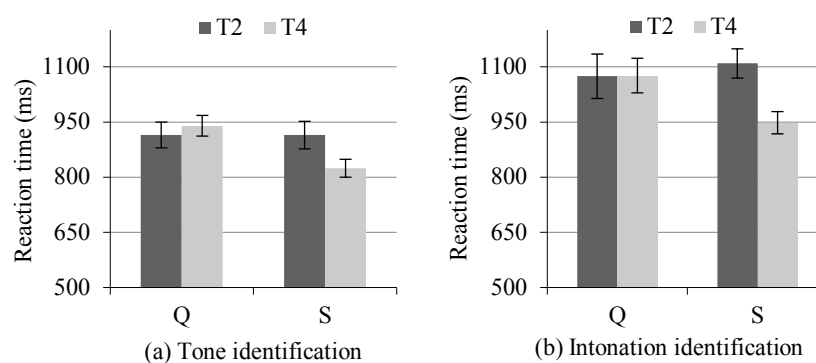


Figure 2. Average reaction time as a function of final lexical tone and sentence intonation for the tone identification task (a) and the intonation identification task (b) in the semantically neutral context. The error bars represent the 95% confidence interval of the means across participants.

Separate models were also constructed for subset data of different tones, the results of which confirmed that there was no Intonation effect for the T2 conditions ($\chi^2(1) = 0.73$, $p = .39$), but there was a significant effect of Intonation for the T4 conditions ($\chi^2(1) = 37.82$, $p < .001$), with shorter reaction times for statements (ST4) than for questions (QT4) regardless of the task types ($\beta = -0.12$, $t = -6.96$, $p < .001$).

Overall, tone identification reached almost ceiling level across all experimental conditions. Participants quickly and accurately identified lexical tones produced with both intonations. However, the identification of intonation, especially question intonation, was much less accurate. Moreover, listeners took longer to identify intonation (as evident in the longer reaction time) than to identify the final lexical tone. Taken together, in a semantically neutral context, participants had great difficulty in perceiving question intonation. This is in line with previous studies (Xu & Mok, 2012a, 2012b; Yuan, 2011). In contrast to these earlier studies, no response accuracy

difference was found for question intonation in questions ending with T2 versus T4.

3.3 Experiment 2

The results of Experiment 1 showed that in the semantically neutral context, question intonation processing is challenging, regardless of the final lexical tone identities. In Experiment 2, we examined whether a semantically constraining context helps resolve the processing difficulty of different F0 encodings of the final lexical tone and sentence intonation.

3.3.1 Method

3.3.1.1 Materials

To avoid learning effects from Experiment 1, an additional set of 40 monosyllables in combination with tone (T2 or T4) was selected for Experiment 2. Each minimal pair of T2_T4 monosyllables were the second syllables of two disyllabic words with comparable word frequency. According to the SUBTLEX-CH frequency list (Cai & Brysbaert, 2010), the average log₁₀ word frequencies were 2.49 for disyllabic words ending in T2 and 2.71 for those ending in T4 ($t(39) = -1.79, p = .08$). These disyllabic words were embedded in the final position of various nine- or ten-syllable natural sentences. The reason for us using the disyllabic word as part of the sentence context frame is that it is the predominant word type in Standard Chinese, and most often used in natural sentences (Duanmu, 2007). Furthermore, previous studies (Xu & Mok, 2012a, 2012b; Yuan, 2011) have embedded disyllabic words sentence-finally in their studies, thus our similar set-up enables a comparison of results.

A cloze probability pretest was conducted to verify that our sentence context provides sufficient constraint on the final syllable. The cloze probability of a word is the percentage of participants who offer that word as a completion for a sentence of which the final word is missing (Van Petten & Luka, 2012). Thirty native Standard Chinese speakers participated in the pretest. They were presented with the sentences but without the final syllable, and were asked to

provide the most likely syllable that fits the given sentence frame. Each final syllable had a cloze probability of at least 70%. This sentence context will be referred to as the semantically constraining context hereafter.

As in Experiment 1, all the sentences were produced with either statement or question intonation, yielding another 160 target sentences (40 Syllables \times 2 Tones \times 2 Intonations, see Table 2 for an example). Fillers were also included (240 sentences).

Table 2. *An example of the experimental design in Experiment 2.*

Condition		Example						
Tone	Intonation	Characters	Pinyin	English				
Tone2	Statement	Characters	这家	旅馆	有	三十间	客房。	
		Pinyin	zhe4jia1	lv3guan3	you3	san1shi2jian1	ke4 fang2	
		English	This hotel has thirty guest rooms.					
Tone2	Question	Characters	这家	旅馆	有	三十间	客房？	
		Pinyin	zhe4jia1	lv3guan3	you3	san1shi2jian1	ke4 fang2	
		English	This hotel has thirty guest rooms?					
Tone4	Statement	Characters	海瑞	故居	将	向	游人	开放。
		Pinyin	Hai3 Rui4	gu4ju1	jiang1	xiang4	you2ren2	kai1 fang4
		English	Hai Rui's former residence will be open to visitors.					
Tone4	Question	Characters	海瑞	故居	将	向	游人	开放？
		Pinyin	Hai3 Rui4	gu4ju1	jiang1	xiang4	you2ren2	kai1 fang4
		English	Hai Rui's former residence will be open to visitors?					

Note. The critical syllables are in bold.

3.3.1.2 Recording and stimuli preparation, participants, procedure and data analysis

Recording and stimuli preparation, participants, procedure and data analysis were the same as in Experiment 1. The same speaker recorded the sentences. Experiment 2 was run after Experiment 1 with the same group of participants.

In a different session, the same five phonetically trained researchers evaluated the typicality of the intonation of the sentences for Experiment 2 on a five-point scale (1 = very typical statement; 5 = very typical question). We applied the same criteria when selecting sentences for data analysis. Despite all

the tokens being included in the perception experiment, only tokens identified as typical of their corresponding intonation category (i.e., score ≤ 1.5 for statements and score ≥ 3.5 for questions) by at least three out of the five researchers were analyzed. Consequently, 13.8% of the data points were excluded. The average typicality rating score for the remaining selected stimuli in the data analysis was 1.0 for statements and 4.4 for questions.

3.3.2 Results

3.3.2.1 Response accuracy

The overall analyses revealed a significant main effect of Task ($\chi^2(1) = 10.41, p = .001$), a two-way interaction of Task \times Intonation ($\chi^2(1) = 21.45, p < .001$), and a three-way interaction of Task \times Tone \times Intonation ($\chi^2(2) = 9.01, p = .01$). The tone identification task showed better performance than the intonation identification task in question sentences with a final T2 ($\beta = 3.14, \zeta = 3.53, p < .001$). However, the response accuracies for final lexical tone and sentence intonation were not significantly different in either question sentences with a final T4 ($\beta = 1.75, \zeta = 1.66, p = .10$), or statement sentences across final lexical tone identities ($\beta = -0.89, \zeta = -0.83, p = .41$) (see Figure 3).

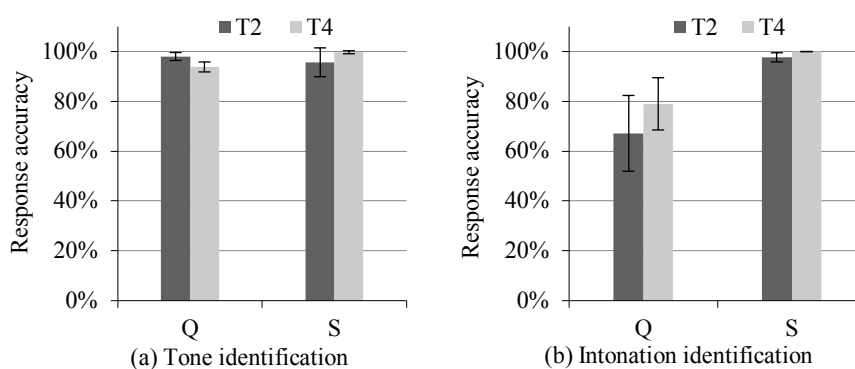


Figure 3. Response accuracy as a function of final lexical tone and sentence intonation for the tone identification task (a) and the intonation identification task (b) in the semantically constraining context. The error bars represent the 95% confidence interval of the means across participants.

Separate models were constructed for subset data of different tasks. For the tone identification task, no effect was found for Tone, Intonation or their interaction (all p s > .05). Regardless of intonation type, T2 and T4 were mostly correctly identified. Again, the identity of lexical tone was not hindered by the intonation information.

Results of the intonation identification task showed a significant main effect of Intonation ($\chi^2(1) = 71.97, p < 0.001$) and a marginal significant interaction of Tone \times Intonation ($\chi^2(1) = 3.69, p = .05$). Specifically, only in sentences with a final T2, did statement intonation identification display an advantage over question intonation ($\beta = 4.37, \zeta = 3.05, p = .002$). And only in question sentences, final tone identity affected intonation identification. Question intonation was more accurately identified in questions ending with T4 than in those ending with T2 ($\beta = 1.06, \zeta = 2.01, p = .045$).

3.3.2.2 Reaction time

Twenty-three trials (1.0%) were identified as outliers and removed from further analysis. The overall analyses of the remaining data points showed a significant interaction of Task \times Intonation ($\chi^2(1) = 4.08, p = .043$). Participants were faster in identifying the final lexical tone than the sentence intonation in question sentences ($\beta = -0.09, t = -2.71, p = .007$), but not in statement sentences ($\beta = -0.01, t = -0.24, p = .81$). With a semantically constraining context, reaction time in the intonation identification task for statements ending with T2 decreased to such a degree that it even became shorter than RTs for the same condition in the tone identification task (see Figure 4).

Separate analyses were performed for subset data of different tasks. For the tone identification task, there was a significant main effect of Tone ($\chi^2(1) = 16.22, p < .001$) and a significant main effect of Intonation ($\chi^2(1) = 11.19, p < .001$). No interaction of Tone \times Intonation was found ($\chi^2(1) = 2.72, p = .10$). Tone identification generally tended to be faster in statements than in questions ($\beta = -0.11, t = -3.49, p < .001$). Moreover, it took less time to identify T4 than T2 across intonation types ($\beta = -0.14, t = -4.75, p < .001$). Consequently, tone identification in ST4 had a significant advantage over the other experimental conditions.

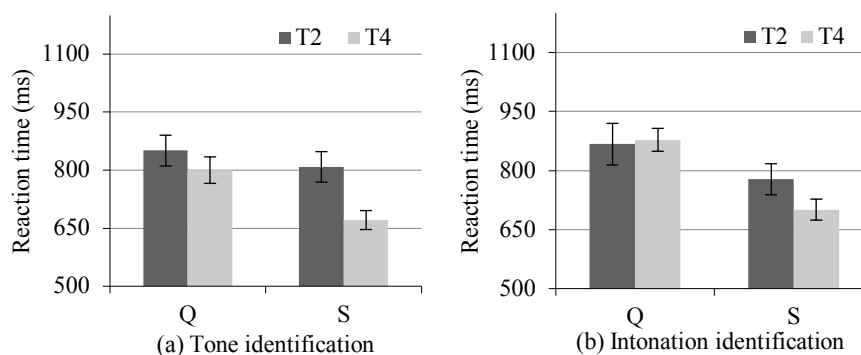


Figure 4. Average reaction time as a function of final lexical tone and sentence intonation for the tone identification task (a) and the intonation identification task (b) in the semantically constraining context. The error bars represent the 95% confidence interval of the means across participants.

For the intonation identification task, all effects reached significance, including a main effect of Tone ($\chi^2(1) = 5.22, p = .02$), a main effect of Intonation ($\chi^2(1) = 26.20, p < .001$) and an interaction of Tone \times Intonation ($\chi^2(1) = 6.48, p = .01$). With a semantically constraining context, the identification of statement intonation was significantly faster compared to the identification of its question counterpart regardless of final lexical tone identities ($\beta = -0.18, t = -5.61, p < .001$). An investigation of the interaction effect revealed shorter reaction times to identify statement intonation in statements ending with T4 relative to statements ending with T2 ($\beta = -0.11, t = -2.57, p = .01$), but when identifying question intonation, no reaction time difference was found between questions ending with T2 and those ending with T4 ($\beta = -0.03, t = -0.71, p = 0.48$).

To sum up, in a semantically constraining context, tone identification maintained its near-ceiling-level identification accuracy across all experimental conditions, and notably with shorter reaction time compared to the semantically neutral context. Similarly, reaction time was consistently shortened for intonation identification. The response accuracy for intonation, in most conditions, increased relative to those in the semantically neutral context except in questions ending with T2, where a slight decrease of response accuracy was found. Consequently, question intonation was better identified in questions

ending with T4 than with T2, as in Xu and Mok (2012a) and Yuan (2011). Detailed comparisons of Experiment 1 and Experiment 2 will be made in the next section.

3.4 Experiment 1 vs. Experiment 2

To further verify the effect of semantic context, trials of Experiments 1 and 2 were merged into one dataset, and Context was added as an additional factor in the analysis.

3.4.1 Response accuracy

The overall combined analyses of response accuracy showed a significant two-way interaction of Context \times Task ($\chi^2(1) = 7.75, p = .005$) and a significant three-way interaction of Context \times Task \times Intonation ($\chi^2(3) = 37.31, p < .001$).

As can be seen from Figure 5(a), tone identification almost reached ceiling level across all experimental conditions. No main effect of Context or any interaction of Context with other factors was found for the tone identification task (all p s $> .05$).

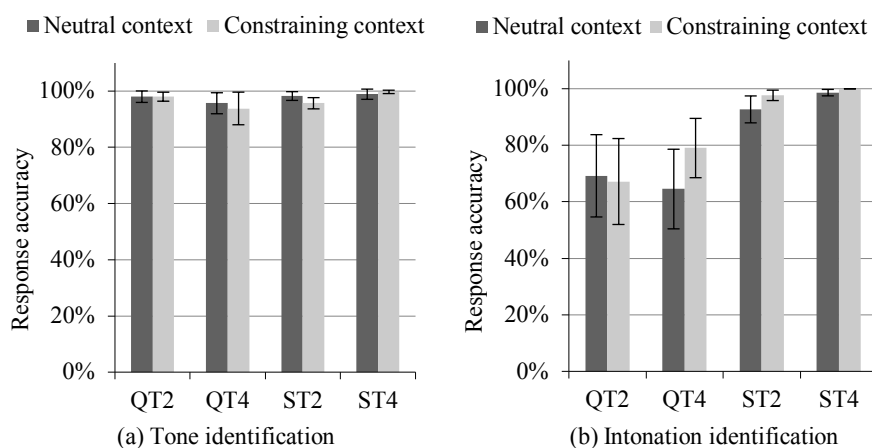


Figure 5. Response accuracy of each experimental condition in the semantically neutral context (dark grey bars) and in the semantically constraining context (light grey bars) for the tone identification task (a) and the intonation identification task (b). The error bars represent the corresponding 95% confidence interval of the means across participants.

For the intonation identification task (Figure 5(b)), we found a significant main effect of Context ($\chi^2(1) = 5.48, p = .02$) and a significant three-way interaction of Context \times Tone \times Intonation ($\chi^2(2) = 6.17, p = .045$). Further analyses confirmed the main effect of Context in statements across final lexical tone identities ($\chi^2(1) = 10.45, p = .001$) and in questions ending with T4 ($\chi^2(1) = 8.06, p = .005$), but not in questions ending with T2 ($\chi^2(1) = 0.07, p = .79$), which suggests that the response accuracy of intonation in the former three conditions increased in the semantically constraining context compared to their semantically neutral counterparts (ST2: 97.7% vs. 92.7%; ST4: 100% vs. 98.6%; QT4: 79.0% vs. 64.5%). In questions ending with T2, the response accuracy of question intonation in the semantically constraining context was inclined to decrease if compared to that in the semantically neutral context (QT2: 67.2% vs. 69.2%).

3.4.2 Reaction time

The overall combined analyses of reaction time showed a significant main effect of Context ($\chi^2(1) = 148.01, p < .001$) and two-way interactions of Context \times Task ($\chi^2(1) = 24.52, p < .001$) as well as Context \times Intonation ($\chi^2(1) = 7.62, p = .006$).

Follow-up analyses for the tone identification task showed a significant main effect of Context ($\chi^2(1) = 55.08, p < .001$) and a significant interaction of Context \times Tone ($\chi^2(1) = 5.08, p = .02$). Clearly, reaction time to identify the final lexical tones was considerably shorter in the semantically constraining context relative to that in the semantically neutral context across all experimental conditions (see Figure 6(a)). Furthermore, the effect of Context was greater for T4 identification compared to its T2 equivalent regardless of the intonation types, as evidenced by the larger reaction time difference between the two semantic contexts for T4 than T2. This suggests that participants benefited more from the semantically constraining context when identifying T4 as compared to T2.

Likewise, follow-up analyses for the intonation identification task revealed a significant main effect of Context ($\chi^2(1) = 117.15, p < .001$) and a significant interaction of Context \times Intonation ($\chi^2(1) = 9.47, p = .002$). Despite the fact

that intonation identification was generally faster in the semantically constraining context than in the semantically neutral context (see Figure 6(b)), semantic context did not affect the identification of different intonation types to the same degree. The semantically constraining context seemed to contribute more to statement intonation identification than to question intonation identification regardless of the final lexical tone identities. Overall, context considerably affected reaction times to identify tone and intonation. The semantically constraining context played a significant role in speeding up the identification of both tone and intonation across the experimental conditions. It shortened reaction times to a larger degree in identifying intonation than tone, in identifying T4 than T2, and also in identifying statement intonation than question intonation.

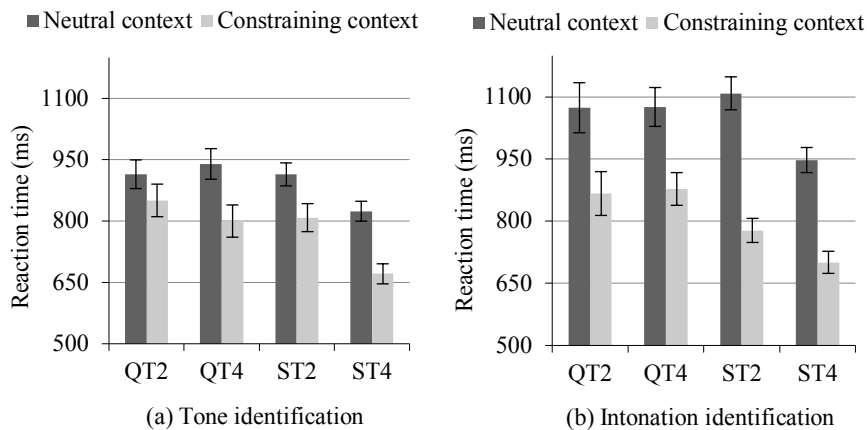


Figure 6. Average reaction time for each experimental condition in the semantically neutral context (dark grey bars) and in the semantically constraining context (light grey bars) for the tone identification task (a) and the intonation identification task (b). The error bars represent the corresponding 95% confidence interval of the means across participants.

3.5 General discussion

To address the question of how top-down information provided by semantic contexts affects tone and intonation processing in Standard Chinese when F0 encodings of the final lexical tone and sentence intonation are in conflict or in

congruency, we examined the identification of tone and intonation in both semantically neutral and constraining contexts. Our results demonstrated that in Standard Chinese, tone identification was seldom affected by intonation information irrespective of semantic contexts, whereas intonation identification, particularly question intonation, was susceptible to the final tone identity and was greatly impeded in the semantically neutral context. A semantically constraining context considerably improved question intonation identification.

In our study, the overall performance of tone identification was better than that of intonation identification regardless of semantic contexts. Evidence was found not only from the response accuracy results, but also from the reaction time patterns. Intonation identification took more time than tone identification regardless of the final lexical tone identities, suggesting that in Standard Chinese, when pitch movements are used to convey post-lexical contrast, its identification becomes a much more difficult decision-making process (Braun & Johnson, 2011). The advantage of tone over intonation is probably because a phonetic dimension (i.e., F0) exploited for one function of the grammar (e.g., lexical tone) limits its effectiveness to cue a different function (e.g., intonation) in the same linguistic system (Liang & Van Heuven, 2007; Nolan, 2006).

Previous studies found reversed patterns of question intonation identification in questions ending with T2 and T4 in normal context (Xu & Mok, 2012a; Yuan, 2011) versus in low-pass filtered context (Xu & Mok, 2012b). It is unclear whether the reversed patterns are due to the different test contexts in semantic or other information. The present study teased apart the effect of semantic context from the other factors by introducing the semantically neutral versus constraining contexts. We found that the semantically neutral context posed greater difficulty to question intonation identification, compared to the semantically constraining context. In the former, questions ending with T2 and T4 were equally poorly identified. In the latter, questions ending with T4 were better identified than those ending with T2, in line with the results in Xu and Mok (2012a) and Yuan (2011). Recall that in low-pass filtered speech, questions ending with T2 even had a higher response accuracy than questions ending with T4 (Xu & Mok, 2012b). From the hierarchy of question intonation identification in T2 and T4 conditions in these

studies, it seems that context plays a role in question intonation identification. The stronger and more informative the linguistic context is (semantically constraining context > semantically neutral context > low-pass filtered context), the better the identification of questions ending with T4. The opposite pattern was observed for questions ending with T2, with better identification of question intonation for weaker and less informative linguistic context. We infer that with less semantic information, the frequency code (Gussenhoven, 2004; Morton, 1994; Ohala, 1983), which holds that high or rising pitch marks questions, and low or falling pitch marks statements, is more likely to be applied to intonation identification, resulting in relatively better identification of questions ending with T2. However, under no circumstance could listeners disentangle question intonation from T2 easily (69.2% vs. 67.2%). When more semantic information is given, questions ending with T4 seem to get more cues of question intonation than questions ending with T2. The reasons for this warrant further investigation.

If response accuracy speaks for context effect only in question intonation identification where overt processing difficulties occur, reaction time lends stronger support to the effect of context in a broader sense. A semantically constraining context speeded up not only intonation identification, but also tone identification compared to the semantically neutral context. It shortened RTs for intonation identification to a larger extent.

In the semantically neutral context, very limited information except bottom-up acoustic information was available to the participants for them to identify tone and intonation. With one and the same carrier sentence for all the final target syllables, participants had no knowledge of what the last syllable would be until they actually heard it. It was therefore not possible for the participants to have predicted the tone of the last syllable before it actually occurred. From the standpoint of information theory (Shannon, 1948), the entropy of the upcoming final syllable was very high. It thus required more resources to process the incoming information (Frank, 2013; Hale, 2001; Levy, 2008), leading to lower response accuracy and longer reaction times. In contrast, in the semantically constraining context, the final target syllables had lower entropy due to their high predictability. Participants could easily predict the target

syllables based on the available context information of the sentence prior to hearing the syllables. Such syllable prediction allows for the pre-activation of both the segment and tone of the syllable (Ye & Connine, 1999). It is therefore not surprising that tone identification was faster in the semantically constraining context relative to the semantically neutral context. With fewer processing resources taken up by the tone identification task, participants could devote more attention and processing resources to identifying the intonation of the sentences, which in turn led to the overall improved response accuracy and shorter reaction times in the intonation identification.

Though a semantically constraining context considerably shortened the reaction time to identify both the final lexical tone and sentence intonation, it did not entirely resolve the processing difficulty in intonation identification. Participants still had difficulty identifying question intonation, especially when the question intonation concurred with T2. The processing difficulty of question intonation here reflects actual difficulties in disentangling pitch information, which signals both lexical tone of the final syllable and intonation of the sentence. This is supported by evidence from an ERP study (Liu et al., 2016b), in which Standard Chinese sentences in a semantically neutral context (the same as in Experiment 1) were auditorily presented to native listeners and their electrophysiological responses were recorded. The participants were asked to pay special attention to the final lexical tone and sentence intonation but were not asked to make any motor-related responses or metalinguistic judgments. A clear P3b effect was found for questions ending with T4 relative to statements ending with T4, whereas no ERP effect was found for questions ending with T2 relative to statements ending with T2. The ERP results thus suggest that the question-statement contrast in the T4 conditions is easier to categorize, whereas categorization of the question-statement contrast in the T2 conditions is much more demanding for Standard Chinese native listeners. Comparing questions ending with T4 with statements ending with T4, the former tends to be more difficult and requires more processing effort as evidenced by the P3b effect in questions ending with T4.

3.6 Conclusion

The results of the two experiments reported here show that tone at the lexical level and intonation at the sentential level in Standard Chinese interact with each other, causing an asymmetrical difficulty of pitch processing at the sentential level. To disentangle intonational information from tonal information more efficiently, listeners not only tune to acoustic cues, but also rely on predictions based on the semantic context. A semantically constraining context considerably improves question intonation identification.

Chapter 4

Tonal mapping of Xi'an Mandarin and Standard Chinese

Abstract

Tonal information can be a determinant of the phonological similarity or difference between some Chinese dialects and Standard Chinese, yet relatively little empirical research has been conducted on the tonal system of other language varieties in Chinese other than Standard Chinese. Among these dialects, Xi'an Mandarin is particularly interesting for its seemingly present one-to-one mapping of tones with Standard Chinese tones. To gain empirical evidence for the mapping, the present study compared the tonal systems of Xi'an Mandarin and Standard Chinese in both tone production and perception. Tones with similar contours from Xi'an Mandarin and Standard Chinese were paired and both tone production and perception experiments were carried out on bi-dialectal speakers of Xi'an Mandarin and Standard Chinese. Acoustic results showed that the F0 difference ranged from no F0 difference (level contour tone pair) through F0 curvature difference (rising contour tone pair) to F0 height difference (falling contour tone pair) and F0 contour difference (low contour tone pair). Except for the falling contour tone pair, all the other tone pairs also exhibited differences in tone duration. The varying acoustic differences in different tone pairs, together with the phonological rule, resulted in varying degrees of tonal similarity in tone perception, but tones with similar contours between the two dialects were basically perceived to be the same. The two experiments together showed the systematic mappings of tones between Xi'an Mandarin and Standard Chinese.

Keywords: Standard Chinese, Xi'an Mandarin, tonal mapping, tone production, tone perception

4.1 Introduction

Chinese is a tonal language where tones are used to distinguish lexical meanings. However, the term “Chinese” refers to a large number of Sinitic language varieties. While numerous studies have been conducted on Standard Chinese (i.e., the official language of China; SC), relatively little attention has been paid to other dialects or language varieties of Chinese. Some of the dialects differ from SC in both segmental and tonal information, whereas others such as dialects within the Mandarin family overlap largely in segmental information with SC. In these latter dialects, tonal information can be important as it determines the phonological similarity or difference between the dialect and SC.

In China, most speakers of SC speak a local dialect (Li & Lee, 2008; Wiener & Ito, 2014). It is therefore of both practical and theoretical value to systematically investigate the tonal similarity or difference between different dialects and SC. Such investigations can be the prerequisite to developing dialect-oriented speech synthesis and speech recognition technology (Czap & Zhao, 2017), to guiding language pedagogy in teaching SC to dialectal speakers (Lam, 2005; Wong & Xiao, 2010), and to addressing issues such as whether the phonological information of one’s two or more dialects are stored separately or integrally (Wu, 2015), or how cross-dialect phonological similarity/difference affects lexical access in the minds of bi-dialectal tonal language speakers.

Currently, relatively little empirical research has been conducted on the tonal system of other language varieties except for SC; even less research compared the tonal system of other language varieties with that of SC. As language varieties within the Mandarin family rely largely on tonal information to make distinctions from SC, the present study aimed to empirically compare the tonal systems of two closely related dialects in the Mandarin family, SC and Xi’an Mandarin (XM).

According to Chappell (2001), there are ten major dialect groups in Chinese (but see Yuan, 1989; Li & Thompson, 1981 which argue for a seven major dialect groups). The Mandarin family is the largest Chinese dialect group. It contains a group of Chinese varieties, which are typically spoken in the northern and southwestern China. The most influential language within the Mandarin family is SC. The other dialects within the Mandarin family share a

common logographic writing system with SC and bear high resemblance with SC as to lexical items and syntactic forms. Some dialects such as XM also exhibit large overlap of segmental features with SC. More interestingly, the tones of XM seem to have one-to-one correspondence with those of SC (Li, 2001; Zhang, 2009). This overall correspondence between the two tonal systems is quite unique and makes XM a very compelling case to study.

XM is a Mandarin dialect typically spoken in the urban areas of Xi'an, the capital of Shaanxi Province. It is the representative dialect of the Guanzhong dialect spoken in the Guanzhong area (Li & Stephen, 1987). XM directly originates from the official language in ancient China and has important historical value. As in SC, there are four tonal categories in XM, and they are referred to as T1, T2, T3 and T4. Here, the terms T1-T4 are adopted to suggest that words which share the same tonal categories across the two dialects are etymologically-related translation equivalents in most cases. Across XM and SC, different tones distinguish lexical meanings for syllables with the same segment. For example, the segment *ma* means *mother*, *bemp*, *horse* and *scold*, respectively, when it is combined with the four lexical tones (tonal category: T1, T2, T3, T4). On the 5-point scale notation system (Chao, 1930; 1968), the pitch value of the SC tones has been established as 55 (T1), 35 (T2), 214 (T3) and 51 (T4) respectively. However, there have been discrepancies among researchers regarding the specific pitch value of each XM tone (see Table 1 for a summary of the representative transcription of XM tones).

The majority of the existing studies on XM tones, including the first six studies listed in Table 1, have been based on impressionistic observation. Pitch values of XM tones in these studies could be susceptible to the subjective pitch sensitivity of the researchers. It is therefore not surprising that these studies vary in pitch value for each tonal category. The remaining studies, such as the last three in Table 1, have attempted to study the pitch value of XM tones with more objective acoustic methods. However, these studies either sampled from a very limited number of speakers (e.g., two in Ma (2005); one in Ren (2012)) or lacked control of the lexical properties of the stimuli used (e.g., Zhang & Shi, 2009). It is not known to what degree these results can represent the typical tonal patterns of XM. The present study thus decided to empirically examine

the acoustic properties of XM tones with a larger sample of speakers and stimuli and better control of lexical properties of the stimuli.

Table 1. *Representative transcription of Xi'an Mandarin tones in previous studies.*

Reference	T1	T2	T3	T4
Bai (1954)	21	24	453	45
Luo & Wang (1981)	31	24	42	55
Yuan (1989)	21	24	53	45
Wang (1996)	21	24	53	44
Peking University (2003)	21	24	53	55
Sun (2007)	31	24	53	55
Ma (2005)	21	24	53	44
Zhang & Shi (2009)	31	24	52	55
Ren (2012)	31	24	52	55

Although the specific pitch value of each tonal category in XM varies among previous studies, the basic tonal contour shape tends to be largely consistent across studies. Generally, the four tonal categories of XM possess the tonal contours of low-falling (T1), mid-rising (T2), high-falling (T3) and high-level (T4), respectively. Interestingly, XM tones display almost the same tonal contours as SC tones. In SC, tonal contours of the four tonal categories are described as high-level (T1, 55), mid-rising (T2, 35), low-falling-rising (T3, 214) and high-falling (T4, 51), respectively. As one can see, both tonal systems of SC and XM contain tones of high-level, mid-rising and high-falling tonal contours, and each of these tone pairs of similar contours share similar pitch values across the two tonal systems, though the similar contours do not necessarily represent the same tonal category in the two tonal systems (see Table 2 for details). Moreover, SC has a tone of low-falling-rising tonal contour, whereas XM has a tone of low-falling tonal contour without the rising tail. The former, however, would lose its rising tail when placed before other tones (Dow, 1972; White, 1980) and become similar to the latter. Overall, each XM tone seems to have a corresponding tone in SC with which it shares similar tonal contour and

pitch value, resulting in a very interesting systematic mapping pattern between the tonal systems of XM and SC.

Table 2. Paired tones with similar contours from SC and XM.

Tone Pair	Standard Chinese (SC)			Xi'an Mandarin (XM)		
	Tonal category	Pitch value	Example	Tonal category	Pitch value	Example
Level contour	SC_T1	55	ma1/妈	XM_T4	55/44/45	ma4/骂
Rising contour	SC_T2	35	ma2/麻	XM_T2	24	ma2/麻
Low contour	SC_T3	214	ma3/马	XM_T1	21/31	ma1/妈
Falling contour	SC_T4	51	ma4/骂	XM_T3	52/53/42	ma3/马

In fact, the mapping of tonal contours between the tonal systems of XM and SC has been proposed in previous studies. Li (2001) introduced the mapping pattern of XM tones and SC tones (similar as in Table 2) and suggested that XM learners of SC utilize their knowledge of XM tones to produce SC tones. Zhang (2009) also claimed the presence of a comparable tonal contour for each XM tone in SC. She further statistically compared the F0 contour of each XM tone with its SC counterpart of a similar tonal contour. The results showed that although the paired tones were similar in tonal contour, there were detailed acoustic differences. Specifically, Xi'an low-falling tone was different from the citation form SC low-falling-rising tone in contour shape; Xi'an rising tone was significantly lower than SC rising tone in F0 height except in the early-middle part; Xi'an falling tone had a similar initial F0 height with but higher final F0 height than the SC falling tone, and Xi'an level tone had overall lower F0 height than SC level tone. Zhang (2009) has made an attempt to reveal the acoustic similarities and differences between the two tonal systems empirically. However, it did not include details on the design of the production experiment, and therefore it is not clear how the tonal patterns were obtained and the comparison did not seem to be made on comparable datasets. The present study thus decided to compare the acoustic properties of XM tones and SC tones in a pairwise fashion with a more balanced and comparable design.

In addition to establishing the acoustic similarities or differences between the paired tones of XM and SC in production, we were also interested to know

whether each tone pair of similar contours from XM and SC is perceived to be similar or different in tone perception by bi-dialectal speakers of XM and SC. The tone production and perception experiments together were expected to confirm the mapping pattern of the two tonal systems. So far, there has not been any perception study comparing XM tones and SC tones. Conventionally, tone discrimination relies on several perceptual cues, among which the most widely adopted and important perceptual cues have proved to be F0 height and F0 contour, according to previous cross-language studies (Gandour, 1983, 1984; Gandour & Harshman, 1978; Francis, Ciocca, Ma, & Fenn, 2008). The relative importance of these two cues, however, varies among listeners of different language background. SC listeners tended to attach more importance to F0 contour than F0 height, whereas Cantonese and English listeners gave more weight to F0 height than F0 contour (Gandour, 1983; 1984). Apart from the F0-related features, other acoustic properties such as duration, amplitude contour and voice quality have also been shown to serve as secondary cues for tone discrimination, especially when the primary F0 information was not available (Liu & Samuel, 2004; Whalen & Xu, 1992; Yang, 2011). Furthermore, phonological rules might play a role in tone discrimination. For example, SC native listeners found it more difficult to discriminate between the rising tone and the low-falling-rising tone than other tone pairs in SC (Huang, 2007), which was partly attributed to the tone sandhi rule that makes the two tones conditioned allophonic tonal variants. Specifically, the low-falling-rising tone would be realized as a rising tone when it precedes another low-falling-rising tone (Duanmu, 2007). The native phonological rule can, sometimes even affect tone discrimination in a non-native language. For example, Cantonese listeners with or without SC experience had difficulty distinguishing between the SC high-level tone (55) and high-falling tone (51) (Hao, 2012; So & Best, 2010). This is because in Cantonese the high-level tone (55) has a free allophonic tonal variant, high-falling tone (53) (Hashimoto, 1972; Yip, 2002), which shows phonetic similarity to SC high-falling tone. In this study, based on the acoustic results in the tone production experiment, we ran a tone perception experiment to see whether each tone pair would be perceived as similar or different by the

bi-dialectal speakers and how the acoustic difference in each tone pair affects tone perception.

To sum up, in the present study, tonal categories with similar contours from XM and SC were paired as in Table 2. Both tone production and perception experiments were carried out to test whether each pair of tones is acoustically and perceptually similar or different. In Experiment 1, we compared the acoustic properties of the paired SC and XM tones produced by a group of highly proficient bi-dialectal speakers of these two dialects and established the acoustic difference of each tone pair. In Experiment 2, we further investigated whether each tone pair would be perceived as similar or different by the bi-dialectal speakers of SC and XM and how the acoustic difference in each tone pair affects tone perception with a five-scale tone judgment task. Results from both experiments were expected to reveal the tonal similarity and confirm the mapping pattern of the two tonal systems.

4.2 Experiment 1

4.2.1 Method

4.2.1.1 Participants

Thirty bi-dialectal speakers of SC and XM (16 males, 14 females) were selected and paid to participate in the experiment. All the selected participants were of high and comparable proficiency in both dialects, judging from their performance on a story reading task and their self-reported language proficiency through an adapted version of the LEAP-Q questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007). They acquired both dialects before the age of 6 and were early XM_SC bi-dialectal speakers with the first dialect (D1) being either XM or SC. All of them were born and raised in the urban areas of Xi'an and had no living experience outside of Xi'an. They were all undergraduate or graduate students at local universities. Their age ranged from 19 to 28 ($M \pm SD$: 22.5 ± 3.2). None of them had reported any speech or hearing disorders.

4.2.1.2 *Stimuli*

Thirty monosyllabic minimal tone sets with full sets of all four tones were selected. The four monosyllables within one minimal tone set were distinguished merely by tone, with the segments being identical. An exemplar of a full minimal tone set was *ma1*, *ma2*, *ma3* and *ma4*. The complete list can be found in Table C1 (see Appendix C). The monosyllables were selected on the condition that no pronunciation difference existed for the segment of each syllable between SC and XM, to avoid any potential effect of segmental pronunciation difference on tones. The monosyllabic items are frequent monosyllabic words with more than 4,500 occurrences in a corpus of 193 million words (Da, 2004). Within each minimal tone set, the monosyllabic words have comparable word frequencies. In total, 120 monosyllabic words (30 Syllables \times 4 Tones) were selected. Some disyllabic words were added as fillers.

4.2.1.3 *Recording*

The recordings took place in a soundproof booth of the behavioral lab at Shaanxi Normal University in Xi'an. Stimuli were randomly presented to the speakers with E-prime 2.0. Each speaker produced all the items in both SC and XM with no repetition in two separate sessions. The order of the sessions was counterbalanced. Half of the speakers did the SC session first and then the XM session, the other half started with the XM session. Each session included one practice block and three experimental blocks. Between each block, there was a 3-minute break. The practice block contained 8 trials, which were not used in the experimental blocks, to familiarize the participants with the specific language mode. An experimental trial started with a 300 ms fixation cross, followed by a 200 ms pause. After that, a stimulus in the form of simplified Chinese character was presented on the screen. Speakers were requested to produce the stimulus in that particular language of the session in a self-paced fashion. They pressed a button to proceed to the next stimulus when finishing producing the current stimulus. The inter-stimulus interval was 500 ms. Instructions were given to the speakers visually on screen in simplified Chinese characters and orally by the experimenter in that particular language before

each session. All the stimuli were recorded at 16-bit resolution and a sampling rate of 44.1 kHz on a laptop via an external digitizer (UA-G1). Altogether, 240 monosyllabic items (30 Syllables \times 4 Tones \times 2 Languages) were elicited from each of the 30 speakers.

4.2.1.4 Data analysis

The F0 and duration of the speech items were analyzed. All the stimuli were manually annotated in Praat (Boersma & Weenink, 2015). A custom-made script was then used to extract ten equally distanced F0 values from the rhyme part of each time-normalized syllable. Gross errors in F0 extractions were manually corrected afterwards. To eliminate between-speaker acoustic differences, the raw F0 values were transformed to Z-score⁴ for each speaker (Rose, 1987), pooling the SC and XM productions.

Statistical analyses of F0 were carried out using the growth curve analysis (Mirman, 2014) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.1.2 (R Core Team, 2015). The overall F0 curves were modeled with up to second-order orthogonal polynomials, given that the most complex F0 contour in this study has a U-shape curve. Three time terms of the models would be of interest: the intercept, the linear slope and the steepness of the quadratic curvature, which indicate the overall F0 mean, the direction of F0 change such as rising or falling, and the steepness of F0 rising or falling, respectively. If tonal contours under investigation are different, we expect statistical differences in at least one of the three time terms. We built separate models for each pair of tones as listed in Table 2. All the models included random effects of Subjects and Items on all time terms. The fixed effects of Language (XM, SC) on all time terms, as well as random effects of Subjects-by-Language and Items-by-Language on all time terms, were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios.

⁴ We used Z-scores instead of *T*-values to normalize tone values because *T*-values can be easily distorted by extreme values like the maximum F0 or the minimum F0.

Statistical analyses of duration were performed using linear mixed-effects regression models with the package *lme4* (Bates et al., 2015) in R version 3.1.2 (R Core Team, 2015). As for F0, we built separate models for the duration of each pair of tones. All the models first included random intercepts of Subjects and Items. The fixed effect of Language and random slopes of Subjects-by-Language and Items-by-Language were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios.

4.2.2 Results

4.2.2.1 F0

Figure 1 presents the mean F0 (Z-score) contours of the four tones in SC and XM. We report the F0 for each pair of tones of similar tonal contours in what follows.

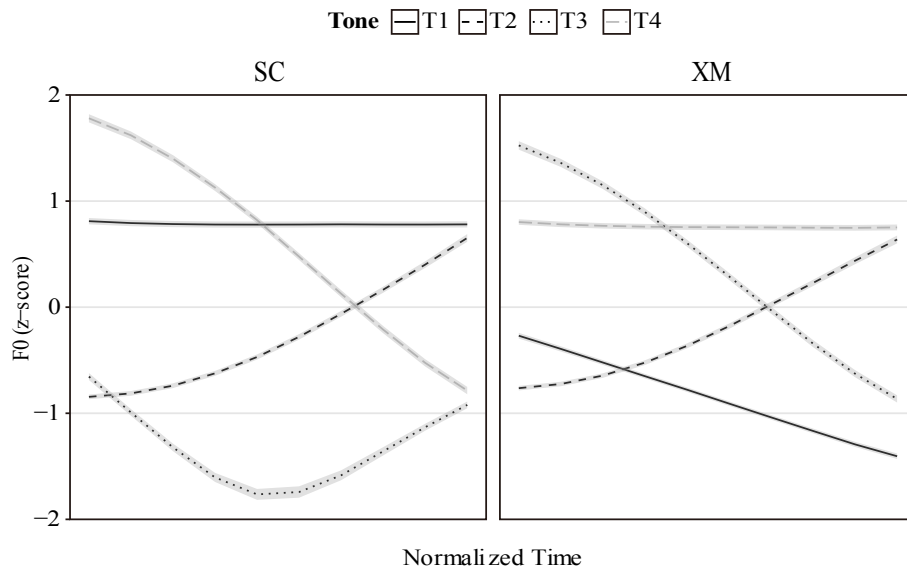


Figure 1. Mean F0 (Z-score) contours of the four tones in SC and XM. The F0 values of each tone were averaged over 30 speakers and 30 monosyllabic items, with the tone of each item represented by ten equally distanced F0 values taken from the rhyme part of the time-normalized item. Grey areas indicate the 95% confidence interval of the corresponding mean.

1) Level contour: SC_T1 vs. XM_T4

Results showed that the effect of Language on the intercept did not improve model fit ($\chi^2(1) = 0.03, p = 0.87$), nor did the effect of Language on the linear term ($\chi^2(1) = 1.54, p = 0.21$) and the effect of Language on the quadratic term ($\chi^2(1) = 0.23, p = 0.63$). Overall, it seems that the F0 contours of SC_T1 and XM_T4 did not differ from each other.

2) Rising contour: SC_T2 vs. XM_T2

Results showed that the effect of Language on the intercept did not improve model fit ($\chi^2(1) = 3.17, p = 0.07$), nor did the effect of Language on the linear term ($\chi^2(1) = 2.73, p = 0.10$). The effect of Language on the quadratic term, however, did improve model fit ($\chi^2(1) = 14.05, p < 0.001$). Therefore, the overall F0 mean and the direction of F0 rising did not differ between the two rising tones from SC and XM, whereas their steepness of rising differed, with XM_T2 having a shallower curvature than SC_T2 ($\beta = -0.08, t = -4.01, p < 0.001$).

3) Low contour: SC_T3 vs. XM_T1

The analyses of SC_T3 and XM_T1 data showed that the effect of Language on the intercept significantly improved model fit ($\chi^2(1) = 58.33, p < 0.001$), as well as the effect of Language on the linear term ($\chi^2(1) = 27.89, p < 0.001$) and the effect of Language on the quadratic term ($\chi^2(1) = 36.00, p < 0.001$). Apparently, SC_T3 was different from XM_T1 in all three time terms. The overall F0 mean of XM_T1 was significantly higher than SC_T3 ($\beta = 0.46, t = 9.13, p < 0.001$). The direction of F0 change was also different between the two tones ($\beta = -0.96, t = -14.06, p < 0.001$), with SC_T3 having a falling-rising contour and XA_T1 having a low-falling contour without the rising tail. Moreover, the F0 curvature of XA_T1 was shallower than that of SC_T3 ($\beta = -1.08, t = -14.23, p < 0.001$).

4) Falling contour: SC_T4 vs. XM_T3

The analyses of SC_T4 and XM_T3 showed that there was a significant effect of Language on the intercept ($\chi^2(1) = 9.06, p = 0.003$). However, no Language effect on the linear term ($\chi^2(1) = 0.50, p = 0.48$) or the quadratic term ($\chi^2(1) =$

0.83, $p = 0.36$) was found. Evidently, the overall F0 mean of XM_T3 was lower than that of SC_T4 ($\beta = -0.19$, $t = -3.08$, $p = 0.002$). The direction of F0 falling and the steepness of F0 falling were not significantly different between the two falling tones.

4.2.2.2 Duration

Figure 2 presents the mean durations of the four tones in SC and XM. The following reports the duration results for each pair of tones of similar tonal contours.

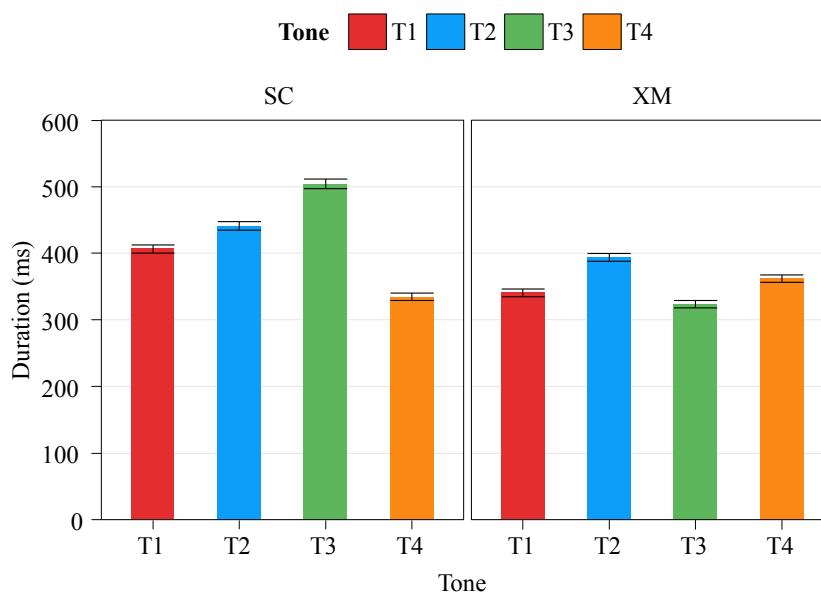


Figure 2. Mean durations with 95% confidence interval of the four tones in SC and XM.

1) Level contour: SC_T1 vs. XM_T4

There was a significant main effect of Language ($\chi^2(1) = 10.91$, $p < 0.001$). SC_T1 was significantly longer (45.09 ms) than XM_T4.

2) Rising contour: SC_T2 vs. XM_T2

The effect of Language significantly improved model fit ($\chi^2(1) = 425.36$, $p < 0.001$). SC_T2 was 47.60 ms longer than XM_T2.

3) Low contour: SC_T3 vs. XM_T1

Not surprisingly, a significant main effect of Language was also found for the durations of this tone pair ($\chi^2(1) = 71.33, p < 0.001$). SC_T3 was considerably longer than XM_T1, with the duration difference reaching up to 166.29 ms.

4) Falling contour: SC_T4 vs. XM_T3

An investigation of the durations of the tone pair of the falling contour revealed no effect of Language ($\chi^2(1) = 0.64, p = 0.42$), indicating that there was no duration difference between SC_T4 and XM_T3.

From the above comparisons of F0 and duration for each pair of tones of similar tonal contours, the acoustic patterns of each tone pair can be summarized as follows. First, the tone pair of level contour did not show any difference in their F0. However, the duration of the tone of level contour in SC was significantly longer than that of its counterpart in XM. Second, the overall F0 mean and the direction of F0 change did not differ between the two tones of rising contour in SC and XM, despite a shallow curvature of the rising F0 contour in XM_T2 relative to SC_T2. In addition, the duration of XM_T2 was considerably shorter than that of SC_T2. Third, the two tones of low contour in SC and XM were significantly different from each other regarding the overall F0 mean, the direction of F0 change and the steepness of F0 change. In fact, their contour shape differed, with the SC tone having a low-falling-rising contour and the XM tone having a low-falling contour without the rising tail. The former also tended to be remarkably longer than the latter. Fourth, having almost parallel F0 contours, the two tones of high-falling contour in SC and XM revealed difference in the overall F0 mean, with an overall higher F0 contour of SC_T4 compared to XM_T3. Their duration nevertheless did not differ.

4.3 Experiment 2

Having established the acoustic difference of each tone pair, the question arises as to whether the acoustic difference in each tone pair is perceivable. In fact, the two tonal systems provide an interesting test case for us to look into the relationship of the production and perception of tones by the bi-dialectal tonal

language speakers. As shown in Experiment 1, the acoustic difference of each tone pair ranged from no F0 difference (level tone pair) through F0 curvature difference (rising tone pair) to F0 height difference (falling tone pair) and F0 contour difference (low tone pair). With this setup of the two tonal systems, we could investigate how different F0 dimensions affect tone perception of the bi-dialectal tone language speakers. In this session, a five-scale tone judgment task was adopted to examine whether the acoustic difference in each tone pair is perceivable. We then compared the tone perception results of different tone pairs to reveal how different F0 dimensions affect tone perception.

4.3.1 Method

4.3.1.1 Participants

Another set of thirty XM_SC bi-dialectal speakers (11 males, 19 females) with high proficiency in both dialects were selected and paid to participate in the experiment. All the selected participants acquired both dialects before the age of 6 and were early XM_SC bi-dialectal speakers with a D1 of either XM or SC. They were born and raised in the urban areas of Xi'an and had no living experience outside of Xi'an. All were undergraduate or graduate students at local universities, between 18 and 29 years old ($M \pm SD$: 21.2 ± 2.6). None of them had reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

4.3.1.2 Stimuli

The four pairs of tones in Experiment 1 were used for tone judgment. Since the two tones in each tone pair were similar in tonal contour overall, two pairs of tones of distinct tonal contours were added as fillers to avoid potential response strategies. Each tone pair was tested with all the 30 root monosyllables in Experiment 1, resulting in 30 trials. The two speech items in a trial always share the same segment so that participants could focus on the tone judgment. For example, a SC_T1 monosyllable (“妈”, ma1, *mother*) was paired with its corresponding XM_T4 monosyllable (“骂”, ma4, *scold*).

Four speakers were recruited to record the stimuli for the perception experiment. They were all university students and in their 20s. Two native speakers (one male, one female) of SC, who were born and raised in Beijing and had no knowledge of any other dialects, recorded the SC monosyllabic sounds. Likewise, two native speakers (one male, one female) of XM, who were born and raised in the urban area of Xi'an and had no living experience outside of Xi'an, recorded the XM monosyllabic sounds. Note that as it is impossible to find monolingual XM speakers nowadays, the two native speakers of XM also speak SC fluently. The recordings took place in Beijing for the Beijing speakers and in Xi'an for the Xi'an speakers. All the speech items were recorded at 16-bit resolution and a sampling rate of 44.1 KHz.

To ensure that the two XM speakers' tone patterns were representative enough of XM, we compared the acoustic properties of their tone patterns with those of the XM tone patterns in Experiment 1 (see Figures 3 and 4) and found no statistical difference in F0 and duration. It was therefore confirmed that the two XM speakers' production of XM tone patterns were representative patterns of XM and suitable for the perception study. We also compared the acoustic properties of the two SC speakers' tone patterns with those of the SC tone patterns in Experiment 1 (see Figures 5 and 6) and did not find statistical difference in F0 and duration either.

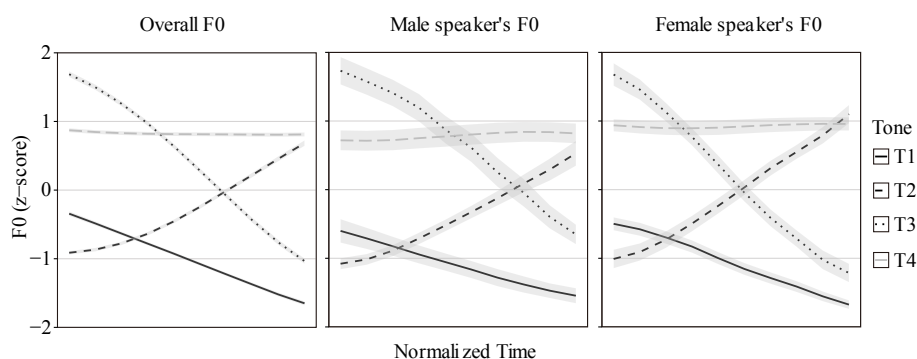


Figure 3. Mean F0 (Z-score) contours of the four tones in XM by the 30 SC and XM bi-dialectal speakers in Experiment 1 (left panel), the male Xi'an speaker (middle panel) and the female Xi'an speaker (right panel) in Experiment 2. Grey areas indicate 95% confidence interval of the corresponding mean.

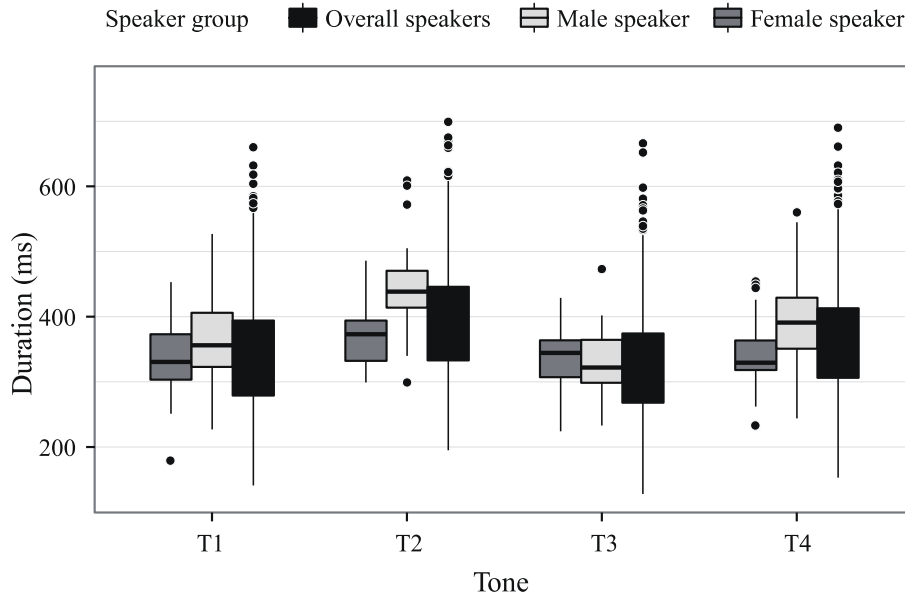


Figure 4. Mean durations with 95% confidence interval of the four tones in XM by the 30 SC and XM bi-dialectal speakers in Experiment 1 (black boxes), the male Xi'an speaker (light grey boxes) and the female Xi'an speaker (dark grey boxes) in Experiment 2.

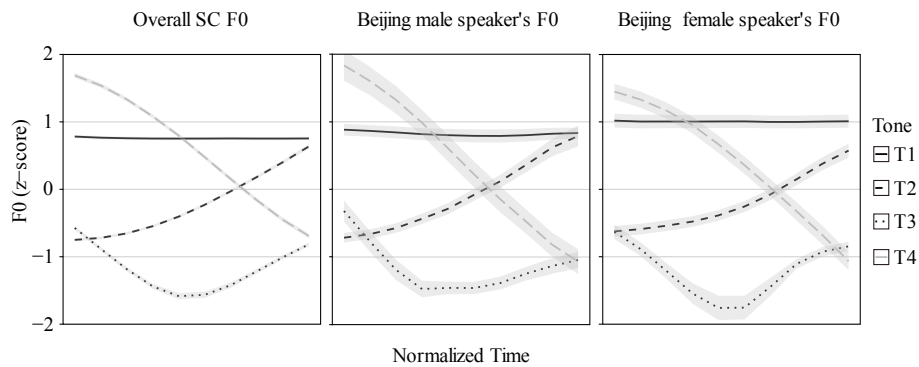


Figure 5. Mean F0 (Z-score) contours of the four tones in SC by the 30 SC and XM bi-dialectal speakers in Experiment 1 (left panel), the male Beijing speaker (middle panel) and the female Beijing speaker (right panel) in Experiment 2. Grey areas indicate 95% confidence interval of the corresponding mean.

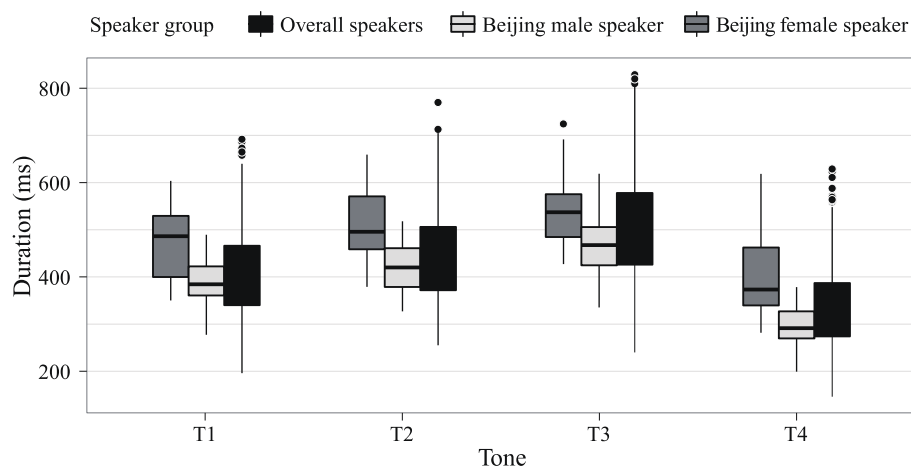


Figure 6. Mean durations with 95% confidence interval of the four tones in SC by the 30 SC and XM bi-dialectal speakers in Experiment 1 (black boxes), the male Beijing speaker (light grey boxes) and the female Beijing speaker (dark grey boxes) in Experiment 2.

After normalizing the amplitude of all the speech items in Praat (Boersma & Weenink, 2015), we paired the Beijing female speaker's speech items with the Xi'an female speaker's corresponding speech items according to tone pairs. The same was done for the two male speakers' speech items. Instead of recording all the speech items by a XM_SC bi-dialectal speaker, we recorded the SC speech items by native speakers of SC and the XM speech items by native speakers of XM. This ensured more typical realizations of SC and XM tones. The inclusion of two groups of speakers could avoid potential speaker bias.

4.3.1.3 Procedure

Participants were tested individually in a soundproof booth of the behavioral lab at Shaanxi Normal University in Xi'an. All the trials (30 Syllables \times 6 Tone pairs \times 2 Speaker groups) were randomly presented to the participants using the E-Prime 2.0 software through headphones at a comfortable listening level.

The experiment included a practice block and four experimental blocks. The practice block contained 6 trials, which were not used in the experimental blocks. Each experimental block contained 90 trials. Between every second block, there was a 3-minute break. An experimental trial started with a 100 ms

warning beep, followed by a 300 ms pause. The first speech item was then presented. After a 600 ms pause, the second speech item was presented. The language order of the two speech items in a trial was counterbalanced for each speaker group of the trials. Half of the trials presented the SC item before its corresponding XM item, while the other half presented the SC item after its corresponding XM item. Participants were requested to judge the similarities of the two tones of the two speech items in a trial on a five-point scale, with 1 indicating “completely different” and 5 indicating “completely the same”. Response accuracy rather than speed was stressed. However, if participants did not make any response from the onset of the second stimulus to 2.5 s after the offset of the second stimulus, the program moved on to the next trial automatically with an inter-trial interval of 500 ms. Instructions were given both visually on screen and orally by the experimenter in SC before the experiment. To eliminate any influence of top-down knowledge on tone judgment, we did not mention the source languages of the auditory stimuli to the participants in the instruction.

4.3.1.4 *Data analysis*

To decide whether each pair of tones was perceived as similar or different, we analyzed the frequency distribution of the responses with chi-square goodness-of-fit test. The observed frequency distribution of the responses was first compared with the expected frequency distribution (null hypothesis: equal proportions) for each tone pair. If the null hypothesis of equal proportions was rejected, the individual response category’s contribution to the overall chi-square statistic was determined by calculating the square of the difference between the observed and expected frequencies for a category, divided by the expected frequency for that category. Generally speaking, categories with a larger difference between the observed and expected frequencies make a larger contribution to the overall chi-square statistic. After recognizing the response category that contributed the most to the overall chi-square statistic, we further conducted several pair-wise goodness-of-fit tests to compare this category’s frequency with that of the other categories. If all the comparisons are statistically significant (p -value adjusted), the category would be considered as

the best indicator of the similarity/difference of the two tones under investigation.

The second analysis concerned how the varying acoustic differences of different tone pairs affect tone perception of XM_SC bi-dialectal speakers. All the four pairs of tones were merged into one dataset and the tone perception results of different tone pairs were compared. Statistical analyses were carried out with the package *ordinal* (Christensen, 2015) in R version 3.1.2 (R Core Team, 2015). Cumulative link mixed models (CLMMs) were constructed for the dependent variable Response (1, 2, 3, 4, 5) with Tone pair (Level, Rising, Low, Falling), Language order (XM before SC; SC before XM), Speaker group (Female, Male), Listener gender (Female, male) and their interactions as fixed factors, and Subjects and Items as random factors. The fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. Post-hoc pairwise comparisons between different tone pairs were conducted using the *lsmeans* package (Lenth, 2016) with single-step *p*-value adjustment.

4.3.2 Results

4.3.2.1 Level contour: SC_T1 vs. XM_T4

The Chi-square goodness-of-fit test showed that the responses were clearly not equally distributed ($\chi^2(4) = 5634.42, p < 0.001$). As can be seen from Table 3, the response category “5” contributed the most to the overall chi-square statistic. Pairwise comparisons showed that the frequency of the response category “5” was significantly higher than that of the other categories (all *ps* < 0.001), indicating that SC_T1 and XM_T4 were mostly judged as 5, i.e., completely the same.

4.3.2.2 Rising contour: SC_T2 vs. XM_T2

The Chi-square goodness-of-fit test showed that the responses were clearly not equally distributed ($\chi^2(4) = 5677.38, p < 0.001$). Again, the response category “5” contributed the most to the overall chi-square statistic (see Table 3). Further pairwise comparisons showed that the frequency of the response category “5”

was significantly higher than that of the other categories (all $ps < 0.001$), indicating that SC_T2 and XM_T2 were mostly judged as 5, i.e., completely the same.

Table 3. *Response counts for each tone pair.*

Tone pair	Measure	Response category				
		1	2	3	4	5
(Level) SC_T1 vs. XM_T4	Observed count	68	22	16	59	1631
	Expected count	359.2	359.2	359.2	359.2	359.2
	Contribution to Chi-Sq	236.1	316.5	327.9	250.9	4503.0
(Rising) SC_T2 vs. XM_T2	Observed count	55	17	21	68	1637
	Expected count	359.6	359.6	359.6	359.6	359.6
	Contribution to Chi-Sq	258.0	326.4	318.8	236.5	4537.7
(Low) SC_T3 vs. XM_T1	Observed count	625	165	102	149	754
	Expected count	359	359	359	359	359
	Contribution to Chi-Sq	197.1	104.8	184.0	122.8	434.6
(Falling) SC_T4 vs. XM_T3	Observed count	116	26	23	86	1539
	Expected count	358	358	358	358	358
	Contribution to Chi-Sq	163.6	307.9	313.5	206.7	3896.0

Note. 1 = “completely different”; 5 = “completely the same”.

4.3.2.3 *Low contour: SC_T3 vs. XM_T1*

The Chi-square goodness-of-fit test showed that the responses were not equally distributed ($\chi^2(4) = 1043.36, p < 0.001$). As shown in Table 3, the response category “5” contributed the most to the overall chi-square statistic. However, the response category “1” also made a relatively large contribution to the overall chi-square statistic. Pairwise comparisons showed that the frequencies of the response categories “1” and “5” were significantly higher than those of the rest categories (all $ps < 0.001$). Moreover, the frequency of the response category “5” was higher than that of the response category “1” ($\chi^2(1) = 12.07, p = 0.0005$). Overall, participants were more likely to perceive SC_T3 and XM_T1 as the same tone, though they also gave slightly fewer but a comparable number of different responses.

4.3.2.4 *Falling contour: SC_T4 vs. XM_T3*

The Chi-square goodness-of-fit test showed that the responses were clearly not equally distributed ($\chi^2(4) = 4887.59, p < 0.001$). The response category “5” contributed the most to the overall chi-square statistic, as demonstrated in Table 3. Pairwise comparisons showed that the frequency of the response category “5” was significantly higher than that of the other categories (all p s < 0.001), indicating that SC_T4 and XM_T3 were mostly judged as 5, i.e., completely the same.

To summarize, the five-scale tone judgment results showed that the tone pair of level contour (SC_T1 and XM_T4) was mostly judged as the same by the XM_SC bi-dialectal speakers. Similarly, the tone pair of rising contour (SC_T2 and XM_T2) and the tone pair of falling contour (SC_T4 and XM_T3) were mostly judged as the same. Different was the tone pair of low contour (SC_T3 and XM_T1), which elicited a comparable number of same and different responses, though the two were statistically different. It seems that participants had a much harder time discriminating between the two tones of low contour in SC and XM.

4.3.2.5 *Comparisons among the four tone pairs*

Statistical results for the models of Response showed a significant main effect of Tone pair ($\chi^2(3) = 253.69, p < 0.001$), indicating that the rating tendency differed significantly among the four tone pairs (see Table 3). There was also a significant main effect of Language order ($\chi^2(1) = 22.61, p < 0.001$) and a significant main effect of Speaker group ($\chi^2(1) = 26.76, p < 0.001$). No effect of Listener gender or interaction effect of the above factors was found (all p s > 0.05). Specifically, when a XM tone was presented before its corresponding SC tone, listeners were more likely to rate higher, i.e., more likely to judge the two tones as being more alike ($\beta = 0.35, \zeta = 5.08, p < .001$) compared to when a SC tone was presented before a XM tone. Likewise, listeners tended to rate higher for the male speaker group’s speech than for the female speaker group’s speech ($\beta = 0.37, \zeta = 5.24, p < .001$).

Post-hoc pairwise comparisons showed that the rating tendency of the tone pair of level contour was not significantly different from that of the tone pair of rising contour ($\beta = -0.002$, $\chi = -0.02$, $p = 1.00$). Both pairs were mostly judged as the same. Their rating tendencies, however, were significantly different from the tone pair of low contour (level vs. low: $\beta = 1.87$, $\chi = 18.95$, $p < 0.001$; rising vs. low: $\beta = 1.87$, $\chi = 19.11$, $p < 0.001$) and the tone pair of falling contour (level vs. falling: $\beta = 0.31$, $\chi = 3.08$, $p = 0.01$; rising vs. falling: $\beta = 0.32$, $\chi = 3.11$, $p = 0.01$). The rating tendencies of the tone pair of low contour and the tone pair of falling contour also showed significant difference ($\beta = -1.55$, $\chi = -16.43$, $p < 0.001$). In summary, the rating tendency of the tone pair of low contour was significantly different from that of the other three tone pairs, with the former being judged as either different or the same (there were slightly more “completely the same” responses than “completely different” responses), whereas the latter three tone pairs were mostly judged as the same, though the tone pair of falling contour elicited more different responses than the tone pairs of level contour and rising contour.

4.4 General discussion

The present study investigated the phonological similarity in tones of two closely related Mandarin dialects, SC and XM. Tones with similar contours from SC and XM were paired and their acoustic properties were compared over properly-controlled large samples produced by a group of highly proficient bi-dialectal speakers of XM and SC. F0 results of the four tone pairs ranged from no F0 difference (level contour tone pair) through F0 curvature difference (rising contour tone pair) to F0 height difference (falling contour tone pair) and F0 contour difference (low contour tone pair). Except the falling contour tone pair, all the other tone pairs also exhibited difference in tone duration and the largest duration difference was found in the low contour tone pair. These tone pairs of varying acoustic differences were then presented to the bi-dialectal speakers of XM and SC for tone perception with a five-scale tone judgment task. Results showed that the rating tendency of the tone pair of low contour was significantly different from that of the other three tone pairs, with the former being judged as either different or the same (there were slightly more

“completely the same” responses than “completely different” responses), whereas the latter three tone pairs were mostly judged as the same, though the tone pair of falling contour elicited more different responses than the tone pairs of level contour and rising contour.

With a balanced comparable design, the present production and perception experiments empirically confirmed the systematic tonal mapping pattern between XM and SC proposed in Li (2001) and Zhang (2009). While there were detailed acoustic differences in tone production, tones with similar contours between the two dialects were basically perceived to be the same, resulting in mapped tone pairs of level contour (SC_T1 vs. XM_T4), of rising contour (SC_T2 vs. XM_T2) and of falling contour (SC_T4 vs. XM_T3). Despite having distinct surface tonal contours, the tone pair of low contour (SC_T3 vs. XM_T1) also showed mapping, though to a lesser degree compared to the other three tone pairs.

The mapping pattern of XM tones and SC tones was initially put forward based on the similarity of the tonal contour and pitch value of XM tones and SC tones represented on the 5-point scale notation system (Chao, 1930; 1968). Different from the established pitch value of SC tones, there have been variances as to the specific pitch value of XM tones in previous studies (e.g., Bai, 1954; Luo & Wang, 1981; Ma, 2005; Peking University, 2003; Ren, 2012; Sun, 2007; Wang, 1996; Yuan, 1989; Zhang & Shi, 2009). However, the basic tonal contour shape of each tone was largely consistent across studies, and it has been noted that each XM tone has a mapped tone in SC with which it shares similar tonal contour and pitch value (Li, 2001; Zhang, 2009). Zhang (2009) tested the mapping pattern of the two tonal systems in tone production, but the tonal comparisons were not made on comparable datasets. The present study thus made more of an effort to empirically test the mapping pattern of the two tonal systems in tone production with more balanced comparable design. Tonal comparisons were made on paired tones of similar tonal contours from the two dialects produced by highly proficient bi-dialectal speakers of SC and XM. Our acoustic results showed that except for the tone pair of level contour, all the other tone pairs showed difference in F0. Specifically, the XM rising tone had a shallower rising F0 curvature than the SC rising tone, with an

overall comparable F0 mean. The XM falling tone had an overall lower F0 height than the SC falling tone. The XM low tone, not surprisingly, had a different F0 contour from the SC low-falling-rising tone. Our results of the specific F0 difference for each tone pair, except for the tone pair of low contour, was different from that in Zhang (2009), showing that the manipulation of a comparable design in this study actually resulted in different tonal realizations. It is therefore important to test on comparable datasets in such cross-dialect investigations. Nevertheless, both studies showed an overall compact tonal space of XM tones than SC tones. Apart from F0 difference, we also found duration difference for each tone pair except for the one of falling contour. All XM tones other than the falling tone tended to be shorter than their respective SC counterparts. Overall, there were acoustic differences for each pair of tones with similar contours from SC and XM.

The acoustically different tone pairs, however, were mostly perceived to be the same, or at least, very similar by the bi-dialectal speakers of SC and XM. In a five-scale tone judgment task, the tone pairs of level contour, rising contour and falling contour were mostly perceived to be completely the same, and the tone pair of low contour was also slightly more likely to be perceived as the same than different. Overall, these tone pairs of similar tonal contours from the two dialects were basically treated as the same during tone perception, despite the presence of acoustic differences. The results of our tone perception experiment confirmed the mapping pattern of XM tones and SC tones proposed in Li (2001) and Zhang (2009), providing new empirical evidence for the mapping of the two tonal systems from a perceptual point of view. Moreover, the mapping pattern seems to be more pronounced in tone perception than in tone production, given that each mapped tone pair was almost perceptually indistinguishable while having acoustically detectable differences.

That the tone pairs of similar tonal contours from XM and SC were basically perceived to be the same by the bi-dialectal speakers does not mean that the participants did not pick up the acoustic differences at all. The mapped tone pairs did vary in the degree to which they were perceived as the same tones. The tone pairs of level contour and rising contour were mostly judged as the

same by the bi-dialectal speakers of XM and SC, followed by the tone pair of falling contour. The tone pair of low contour elicited more different responses relative to the other three tone pairs. As only acoustic information was available to the participants during tone judgment, it is reasonable to assume that the different perceptual results for all the tone pairs resulted from their acoustic differences, in some way. Our acoustic analyses demonstrated that the four tone pairs showed variance in different F0 dimensions, ranging from no F0 difference (level contour tone pair) through F0 curvature difference (rising contour tone pair) to F0 height difference (falling contour tone pair) and F0 contour difference (low contour tone pair). This varying acoustic difference in different F0 dimensions seems to have affected the tone perception results of each tone pair to varying degrees. Compared to the level contour tone pair with no F0 difference, the rising contour tone pair with F0 curvature difference was not perceived any different, seemingly indicating that the bi-dialectal speakers of SC and XM were not sensitive to the F0 curvature difference between the two rising tones. This is not surprising, as F0 curvature has not been identified as a strong perceptual cue for tone discrimination. In contrast, the tone pair of falling contour with F0 height difference was perceived to be less similar than the tone pair of level contour with no F0 difference, suggesting that F0 height difference contributed to the discrimination of the two falling tones. This is consistent with the previous cross-language finding that F0 height is an important perceptual cue for tone discrimination (Gandour, 1983, 1984; Gandour & Harshman, 1978; Francis, Ciocca, Ma, & Fenn, 2008). Lastly, the tone pair of low contour with F0 contour difference was perceived to be much more different than the tone pair of level contour with no F0 difference, as well as than the tone pair of falling contour with F0 height difference. Obviously, F0 contour difference significantly affected the discrimination between the two low tones. Also, the bi-dialectal speakers tended to be more sensitive to the dimension of F0 contour than F0 height in tone discrimination, as has been found by Gandour (1983, 1984) for SC speakers.

Note that although the duration property of each tone pair was maintained in the speech stimuli, participants did not seem to make fully use of it in tone perception, if they used it at all. Duration difference was found in all the tone

pairs except in the falling tone pair. If the participants did use the duration cue for tone perception, with a duration difference of nearly 50 ms, the tone pair of level contour as well as the tone pair of rising contour should have been judged as different tones rather than similar tones. If this is not convincing, a duration difference of about 166 ms in the tone pair of low contour should be certainly salient enough to rule out the possibility that the two tones were judged as similar. However, the pair of low contour tones ended up eliciting even slightly more same responses than different responses. Overall, duration was not adopted as a valid perceptual cue for tone discrimination by the bi-dialectal speakers of SC and XM. They relied primarily on F0 information to make tone judgments.

Acoustic information, especially F0 information, is not the only perceptual cue that listeners employ during tone discrimination. Phonological rules can sometimes play a role in the process, too (Hao, 2012; Huang, 2007; So, & Best, 2010). In this study, the tone pair of low contour (SC_T3 vs. XM_T1) had distinct F0 contours. SC_T3 has a low-falling-rising contour and XM_T1 has a low-falling contour. If participants made tone perceptions purely based on acoustic information, the two low tones would have been judged as different. Instead, the two tones were perceived as either different or similar, with even slightly more same responses than different responses. This could presumably be attributed to the phonological rule of SC_T3. SC_T3 has a low-falling-rising contour when it is in citation form or in the final position of an utterance. When placed before other tones or in a context, it loses its rising tail and becomes a low-falling contour (Dow, 1972; White, 1980), which shows phonetic similarity to the XM low-falling tone. Participants seem to have applied the phonological rule of SC_T3 and used the context form SC_T3 to mediate between the citation form SC_T3 and XM_T1, and therefore classified the citation form SC_T3 and XM_T1 as similar tones. Recall that we did not mention the source languages of the speech stimuli to the participants. It is therefore not clear where and how the context form SC_T3 came into play. There might be two scenarios. One is that XM_T1 here were considered as a representation of the context form SC_T3. Participants then made a comparison between the citation form SC_T3 with the context form SC_T3,

which were judged as similar (citation form SC_T3 vs. (XM_T1 → context form SC_T3)). The other scenario is that when presented with the citation form SC_T3 and XM_T1, participants activated the corresponding context form SC_T3, and they compared the context form SC_T3 with XM_T1, the latter being considered either as a representation of XM_T1 or context form SC_T3, resulting in similar response ((citation form SC_T3 → context form SC_T3) vs. XM_T1). In either scenario, the tone pair of low contour should be judged as similar tones as a result of the phonological rule. In our result, the two low tones from SC and XM elicited a comparable number of same and different responses, suggesting that both the acoustic information and phonological rule played roles in the tone discrimination process, and the two effects seem to counterbalance each other.

4.5 Conclusion

To conclude, the present study investigated the phonological similarity in tones of two closely related Mandarin dialects, SC and XM. Through production and perception experiments, it was established that there is systematic mapping of tones between XM and SC. The degree of the similarity of the mapped tone pair in tone perception was largely dependent on the acoustic phonetic similarity between the tones in tone production, with the phonological rule playing a role in certain circumstance.

Chapter 5

**Effects of cross-dialect phonological
similarity in segment and tone on
bi-dialectal auditory word recognition:
Evidence from Xi'an Mandarin and
Standard Chinese**

Abstract

The present study examined if and how cross-dialect phonological similarity in segment and tone affects bi-dialectal listeners' lexical access during spoken word recognition. Balanced bi-dialectal speakers of Xi'an Mandarin and Standard Chinese took part in an auditory-auditory priming experiment with a generalized lexical decision task. The primes were monosyllabic homophones from either Xi'an Mandarin or Standard Chinese while the targets were disyllabic Xi'an Mandarin or Standard Chinese words. Primes and the first syllable of the target words overlapped in both segment and tone within a dialect (identical) or across two dialects (interdialectal homophones), or they overlapped in segment only within a dialect or across two dialects. In addition, a control condition was included where primes and targets shared neither tone nor segment. Results showed that Standard Chinese primes did not yield significant priming effects for within- or cross-dialect segment-only overlap targets. Standard Chinese primes did not produce significant priming effects for within-dialect identical targets either. However, they did yield significant inhibitory priming effects for cross-dialect homophone targets. This overall pattern was reversed for Xi'an Mandarin primes because these primes were not treated differently from their interdialectal homophonous primes in the current mixed dialect setting. These results suggest that cross-dialect phonological similarity in segment alone does not affect lexical access in bi-dialectal auditory word recognition while cross-dialect phonological similarity in both segment and tone poses a threat to the recognition system of bi-dialectal listeners. We conclude that tonal information plays a significant role in constraining word activation in bi-dialectal auditory word recognition.

Keywords: bi-dialect, Standard Chinese, Xi'an Mandarin, segment, tone, cross-dialect homophones, auditory word recognition

5.1 Introduction

Spoken word recognition is a key aspect of language comprehension. To arrive at the correct recognition of a spoken word, listeners constantly map the incoming speech signal onto possible lexical representations in their mental lexicon until the best-matched candidate is found (Luce & Pisoni, 1998; Marslen-Wilson, 1987; McClelland & Elman, 1986; Norris, 1994). This process usually takes place in just a few hundred milliseconds and seems effortless. However, when the incoming speech signal becomes ambiguous to the listener, for example when there are phonologically similar words to the intended target within one language system or across two language systems mastered by the listener, the recognition system can be challenged (Dijkstra, Grainger, & Van Heuven, 1999) and additional processing costs might be incurred. How an ambiguous speech input due to phonological similarities across two closely related linguistic systems affects spoken word recognition is the focus of interest in this study.

An extreme case of phonological similarity within a language system is homophony. Intralingual homophones have been found to be more difficult to process than non-homophone controls in tasks such as sentence verification (Coltheart, Avons, Masterson, & Laxon, 1991), semantic categorization (Van Orden, 1987), proofreading (Daneman & Stainton, 1991), eye movements (Daneman, Reingold, & Davidson, 1995; Jared, Levy, & Rayner, 1999) and lexical decision (Ferrand & Grainger, 2003; Newman, 2012; Pexman, Lupker, & Jared, 2001; Unsworth & Pexman, 2003). These results have been taken as evidence for parallel activation of multiple lexical candidates within a language.

Studies on homophones across two language systems are mainly concerned with bilingual word recognition. Interlingual homophones have been found to be processed much more slowly and less accurately compared to monolingual controls by bilingual speakers in tasks such as language-specific/generalized lexical decision (Dijkstra et al., 1999; Doctor & Klein, 1992; Lagrou, Hartsuiker, & Duyck, 2011; Nas, 1983), gating (Grosjean, 1988; Schulpen, Dijkstra, Schriefers, & Hasper, 2003) and semantic-relatedness decision (Luo, Johnson, & Gallo, 1998; Ota, Hartsuiker, & Haywood, 2009). This holds across both the visual (Dijkstra et al., 1999) and the auditory modalities (Lagrou et al., 2011). In

cross-modal form priming tasks, bilinguals also showed longer visual lexical decision times for targets preceded by the auditory interlingual homophone primes than those by monolingual control primes (Schulpen et al., 2003). Overall, these studies suggest parallel activation of homophone candidates from both languages and an interference effect of cross-language phonological similarity on word recognition.

Most studies on phonological similarity across two language systems have been conducted in Indo-European languages with an alphabetic writing system, such as Dutch-English. While there are interlingual homophones across these languages in the visual domain, within the auditory domain many of them differ in subphonemic features in their actual pronunciation. This difference might partly account for the neglected role of phonology in bilingual studies (Dijkstra et al., 1999).

About 70% of the world languages are tonal languages (Yip, 2002). Some of these languages differentiate lexical meanings via pitch changes. For example, in Standard Chinese, the same segment *ma* can mean *mother*, *hemp*, *horse* and *scold* with different pitch contours, known as lexical tones. Little, however, do we know about how lexical tones constrain auditory word recognition in bilingual tonal language speakers, when interlingual phonological similarities can be due to overlap in segment and/or tone.

It is also worth noting that under the cover term Chinese, there are many language varieties spoken in China, which share an abundant number of homophones. The majority of speakers are proficient in at least two varieties: the national language Standard Chinese and their regional native dialect. Little attention has been paid to the possible co-activation of homophones in bi-dialectal word recognition. The question that arises is whether in bi-dialectal lexical processing, homophones co-activate and interfere, as in the bilingual situation. Furthermore, for tonal language speakers, what role does tone play in the activation and processing of bi-dialectal lexical representations during spoken word recognition?

Studies on the role of tone in lexical processing have mostly been conducted on Chinese varieties (mostly Standard Chinese and Cantonese) in a monolingual context. The general consensus is that tonal information is used in recognition

(Ching, 1985; Fox & Unkefer, 1985). However, in sublexical tasks such as homophone decision (Taft & Chen, 1992), vowel-tone monitoring (Ye & Connine, 1999), same-different judgment and non-word identification (Cutler & Chen, 1997), the processing of tonal information has been shown to be more error-prone than segmental information. It was therefore concluded that tone plays a weaker role in word recognition than segment at the pre-lexical processing stage. In contrast, an argument for the relatively strong role of tone at the lexical processing stage, comparable to that of the segment, has gained support from lexical tasks in an ERP study (Schirmer, Tang, Penney, Gunter, & Chen, 2005) and an eye-tracking study (Malins & Joanisse, 2010).

An auditory-auditory priming paradigm has also been adopted to investigate the role of tone and segment in word recognition at the lexical processing stage. Contradictory results were obtained as to whether tonal information constrains lexical activation. Lee (2007) investigated monosyllabic word recognition in Standard Chinese. Primes and targets overlapped only in segment (e.g., *lou3-lou2*), only in tone (e.g., *cang2-lou2*), both in segment and tone (e.g., *lou2-lou2*), or – in the baseline condition – neither in segment nor tone (e.g., *pan1-lou2*). Standard Chinese listeners were asked to make lexical decisions over the targets. A facilitatory priming effect was only found when primes and targets overlapped in both segment and tone. Segment-only overlap (minimal tone pair) or tone-only overlap did not produce any priming effect, comparable to the baseline condition. Lee (2007) interpreted the absence of priming in the minimal tone pair in Standard Chinese as the use of tonal information to constrain lexical activation. Sereno and Lee (2015), however, raised the concern that Lee (2007) did not control for the tonal similarity of the prime-target pairs. They conducted a follow-up study with balanced tonal distribution in the prime-target pairs and replicated the identity priming effect in Lee (2007) for the segment and tone overlap condition. In addition, they found a segment-only overlap facilitation effect, though smaller than the identity priming effect. Tone-only overlap, on the other hand, produced significant inhibition.

Given the conflicting results, more research is clearly needed to establish the role of lexical tone in auditory word recognition in Standard Chinese. It is also important to note that most speakers of Standard Chinese are bi-dialectal

speakers. Existing studies have not controlled for participants' dialect background, which could be a potential cause of the different roles of tone and segment found in the literature. This study therefore set out to tap directly into their role(s) in bi-dialectal speakers' lexical processing. Specifically, we investigated the effect of cross-dialect phonological similarity in segment and tone on auditory word recognition in a bi-dialectal context using the auditory-auditory priming paradigm.

5.2 The present study

The two dialects we examined in this study are Standard Chinese (SC) and Xi'an Mandarin (XM). They both belong to the Mandarin family, which is the largest of the ten major Chinese dialect groups, following Chappell (2001) (but see Li & Thompson, 1981 which argues for a seven major dialect groups). SC is the most influential language within the Mandarin family. It is the official language of China and the medium of education. XM, on the other hand, is a local dialect spoken in the urban areas of Xi'an, the capital of Shaanxi Province. XM shares a common logographic writing system with SC and bears high resemblance with SC in terms of lexical items and syntactic forms. Moreover, XM exhibits large overlap of segmental features and tonal inventories with SC.

There are four lexical tones in SC, commonly known as the high-level (T1), mid-rising (T2), low-dipping (T3) and high-falling (T4) tones. XM also has four lexical tones, which are typically described as low-falling (T1), mid-rising (T2), high-falling (T3) and high-level (T4). Here, the terms T1-T4 are adopted to suggest that words which share the same tonal categories across the two dialects are etymologically-related translation equivalents in most cases. And this is consistently the case in our stimuli. Results of a tone production and a perception experiment (reported in Chapter 4) showed that tones with similar contours between the two dialects are basically perceived to be the same (see Table 1 for a summary). Specifically, tonal pairs of level contour (SC_T1 vs. XM_T4), rising contour (SC_T2 vs. XM_T2) and falling contour (SC_T4 vs. XM_T3) are perceived to be the same by the bi-dialectal speakers of XM and SC. Also, the tonal pair of low contour (SC_T3 vs. XM_T1) is generally perceived to be the same because XM_T1 sounds like an allotone of the

citation form SC_T3. This systematic mapping of tones between SC and XM, together with the large overlap of segmental features between the two dialects, makes cross-dialect homophones prevalent in SC and XM. Such cross-dialect homophones exist in a consistent tonal mapping fashion. For example, with the segments being identical, a SC_T1 monosyllable (e.g., SC_T1, ma1/妈, “mother”, high-level contour) is homophonous with a XM_T4 monosyllable (e.g., XM_T4, ma4/骂, “scold”, high-level contour). Cross-dialect minimal tone pairs are also common in SC and XM. For example, a SC_T1 monosyllable (e.g., SC_T1, ma1/妈, “mother”, high-level contour) shares the segmental structure but not the tonal contour with a XM_T1 monosyllable (e.g., XM_T1, ma1/妈, “mother”, low-falling contour).

Table 1. Paired tones with similar contours from Standard Chinese and Xi’an Mandarin.

Tone pair	Standard Chinese (SC)			Xi’an Mandarin (XM)		
	Tonal category	Pitch value	Example	Tonal category	Pitch value	Example
Level contour	SC_T1	55	ma1/妈	XM_T4	55/44/45	ma4/骂
Rising contour	SC_T2	35	ma2/麻	XM_T2	24	ma2/麻
Low contour	SC_T3	214	ma3/马	XM_T1	21/31	ma1/妈
Falling contour	SC_T4	51	ma4/骂	XM_T3	52/53/42	ma3/马

Most monolingual priming studies on the role of segment and tone in auditory word recognition used monosyllabic primes and monosyllabic targets in their setup (Lee, 2007; Sereno & Lee, 2015; Yip, 2001). The present cross-dialect priming study instead used monosyllabic primes and disyllabic targets. This is because the dialect membership information is critical for lexical decision, and it is difficult for bi-dialectal listeners to determine whether a monosyllabic target belongs to SC or XM due to their great phonetic similarity. It was also possible to use disyllabic primes and disyllabic targets. However, Cutler and Chen (1995) have reported a positional effect of segmental and tonal similarity on disyllabic spoken word recognition in Cantonese. Segment and tone overlap in the first syllables produced inhibition (e.g., *ji6liu4-ji6liu5/to4fa1-to4foo1*), whereas segment and tone overlap in the second syllables generated facilitation (e.g., *to4va6-to2wa6/si6yip6-sue6yip6*). Given their findings, we chose

monosyllabic primes and disyllabic targets to illuminate the pattern of bi-dialectal co-activation with the presentation of the prime and how the phonological similarities between the prime and target facilitate or inhibit target processing.

We constructed five types of relationships between the monosyllabic prime and the first syllable of the disyllabic target. They overlapped in both segment and tone within a dialect (identical) or across two dialects (interdialectal homophones). They overlapped in segment only within a dialect (within-dialect minimal tone pair) or across two dialects (cross-dialect minimal tone pair). The baseline condition was that they overlapped neither in tone nor segment within a dialect.

One factor that is known to influence bilingual lexical access is language proficiency (Jared & Kroll, 2001). Bilingual speakers often display high proficiency in one language over the other, with one language being more dominant. Previous cross-language phonological similarity studies have consistently reported an interference effect of phonology from the more dominant language to the less dominant language (e.g., Spivey & Marian, 1999, Marian & Spivey, 2003; Marian, Blumenfeld, & Boukrina, 2008), whereas there was mixed evidence of an interference effect of phonology (Lagrou et al., 2011) and a null effect (Weber & Cutler, 2004; Haigh & Jared, 2007) from the less dominant language to the more dominant language. We minimized language dominance influence by testing balanced, highly proficient bi-dialectal speakers of XM and SC. Furthermore, we adopted a generalized lexical decision task, similar to that in Doctor and Klein (1992), to ensure that the direction effect, if present, would not be biased by the target dialect of the task (Lemhöfer & Dijkstra, 2004). Since it can be difficult to control for the absolute proficiency level of bi-dialectals, we included primes and targets in both dialects by pairing all the target types with two prime types, one from SC (i.e., SC_T1), one from XM (i.e., XM_T4), which are cross-dialect homophones. In this way, if there is a language dominance effect, we are still able to detect it.

We expect that cross-dialect phonological similarity in both segment and tone would affect lexical access in bi-dialectal auditory word recognition. A facilitatory priming effect of the prime would be found not only for the within-

dialect identical targets, but also for the cross-dialect homophone targets relative to the unrelated control targets. However, the prime might activate lexical candidates of homophones in both dialects, which could cause interference effects during the lexical decision of the within-dialect identical targets and cross-dialect homophone targets. The interference effect would in turn reduce the facilitatory priming effect of the primes for the within-dialect identical targets and for the cross-dialect homophone targets relative to the unrelated control targets. Consequently, the identity priming effect in the bi-dialect context might not be as strong as that in the monolingual context. Moreover, if tone plays a role in constraining lexical activation in word recognition, the segment-only overlap primes and targets (the minimal tone pair) would not prime each other. No priming effect would be found for either the within-dialect minimal tone pairs or cross-dialect minimal tone pairs. If tone does not play a role in constraining lexical activation in word recognition, the minimal tone pair would prime each other, but to a lesser extent than the priming effect between the identical prime and target. Since we tested balanced bi-dialectal listeners of XM and SC, we did not expect a direction difference in any priming effect. Any effect should be found in both directions, from SC to XM and from XM to SC.

5.3 Method

5.3.1 Participants

One-hundred balanced XM_SC bi-dialectal speakers (41 males, 59 females) were selected and paid to participate in the experiment. To assess their language proficiency in the two dialects, we asked participants to read the story “the North Wind and the Sun” in both SC and XM. In addition, an adapted version of the LEAP-Q questionnaire (Marian, Blumenfeld, & Kaushanskaya, 2007) was used to thoroughly check their language background and language proficiency. As can be seen in Table 2, all the selected participants were of high and comparable speaking proficiency (XM vs. SC: 7.8 vs. 8.1, $t(99) = -1.05$, $p = .30$) and spoken language comprehension skills (XM vs. SC: 8.3 vs. 8.6, $t(99) = -1.11$, $p = .27$) in the two dialects. They were born and raised in the urban

areas of Xi'an and had no living experience outside of Xi'an. All were undergraduate or graduate students at local universities, with an age range from 19 to 28 ($M \pm SD$: 21.7 ± 3.2). None of them reported any speech or hearing disorders. Informed consent was obtained from all the participants before the experiment.

Table 2. *Self-ratings of language skills of Xi'an Mandarin and Standard Chinese by the participants.*

Measure	Xi'an Mandarin		Standard Chinese	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Speaking proficiency	7.8	1.6	8.1	1.4
Level of spoken comprehension	8.3	1.5	8.6	1.0
Age onset speaking (in years)	4.0	4.5	5.6	9.4
Usage language interacting with family (in years)	16.5	8.1	8.3	10.2
Usage language in school/working settings (in years)	5.6	6.1	12.6	6.3
Percentage of current exposure time	29.2%	18.5%	58.9%	19.3%

Note. $N = 100$. Scale: 0 = None; 10 = Perfect.

Most of the selected participants learned to speak XM as the first dialect (D1) from their parents and then acquired SC from the age of 6 when they started to receive education at school. Some other participants learned to speak XM and SC almost simultaneously when they were young and had difficulty deciding whether SC or XM is their D1. Very few participants learned SC as D1 and started to speak XM afterwards. However, due to the fact that these participants have been exposed to XM-speaking settings by at least one of their parents from birth, they mastered XM as well as their SC despite starting to speak XM relatively later. All participants were thus selected as balanced XM_SC bi-dialectal speakers due to their language proficiency. For clarity purposes, we will not use the terminologies D1 and D2 but rather XM and SC in the present study. Overall, the average age at which participants began speaking XM and SC was 4.0 (SD : 4.5) and 5.6 (SD : 5.6), with the former being significantly younger than the latter ($t(99) = 5.09$, $p < .001$). It should also be noted that XM was mainly used for interacting with family ($M \pm SD$: 16.5 ± 8.1), whereas SC was mainly used in school/working settings ($M \pm SD$: $12.6 \pm$

6.3) across participants. At the time of testing, the percentage of time they were exposed to XM and SC in their daily life was 29.2% (*SD*: 18.5%) and 58.9% (*SD*: 19.3%), respectively. Apparently, despite their balanced linguistic competence in the two dialects and their relatively earlier exposure to XM, the participants had more access to SC compared to XM at the time of testing.

5.3.2 Stimuli

Forty SC_T1 monosyllables and their corresponding interdialectal homophonous XM_T4 monosyllables were selected as primes. Each monosyllabic prime was paired with five disyllabic targets, resulting in 400 prime-target trials (40×2 Prime types $\times 5$ Target types = 400).

The first syllable of the disyllabic targets is our focus of interest. We therefore differentiated the five disyllabic target types according to property of their first syllable. The second syllable of the disyllabic targets always bears a T2, because T2 shows great resemblance in acoustic realization between the two dialects, i.e., SC_T2 maps onto XM_T2 both categorically and acoustically (see Table 1), and thus lends no ambiguity to the dialect membership of the disyllabic word by itself. T2 syllables, however, cue the dialect membership information of the disyllabic word together with the first syllable. For each prime type (e.g., SC_T1, “bang1/帮”, “help”), the five disyllabic target types included a within-dialect segment and tone overlap target (Identical: e.g., SC_T1 target, “bang1mang2/帮忙”, “help”), a within-dialect segment-only overlap target (D+Seg: e.g., SC_T4 target, “bang4qiu2/棒球”, “baseball”), a cross-dialect segment-only overlap target (D–Seg: e.g., XM_T1 target, “bang1mang2/帮忙”, “help”), a cross-dialect segment and tone overlap, i.e., an interlingual homophone, target (D–Homophone: e.g., XM_T4 target, “bang4qiu2/棒球”, “baseball”), and a within-dialect control target which had neither segment nor tone overlap with the prime and served as baseline (Baseline: e.g., control target, “wan2cheng2/完成”, “finish”).

Note that the target words in the first four conditions were restricted to words that showed no pronunciation difference in the segment of the words between SC and XM. Also, there was no pronunciation difference between the segment of the prime and that of the first syllable of these targets. Regarding

the control items, we ideally would have used words belonging exclusively to the dialect of the prime type. However, there are very few such words as the vocabularies largely overlap between the two dialects. What is possible is to narrow down the control words to those that show a pronunciation difference in terms of segment between the two dialects. An effort was made to obtain such control words for each prime type. Different control target words were selected for the SC primes and XM primes. Table 3 shows an example of the experimental design. The complete set of word stimuli is listed in Table D1 (see Appendix D).

Table 3. *An example of the experimental design for the prime-word target stimuli.*

	Prime type	Word target type				
		Identical	D+Seg	D-Seg	D-Homophone	Baseline
Tonal category	SC_T1	SC_T1	SC_T4	XM_T1	XM_T4	SC_T2
Tonal contour	level	level	falling	low	level	rising
Character	帮	帮忙	棒球	帮忙	棒球	完成
Pinyin	bang1	bang1mang2	bang4qiu2	bang1mang2	bang4qiu2	wan2cheng2
English	<i>help</i>	<i>help</i>	<i>baseball</i>	<i>help</i>	<i>baseball</i>	<i>finish</i>
Tonal category	XM_T4	XM_T4	XM_T1	SC_T4	SC_T1	XM_T2
Tonal contour	level	level	low	falling	level	rising
Character	棒	棒球	帮忙	棒球	帮忙	成熟
Pinyin	bang4	bang4qiu2	bang1mang2	bang4qiu2	bang1mang2	cheng2shu2
English	<i>stick</i>	<i>baseball</i>	<i>help</i>	<i>baseball</i>	<i>help</i>	<i>mature</i>

Note. We only listed the “Tonal category” and “Tonal contour” for the first syllable of the disyllables.

For all selected items, their listed frequency was controlled. Since there was no specific frequency reference available for XM, and XM shares most of its vocabulary with SC, we took the frequency of the translation equivalent in SC (e.g., SC_T4, “bang4qiu2/棒球”, “baseball”) as the frequency of the selected XM items (e.g., XM_T4, “bang4qiu2/棒球”, “baseball”). In this study, the monosyllabic primes are frequent monosyllabic words with more than 4,500 occurrences in a corpus of 193 million words (Da, 2004). We ensured that each SC_T1 prime had a comparable word frequency to the matched XM_T4 prime

($t(39) = -1.85, p = .07$). The disyllabic targets were also of comparable frequency among conditions ($F(3, 156) = .52, p = .67$). According to the SUBTLEX-CH frequency list (Cai & Brysbaert, 2010), the average log₁₀ word frequencies for the disyllabic targets were 2.45 (SC_T1/XM_T1), 2.34 (SC_T4/XM_T4), 2.46 (SC control) and 2.36 (XM control), respectively. None of the averaged word frequency was significantly different between target types.

In addition to the SC_T1 and XM_T4 monosyllabic primes and the disyllabic word targets, the same number of SC_T3 and XM_T1 monosyllabic primes and the disyllabic nonword targets were paired and added in the experiment (40×2 Prime types $\times 5$ Target types = 400) based on similar logic. The SC_T3 and XM_T1 monosyllabic primes are legitimate monosyllables, and also interdialectal near-homophones, whereas the disyllabic nonword targets are not legitimate words in either SC or XM. The latter were made up of two permitted monosyllables, the second syllable of which bears a T2. For each prime type (e.g., SC_T3, “ma3/马”), the five disyllabic nonword target types included a within-dialect segment and tone overlap nonword target (Identical: e.g., SC_T3 target, “ma3duo2/马夺”), a within-dialect segment-only overlap nonword target (D+Seg: e.g., SC_T1 target, “ma1miao2/妈苗”), a cross-dialect segment-only overlap nonword target (D-Seg: e.g., XM_T3 target, “ma3duo2/马夺”), a cross-dialect segment and tone overlap, i.e., an interlingual homophone, nonword target (D-Homophone: e.g., XM_T1 target, “ma1miao2/妈苗”), and a within-dialect control nonword target which had neither segment nor tone overlap with the prime and served as baseline (Baseline: e.g., control target, “zhe4zuo2/这昨”).

Again, we selected different control nonword targets for the SC_T3 primes and XM_T1 primes. The segment constraints between the primes and nonword targets were identical with those between the primes and word targets. Furthermore, none of the segments of the SC_T3 and XM_T1 primes and the nonword targets was ever used in the prime-word target conditions. Table 4 shows an example of the experimental design for the prime-nonword target stimuli. The complete set is listed in Table D2 (see Appendix D). The SC_T3 and XM_T1 monosyllabic primes are frequent monosyllabic words with more than 4,500 occurrences in a corpus of 193 million words (Da, 2004). Each

SC_T3 prime was ensured to have comparable word frequency to the matched XM_T1 prime ($t(39) = -0.52, p = .6$).

Table 4. *An example of the experimental design for the prime-nonword target stimuli.*

	Prime type	Nonword target type				
		Identical	D+Seg	D-Seg	D-Homophone	Baseline
Tonal category	SC_T3	SC_T3	SC_T1	XM_T3	XM_T1	SC_T4
Tonal contour	low-rising	low-rising	level	falling	low	falling
Character	马	马夺	妈苗	马夺	妈苗	这昨
Pinyin	ma3	ma3duo2	ma1miao2	ma3duo2	ma1miao2	zhe4zuo2
English	<i>help</i>	<i>N.A.</i>	<i>N.A.</i>	<i>N.A.</i>	<i>N.A.</i>	<i>N.A.</i>
Tonal category	XM_T1	XM_T1	XM_T3	SC_T1	SC_T3	XM_T3
Tonal contour	low	low	falling	level	low-rising	falling
Character	妈	妈苗	马夺	妈苗	马夺	厂闲
Pinyin	ma1	ma1miao2	ma3duo2	ma1miao2	ma3duo2	chang3xian2
English	<i>mother</i>	<i>N.A.</i>	<i>N.A.</i>	<i>N.A.</i>	<i>N.A.</i>	<i>N.A.</i>

Note. We only listed the ‘‘Tonal category’’ and ‘‘Tonal contour’’ for the first syllable of the disyllables.

In total, there were 400 prime-word target trials and 400 prime-non-word target trials. Across all the items in the two dialects, the frequency of occurrence was comparable for the level-contour tones, low-contour tones and falling-contour tones.

5.3.3 Stimuli recording

A balanced XM_SC bi-dialectal male speaker was recruited to produce the stimuli in two separate blocks for the two dialects. This speaker was born and raised in the urban area of Xi’an and had no living experience out of Xi’an. He learned XM and SC simultaneously when he was young and was of high and comparable proficiency in the two dialects. He was an undergraduate student at a local university and used the two dialects equally frequently in his daily life. All the stimuli were recorded by him in a soundproof room at 16-bit resolution

and a sampling rate of 44.1 kHz on a laptop via an external digitizer (UA-G1). The recorded stimuli were trimmed of silence and normalized amplitude for perception using Praat (Boersma & Weenink, 2015).

5.3.4 Procedure

We adopted an auditory-auditory priming paradigm in the perception experiment, with a monosyllabic prime preceding a disyllabic target in each trial. All the trials were distributed in a Latin Square design, so that participants only heard the same stimulus (prime or target) once during the experiment. Consequently, all the stimuli were divided into five lists. Each list contained both prime types and all the five word target types. For a given prime type, only one of the five target types occurred in every list. If a target (e.g., SC_T1 target, “bang1mang2/帮忙”, “help”) has already been assigned to the SC_T1 prime (e.g., SC_T1, “bang1/帮”), a different target (e.g., SC_T4 target, “bang4qiu2/棒球”, “baseball”) would be assigned to the corresponding XM_T4 prime (e.g., XM_T4, “bang4/棒”) in the same list. The prime-nonword target trials were constructed in the same way in the list. In sum, each list included 80 prime-word target trials (40×2 Prime types \times 1 Target type) and 80 prime-nonword target trials (40×2 Prime types \times 1 Target type) with the five types of targets equally distributed.

Participants were tested individually on one list only in a soundproof booth of the behavioral lab at Shaanxi Normal University in Xi’an. Across all participants, the five lists were presented equally often (20 participants/list). All the trials in each list were presented to the participants using the E-Prime 2.0 software through headphones at a comfortable listening level. Trials were pseudo-randomized with the restriction that the shortest distance between the two interdialectal homophone primes was 9 trials and the shortest distance between two targets of the same type was 3 trials.

The experiment included a practice block and two experimental blocks. The practice block contained 10 trials to familiarize the participants with the task. These trials were not used in the experimental blocks. Each experimental block contained 80 trials. Between each block, there was a 3-minute break. Each trial started with a 100 ms warning beep, followed by a 300 ms pause. Participants

then heard a pair of speech items separated by a 250 ms interval. The first item was a monosyllabic prime, and the second was a disyllabic target. Participants were asked to perform a generalized lexical decision task on the target as accurately and as quickly as possible, i.e., press the button labeled “yes” on the keyboard if the second item is either a SC word or a XM word, and press the button labeled “no” if the second item is neither a word in SC nor in XM. Button-press latencies were measured from the target offset. They were given up to 3 seconds after target offset to respond. Instructions were given both visually on screen in simplified Chinese characters and orally by the experimenter in mixed fashion of the two dialects (both SC and XM) before the experiment.

5.3.5 Data analysis

We restricted our analyses to the prime-word targets trials and discarded the prime-nonword target trials. The dependent variables included response accuracy and reaction time. Response accuracy was defined as the percentage of correct judgments of the word targets in the lexical decision task. Reaction time was defined as the response time relative to the offset of the word targets which were correctly responded to. To normalize the distribution, raw reaction times were transformed using the natural logarithm.

Statistical analyses were carried out with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.1.2 (R Core Team, 2015). Analysis of response accuracy was performed using binomial logistic regression models, and analysis of reaction time was performed using linear mixed-effects regression models. The models included Prime type (SC_T1, XM_T4), Target type (Identical, D+Seg, D-Seg, D-Homophone, Baseline) and their interactions as fixed factors, and Subjects and Items as random factors. The fixed factors were added in a stepwise fashion and their effects on model fits were evaluated via model comparisons based on log-likelihood ratios. For Target type, all the conditions were first compared with the baseline condition. Post-hoc pairwise comparisons between different target conditions were conducted using *lsmeans* package (Lenth, 2016) with single-step *p*-value adjustment. For models of reaction time, trials with absolute standardized

deviations exceeding 2.5 from the mean were considered as outliers and removed from further analysis.

5.4 Results

5.4.1 Response accuracy

Figure 1 presents the response accuracy for different target types preceded by SC_T1 primes and XM_T4 primes (see also Table D3 in Appendix D for details). Results showed a significant main effect of Target type ($\chi^2(4) = 18.73$, $p < .001$) and a significant two-way interaction of Prime type \times Target type ($\chi^2(4) = 29.28$, $p < .001$). No main effect of Prime type was found ($\chi^2(1) = 0.40$, $p = .52$).

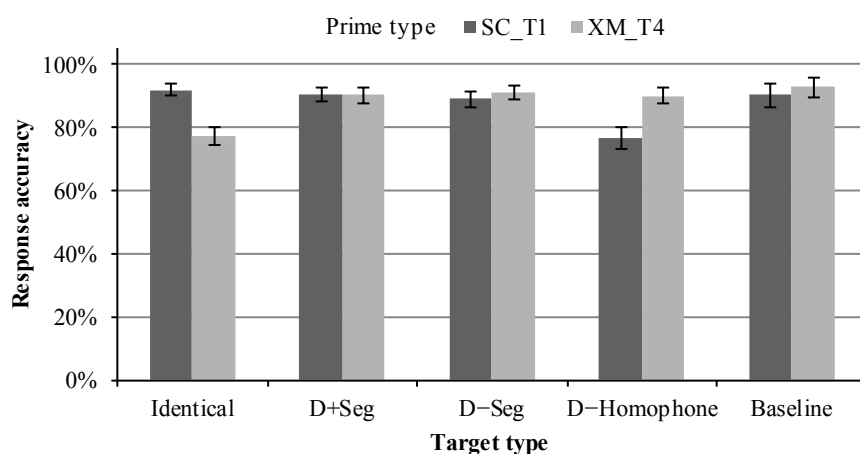


Figure 1. Response accuracy for different target types preceded by SC_T1 primes (dark grey bars) and XM_T4 primes (light grey bars).

Separate models were constructed for subset data of different prime types. When the prime was SC_T1, there was a significant main effect of Target type ($\chi^2(4) = 20.19$, $p < .001$), indicating that the response accuracy differed significantly among different target types. Further multiple pairwise comparisons showed that the response accuracy for the interdialectal homophone target was significantly lower than that for the other four target

types (all p s < .05). No difference was found for any other pair of target types (all p s > .05). Overall, a SC_T1 prime made the recognition of its interdialectal homophone target XM_T4 words more erroneous for the balanced XM_SC bi-dialectal listeners.

When the prime was XM_T4, there was also a significant main effect of Target type ($\chi^2(4) = 27.88, p < .001$). Surprisingly, multiple pairwise comparisons showed that the response accuracy for the identical target, rather than the interdialectal homophone target from SC, was significantly lower than that for the other four conditions (all p s < .05). No difference was found for any other pair of conditions (all p s > .05). In other words, when a XM_T4 monosyllabic prime preceded a XM_T4 disyllabic target, the recognition of the latter became an error-prone process. Taken together, irrespective of whether the prime was the SC version or XM version of the homophone, the balanced XM_SC bi-dialectal listeners recognized the XM disyllabic target less accurately than the other target types.

5.4.2 Reaction time

2.3% of the data points were identified as outliers and removed from further analysis. Figure 2 presents the average reaction time for different target types preceded by SC_T1 primes and XM_T4 primes (see also Table D3 in Appendix D for details). The overall analyses showed a significant main effect of Target type ($\chi^2(4) = 15.22, p = .004$) and a significant two-way interaction of Prime type \times Target type ($\chi^2(4) = 75.29, p < .001$). No main effect of Prime type was found ($\chi^2(1) = 0.09, p = .77$).

Separate models were constructed for subset data of different prime types. When the prime was SC_T1, there was a significant main effect of Target type ($\chi^2(4) = 46.05, p < .001$). Pairwise comparisons showed a non-significant facilitatory priming effect by the SC_T1 prime for the identical target compared to the baseline target ($\beta = -0.07, t = -1.31, p = .69$). Likewise, there was a non-significant facilitatory priming effect by the SC_T1 prime for the within-dialect segment-only overlap target (D+Seg) compared to the baseline target ($\beta = -0.07, t = -1.33, p = .67$). While it was quite unexpected that the identity priming and the within-dialect segment alone overlap priming did not reach

significance, the priming trend for these two target types was indeed consistent with the priming effects reported in Sereno and Lee (2015). In addition, a comparison between the reaction time of these two target types showed almost no difference ($\beta = -0.001$, $t = -0.02$, $p = 1.00$), which is in contrast with the previous finding that the priming effect for the segmental and tonal overlap (identical) prime and target was stronger than that for the segment-only overlap prime and target (Sereno & Lee, 2015).

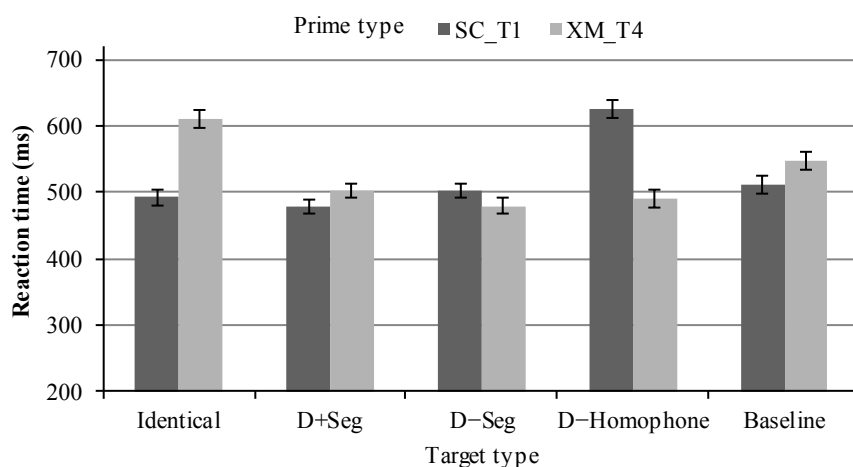


Figure 2. Average reaction time for different target types preceded by SC_T1 primes (dark grey bars) and XM_T4 primes (light grey bars). The error bars represent the 95% confidence interval of the means across participants.

As for the cross-dialect segment-only overlap target (D-Seg), its reaction time was comparable to that of the baseline target. No priming effect by the SC_T1 prime was therefore found ($\beta = 0.03$, $t = -0.61$, $p = .97$), seemingly indicating that cross-dialect segment-only overlap between primes and targets was not enough to yield priming between two tonal languages.

Regarding the interdialectal homophone target (D-Homophone), a significant inhibitory priming effect by the SC_T1 prime was found for this condition compared to the baseline target ($\beta = 0.25$, $t = 4.80$, $p < .001$). In fact, the reaction time for the interdialectal homophone target was not only longer than that for the baseline target, but also longer than that for all of the other

target types (all p s < .001). The SC_T1 monosyllabic primes considerably slowed down the recognition of the interdialectal homophonous XM_T4 disyllabic targets.

Similar analyses were conducted for the data of XM_T4 primes, where a reversed pattern of reaction time was found compared to the SC_T1 prime. There was a significant main effect of Target type ($\chi^2(4) = 44.22, p < .001$). Contrary to the facilitatory priming trend of the SC_T1 prime for the identical target, the XM_T4 prime showed a significant inhibitory priming effect for the identical target ($\beta = 0.15, t = 2.85, p = .04$). The reaction time for the identical target was not only longer than that for the baseline target, it was also longer than that for all the other target types (all p s < .001). Overall, the XM_T4 monosyllabic primes considerably slowed down the recognition of the identical XM_T4 disyllabic targets. This result was rather surprising, considering that a complete overlap of both segment and tone between primes and targets (identical) has almost always shown a facilitatory priming effect (see Lee, 2007; Sereno & Lee, 2015).

The comparison between the reaction times of the within-dialect segment-only overlap targets and the baseline targets was carried out next. No priming effect by the XM_T4 prime was found for the within-dialect segment-only overlap target compared to the baseline target ($\beta = -0.07, t = -1.41, p = .62$).

Next was the comparison between the reaction times of the cross-dialect segment-only overlap targets and that of the baseline targets. There was a significant facilitatory priming effect of the XM_T4 prime for the cross-dialect segment-only overlap target relative to for the baseline target ($\beta = -0.16, t = -3.14, p = .02$).

Lastly, we compared the reaction time of the interdialectal homophone target and that of the baseline target following the XM_T4 primes. The former was significantly shorter than the latter ($\beta = -0.15, t = -2.98, p = .03$), suggesting an evident facilitatory priming effect of the XM_T4 prime for the interdialectal homophone target. Different from the results in the SC_T1 prime data showing that the facilitatory priming trend was found when the primes and targets belonged to the same dialect (i.e., in the SC_T1 prime-Identical target condition and the SC_T1 prime-D+Seg target condition), in the XM_T4 prime

data we found a facilitatory priming effect when primes and targets belonged to different dialects (i.e., in the XM_T4 prime-D–Seg target condition and the XM_T4 prime-D–homophone target condition). A comparison between the reaction time of the interdialectal homophone targets and that of the cross-dialect segment-only overlap targets barely showed any difference ($\beta = 0.01$, $t = 0.16$, $p = 1.00$).

So far, we have found that when the prime was SC_T1, there was a similar facilitatory priming trend for the identical target and for the within-dialect segment-only overlap target. This facilitatory priming trend, however, did not hold for the cross-dialect segment-only overlap target. If the target was an interdialectal homophone of the prime, a significant inhibitory priming effect emerged from the prime to the target. The overall pattern was reversed for the XM_T4 prime data. When the prime was XM_T4, there was a significant inhibitory priming effect for the identical targets and a null effect for the within-dialect segment-only overlap targets. The cross-dialect segment-only overlap targets and the interdialectal homophone targets, on the other hand, showed similar facilitatory priming effects by the XM_T4 prime.

The pattern for the XM_T4 prime data was counterintuitive. If the SC_T1 prime and XM_T4 prime were represented equally well in the mental lexicon of the XM_SC bi-dialectal speakers, we would expect that the two prime types performed similarly on each target type. Yet the SC_T1 prime and XM_T4 prime behaved in a reversed fashion on different target types. What could be the possible reason for this?

One alternative way of viewing the reversed pattern of the XM_T4 prime data is that it could be a rearrangement of the pattern of the SC_T1 prime data, which led us to make the assumption that the XM_T4 prime might not be taken as XM_T4 itself, but as its interdialectal homophone counterpart, i.e. SC_T1. Since the two prime types were interdialectal homophones with almost no pronunciation difference, it is very likely that participants did not recognize the dialect membership of the XM_T4 prime and treated it as the SC_T1 prime. To test the validity of this assumption, we replotted Figure 2 according to the tonal category of the targets rather than the prime-target relationship. This is possible because the two prime types had targets of the same tonal categories

which corresponded to different target types based on the prime-target relationship as stated earlier. The replotted results are presented in Figure 3. Just as what we have assumed, the pattern of the XM_T4 prime was very similar to that of the SC_T1 prime.

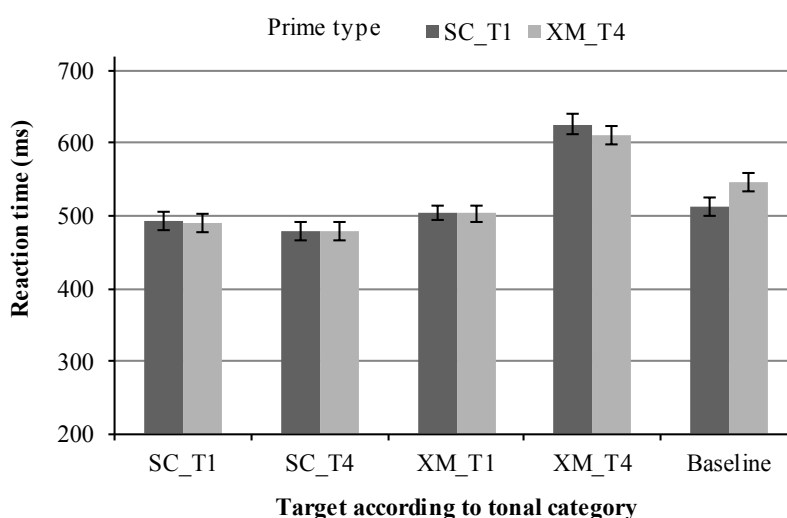


Figure 3. Average reaction time for different targets arranged according to tonal category preceded by SC_T1 primes (dark grey bars) and XM_T4 primes (light grey bars). The error bars represent the 95% confidence interval of the means across participants.

We also ran statistical analyses for the rearranged data. Linear mixed-effects regression models were built for reaction time with the fixed factors Prime type (SC_T1, XM_T4) and Target tonal category (SC_T1, SC_T4, XM_T1, XM_T4, Baseline). We only found a significant main effect of Target tonal category ($\chi^2(4) = 88.16, p < .001$). Neither a main effect of Prime type ($\chi^2(1) = 0.06, p = .81$) nor a significant two-way interaction of Prime type \times Target tonal category ($\chi^2(4) = 2.85, p = .58$) was found. The null effect of the Prime type showed that the SC_T1 prime and XM_T4 prime were not treated differently. Since the pattern of the SC_T1 prime was more in line with previous results, we are tempted to claim that the XM_T4 monosyllabic primes were treated as their interdialectal homophonous SC_T1 equivalents in the current mixed dialect context.

5.5 General discussion

The present study investigated if and how cross-dialect phonological similarity in segment and tone affects lexical access in bi-dialectal auditory word recognition of balanced bi-dialectal tonal language listeners. In an auditory-auditory priming experiment with a generalized lexical decision task, we found that when the prime was in Standard Chinese, there was a non-significant within-dialect facilitatory priming trend for targets overlapping in both segment and tone with the prime, and also for targets overlapping only in segment with the prime. Both priming trends were of similar magnitude. The Standard Chinese prime did not produce any cross-dialect priming effect for the cross-dialect segment-only overlap target. It, however, produced a significant inhibitory priming effect, as evidenced by the lower response accuracy and longer reaction time, for the interdialectal homophone target relative to the unrelated Standard Chinese control target. The overall pattern was reversed when the prime was in Xi'an Mandarin, because the Xi'an Mandarin prime was treated as its interdialectal homophonous Standard Chinese prime in the current mixed dialect setting. It seems that cross-dialect phonological similarity in segment alone does not affect lexical access in bi-dialectal auditory word recognition while a cross-dialect phonological similarity in both segment and tone (cross-dialect homophones) does pose a threat to the recognition system of the bi-dialectal tonal language listeners. Tonal information plays a significant role in constraining word activation in bi-dialectal auditory word recognition.

The present study extended the investigation of the role of segment and tone in Standard Chinese auditory word recognition from a monolingual context to a bi-dialectal context. In the monolingual Standard Chinese context, a significant facilitatory priming effect has been consistently found for the identical primes and targets which overlap in both segment and tone. Also, a complete overlap in segment and tone between the prime and target has always shown more facilitation than segment-only overlap between primes and targets (Lee, 2007; Sereno & Lee, 2015). In the bi-dialectal context, we found that a complete overlap in segment and tone between the Standard Chinese primes and targets showed a non-significant facilitatory priming trend, and so did the Standard Chinese primes and targets overlapping in segment only. No

magnitude difference was found between the priming trends of the two conditions. Our results contrast with the results in the monolingual context. The identity priming effect of Standard Chinese primes and targets in the current bi-dialectal context was not as strong as that in the monolingual context. It shrunk in size and did not reach significance, but the facilitatory priming trend was maintained. This result was in line with our hypothesis. As has been shown in previous bilingual studies, a homophone representation activated lexical candidates from both languages in a bilingual context (Schulpen et al., 2003; Lagrou et al., 2011). In the current bi-dialectal context, the presentation of the Standard Chinese version of the homophone prime seemed to have activated both the lexical candidates from Standard Chinese and its homophonous lexical candidates from Xi'an Mandarin, which caused an interference effect and reduced the identity facilitatory priming effect. The identical Standard Chinese primes and targets sharing phonological similarity in both segment and tone showed an overall facilitatory priming trend, despite the fact that the facilitation was reduced by the interference effect resulting from the coactivation of lexical candidates from both dialects by the Standard Chinese prime. A similar result was expected for primes and targets which share cross-dialect phonological similarity in both segment and tone (i.e., cross-dialect homophones). However, a significant inhibitory priming effect was found for the latter. The Standard Chinese prime considerably slowed down the recognition of interdialectal homophone targets with significantly more errors than the unrelated control target. As inhibitory priming has generally been taken as evidence of competition between lexical candidates activated by the prime and the target (Dufour & Peereman, 2003; Radeau, Morais, & Dewier, 1989; Slowiaczek & Hamburger, 1992), it seems that there was competition among the lexical candidates activated by the SC_T1 prime and the XM_T4 target. The inhibitory priming effects might have taken place under one of the following two scenarios. The first scenario is that the input signal of the auditory monosyllabic SC_T1 prime only activated the lexical representations of SC_T1, the presence of the XM_T4 target contrasts with the activated SC_T1 prime in dialect membership and the bi-dialectal listeners had to make a switch to activate the XM_T4 target to perform the task. Since dialect membership

information can only be determined by the second syllable of the target together with the first syllable, the participants could not switch to Xi'an Mandarin until the presence of the second syllable, which caused longer reaction times and lower response accuracy in lexical decisions on XM_T4 target. The second scenario is that the input signal of the auditory monosyllabic SC_T1 prime activated the lexical representations of both SC_T1 and XM_T4, but with a stronger activation of SC_T1 and a weaker spreading activation of XM_T4. The coactivation of the SC_T1 and XM_T4 lexical representations caused an interference effect when the XM_SC bi-dialectal listeners made lexical decisions on the XM_T4 target, and the interference from SC_T1 would be much stronger than that from XM_T4, resulting in strong inhibition of the former to activate the latter. The lexical decision for the XM_T4 target was thus more time-consuming and error-prone compared to that for other target types. The first scenario advocates for a selective activation of Standard Chinese lexical representations by the SC_T1 prime, whereas the second scenario speaks for a non-selective parallel activation of both Standard Chinese and Xi'an Mandarin lexical representations with a stronger activation of the former than the latter by the SC_T1 prime. Both scenarios seem plausible to account for the inhibitory priming effect of the Standard Chinese prime for the Xi'an Mandarin homophone target. However, only the second scenario could account for the different priming results in the current bi-dialectal context with those in the monolingual context when the prime and target were both from Standard Chinese. Therefore, it is more likely that the SC_T1 prime activated the lexical representations of both Standard Chinese and Xi'an Mandarin with the former being more strongly activated than the latter.

A Standard Chinese prime produced an inhibitory priming effect for the cross-dialect segment and tone overlap target. It, however, did not produce any priming effect for the cross-dialect segment-only overlap target. Recall that there was also no significant priming effect of the Standard Chinese prime for the within-dialect segment-only overlap target. This lack of significant facilitatory priming in the minimal tone pair within and across dialects suggests that the members of the minimal tone pair were not treated as homophones. Tonal information was indeed used to constrain lexical activation in spoken

word recognition, as claimed in Lee (2007). This finding echoes the previous findings that pitch accent in Japanese (Cutler & Otake, 1999) and stress in English (Cooper, Cutler, & Wales, 2002) and Dutch (Cutler & Van Donselaar, 2001) could be used to constrain lexical activations, together indicating that prosodic information might be universally adopted to constrain lexical activation in spoken word recognition. It should also be noted that the priming results for the within-dialect and cross-dialect minimal tone pair were not entirely the same. The Standard Chinese prime made it relatively easier (reflected in the shorter reaction time) for the XM_SC bi-dialectal listeners to recognize the Standard Chinese segment-only overlap targets than the Xi'an Mandarin segment-only overlap targets, though the reaction time of the two target types was not statistically different. The weaker effect of the prime for the cross-dialect segment-only overlap target than for the within-dialect segment-only overlap target again seems to be indicative of a weaker activation of the XM than SC representations by the Standard Chinese prime in the minds of the balanced XM_SC bi-dialectal listeners.

The above discussions were all for targets preceded by the SC_T1 primes. We also investigated the priming effects for all the target types preceded by the XM_T4 primes, which are homophonous with SC_T1. It was found that the overall pattern was reversed for all the target types with the XM_T4 primes compared to with the SC_T1 primes. Further analyses have shown that the Xi'an Mandarin primes were not treated differently from interdialectal homophonous SC_T1 primes in the current mixed dialect setting. The two primes were actually represented as the same in the minds of the balanced XM_SC bi-dialectal listeners. Like the SC_T1 prime, the XM_T4 prime activated the lexical representations of both SC_T1 and XM_T4, but with a stronger activation of SC_T1 and a weaker spreading activation of XM_T4 representations.

The fact that the SC_T1 prime and XM_T4 prime were not treated differently indicates that the balanced XM_SC bi-dialectal listeners did not perceive any subphonemic difference between the two primes. Both homophone primes were more strongly associated with the Standard Chinese representations than the Xi'an Mandarin representations under the current

mixed dialect context. For balanced XM_SC bi-dialectal listeners, one might expect that both dialects were activated to similar degrees. The resulting stronger activation of the Standard Chinese representations relative to the Xi'an Mandarin representations by the homophone prime in the present study indicates that the two dialects were not represented to similar degrees in the minds of these XM_SC bi-dialectal listeners. In bilingual studies, Grosjean (1988; 1997) proposed that in the bilingual language mode, bilingual speakers choose a base language (the main language for communication) and call upon the other language (guest language) when necessary. The present study seems to show similar mechanisms for bi-dialectal word recognition in the bi-dialectal mode. Based on the stronger activation of Standard Chinese presentations than Xi'an Mandarin representations, the bi-dialectal listeners here appeared to choose Standard Chinese as their base language. This could possibly reflect that though the XM_SC bi-dialectal listeners we recruited were comparable in their language competence of the two dialects (see Table 2), perhaps they were not balanced XM_SC bi-dialectal listeners after all. They more likely had overall language dominance in Standard Chinese rather than in Xi'an Mandarin. One might argue that the balanced XM_SC bi-dialectal listeners we investigated were a mixture of bi-dialectal speakers with different orders of learning of the two dialects. It may be unfair to talk about language dominance for such a mixed group of participants without concerning their order of learning of the two dialects, as they might have performed differently when processing the two dialects. We admit that better control of the language background of the participants should have been made. However, the different orders of learning of the two dialects could hardly be a factor which has affected the priming pattern. A closer analysis showed that the bi-dialectal listeners with Standard Chinese as their D1 did not behave differently from those with Xi'an Mandarin as their D1. All the bi-dialectal listeners tended to be Standard Chinese dominant regardless of their D1. This is understandable given that Standard Chinese is the medium of education and it is more frequently used in campus life than Xi'an Mandarin by the bi-dialectal listeners.

In the field of spoken word recognition, previous cross-language phonological similarity studies are mostly concerned with bilingual word

recognition in two Indo-European languages. The present study shifted the focus to cross-dialect phonological similarity effects in two tonal dialects. The results extend our understanding of the role of segment and tone in auditory word recognition in tonal languages from the monolingual context to the bi-dialectal context. The investigation of the cross-dialect homophone effect in bi-dialectal auditory word recognition also allows us to compare the effect of cross-dialect phonological similarity on bi-dialectal auditory word recognition with the effect of cross-language phonological similarity on bilingual auditory word recognition. In bilingual auditory word recognition, it has been found that the presentation of a homophone prime activated homophone representations from both languages; homophone primes of both languages facilitated the recognition of the L2 visual targets, yet there was competition between the two interlingual homophone representations compared to the monolingual control prime and target words (Schulpen et al., 2003). In the current bi-dialectal auditory word recognition, homophone primes of both dialects yielded a stronger activation of the lexical representations in Standard Chinese and a much weaker activation of the lexical representations in Xi'an Mandarin. There was intense competition of the activated representations of the homophone with the Xi'an Mandarin target, resulting in a significant inhibitory priming effect of the homophone prime for the Xi'an Mandarin target. Overall, the coactivation of the lexical representations in both dialects by the homophone prime in bi-dialectal auditory word recognition is in line with the coactivation of the lexical representations in both languages by the homophone prime in bilingual auditory word recognition, suggesting a non-selective processing mechanism in both bilingual lexical access and bi-dialectal lexical access during auditory word recognition.

The priming results in our bi-dialectal study also differed from that in the bilingual studies. For example, in the bilingual study, the homophone prime showed facilitatory priming for the L2 target (Schulpen et al., 2003), whereas in our bi-dialectal study, the homophone prime showed inhibitory priming for the Xi'an Mandarin target. This difference presumably results from the different strength of activation of the lexical representations in each language system due to difference in tasks, participants, and stimulus features between the two

studies. More bi-dialectal research is clearly needed to understand the similarities and differences between bilingual and bi-dialectal word recognition.

5.6 Conclusion

To conclude, the present study showed that phonological similarity in segment alone did not affect lexical access in bi-dialectal auditory word recognition, whereas phonological similarity in both segment and tone (cross-dialect homophones) posed a threat to the recognition system of the bi-dialectal tonal language listeners due to coactivation of the lexical representations in both dialects. Tonal information played a significant role in constraining word activation in bi-dialectal auditory word recognition. The results extends our understanding of the role of segment and tone in auditory word recognition in tonal languages from the monolingual context to the bi-dialectal context, and reveals a non-selective processing mechanism in bi-dialect lexical access during auditory word recognition, as in bilingual lexical access.

Chapter 6

General discussion

This dissertation investigated how ambiguous acoustic signals representing different prosodic information affect spoken language processing. Specifically, it investigated how pitch is processed within a linguistic system (i.e., Standard Chinese, Chapters 2 and 3) and across two linguistic systems (i.e., Standard Chinese and Xi'an Mandarin, Chapters 4 and 5) when the same pitch contour cues different linguistic functions (i.e., tone and intonation) or different categories of the same linguistic function (i.e., tone). Tone and intonation in Standard Chinese both adopt F0 as their primary acoustic correlate and therefore result in pitch processing difficulties. Chapter 2 tapped into the neural correlates of tone and intonation processing in Standard Chinese and presented ERP evidence for pitch processing costs due to the interaction of tone and intonation. Chapter 3 examined the role of semantic context in resolving pitch processing difficulties in tone and intonation processing in Standard Chinese. Chapters 4 and 5 went beyond Standard Chinese, and investigated how the tonal system of a closely related dialect of Standard Chinese (i.e., Xi'an Mandarin) affects tone processing (Chapter 4) and lexical access (Chapter 5) of bi-dialectal tonal language speakers. Together, these chapters revealed two of the most prominent pitch processing difficulties tonal language speakers encounter from within and across languages, advancing our current understanding of pitch processing from various aspects.

Chapter 2 investigated the online processing mechanisms of tone and intonation in Standard Chinese at the attentive stage using the event-related potential technique. Native Standard Chinese listeners were presented with semantically neutral Standard Chinese sentences, which contrast in final tones (rising T2 or falling T4) and intonations (Question or Statement). Their behavioral and electrophysiological responses were recorded. The behavioral results showed that, while the identification of tone was not hindered by intonation, the identification of intonation was greatly impeded due to the interference of tone. In the T4 conditions, question intonation was rather difficult to identify correctly, whereas the identification of statement intonation presented almost no difficulty. In the T2 conditions, question intonation was still difficult to identify, while identifying statement intonation also tended to be problematic. Regarding the ERP results, a clear P300 effect was observed for

the question-statement contrast in sentences ending with T4, but no ERP effect was found for the question-statement contrast in sentences ending with T2. These results provide ERP evidence for the interaction of tone and intonation in Standard Chinese, confirming the findings from behavioral metalinguistic data that native Standard Chinese listeners can distinguish between question and statement intonation when the intonation is associated with a final T4, but fail to do so when the intonation is associated with a final T2.

The ERP results of tone and intonation processing at the attentive processing stage are highly consistent with findings for tone and intonation processing at the pre-attentive processing stage. Ren et al. (2009, 2013) reported that listeners are able to perceive the difference between question and statement intonation when the final tone is T4 (reflected in an MMN effect), but they cannot make a distinction between question and statement intonation when the final tone is T2 (reflected by the absence of an MMN effect) at the pre-attentive stage. These two studies used one-syllable sentences, while our study extended the length of the utterances from one syllable to five syllables. Chapter 2 therefore extends our understanding of online processing of tone and intonation 1) from the pre-attentive stage to the attentive stage; and 2) within a larger domain (i.e., multi-word utterances) than a single word utterance.

Chapter 3 tapped into the role of semantic context in resolving pitch processing difficulties in tone and intonation processing in Standard Chinese. In Standard Chinese, the F0 encodings of final lexical tone and sentence intonation can sometimes be in conflict (e.g., statements ending with a rising T2/questions ending with a falling T4) and sometimes be in congruency (e.g., statements ending with a falling T4/questions ending with a rising T2). Chapter 3 investigated how tone and intonation, with F0 encodings in conflict or in congruency, are processed and how semantic context may facilitate or inhibit their processing. Tone and intonation identification experiments were conducted in both semantically neutral and constraining contexts with the same group of native speakers of Standard Chinese. The resulting measurements included the commonly-reported response accuracy, as well as an additional measurement, reaction time.

Results showed that the overall performance of tone identification was better than that of intonation. Tone identification was seldom affected by intonation information, irrespective of semantic contexts. Participants were able to perceive tones accurately and quickly in both question and statement intonation in both contexts. Intonation identification, particularly question intonation, however, was susceptible to the final lexical tone identity and was greatly affected by the semantic context. Specifically, in the semantically neutral context, questions were difficult to identify, as evidenced in the lower response accuracy and longer reaction time, regardless of the lexical tone identity. In the semantically constraining context, both intonations took significantly less time to be identified than in the semantically neutral context. Moreover, questions ending with a falling tone were more accurately identified than questions ending with a rising tone. These results suggest that top-down information provided by the semantically constraining context can play a facilitating role for listeners to disentangle intonational information from tonal information, especially in sentences with a lexical falling tone in final position.

Chapter 3 provides strong evidence for the role of semantic context in resolving pitch processing difficulties in Standard Chinese, particularly from the reaction time patterns, which have not been reported in earlier studies. The results reported in Chapter 3 also resolved the puzzle of the reversed patterns of question intonation identification in sentences ending with T2 and T4 in normal natural context (Xu & Mok, 2012a; Yuan, 2011) versus in low-pass filtered context (Xu & Mok, 2012b). In normal natural context, questions ending with T4 were more accurately identified than questions ending with T2 (Xu & Mok, 2012a; Yuan, 2011), whereas in low-pass filtered context, questions ending with T4 were less accurately identified than questions ending with T2 (Xu & Mok, 2012b). These reversed patterns could be due to different factors, and the results in Chapter 3 showed that context is a significant factor. It was found that the stronger and more informative the linguistic context is (semantically constraining context > semantically neutral context > low-pass filtered context), the better the identification of questions ending with T4. The opposite pattern holds for questions ending with T2.

Chapter 4 was concerned with cross-linguistic pitch processing. One long-neglected fact in linguistic research on Standard Chinese is that most speakers of Standard Chinese also speak a local dialect (Li & Lee, 2008; Wiener & Ito, 2014), which may share phonological features with Standard Chinese. Tonal information can be a determinant of the phonological similarities or differences between Standard Chinese and regional dialects, yet relatively little empirical research has been conducted on the tonal system of other language varieties spoken in China aside from Standard Chinese. Among these dialects, Xi'an Mandarin is particularly interesting for the seemingly simple, yet intricate mappings between its lexical tones with those in Standard Chinese (Li, 2001; Zhang, 2009).

In Chapter 4, the tonal systems of Xi'an Mandarin and Standard Chinese were compared empirically. Tones with similar contours from Xi'an Mandarin and Standard Chinese were paired, and both tone production and perception experiments were carried out on highly proficient bi-dialectal speakers of Xi'an Mandarin and Standard Chinese. Acoustic results showed that the F0 difference ranged from no F0 difference (level contour tone pair) through F0 curvature difference (rising contour tone pair) to F0 height difference (falling contour tone pair) and F0 contour difference (low contour tone pair). Except for the falling contour tone pair, all the other tone pairs also exhibited differences in tone duration. The varying acoustic differences in different tone pairs, together with the phonological rule, resulted in varying degrees of tonal similarity in tone perception. Specifically, the rating tendency of the tone pair of low contour was significantly different from that of the other three tone pairs. The low contour pair was judged as either different or the same, with slightly more same responses than different responses, whereas the latter three tone pairs were mostly judged as the same. That said, the tone pair of falling contour elicited more different responses than the tone pairs of level contour and rising contour. Overall, tones with similar contours between the two dialects were basically perceived to be the same. The two experiments together showed that there are indeed systematic mappings of tones between Xi'an Mandarin and Standard Chinese.

Chapter 4 compared the two systems in tone production with a more balanced design compared to the only previous acoustic study (Zhang, 2009). Moreover, it provides new empirical evidence for the mapping of the two tonal systems from a perceptual point of view. It also allows for an investigation of the relationship between tone production and perception in bi-dialectal tonal language speakers.

Chapter 5 further examined the effects of cross-dialect phonological similarity in segment and tone on bi-dialectal lexical access in spoken word recognition. The systematic mapping of tones between Standard Chinese and Xi'an Mandarin, as shown in Chapter 4, together with the large overlap of segmental features between the two dialects, makes cross-dialect homophones prevalent in the two languages. Cross-dialect minimal tone pairs (i.e., syllables sharing the segmental structure but not tonal contour) are also common in Standard Chinese and Xi'an Mandarin. Using an auditory-auditory priming paradigm, Chapter 5 investigated the effects of cross-dialect phonological similarity in segment and tone on auditory word recognition in a bi-dialectal context (i.e., Standard Chinese and Xi'an Mandarin).

Balanced bi-dialectal speakers of Xi'an Mandarin and Standard Chinese took part in an auditory-auditory priming experiment with a generalized lexical decision task in Chapter 5. The primes were monosyllabic homophones from either Xi'an Mandarin or Standard Chinese, while the targets were disyllabic Xi'an Mandarin or Standard Chinese words. Primes and the first syllable of the target words had five configurations. They either overlapped in both segment and tone within a dialect (identical) or across two dialects (interdialectal homophones), or they overlapped in segment only within a dialect or across two dialects. The baseline condition was that they overlapped neither in tone nor segment within a dialect. Results showed that Standard Chinese primes did not yield significant priming effects for within- or cross-dialect segment-only overlap targets. Standard Chinese primes did not produce significant priming effects for within-dialect identical targets either. However, they did yield significant inhibitory priming effects for cross-dialect homophone targets. This overall pattern was reversed for Xi'an Mandarin primes because these primes were not treated differently from their interdialectal homophonous primes in

the current mixed dialect setting. These results suggest that cross-dialect phonological similarity in segment alone does not affect lexical access in bi-dialectal auditory word recognition, while cross-dialect phonological similarity in both segment and tone poses a threat to the recognition system of bi-dialectal listeners. It is clear that tonal information plays a significant role in constraining word activation in bi-dialectal auditory word recognition.

The results reported in Chapter 5 extend our understanding of the role of segment and tone in auditory word recognition in tonal languages from the monolingual context to the bi-dialectal context, and reveal a non-selective processing mechanism in bi-dialectal lexical access during auditory word recognition, as has been demonstrated for bilingual lexical access.

In summary, this dissertation has demonstrated that pitch processing in Standard Chinese is subject to both within- and cross-linguistic influences. The ambiguous acoustic signals due to dual functions of the F0 channel in signalling tone and intonation in Standard Chinese cause pitch processing difficulty at the sentential level. This pitch processing difficulty has a neural correlate and can be resolved via top-down information provided by a constraining semantic context. Acoustic ambiguities in Standard Chinese can also arise from a closely related Chinese dialect that shares tonal similarities with Standard Chinese, here Xi'an Mandarin. The cross-dialect tonal similarities affect tone processing and further interfere in lexical access during spoken word recognition in bi-dialectal tonal language speakers.

The present research on tone and intonation processing in Standard Chinese may contribute to the potential typology of the interaction between tone and intonation in tonal languages. Simply comparing the results of the present study with those Cantonese studies has demonstrated different mechanisms of tone and intonation interaction. As shown from the neural correlates, in Standard Chinese, the interaction of tone and intonation leads to difficulties in intonation processing (Liu et al., 2016b), whereas in Cantonese, it is tone processing rather than intonation processing that is problematic for native listeners (Kung et al., 2014). It seems that in tonal languages when tone and intonation interact, whether tone or intonation causes processing difficulties can be language-dependent. Nevertheless, the pitch processing difficulties in both languages

(tone or intonation) can be resolved via top-down information provided by a constraining semantic context.

The pitch influence of a closely related dialect on Standard Chinese has received little attention in previous studies. In this study, it was found that bi-dialectals of Standard Chinese and Xi'an Mandarin activate lexical candidates from both dialects when presented with a cross-dialect homophone, but with different activation levels. The results suggest that future studies on Standard Chinese tones may have to control for participants' dialect background to avoid potential tonal influence from dialects. Moreover, although the bi-dialectals we recruited are comparable in their language competence of the two dialects, they tend to be Standard Chinese dominant after all. Since it is almost unlikely to find more balanced bi-dialectals than the participants investigated here, the result suggests that truly balanced bilinguals or bi-dialectals may never exist. Language users always have a preference for one or the other language or dialect. Overall, this dissertation investigated two of the most prominent pitch processing difficulties that tonal language speakers encounter from within and across languages. Several implications can be made for future research.

First, Chapters 2 and 3 exclusively investigated the interaction of tone and intonation in Standard Chinese when the final tone is T2 or T4. It would be instructive to include other tones to gain a fuller picture of the interaction of tone and intonation in Standard Chinese. Some research questions worth investigation include how intonation modulates acoustically similar tones (e.g., T2 and T3) in sentence-final position, how such modulation affects tone and intonation processing, and what the underlying neural mechanisms are.

Second, apart from tone and intonation, F0 also cues other linguistic functions, such as focus, in Standard Chinese. Future research might consider investigating how other linguistic functions of F0 affect the interaction of tone and intonation in Standard Chinese, and how the different information is encoded in the minds of native speakers. Such investigations could enable us to better understand the F0 capacity Standard Chinese employs for each of these prosodic functions, which, in turn, may shed light on the acoustic modelling of Standard Chinese and neurobiological studies of language prosody in general.

Third, Chapter 4 and Chapter 5 investigated cross-dialect tone processing and bi-dialectal word recognition, which is still an understudied research area. More parallel studies from other tonal dialects should be conducted to verify and deepen current understanding. Future research should also examine bi-dialectal word recognition in more detail. Comparative studies of monolingual, bi-dialectal and bilingual word recognition should be carried out to reveal their similarities and differences. Moreover, models of spoken word recognition should take the bi-dialectal situation into account.

Appendix A

Table A1. Identification rate (%; SEs in parentheses) for each experimental condition under the tone identification task and the intonation identification task in Chapter 2.

Condition	Task	
	Tone	Intonation
QT2	98.0 (1.0)	64.0 (4.9)
QT4	92.7 (4.3)	68.0 (5.7)
ST2	98.0 (1.0)	91.3 (2.3)
ST4	99.0 (0.7)	99.3 (0.5)

Table A2. Time course analyses for the Midline Electrodes. *F*-values are reported for the main effects of T (Tone), I (Intonation) and interactions of T (Tone), I (Intonation) with R (Region).

Time window (ms)	T	I	T×I	T×R	I×R	T×I×R
0-50	8.87*	0.23	0.58	1.37	1.12	0.03
50-100	0.00	1.13	0.00	0.66	1.44	1.71
100-150	0.03	1.05	0.04	1.28	5.19*	0.49
150-200	4.04	2.42	0.48	1.00	1.14	2.20
200-250	0.63	3.93	0.42	1.56	3.13	5.71*
250-300	3.39	8.33*	1.57	0.26	1.48	3.68*
300-350	1.35	7.90*	2.13	1.22	4.49*	1.37
350-400	0.43	5.53*	0.57	0.13	2.63	6.21*
400-450	0.11	3.69	1.15	0.95	0.56	2.70
450-500	0.03	2.05	1.59	0.00	1.40	2.48*
500-550	9.19**	6.63*	2.26	0.66	1.09	1.83
550-600	1.34	2.78	3.03	3.42	2.66	4.21*
600-650	0.20	12.58**	1.30	0.79	0.09	2.28
650-700	0.97	3.33	0.61	1.59	3.43	0.79
700-750	1.25	9.73**	0.18	0.86	1.27	2.13
750-800	0.62	0.25	0.71	0.07	5.17*	1.76

Note. * $p < .05$; ** $p < .01$.

Table A3. Time course analyses for the Lateral Electrodes. F-values are reported for the main effect of T (Tone), I (Intonation) and interactions of T (Tone), I (Intonation) with R (Region) and H (Hemisphere).

Time window (ms)	T	I	T×I	T×R	I×R	T×H	I×H	T×I×R	T×I×H	T×R×H	I×R×H	T×I×R×H
0-50	7.31*	0.11	0.47	4.15*	0.93	1.61	0.82	0.05	0.27	2.04	1.34	0.01
50-100	0.00	1.59	0.01	0.03	2.66	0.05	0.53	0.36	1.07	6.63**	0.19	5.07*
100-150	0.00	1.74	0.13	0.56	6.98**	0.67	0.63	0.29	0.04	0.11	2.18	0.21
150-200	3.62	3.13	0.29	0.99	1.98	0.99	6.24*	0.37	0.05	3.08	0.73	1.90
200-250	0.49	4.29	0.39	1.45	3.91	0.11	0.87	1.80	1.02	0.34	0.92	0.68
250-300	3.51	8.57*	1.32	0.30	1.15	0.26	5.11*	0.76	0.22	1.19	1.50	2.39
300-350	1.31	8.70*	1.78	1.32	4.31*	0.39	0.00	0.26	0.02	0.60	0.80	0.55
350-400	0.62	7.52*	0.63	0.04	3.74	1.83	0.50	3.24	0.00	2.19	0.42	0.98
400-450	0.15	4.63*	1.08	1.07	1.89	1.11	0.08	1.16	0.49	0.30	0.14	0.39
450-500	0.00	3.66	1.75	0.02	3.99*	1.24	0.08	0.86	0.50	2.55	0.59	2.05
500-550	7.41*	8.02*	2.21	0.54	1.96	0.59	0.42	0.24	0.17	0.03	0.11	0.18
550-600	1.18	3.98	3.36	2.10	5.66*	0.04	0.13	1.56	0.21	0.93	0.17	0.69
600-650	0.11	13.43**	1.31	0.22	0.51	0.36	0.18	1.14	0.56	0.91	0.21	0.00
650-700	0.76	4.04	0.47	0.51	10.24**	1.17	7.35*	0.23	0.03	1.05	0.12	0.24
700-750	1.25	11.37**	0.24	0.46	1.73	1.69	0.04	1.37	0.15	0.94	1.13	0.25
750-800	0.83	0.53	0.69	0.03	8.53**	1.98	2.81	0.43	0.25	1.25	0.51	1.70

Note. * $p < .05$; ** $p < .01$.

Appendix B

Table B1. Response accuracy and average reaction time (SEs in parentheses) for each experimental condition in different tasks under different contexts in Chapter 3.

Condition	Response accuracy (%)						Reaction time (ms)					
	Neutral semantic context		Constraining semantic context		Neutral semantic context		Neutral semantic context		Constraining semantic context		Constraining semantic context	
	Tone	Intonation	Tone	Intonation	Tone	Intonation	Tone	Intonation	Tone	Intonation	Tone	Intonation
QT2	98.0 (1.0)	69.2 (6.9)	98.0 (0.8)	67.2 (7.2)	915 (18)	1074 (31)	851 (20)	1074 (31)	851 (20)	867 (27)	867 (27)	867 (27)
QT4	95.7 (1.8)	64.5 (6.7)	93.8 (2.8)	79.0 (5.0)	940 (19)	1076 (24)	800 (20)	1076 (24)	800 (20)	878 (20)	878 (20)	878 (20)
ST2	98.2 (0.7)	92.7 (2.3)	95.7 (1.0)	97.7 (0.9)	914 (14)	1109 (20)	808 (17)	1109 (20)	808 (17)	778 (15)	778 (15)	778 (15)
ST4	98.9 (0.9)	98.6 (0.5)	99.7 (0.3)	100.0 (0.0)	824 (12)	948 (15)	671 (13)	948 (15)	671 (13)	701 (14)	701 (14)	701 (14)

Appendix C

Table C1. *The full monosyllabic words used in the two experiments in Chapter 4.*

Syllable ID	Tone			
	T1	T2	T3	T4
1	逼 (bī1, force)	鼻 (bí2, nose)	比 (bǐ3, compare)	闭 (bì4, close)
2	搭 (dā1, travel by)	达 (dá2, reach)	打 (dǎ3, hit)	大 (dà4, big)
3	低 (dī1, low)	敌 (dí2, enemy)	底 (dǐ3, bottom)	弟 (dì4, younger brother)
4	督 (dū1, supervise)	毒 (dú2, poison)	赌 (dǔ3, gamble)	度 (dù4, extent)
5	涛 (tāo1, great waves)	桃 (tāo2, peach)	讨 (tǎo3, ask for)	套 (tào4, encase)
6	梯 (tī1, ladder)	题 (tí2, question)	体 (tǐ3, body)	替 (tì4, substitute)
7	通 (tōng1, through)	铜 (tóng2, copper)	桶 (tǒng3, bucket)	痛 (tòng4, pain)
8	方 (fāng1, square)	房 (fāng2, house)	访 (fǎng3, visit)	放 (fàng4, put)
9	风 (fēng1, wind)	冯 (féng2, a surname)	讽 (fēng3, irony)	凤 (fèng4, phoenix)
10	呼 (hū1, call)	湖 (hú2, lake)	虎 (hǔ3, tiger)	户 (hù4, household)
11	灰 (huī1, gray)	回 (huí2, go back)	毁 (huǐ3, destroy)	会 (huì4, able)
12	吸 (xī1, absorb)	席 (xí2, mat)	洗 (xǐ3, wash)	戏 (xì4, drama)
13	歇 (xiē1, rest)	邪 (xié2, evil)	写 (xiě3, write)	谢 (xiè4, thank)
14	星 (xīng1, star)	形 (xíng2, shape)	醒 (xǐng3, awake)	姓 (xìng4, a surname)
15	虚 (xū1, virtual)	徐 (xú2, slowly)	许 (xǔ3, allow)	续 (xù4, continue)

Syllable ID	Tone			
	T1	T2	T3	T4
16	抽 (chou1, pump)	仇 (chou2, hatred)	丑 (chou3, ugly)	臭 (chou4, smelly)
17	鸡 (ji1, chicken)	急 (ji2, urgent)	挤 (ji3, crowded)	记 (ji4, remember)
18	街 (jie1, street)	杰 (jie2, hero)	姐 (jie3, elder sister)	戒 (jie4, quit)
19	居 (ju1, residence)	局 (ju2, bureau)	举 (ju3, lift)	句 (ju4, sentence)
20	期 (qi1, period)	旗 (qi2, flag)	起 (qi3, start)	气 (qi4, gas)
21	轻 (qing1, light)	晴 (qing2, sunny)	请 (qing3, invite)	庆 (qing4, celebration)
22	妈 (ma1, mom)	麻 (ma2, hemp)	马 (ma3, horse)	骂 (ma4, scold)
23	汪 (wang1, a surname)	王 (wang2, king)	网 (wang3, network)	旺 (wang4, prosperous)
24	屋 (wu1, house)	吴 (wu2, a surname)	五 (wu3, five)	误 (wu4, mistake)
25	优 (you1, excellent)	油 (you2, oil)	有 (you3, have)	右 (you4, right)
26	威 (wei1, prestige)	围 (wei2, surround)	伟 (wei3, great)	胃 (wei4, stomach)
27	敲 (qiao1, knock)	桥 (qiao2, bridge)	巧 (qiao3, skillful)	俏 (qiao4, pretty)
28	荒 (huang1, uncultivated)	黄 (huang2, yellow)	谎 (huang3, lies)	晃 (huang4, sway)
29	秃 (tu1, bald)	图 (tu2, picture)	土 (tu3, dust)	兔 (tu4, rabbit)
30	汤 (tang1, soup)	唐 (tang2, Tang dynasty)	躺 (tang3, lie down)	烫 (tang4, scalding)

Appendix D

Table D1. The complete set of word target stimuli in Chapter 5. For each ID, the first row is with a SC_T1 prime; the second row is with a XM_T4 prime.

ID	Prime	Target				
		SC_T1	XM_T4	XM_T1	SC_T4	Baseline
1	冰 bing1 <i>ice</i>	冰球 bing1qiu2 <i>ice hockey</i>	病情 bing4qing2 <i>state of an illness</i>	冰球 bing1qiu2 <i>ice hockey</i>	病情 bing4qing2 <i>state of an illness</i>	融合 rong2he2 <i>mix together</i>
	病 bing4 <i>illness</i>	冰球 bing1qiu2 <i>ice hockey</i>	病情 bing4qing2 <i>state of an illness</i>	冰球 bing1qiu2 <i>ice hockey</i>	病情 bing4qing2 <i>state of an illness</i>	岩石 yan2shi2 <i>rock</i>
2	思 si1 think	思维 si1wei2 thinking	四十 si4shi2 forty	思维 si1wei2 thinking	四十 si4shi2 forty	人格 ren2ge2 personality
	四 si4 <i>four</i>	思维 si1wei2 thinking	四十 si4shi2 forty	思维 si1wei2 thinking	四十 si4shi2 forty	偿还 chang2huan2 <i>pay back</i>
3	加 jia1 <i>add</i>	加油 jia1you2 <i>make an extra effort</i>	价值 jia4zhi2 <i>value</i>	加油 jia1you2 <i>make an extra effort</i>	价值 jia4zhi2 <i>value</i>	男孩 nan2hai2 <i>boy</i>
	价 jia4 <i>price</i>	加油 jia1you2 <i>make an extra effort</i>	价值 jia4zhi2 <i>value</i>	加油 jia1you2 <i>make an extra effort</i>	价值 jia4zhi2 <i>value</i>	厨房 chu2fang2 <i>kitchen</i>

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
4	知	知名	制服	知名	制服	才能
	zhi1	zhi1ming2	zhi4fu2	zhi1ming2	zhi4fu2	cai2neng2
	<i>know</i>	<i>well-known</i>	<i>uniform</i>	<i>well-known</i>	<i>uniform</i>	<i>talent</i>
	制	知名	制服	知名	制服	提前
5	zhi4	zhi1ming2	zhi4fu2	zhi1ming2	zhi4fu2	t2qian2
	<i>system</i>	<i>well-known</i>	<i>uniform</i>	<i>well-known</i>	<i>uniform</i>	<i>beforehand</i>
	街	街头	戒毒	街头	戒毒	隔离
	jie1	jie1tou2	jie4du2	jie1tou2	jie4du2	ge2li2
6	<i>street</i>	<i>street</i>	<i>detoxification</i>	<i>street</i>	<i>detoxification</i>	<i>isolation</i>
	戒	街头	戒毒	街头	戒毒	民族
	jie4	jie1tou2	jie4du2	jie1tou2	jie4du2	min2zu2
	<i>give up</i>	<i>street</i>	<i>detoxification</i>	<i>street</i>	<i>detoxification</i>	<i>nation</i>
7	惊	惊奇	镜头	惊奇	镜头	荷兰
	jing1	jing1qi2	jing4tou2	jing1qi2	jing4tou2	he2lan2
	<i>surprise</i>	<i>surprise</i>	<i>camera lens</i>	<i>surprise</i>	<i>camera lens</i>	<i>Netherlands</i>
	镜	惊奇	镜头	惊奇	镜头	联盟
8	jing4	jing1qi2	jing4tou2	jing1qi2	jing4tou2	lian2meng2
	<i>mirror</i>	<i>surprise</i>	<i>camera lens</i>	<i>surprise</i>	<i>camera lens</i>	<i>union</i>
	医	医疗	意图	医疗	意图	全国
	yi1	yi1liao2	yi4tu2	yi1liao2	yi4tu2	quan2guo2
9	<i>doctor</i>	<i>medical treatment</i>	<i>intention</i>	<i>medical treatment</i>	<i>intention</i>	<i>entire country</i>
	意	医疗	意图	医疗	意图	船员
	yi4	yi1liao2	yi4tu2	yi1liao2	yi4tu2	chuan2yuan2
	<i>meaning</i>	<i>medical treatment</i>	<i>intention</i>	<i>medical treatment</i>	<i>intention</i>	<i>sailor</i>

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
8	拘	拘留	距离	拘留	距离	童年
	ju1	ju1liu2	ju4li2	ju1liu2	ju4li2	tong2nian2
	<i>restrain</i>	<i>detention</i>	<i>distance</i>	<i>detention</i>	<i>distance</i>	<i>childhood</i>
	距	拘留	距离	拘留	距离	贫穷
9	孤	孤独	固执	孤独	固执	德国
	gu1	gu1du2	gu4zhi2	gu1du2	gu4zhi2	de2guo2
	<i>lonely</i>	<i>lonely</i>	<i>stubborn</i>	<i>lonely</i>	<i>stubborn</i>	<i>Germany</i>
	固	孤独	固执	孤独	固执	时髦
10	辉	辉煌	绘图	辉煌	绘图	门铃
	hui1	hui1huang2	hui4tu2	hui1huang2	hui4tu2	men2ling2
	<i>bright</i>	<i>glorious</i>	<i>draw</i>	<i>glorious</i>	<i>draw</i>	<i>doorbell</i>
	绘	辉煌	绘图	辉煌	绘图	轮船
11	虚	虚无	继续	虚无	继续	河豚
	xu1	xu1wu2	xu4ji2	xu1wu2	xu4ji2	he2tun2
	<i>empty</i>	<i>nothingness</i>	<i>sequel</i>	<i>nothingness</i>	<i>sequel</i>	<i>puffer fish</i>
	续	虚无	继续	虚无	继续	游轮
	xu4	xu1wu2	xu4ji2	xu1wu2	xu4ji2	you2lun2
	<i>continue</i>	<i>nothingness</i>	<i>sequel</i>	<i>nothingness</i>	<i>sequel</i>	<i>cruise</i>

ID	Prime	Target				
		SC_T1	XM_T4	XM_T1	SC_T4	Baseline
12	吸	吸毒	戏迷	吸毒	戏迷	年龄
	xi1	xi1du2	xi4mi2	xi1du2	xi4mi2	nian2ling2
	suck	take drugs	theater fan	take drugs	theater fan	age
	戏	吸毒	戏迷	吸毒	戏迷	和平
13	xi4	xi1du2	xi4mi2	xi1du2	xi4mi2	he2ping2
	drama	take drugs	theater fan	take drugs	theater fan	peace
	帮	帮忙	棒球	帮忙	棒球	完成
	bang1	bang1mang2	bang4qiu2	bang1mang2	bang4qiu2	wan2cheng2
14	help	help	baseball	help	baseball	complete
	棒	帮忙	棒球	帮忙	棒球	成熟
	bang4	bang1mang2	bang4qiu2	bang1mang2	bang4qiu2	cheng2shu2
	stick	help	baseball	help	baseball	mature
15	归	规模	贵族	规模	贵族	前台
	gui1	gui1mo2	gui4zu2	gui1mo2	gui4zu2	qian2tai2
	return	scale	noble	scale	noble	front desk
	贵	规模	贵族	规模	贵族	文明
16	gui4	gui1mo2	gui4zu2	gui1mo2	gui4zu2	wen2ming2
	expensive	scale	noble	scale	noble	civilization
	香	香肠	象棋	香肠	象棋	丛林
	xiang1	xiang1chang2	xiang4qi2	xiang1chang2	xiang4qi2	cong2lin2
17	fragrant	sausage	chess	sausage	chess	jungle
	象	香肠	象棋	香肠	象棋	垂直
	xiang4	xiang1chang2	xiang4qi2	xiang1chang2	xiang4qi2	chui2zhi2
	elephant	sausage	chess	sausage	chess	vertical

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
16	征	征服	证实	征服	证实	寻常
	zheng1	zheng1fu2	zheng4shi2	zheng1fu2	zheng4shi2	xun2chang2
	<i>expedition</i>	<i>conquer</i>	<i>confirm</i>	<i>conquer</i>	<i>confirm</i>	<i>usual</i>
	证	征服	证实	征服	证实	迷人
17	zheng4	zheng1fu2	zheng4shi2	zheng1fu2	zheng4shi2	mi2ren2
	<i>proof</i>	<i>conquer</i>	<i>confirm</i>	<i>conquer</i>	<i>confirm</i>	<i>charming</i>
	伤	伤亡	上楼	伤亡	上楼	雷达
	shang1	shang1wang2	shang4lou2	shang1wang2	shang4lou2	lei2da2
18	<i>wound</i>	<i>casualties</i>	<i>go upstairs</i>	<i>casualties</i>	<i>go upstairs</i>	<i>radar</i>
	上	伤亡	上楼	伤亡	上楼	为难
	shang4	shang1wang2	shang4lou2	shang1wang2	shang4lou2	wei2nan2
	<i>on top</i>	<i>casualties</i>	<i>go upstairs</i>	<i>casualties</i>	<i>go upstairs</i>	<i>feel awkward</i>
19	交	交流	教堂	交流	教堂	符合
	jiao1	jiao1liu2	jiao4tang2	jiao1liu2	jiao4tang2	fu2he2
	<i>deliver</i>	<i>communication</i>	<i>church</i>	<i>communication</i>	<i>church</i>	<i>accord with</i>
	教	交流	教堂	交流	教堂	灵魂
20	jiao4	jiao1liu2	jiao4tang2	jiao1liu2	jiao4tang2	ling2hun2
	<i>teaching</i>	<i>communication</i>	<i>church</i>	<i>communication</i>	<i>church</i>	<i>soul</i>
	失	失明	世俗	失明	世俗	残疾
	shi1	shi1ming2	shi4su2	shi1ming2	shi4su2	can2ji2
21	<i>lose</i>	<i>go blind</i>	<i>secular</i>	<i>go blind</i>	<i>secular</i>	<i>disabled</i>
	世	失明	世俗	失明	世俗	油田
	shi4	shi1ming2	shi4su2	shi1ming2	shi4su2	you2tian2
	<i>generation</i>	<i>go blind</i>	<i>secular</i>	<i>go blind</i>	<i>secular</i>	<i>oilfield</i>

ID	Prime	Target					Baseline
		SC_T1	XM_T4	XM_T1	SC_T4		
20	升	升学	剩余	升学	剩余	循环	
	sheng1	sheng1xue2	sheng4yu2	sheng1xue2	sheng4yu2	xun2huan2	
	rise	enter a higher school	remainder	enter a higher school	remainder	cycle	
	剩	升学	剩余	升学	剩余	难题	
21	sheng4	sheng1xue2	sheng4yu2	sheng1xue2	sheng4yu2	nan2ti2	
	remainder	enter a higher school	remainder	enter a higher school	remainder	problem	
	推	推移	退学	推移	退学	纯洁	
	tui1	tui1yi2	tui4xue2	tui1yi2	tui4xue2	chun2jie2	
22	push	elapse	drop out	elapse	drop out	pure	
	退	推移	退学	推移	退学	频繁	
	tui4	tui1yi2	tui4xue2	tui1yi2	tui4xue2	pin2fan2	
	retreat	elapse	drop out	elapse	drop out	frequent	
23	搭	搭乘	大桥	搭乘	大桥	延迟	
	da1	da1cheng2	da4qiao2	da1cheng2	da4qiao2	yan2chi2	
	build	travel by	bridge	travel by	bridge	delay	
	大	搭乘	大桥	搭乘	大桥	学员	
24	da4	da1cheng2	da4qiao2	da1cheng2	da4qiao2	xue2yuan2	
	big	travel by	bridge	travel by	bridge	student	
	花	花瓶	化疗	花瓶	化疗	同谋	
	hua1	hua1ping2	hua4hiao2	hua1ping2	hua4hiao2	tong2mou2	
25	flower	nase	chemotherapy	nase	chemotherapy	conspirator	
	化	花瓶	化疗	花瓶	化疗	传闻	
	hua4	hua1ping2	hua4hiao2	hua1ping2	hua4hiao2	chuan2wen2	
	transform	nase	chemotherapy	nase	chemotherapy	rumor	

ID	Prime	Target				
		SC_T1	XM_T4	XM_T1	SC_T4	Baseline
24	飞	飞翔	费城	飞翔	费城	临床
	fei1	fei1xiang2	fei4cheng2	fei1xiang2	fei4cheng2	lin2chuang2
	fly	fly	Philadelph ia	fly	Philadelph ia	clinical
	费	飞翔	费城	飞翔	费城	才华
fei4	fei1xiang2	fei4cheng2	fei1xiang2	fei4cheng2	cai2hua2	
fee	fly	Philadelph ia	fly	Philadelph ia	talent	
25	低	低头	地牢	低头	地牢	拳王
	di1	di1tou2	di4lao2	di1tou2	di4lao2	quan2wang2
	low	lover one's head	dungeon	lover one's head	dungeon	boxing champion
	地	低头	地牢	低头	地牢	划船
di4	di1tou2	di4lao2	di1tou2	di4lao2	hua2c2huan2	
ground	lover one's head	dungeon	lover one's head	dungeon	boating	
26	激	激活	继承	激活	继承	阳台
	ji1	ji1huo2	ji4cheng2	ji1huo2	ji4cheng2	yang2tai2
	stimulate	activate	inherit	activate	inherit	balcony
	技	激活	继承	激活	继承	连环
ji4	ji1huo2	ji4cheng2	ji1huo2	ji4cheng2	lian2huan2	
skill	activate	inherit	activate	inherit	serial	
27	枯	枯竭	酷刑	枯竭	酷刑	排查
	gu1	ku1jie2	ku4xing2	ku1jie2	ku4xing2	pai2cha2
	withered	exhausted	cruel torture	exhausted	cruel torture	troubleshot
	酷	枯竭	酷刑	枯竭	酷刑	缠绵
ku1	ku1jie2	ku4xing2	ku1jie2	ku4xing2	chan2mian2	
cool	exhausted	cruel torture	exhausted	cruel torture	lingering	

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
28	欺	欺凌	汽笛	欺凌	汽笛	红茶
	qi1	qi1ling2	qi4di2	qi1ling2	qi4di2	hong2cha2
	bully	<i>bully and humiliate</i>	<i>siren</i>	<i>bully and humiliate</i>	<i>siren</i>	<i>black tea</i>
	汽	欺凌	汽笛	欺凌	汽笛	闲谈
29	招	招摇	照明	招摇	照明	牛群
	zhao1	zhao1yao2	zhao4ming2	zhao1yao2	zhao4ming2	niu2qun2
	recruit	<i>flaunt</i>	<i>illuminate</i>	<i>flaunt</i>	<i>illuminate</i>	<i>cattle</i>
	照	招摇	照明	招摇	照明	石棉
30	优	优良	幼童	优良	幼童	顽皮
	you1	you1liang2	you4tong2	you1liang2	you4tong2	wan2pi2
	excellent	<i>good</i>	<i>young children</i>	<i>good</i>	<i>young children</i>	<i>naughty</i>
	幼	优良	幼童	优良	幼童	云层
31	松	松弛	送别	松弛	送别	翱翔
	song1	song1chi2	song4bie2	song1chi2	song4bie2	ao2xiang2
	pine	<i>loose</i>	<i>farewell</i>	<i>loose</i>	<i>farewell</i>	<i>hover</i>
	送	松弛	送别	松弛	送别	繁华
send	song1chi2	song4bie2	song1chi2	song4bie2	fan2hua2	
	<i>loose</i>	<i>farewell</i>	<i>loose</i>	<i>farewell</i>	<i>prosperous</i>	

ID	Prime	Target					Baseline
		SC_T1	XM_T4	XM_T1	SC_T4		
32	包	包围	报仇	包围	报仇	来源	
	bao1	bao1wei2	bao4chou2	bao1wei2	bao4chou2	lai2yuan2	
	bag	surround	revenge	surround	revenge	source	
	报	包围	报仇	包围	报仇	平民	
33	bao4	bao1wei2	bao4chou2	bao1wei2	bao4chou2	ping2min2	
	newspaper	surround	revenge	surround	revenge	cinlian	
	高	高潮	告别	高潮	告别	着迷	
	gao1	gao1chao2	gao4bie2	gao1chao2	gao4bie2	zhao2mi2	
34	high	climax	bid farewell to	climax	bid farewell to	fascinated	
	告	高潮	告别	高潮	告别	牛排	
	gao4	gao1chao2	gao4bie2	gao1chao2	gao4bie2	niu2pai2	
	tell	climax	bid farewell to	climax	bid farewell to	steak	
35	偷	偷袭	透明	偷袭	透明	弹琴	
	tou1	tou1xi2	tou4ming2	tou1xi2	tou4ming2	tan2qin2	
	steal	sneak attack	transparent	sneak attack	transparent	play the piano	
	透	偷袭	透明	偷袭	透明	神灵	
36	tou4	tou1xi2	tou4ming2	tou1xi2	tou4ming2	shen2ling2	
	penetrate	sneak attack	transparent	sneak attack	transparent	gods	
	纠	纠结	救活	纠结	救活	摇篮	
	jiu1	jiu1jie2	jiu4huo2	jiu1jie2	jiu4huo2	yao2lan2	
37	correct	conflicted	bring sb back to life	conflicted	bring sb back to life	cradle	
	救	纠结	救活	纠结	救活	重逢	
	jiu4	jiu1jie2	jiu4huo2	jiu1jie2	jiu4huo2	chong2feng2	
	save	conflicted	bring sb back to life	conflicted	bring sb back to life	reunion	

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
36	风 feng1 <i>wind</i>	风情 feng1qing2 <i>flirtatious expression</i>	奉陪 feng4pei2 <i>keep sb company</i>	风情 feng1qing2 <i>flirtatious expression</i>	奉陪 feng4pei2 <i>keep sb company</i>	埋藏 mai2cang2 <i>bury</i>
	奉 feng4 <i>wait on</i>	风情 feng1qing2 <i>flirtatious expression</i>	奉陪 feng4pei2 <i>keep sb company</i>	风情 feng1qing2 <i>flirtatious expression</i>	奉陪 feng4pei2 <i>keep sb company</i>	查询 cha2xun2 <i>inquire</i>
	星 xing1 <i>star</i>	星球 xing1qiu2 <i>planet</i>	幸福 xing4fu2 <i>happiness</i>	星球 xing1qiu2 <i>planet</i>	幸福 xing4fu2 <i>happiness</i>	神奇 shen2qi2 <i>miracle</i>
	幸 xing4 <i>lucky</i>	星球 xing1qiu2 <i>planet</i>	幸福 xing4fu2 <i>happiness</i>	星球 xing1qiu2 <i>planet</i>	幸福 xing4fu2 <i>happiness</i>	随时 sui2shi2 <i>at any time</i>
37	公 gong1 <i>public</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	来临 lai2lin2 <i>approach</i>
	共 gong4 <i>all together</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	传奇 chuan2qi2 <i>legend</i>
	歌 ge1 <i>song</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	延长 yan2chang2 <i>extend</i>
	个 ge4 <i>individual</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	条纹 tiao2wen2 <i>stripe</i>
38	歌 ge1 <i>song</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	延长 yan2chang2 <i>extend</i>
	个 ge4 <i>individual</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	条纹 tiao2wen2 <i>stripe</i>
	公 gong1 <i>public</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	来临 lai2lin2 <i>approach</i>
	共 gong4 <i>all together</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	传奇 chuan2qi2 <i>legend</i>
39	歌 ge1 <i>song</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	延长 yan2chang2 <i>extend</i>
	个 ge4 <i>individual</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	歌迷 ge1mi2 <i>fan</i>	条纹 tiao2wen2 <i>stripe</i>
	公 gong1 <i>public</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	来临 lai2lin2 <i>approach</i>
	共 gong4 <i>all together</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	公爵 gong1jue2 <i>duke</i>	传奇 chuan2qi2 <i>legend</i>

ID	Prime	Target					Baseline
		SC_T1	XM_T4	XM_T1	SC_T4		
40	窝	窝藏	卧房	窝藏	卧房	茶壶	
	wol	wo1cang2	wo4fang2	wo1cang2	wo4fang2	cha2hu2	
	nest	shelter	bedroom	shelter	bedroom	teapot	
	卧	窝藏	卧房	窝藏	卧房	存折	
	wol	wo1cang2	wo4fang2	wo1cang2	wo4fang2	cun2zhe2	
	lie	shelter	bedroom	shelter	bedroom	deposit book	

Table D2. The complete set of nonword target stimuli in Chapter 5. For each ID, the first row is with a SC_T3 prime; the second row is with a XM_T4 prime.

ID	Prime	Target				
		SC_T1	XM_T4	XM_T1	SC_T4	Baseline
1	马	马夺	妈苗	马夺	妈苗	这昨
	ma3(horse)	ma3duo2	ma1miao2	ma3duo2	ma1miao2	zh4zuo2
	妈	马夺	妈苗	马夺	妈苗	厂闲
	ma1(mother)	ma3duo2	ma1miao2	ma3duo2	ma1miao2	chang3xian2
2	寡	寡杂	瓜敌	寡杂	瓜敌	那贫
	gua3(cu)	gua3za2	gua1di2	gua3za2	gua1di2	na4pin2
	瓜	寡杂	瓜敌	寡杂	瓜敌	鸟盐
	gua1(melon)	gua3za2	gua1di2	gua3za2	gua1di2	niao3yan2
3	瓦	瓦粮	挖潮	瓦粮	挖潮	纵格
	wa3(tile)	wa3liang2	wa1chao2	wa3liang2	wa1chao2	zong4ge2
	挖	瓦粮	挖潮	瓦粮	挖潮	打谈
	wa1(dig)	wa3liang2	wa1chao2	wa3liang2	wa1chao2	da3tan2
4	扯	扯胡	车茄	扯胡	车茄	怕群
	che3(pull)	che3hu2	che1qie2	che3hu2	che1qie2	pa4qun2
	车	扯胡	车茄	扯胡	车茄	响棉
	che1(car)	che3hu2	che1qie2	che3hu2	che1qie2	xiang3mian2
5	躲	躲随	多卢	躲随	多卢	跨垂
	duo3(hide)	duo3sui2	duo1lu2	duo3sui2	duo1lu2	kua4chui2
	多	躲随	多卢	躲随	多卢	火如
	duo1(many)	duo3sui2	duo1lu2	duo3sui2	duo1lu2	huo3ru2

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
6	锁	锁停	缩曹	锁停	缩曹	赤除
	suo3(lock)	suo3ting2	suo1cao2	suo3ting2	suo1cao2	chi4chu2
	缩	锁停	缩曹	锁停	缩曹	写床
7	suo1(shrink)	suo3ting2	suo1cao2	suo3ting2	suo1cao2	xie3chuang2
	野	野宁	耶霞	野宁	耶霞	醋阜
	ye3(mild)	ye3ning2	ye1xia2	ye3ning2	ye1xia2	cu4zhuo2
8	耶	野宁	耶霞	野宁	耶霞	死瞞
	ye1(jesus)	ye3ning2	ye1xia2	ye3ning2	ye1xia2	si3man2
	紫	紫渠	姿泥	紫渠	姿泥	滤熬
9	zi3(purple)	zi3qu2	zi1ni2	zi3qu2	zi1ni2	lv4ao2
	姿	紫渠	姿泥	紫渠	姿泥	猛牙
	zi1(posture)	zi3qu2	zi1ni2	zi3qu2	zi1ni2	meng3ya2
10	彼	彼叠	逼黄	彼叠	逼黄	妙残
	bi3(another)	bi3die2	bi1huang2	bi3die2	bi1huang2	miao4can2
	逼	彼叠	逼黄	彼叠	逼黄	妥孩
11	bi1(force)	bi3die2	bi1huang2	bi3die2	bi1huang2	tuo3hai2
	普	普眉	扑虎	普眉	扑虎	货邻
	pu3(universal)	pu3mei2	pu1pang2	pu3mei2	pu1pang2	huo4lin2
12	扑	普眉	扑虎	普眉	扑虎	朗来
	pu1(rub at)	pu3mei2	pu1pang2	pu3mei2	pu1pang2	lang3lai2
	腐	腐狂	肤楼	腐狂	肤楼	颯宅
13	fu3(rotten)	fu3kuang2	fu1lou2	fu3kuang2	fu1lou2	ni4zhai2
	肤	腐狂	肤楼	腐狂	肤楼	果鞞
	fu1(skin)	fu3kuang2	fu1lou2	fu3kuang2	fu1lou2	guo3han2

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
12	祖	祖绳	租罚	祖绳	租罚	利荷
	zu3(ancestor)	zu3sheng2	zu1fa2	zu3sheng2	zu1fa2	li4he2
	租	祖绳	租罚	祖绳	租罚	补匣
	zu1(rent)	zu3sheng2	zu1fa2	zu3sheng2	zu1fa2	bu3xia2
13	取	取朋	区仍	取朋	区仍	屁德
	qu3(fetch)	qu3peng2	qu1reng2	qu3peng2	qu1reng2	pi4de2
	区	取朋	区仍	取朋	区仍	顶桌
	qu1(distrid)	qu3peng2	qu1reng2	qu3peng2	qu1reng2	ding3quan2
14	伟	伟仇	威平	伟仇	威平	育浑
	wei3(great)	wei3chou2	wei1ping2	wei3chou2	wei1ping2	yu4hun2
	威	伟仇	威平	伟仇	威平	等虫
	wei1(prestige)	wei3chou2	wei1ping2	wei3chou2	wei1ping2	deng3chong2
15	跑	跑茄	抛藏	跑茄	抛藏	憎白
	pao3(run)	pao3qie2	pao1cang2	pao3qie2	pao1cang2	zeng4bai2
	抛	跑茄	抛藏	跑茄	抛藏	米才
	pao1(thron)	pao3qie2	pao1cang2	pao3qie2	pao1cang2	mi3cai2
16	岛	岛拔	刀吴	岛拔	刀吴	胖鞋
	dao3(island)	dao3ba2	dao1wu2	dao3ba2	dao1wu2	pang4xie2
	刀	岛拔	刀吴	岛拔	刀吴	体谋
	dao1(knife)	dao3ba2	dao1wu2	dao3ba2	dao1wu2	ti3mei2
17	讨	讨驴	涛红	讨驴	涛红	趣民
	tao3(ask for)	tao3lv2	tao1hong2	tao3lv2	tao1hong2	qu4min2
	涛	讨驴	涛红	讨驴	涛红	老存
	tao1(big wave)	tao3lv2	tao1hong2	tao3lv2	tao1hong2	lao3cun2

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
18	早	早挪	遭牛	早挪	遭牛	对男
	zao3(early)	zao3nuo2	zao1niu2	zao3nuo2	zao1niu2	dui4nan2
	遭	早挪	遭牛	早挪	遭牛	努环
	zao1(suffer)	zao3nuo2	zao1niu2	zao3nuo2	zao1niu2	nu3huan2
	草	草痢	操强	草痢	操强	夏臀
19	cao3(grass)	cao3que2	cao1qiang2	cao3que2	cao1qiang2	xia4tun2
	操	草痢	操强	草痢	操强	楼坟
	cao1(exercise)	cao3que2	cao1qiang2	cao3que2	cao1qiang2	lou3fen2
	少	少爬	烧昨	少爬	烧昨	用鞅
	shao3(little)	shao3pa2	shao1zuo2	shao3pa2	shao1zuo2	yong4e2
20	烧	少爬	烧昨	少爬	烧昨	母文
	shao1(burn)	shao3pa2	shao1zuo2	shao3pa2	shao1zuo2	mu3wen2
	表	表狼	标娃	表狼	标娃	配付
	biao3(table)	biao3lang2	biao1wa2	biao3lang2	biao1wa2	pei4fu2
	标	表狼	标娃	表狼	标娃	虎鞋
21	biao1(sign)	biao3lang2	biao1wa2	biao3lang2	biao1wa2	hu3xie2
	丑	丑瓢	抽霞	丑瓢	抽霞	暴旋
	chou3(ugly)	chou3piao2	chou1xia2	chou3piao2	chou1xia2	bao4xuan2
	抽	丑瓢	抽霞	丑瓢	抽霞	冷元
	chou1(pump)	chou3piao2	chou1xia2	chou3piao2	chou1xia2	leng3yuan2
23	手	手苗	收葵	手苗	收葵	梦年
	shou3(hand)	shou3miao2	shou1kui2	shou3miao2	shou1kui2	meng4nian2
	收	手苗	收葵	手苗	收葵	扭神
	shou1(receive)	shou3miao2	shou1kui2	shou3miao2	shou1kui2	niu3shen2

ID	Prime	Target					Baseline
		SC_T1	XM_T4	XM_T1	SC_T4		
24	狗	狗妮	沟浮	狗妮	沟浮	漏林	
	<i>gou3(dog)</i>	<i>gou3ni2</i>	<i>gou1fu2</i>	<i>gou3ni2</i>	<i>gou1fu2</i>	<i>lou4lin2</i>	
	沟	狗妮	沟浮	狗妮	沟浮	涨秦	
	<i>gou1(ditch)</i>	<i>gou3ni2</i>	<i>gou1fu2</i>	<i>gou3ni2</i>	<i>gou1fu2</i>	<i>zhang3qin2</i>	
	访	访驼	方聊	访驼	方聊	另银	
25	<i>fang3(visi)</i>	<i>fang3tuo2</i>	<i>fang1liao2</i>	<i>fang3tuo2</i>	<i>fang1liao2</i>	<i>ling4yin2</i>	
	方	访驼	方聊	访驼	方聊	考繁	
	<i>fang1(square)</i>	<i>fang3tuo2</i>	<i>fang1liao2</i>	<i>fang3tuo2</i>	<i>fang1liao2</i>	<i>kao3fan2</i>	
	党	党姚	当鱼	党姚	当鱼	胜埋	
	<i>dang3(party)</i>	<i>dang3yao2</i>	<i>dang1yu2</i>	<i>dang3yao2</i>	<i>dang1yu2</i>	<i>sheng4mai2</i>	
26	当	党姚	当鱼	党姚	当鱼	可颜	
	<i>dang1(serve as)</i>	<i>dang3yao2</i>	<i>dang1yu2</i>	<i>dang3yao2</i>	<i>dang1yu2</i>	<i>ke3yan2</i>	
	耢	耢矛	汤伐	耢矛	汤伐	抗盆	
	<i>tang3(he down)</i>	<i>tang3mao2</i>	<i>tang1fa2</i>	<i>tang3mao2</i>	<i>tang1fa2</i>	<i>kang4pen2</i>	
	汤	耢矛	汤伐	耢矛	汤伐	走盘	
27	<i>tang1(soup)</i>	<i>tang3mao2</i>	<i>tang1fa2</i>	<i>tang3mao2</i>	<i>tang1fa2</i>	<i>zou3pan2</i>	
	港	港冯	刚详	港冯	刚详	岁茶	
	<i>gang3(harbor)</i>	<i>gang3feng2</i>	<i>gang1xiang2</i>	<i>gang3feng2</i>	<i>gang1xiang2</i>	<i>sui4cha2</i>	
	刚	港冯	刚详	港冯	刚详	扰然	
	<i>gang1(just)</i>	<i>gang3feng2</i>	<i>gang1xiang2</i>	<i>gang3feng2</i>	<i>gang1xiang2</i>	<i>rao3ran2</i>	
29	讲	讲衙	江奴	讲衙	江奴	妹群	
	<i>jiang3(speech)</i>	<i>jiang3ya2</i>	<i>jiang1nu2</i>	<i>jiang3ya2</i>	<i>jiang1nu2</i>	<i>mei4qun2</i>	
	江	讲衙	江奴	讲衙	江奴	旅田	
	<i>jiang1(river)</i>	<i>jiang3ya2</i>	<i>jiang1nu2</i>	<i>jiang3ya2</i>	<i>jiang1nu2</i>	<i>lv3tian2</i>	

ID	Prime	Target				Baseline
		SC_T1	XM_T4	XM_T1	SC_T4	
30	枪	枪毫	枪唯	枪毫	枪唯	翠时
	qiang3(rob)	qiang3hao2	qiang1wei2	qiang3hao2	qiang1wei2	cui4shi2
	枪	枪毫	枪唯	枪毫	枪唯	扫怀
	qiang1(gun)	qiang3hao2	qiang1wei2	qiang3hao2	qiang1wei2	sao3huai2
	养	养拿	央曾	养拿	央曾	挂云
31	yang3(raise)	yang3na2	yang1ceng2	yang3na2	yang1ceng2	gua4yun2
	央	养拿	央曾	养拿	央曾	惹容
	yang1(center)	yang3na2	yang1ceng2	yang3na2	yang1ceng2	re3rong2
	广	广罗	光娘	广罗	光娘	坐季
	guang3(wide)	guang3luo2	guang1niang2	guang3luo2	guang1niang2	zuo4luan2
32	光	广罗	光娘	广罗	光娘	女门
	guang1(light)	guang3luo2	guang1niang2	guang3luo2	guang1niang2	nv3men2
	谎	谎佛	慌哲	谎佛	慌哲	翁牌
	huang3(lie)	huang3fo2	huang1zhe2	huang3fo2	huang1zhe2	weng4pen2
	慌	谎佛	慌哲	谎佛	慌哲	巧联
33	huang1(nervous)	huang3fo2	huang1zhe2	huang3fo2	huang1zhe2	qiao3lian2
	等	等狂	登疼	等狂	登疼	闭熟
	deng3(wait)	deng3kuang2	deng1teng2	deng3kuang2	deng1teng2	bi4shou2
	登	等狂	登疼	等狂	登疼	脑唇
	deng1(climb)	deng3kuang2	deng1teng2	deng3kuang2	deng1teng2	nao3chun2
34	挺	挺梨	厅绳	挺梨	厅绳	栗国
	ting3	ting3li2	ting1sheng2	ting3li2	ting1sheng2	piao4guo2
	挺(rubber)	ting3li2	厅绳	挺梨	厅绳	雅寻
	厅	挺梨	厅绳	挺梨	厅绳	ya3xun2
	ting1(ball)	ting3li2	ting1sheng2	ting3li2	ting1sheng2	

ID	Prime	Target				
		SC_T1	XM_T4	XM_T1	SC_T4	Baseline
36	请	请奴	清持	请奴	清持	二竹
	qing3(request)	qing3nu2	qing1chi2	qing3nu2	qing1chi2	er4zhu2
	清	请奴	清持	请奴	清持	美坟
	qing1(clear)	qing3nu2	qing1chi2	qing3nu2	qing1chi2	mei3fen2
37	懂	懂柔	冬麻	懂柔	冬麻	次谈
	dong3(understand)	dong3rou2	dong1ma2	dong3rou2	dong1ma2	ci4tan2
	冬	懂柔	冬麻	懂柔	冬麻	雨槐
	dong1(winter)	dong3rou2	dong1ma2	dong3rou2	dong1ma2	yu3huai2
38	统	统猴	通陶	统猴	通陶	夜赎
	tong3(unite)	tong3hou2	tong1tao2	tong3hou2	tong1tao2	ye4shu2
	通	统猴	通陶	统猴	通陶	嘴抬
	tong1(go through)	tong3hou2	tong1tao2	tong3hou2	tong1tao2	zui3tai2
39	总	总梅	宗揉	总梅	宗揉	暮缠
	zong3(total)	zong3mei2	zong1rou2	zong3mei2	zong1rou2	mu4chan2
	宗	总梅	宗揉	总梅	宗揉	苦蛮
	zong1(action)	zong3mei2	zong1rou2	zong3mei2	zong1rou2	ku3man2
40	勇	勇棚	拥狭	勇棚	拥狭	亮儒
	yong3(brave)	yong3peng2	yong1xia2	yong3peng2	yong1xia2	liang4ru2
	拥	勇棚	拥狭	勇棚	拥狭	囤钱
	yong1(hug)	yong3peng2	yong1xia2	yong3peng2	yong1xia2	jun2qian2

Table D3. *Response accuracy and average reaction time (SEs in parentheses) for different target types preceded by SC_T1 primes and XM_T4 primes in Chapter 5.*

Target type	Response accuracy (%)		Reaction time (ms)	
	SC_T1 prime	XM_T4 prime	SC_T1 prime	XM_T4 prime
Identical	92.0 (1.0)	77.3 (1.5)	492 (13)	611 (14)
D+Seg	90.4 (1.1)	90.3 (1.2)	479 (12)	503 (11)
D-Seg	89.0 (1.1)	91.3 (1.2)	504 (11)	479 (12)
D-Homophone	76.8 (1.7)	90.0 (1.3)	626 (14)	489 (13)
Baseline	90.4 (1.9)	92.8 (1.7)	511 (13)	547 (13)

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition: International Journal of Cognitive Science*, 73(3), 247-264.
[doi:http://dx.doi.org/10.1016/S0010-0277\(99\)00059-1](http://dx.doi.org/10.1016/S0010-0277(99)00059-1)
- Azizian, A., Freitas, A. L., Watson, T. D., & Squires, N. K. (2006). Electrophysiological correlates of categorization: P300 amplitude as index of target similarity. *Biological Psychology*, 71(3), 278-288.
[doi:http://dx.doi.org/10.1016/j.biopsycho.2005.05.002](http://dx.doi.org/10.1016/j.biopsycho.2005.05.002)
- Bai D. (1954). *Guanzhong fangyan diaocha baogao* [A report on Guanzhong dialect]. Beijing: Chinese Academy of Sciences. (in Chinese)
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. Retrieved from <https://CRAN.R-project.org/package=lme4>
- Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11(1-2), 17-68. doi:10.1080/016909696387213
- Bishop, J. (2012). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (pp. 239-270). Berlin: De Gruyter Mouton.
- Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer (Version 5.4.2) [Computer program]. Retrieved from <http://www.praat.org/>
- Bogacz, R., Brown, E., Moehlis, J., & D.Cohen, J. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, 113(4), 700-765.
- Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics*, 39(4), 585-594.
[doi:http://dx.doi.org/10.1016/j.wocn.2011.06.002](http://dx.doi.org/10.1016/j.wocn.2011.06.002)
- Buxó-Lugo, A., & Watson, D. G. (2016). Evidence for the influence of syntax on prosodic parsing. *Journal of Memory and Language*, 90, 1-13.
[doi:http://dx.doi.org/10.1016/j.jml.2016.03.001](http://dx.doi.org/10.1016/j.jml.2016.03.001)

- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5(6), e10729. doi:10.1371/journal.pone.0010729
- Cao, J. (2004). Intonation structure of spoken Chinese: Universality and specificity. *Report of Phonetic Research*, 31-38.
- Chao, Y. R. (1929). Beijing yudiao de yanjiu [On intonation in Beijing Mandarin]. In Z. Wu & X. Zhao (Eds.), *Proceedings of Chao Yuen-Ren's linguistic papers* (pp. 253-271). Beijing: The Commercial Press. (in Chinese)
- Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique*, 45, 24-27.
- Chao, Y. R. (1933). Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology*, 4, 121-134.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chappell, H. (2001). *Sinitic Languages of China. Typological Descriptions*. Berlin, Boston: De Gruyter Mouton.
- Chen, H.-C., Vaid, J., & Wu, J.-T. (2009). Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes*, 24(7-8), 967-982. doi:10.1080/01690960902804515
- Ching, Y. C. (1985). Lipreading Cantonese with voice pitch. *Journal of the Acoustical Society of America*, 77, S39-40.
- Christensen, R. H. B. (2015). Ordinal — Regression models for ordinal data. R package version 2015.6.28. Retrieved from <http://www.cran.r-project.org/package=ordinal/>.
- Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, 1(2), 425-452. doi:10.1515/labphon.2010.022
- Coltheart, V., Avons, S. E., Masterson, J., & Laxon, V. J. (1991). The role of assembled phonology in reading comprehension. *Memory & Cognition*, 19, 387-400.
- Connell, B. A., Hogan, J. T., & Rozsypal, A. J. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics*, 11(4), 337-351.

- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*(3), 207-228.
doi:10.1177/00238309020450030101
- Courchesne, E., Hillyard, S. A., & Courchesne, R. Y. (1977). P3 waves to the discrimination of targets in homogeneous and heterogeneous stimulus sequences. *Psychophysiology, 14*(6), 590-597.
- Cutler, A., & Chen, H.-C. (1995). *Phonological similarity effects in Cantonese word recognition*. Paper presented at the ICPs, Stocklom, Sweden.
- Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59*(2), 165-179.
doi:10.3758/BF03211886
- Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America, 105*, 1977-1988.
- Cutler, A., & Van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*(2), 171-195. doi:10.1177/00238309010440020301
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language & Speech, 40*(2), 141-201. doi:10.1177/002383099704000203
- Czap, L., & Zhao, L. (2017). *Phonetic aspects of Chinese Shaanxi Xi'an dialect*. Paper presented at the 2017 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom).
- Da, J. (2004). A corpus-based study of character and bigram frequencies in Chinese e-texts and its implications for Chinese language instruction. In P. Zhang, T. Xie & J. Xu (Eds.), *Proceedings of 4th International Conference on New Technologies in Teaching and Learning Chinese* (pp. 501-511). Beijing: The Tsinghua University Press.
- Daneman, M., & Stainton, M. (1991). Phonological recoding in silent reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 618-632.
- Daneman, M., Reingold, E. M., & Davidson, M. (1995). Time course of phonological activation during reading: Evidence from eye fixations.

- Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21(4), 884-898. doi:<http://dx.doi.org/10.1037/0278-7393.21.4.884>
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117-1121. doi:http://www.nature.com/neuro/journal/v8/n8/suppinfo/n1504_S1.html
- Dewaele, J.-M. (1999). Word order variation in interrogative structures of native and non-native French. *International Journal of Applied Linguistics*, 123, 161-180.
- Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41(4), 496-518. doi:<http://dx.doi.org/10.1006/jmla.1999.2654>
- Doctor, E. A., & Klein, D. (1992). Phonological processing in bilingual word recognition. *Advances in Psychology*, 83, 237-252. doi:[http://dx.doi.org/10.1016/S0166-4115\(08\)61498-3](http://dx.doi.org/10.1016/S0166-4115(08)61498-3)
- Dornisch, E. (1998). *Multiple-wh-questions in Polish: The interactions between wh-phrases and clitics* (PhD dissertation). Cornell University.
- Dow, F. D. M. (1972). A discussion on tone sandhi problems in Chinese. *Journal of the International Phonetic Association*, 2(1), 13-19. doi:10.1017/S0025100300000396
- Duanmu, S. (2007). *The phonology of Standard Chinese*. Oxford: OUP Oxford.
- Dufour, S., & Peereman, R. (2003). Lexical competition in phonological priming: Assessing the role of phonological match and mismatch lengths between primes and targets. *Memory & Cognition*, 31(8), 1271-1283. doi:10.3758/BF03195810
- Durrell, M. (2011). *Hammer's German grammar and usage (5th Edition)*. Routledge, New York: McGraw-Hill.
- Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20(6), 641-655. doi:[https://doi.org/10.1016/S0022-5371\(81\)90220-6](https://doi.org/10.1016/S0022-5371(81)90220-6)

- Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology*, *44*(4), 491-505.
doi:10.1111/j.1469-8986.2007.00531.x
- Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Cognitive Brain Research*, *8*(3), 373-392.
[doi:http://dx.doi.org/10.1016/S0926-6410\(99\)00036-1](http://dx.doi.org/10.1016/S0926-6410(99)00036-1)
- Ferrand, L., & Grainger, J. (2003). Homophone interference effects in visual word recognition. *The Quarterly Journal of Experimental Psychology*, *56*(3), 403-419. doi:10.1080/02724980244000422
- Fok-Chan, Y.-Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: University of Hong Kong Press.
- Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics*, *13*, 69-90.
- Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268-294.
[doi:http://dx.doi.org/10.1016/j.wocn.2007.06.005](http://dx.doi.org/10.1016/j.wocn.2007.06.005)
- Frank, S. L. (2013). Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science*, *5*(3), 475-494.
doi:10.1111/tops.12025
- Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials. *Journal of Speech, Language and Hearing Research*, *48*, 1496-1510.
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *TRENDS in Cognitive Sciences*, *6*(2), 78-84.
- Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention — Focusing the searchlight on sound. *Current Opinion in Neurobiology*, *17*(4), 437-455.
[doi:http://dx.doi.org/10.1016/j.conb.2007.07.011](http://dx.doi.org/10.1016/j.conb.2007.07.011)
- Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149-175.

- Gandour, J. T. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, 12(2), 235-261.
- Gandour, J. T., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*, 21(1), 1-33. doi:10.1177/002383097802100101
- Gandour, J., Ponglorpisit, S., Khunadorn, F., Dechongkit, S., Boongird, P., Boonklam, R., & Potisuk, S. (1992). Lexical tones in Thai after unilateral brain damage. *Brain and Language*, 43(2), 275-307.
doi:http://dx.doi.org/10.1016/0093-934X(92)90131-W
- Gårding, E. (1987). Speech act and tonal pattern in Standard Chinese: Constancy and variation. *Phonetica: International Journal of Speech Science*, 44(1), 13-29. Retrieved from
<http://www.karger.com/DOI/10.1159/000261776>
- Garellek, M., Keating, P., Esposito, C. M., & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *The Journal of the Acoustical Society of America*, 133(2), 1078-1089. doi:http://dx.doi.org/10.1121/1.4773259
- Grosjean, F. (1988). Exploring the recognition of guest words in bilingual speech. *Language and Cognitive Processes*, 3(3), 233-274.
doi:10.1080/01690968808402089
- Grosjean, F. (1997). Processing mixed language: Issues, findings and models. In A. M. B. d. Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives*. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Gussenhoven, C., & Chen, A. (2000). Universal and language-specific effects in the perception of question intonation. In: *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP 2000)*, 91-94.
- Haigh, C. A., & Jared, D. (2007). The activation of phonological representations by bilinguals while reading silently: Evidence from interlingual homophones. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(4), 623-644.

- Hale, J. (2001). *A probabilistic earley parser as a psycholinguistic model*. Paper presented at the NAACL, Pittsburgh, Pennsylvania.
- Hao, Y.-C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269-279. doi:10.1016/j.wocn.2011.11.001
- Hashimoto, A. O.-k. Y. (1972). *Phonology of Cantonese*. Cambridge: Cambridge University Press.
- Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, 182, 177-180.
- Ho, A. T. (1977). Intonation variation in a Mandarin sentence for three expressions: Interrogative, exclamatory and declarative. *Phonetica*, 34(6), 446-457. Retrieved from <http://www.karger.com/DOI/10.1159/000259916>
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Hu, M. (1987). *Guanyu Beijinghua de yudiao wenti* [On the intonation of Beijing Mandarin]. Beijing: The Commerical Press. (in Chinese)
- Huang, T. (2007). *Perception of Mandarin tones by Chinese- and English-speaking listeners*. Paper presented at the the 16th International Congress of Phonetic Sciences, Saarbrücken, Germany.
- Isel, F., Alter, K., & Friederici, A. D. (2005). Influence of prosodic information on the processing of split particles: ERP evidence from spoken German. *Journal of Cognitive Neuroscience*, 17(1), 154-167.
- Isreal, J. B., Chesney, G. L., Wickens, C. D., & Donchin, E. (1980). P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. *Psychophysiology*, 17(3), 259-273. doi:10.1111/j.1469-8986.1980.tb00146.x
- Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words? *Journal of Memory and Language*, 44(1), 2-31. doi:<http://dx.doi.org/10.1006/jmla.2000.2747>
- Jared, D., Levy, B. A., & Rayner, K. (1999). The role of phonology in the activation of word meanings during reading: Evidence from

- proofreading and eye movements. *Journal of Experimental Psychology: General*, 128(3), 219-264. doi:10.1037/0096-3445.128.3.219
- Johnson, R. (1984). P300: A model of the variables controlling its amplitude. *Annals of the New York Academy of Sciences*, 425, 223-229. doi:10.1111/j.1749-6632.1984.tb23538.x
- Johnson, R. (1986). A triarchic model of P300 amplitude. *Psychophysiology*, 23(4), 367-384. doi:10.1111/j.1469-8986.1986.tb00649.x
- Johnson, R., & Donchin, E. (1980). P300 and stimulus categorization: Two plus one is not so different from one plus one. *Psychophysiology*, 17(2), 167-178.
- Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in Mandarin Chinese and English speakers. *NeuroImage*, 13(4), 646-653. doi:http://dx.doi.org/10.1006/nimg.2000.0738
- Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, 38(3), 557-577. doi:10.1017/S0048577201990559
- Kotchoubey, B., & Lang, S. (2001). Event-related potentials in an auditory semantic oddball task in humans. *Neuroscience Letters*, 310, 93-96.
- Koutsoudas, A. (1968). On wh-words in English. *Journal of Linguistics*, 4(2), 267-273. doi:https://doi.org/10.1017/S0022226700001912
- Kratochvil, P. (1998). Intonation in Beijing Chinese. In D. Hirst & A. D. Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 417-431). Cambridge: Cambridge University Press.
- Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia*, 53, 293-309. doi:http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.020
- Kuong, I.-K. J. (2008). *Yes/no question particles revisited: The grammatical functions of mo4, me1, and maa3*. Paper presented at the Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20), Ohio.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32-59. doi:10.1080/23273798.2015.1102299

- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*, 161.
doi:10.1038/307161a0
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190-207): Oxford Scholarship Online.
- Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science*, *197*, 792-795.
- Ladd, D. R. (2008). *Intonational phonology*. Cambridge: Cambridge University Press.
- Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(4), 952-965. doi:10.1037/a0023217
- Lam, A. (2005). Language learning in China: The experience of four learners. *Reflections on English Language Teaching*, *4*, 1-14.
- Lee, C.-Y. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language & Speech*, *50*(1), 101-123. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=ufh&AN=24655341&site=ehost-live>
- Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, *32*(4), 533-550.
doi:10.3758/BF03195845
- Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, *69*(1), 33. doi:10.18637/jss.v069.i01
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition: International Journal of Cognitive Science*, *106*(3), 1126-1177.
doi:<https://doi.org/10.1016/j.cognition.2007.05.006>
- Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

- Li, D. C. S., & Lee, S. (2005). Bilingualism in East Asia. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 742-779). Malden, MA: Blackwell.
- Li, P. (2001). Xi'anhua yu Putonghua de yuyin duiying guilv [The correspondence pattern between Xi'an dialect and Putonghua in production]. *Journal of Xi'an Educational College*, 16(2), 57-61. (in Chinese)
- Li, R., & Stephen, W. (1987). *Language atlas of China*. Hong Kong: Longman.
- Li, X., Chen, Y., & Yang, Y. (2011). Immediate integration of different types of prosodic information during on-line spoken language comprehension: An ERP study. *Brain Research*, 1386, 139-152.
doi:<http://dx.doi.org/10.1016/j.brainres.2011.02.051>
- Liang, J., & Van Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes: Experimental studies in word and sentence prosody* (Vol. 12-2, pp. 27-61). Berlin/New York: Mouton de Gruyter.
- Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica*, 62(2-4), 70-87. Retrieved from <http://www.karger.com/DOI/10.1159/000090090>
- Liu, M., Chen, Y., & Schiller, N. O. (2016a). Context effects on tone and intonation processing in Mandarin. In: *Proceedings of Speech Prosody 2016*, 1056-1060. doi:10.21437/SpeechProsody.2016-217
- Liu, M., Chen, Y., & Schiller, N. O. (2016b). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia*, 91, 307-317. doi:<http://dx.doi.org/10.1016/j.neuropsychologia.2016.08.025>
- Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when F0 information is neutralized. *Language and Speech*, 47(2), 109-138.
doi:10.1177/00238309040470020101
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and hearing*, 19(1), 1-36.
- Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- Luo C., & Wang J. (1981). *Putong yuyanxue gangyao* [An introduction to general linguistics]. Beijing: The Commercial Press. (in Chinese)

- Luo, C. R., Johnson, R. A., & Gallo, D. A. (1998). Automatic activation of phonological information in reading: Evidence from the semantic relatedness decision task. *Memory & Cognition*, *26*(4), 833-843. doi:10.3758/BF03211402
- Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America*, *120*(6), 3978-3987. doi:http://dx.doi.org/10.1121/1.2363927
- Ma, J. K., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *Journal of the Acoustical Society of America*, *129*(2), 1012-1023. doi:10.1121/1.3531840
- Ma, M. (2005). Shaanxi Xi'an fangyan danziyin shengdiao shengxue shiyan yanjiu [Acoustic study of the tones of Xi'an dialect]. *Journal of Yan'an University (Social Science)*, *27*(4), 110-112. (in Chinese)
- Malins, J. G., & Joannis, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407-420. doi:10.1016/j.jml.2010.02.004
- Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97-115. doi:10.1017/S1366728903001068
- Marian, V., Blumenfeld, H. K., & Boukrina, O. V. (2008). Sensitivity to phonological similarity within and across Languages. *Journal of Psycholinguistic Research*, *37*(3), 141-170. doi:10.1007/s10936-007-9064-9
- Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*, 940-967. doi:10.1044/1092-4388(2007/067)
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition: International Journal of Cognitive Science*, *25*(1), 71-102. doi:[https://doi.org/10.1016/0010-0277\(87\)90005-9](https://doi.org/10.1016/0010-0277(87)90005-9)
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1-86. doi:http://dx.doi.org/10.1016/0010-0285(86)90015-0

- Mirman, D. (2008). Mechanisms of semantic ambiguity resolution: Insights from speech perception. *Research on Language and Computation*, 6(3), 293-309. doi:10.1007/s11168-008-9055-5
- Mirman, D. (2016). *Growth curve analysis and visualization using R*. Boca Raton: CRC.
- Morton, E. S. (1994). Sound symbolism and its role in non-human vertebrate communication. In L. Hinton, J. Nichols, & J. J. Ohala (Eds.), *Sound symbolism* (pp. 348–365). Cambridge: Cambridge University Press.
- 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), 1-21.
- Nas, G. (1983). Visual word recognition in bilinguals: Evidence for a cooperation between visual and sound based codes during access to a common lexical store. *Journal of Verbal Learning and Verbal Behavior*, 22(5), 526-534. doi:https://doi.org/10.1016/S0022-5371(83)90319-5
- Newman, S. D. (2012). The homophone effect during visual word recognition in children: An fMRI study. *Psychological Research*, 76(3), 280-291. doi:10.1007/s00426-011-0347-2
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychological Bulletin*, 131(4), 510-532. doi:10.1037/0033-2909.131.4.510
- Nolan, F. (2006). Intonation. In B. Aarts & A. McMahon (Eds.), *The handbook of English linguistics* (pp. 433-457). Malden, USA: Blackwell Publishing.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition: International Journal of Cognitive Science*, 52(3), 189-234. doi:http://dx.doi.org/10.1016/0010-0277(94)90043-4
- Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, 40(1), 1-18. Retrieved from <http://www.karger.com/DOI/10.1159/000261678>
- Ota, M., Hartsuiker, R. J., & Haywood, S. L. (2009). The KEY to the ROCK: Near-homophony in nonnative visual word recognition. *Cognition: International Journal of Cognitive Science*, 111(2), 263-269. doi:https://doi.org/10.1016/j.cognition.2008.12.007

- Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: Selected studies of the event-related potential. *International Journal of Medical Sciences*, 2(4), 147-154. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1252727/>
- Patro, C., & Mendel, L. L. (2016). Role of contextual cues on the perception of spectrally reduced interrupted speech. *The Journal of the Acoustical Society of America*, 140(2), 1336-1345.
[doi:doi:http://dx.doi.org/10.1121/1.4961450](http://dx.doi.org/10.1121/1.4961450)
- Peking University (1989). *Hanyu fangyin zihui (second version)* [A dictionary of pronunciations of characters in Chinese dialects]. Beijing: Wenzhi Gaige Chubanshe. (in Chinese)
- Peng, S.-h., Chan, M. K. M., Tseng, C.-y., Huang, T., Lee, O. J., & Beckman, M. E. (2005). Towards a pan-Mandarin system for prosodic transcription. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 230-270). Oxford: Oxford University Press.
- Peters, B., & Pfitzinger, H. R. (2008). Duration and F0 interval of utterance-final intonation contours in the perception of German sentence modality. In: *Proceedings of INTERSPEECH 2008*, 65-68.
- Pexman, P. M., Lupker, S. J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(1), 139-156. doi:10.1037/0278-7393.27.1.139
- Pfefferbaum, A., Ford, J., Johnson, R., Wenegrat, B., & Kopell, B. S. (1983). Manipulation of P3 latency: Speed vs. accuracy instructions. *Electroencephalography and Clinical Neurophysiology*, 55, 188-197.
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128-2148.
[doi:http://dx.doi.org/10.1016/j.clinph.2007.04.019](http://dx.doi.org/10.1016/j.clinph.2007.04.019)
- 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(1), 49-71.
[doi:10.1016/j.pneurobio.2006.04.004](http://dx.doi.org/10.1016/j.pneurobio.2006.04.004)
- Quirk, R., Greenbaum, S., & Leech, G. (1972). *A grammar of contemporary English*. London UK: Longman.

- R Core Team (2015). R: A language and environment for statistical computing. In R *Foundation for Statistical Computing*. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition: Task effects. *Memory & Cognition*, 17(5), 525-535. doi:10.3758/BF03197074
- Ren, G., Tang, Y., Li, X., & Sui, X. (2013). Pre-attentive processing of Mandarin tone and intonation: Evidence from event-related potentials. In F. Signorelli & D. Chirchiglia (Eds.), *Functional brain mapping and the endeavor to understand the working brain* (pp. 95-108): InTech.
- Ren, G., Yang, Y., & Li, X. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience*, 162(1), 87-95. doi:10.1016/j.neuroscience.2009.04.021
- Ren, J. (2012). *Jiyu EGG de Guangzhou/Jixi/Xi'an fangyan shengdiao shiyan yanjiu* [Investigations of Guangzhou, Jixi and Xi'an tones based on EGG] (M.A thesis). Nanjing Normal University. (in Chinese)
- Rohrbaugh, J., Donchin, E., & Eriksen, C. (1974). Decision making and the P300 component of the cortical evoked response. *Perception & Psychophysics*, 15, 368-374.
- Rojina, N. (2004). The acquisition of wh-questions in Russian. *Nordlyd: Tromsø University Working Papers on Language & Linguistics*, 32(1), 68-87. doi:<http://dx.doi.org/10.7557/12.59>
- Rose, P. (1987). Considerations in the normalisation of the fundamental frequency of linguistic tone. *Speech Communication*, 6(4), 343-352. doi:[http://dx.doi.org/10.1016/0167-6393\(87\)90009-4](http://dx.doi.org/10.1016/0167-6393(87)90009-4)
- Salisbury, D. F., Rutherford, B., Shenton, M. E., & McCarley, R. W. (2001). Button-pressing affects P300 amplitude and scalp topography. *Clinical Neurophysiology*, 112(9), 1676-1684.
- 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), 1-12. doi:10.1162/0898929052880057

- Schneider, K., Dogil, G., & Möbius, B. (2011). *Reaction time and decision difficulty in the perception of intonation*. Paper presented at the Proceedings of Interspeech 2011, Florence, Italy.
- Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, 29(6), 1155-1178. doi:10.1037/0096-1523.29.6.1155
- Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language & Speech*, 58(2), 131-151. doi:10.1177/0023830914522956
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(4), 623-656. doi:10.1002/j.1538-7305.1948.tb00917.x
- Sheldon, S., Pichora-Fuller, M. K., & Schneider, B. A. (2008). Priming and sentence context support listening to noise-vocoded speech by younger and older adults. *The Journal of the Acoustical Society of America*, 123(1), 489-499. doi:[doi:http://dx.doi.org/10.1121/1.2783762](http://dx.doi.org/10.1121/1.2783762)
- Shen, J. (1985). Beijinghua shengdiao de yinyu he yudiao [The tonal range and intonation in Mandarin]. In T. Lin & L. Wang (Eds.), *Experimental analyses on Beijing Mandarin* (pp. 73-125). Beijing: Peking University Press. (in Chinese)
- Shen, X. S. (1989). *The prosody of Mandarin Chinese* (Vol. 118). Berkeley, Los Angeles and California: University of California Press.
- Shi, P. (1980). Sizhong juzi de yudiao bianhua [Intonation variations in four sentence types]. *Language Teaching and Linguistic Studies*(2), 71-81. (in Chinese)
- Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation: Analysis, modelling and technology* (pp. 243-268). Dordrecht: Springer.
- Simpson, G. B. (1984). Lexical ambiguity and its role in models of word recognition. *Psychological Bulletin*, 96(2), 316-340. doi:<http://dx.doi.org/10.1037/0033-2909.96.2.316>

- Slowiaczek, L. M., & Hamburger, M. (1992). Prelexical facilitation and lexical interference in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(6), 1239-1250.
- Smith, D. B. D., Donchin, E., Cohen, L., & Starr, A. (1970). Auditory averaged evoked potentials in man during selective binaural listening. *Electroencephalography and Clinical Neurophysiology*, 28(2), 146-152.
doi:[http://dx.doi.org/10.1016/0013-4694\(70\)90182-3](http://dx.doi.org/10.1016/0013-4694(70)90182-3)
- So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language & Speech*, 53(2), 273-293. doi:10.1177/0023830909357156
- Speer, S., & Blodgett, A. (2006). Prosody. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics (2nd Edition)* (pp. 505-537): Academic Press.
- Spivey, M. J., & Marian, V. (1999). Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10(3), 281-284. doi:10.1111/1467-9280.00151
- Sun L. (2007). *Xi'an fangyan yanjiu* [A study on Xi'an dialect]. Xi'an: Xi'an Publishing House. (in Chinese)
- Taft, M., & Chen, H.-C. (1992). Judging homophony in Chinese: The influence of tones. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Advances in Psychology* (Vol. 90, pp. 151-172): North-Holland.
- Tseng, C.-y., & Su, C.-y. (2014). From ripples to waves, tides and beyond. In C.-T. J. Huang & F.-h. Liu (Eds.), *Peaches and plums* (pp. 257-278). Taipei, Taiwan: Institute of Linguistics, Academia Sinica.
- Tsuchihashi, M. (1983). The speech act continuum: An investigation of Japanese sentence final particles. *Journal of Pragmatics*, 7(4), 361-387.
[doi:http://dx.doi.org/10.1016/0378-2166\(83\)90024-3](http://dx.doi.org/10.1016/0378-2166(83)90024-3)
- Ulan, R. (1978). Interrogative systems. In J. H. Greenberg (Ed.), *Universals of human language* (Vol. 4, pp. 211-248). Stanford: Stanford University Press.
- Unsworth, S. J., & Pexman, P. M. (2003). The impact of reader skill on phonological processing in visual word recognition. *The Quarterly Journal of Experimental Psychology Section A*, 56(1), 63-81.
doi:10.1080/02724980244000206

- Vaissière, J. (2008). Perception of intonation. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 236-263). Oxford, UK: Blackwell Publishing Ltd.
- Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443-467. doi:<http://dx.doi.org/10.1037/0278-7393.31.3.443>
- Van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and "tone": Evidence from Thai. *Journal of Phonetics*, *1*, 101-109.
- Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*(3), 181-198. doi:10.3758/BF03197716
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190. doi:<http://dx.doi.org/10.1016/j.ijpsycho.2011.09.015>
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394-417. doi:<http://dx.doi.org/10.1037/0278-7393.25.2.394>
- Verleger, R., Gasser, T., & Möcks, J. (1985). Short term changes of event related potentials during concept learning. *Biological Psychology*, *20*(1), 1-16. doi:[http://dx.doi.org/10.1016/0301-0511\(85\)90036-5](http://dx.doi.org/10.1016/0301-0511(85)90036-5)
- Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology*, *19*(3), 165-181. doi:10.1027/0269-8803.19.3.165
- Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes*, *25*(7-9), 905-945. doi:10.1080/01690961003589492
- Wang J. (1996). *Xi'an fangyan cidian* [A dictionary of Xi'an dialect]. Nanjing: Phoenix Publishing & Media Network. (in Chinese)
- Wang, Y., Sereno, J. A., Jongman, A., & Hirsch, J. (2003). fMRI evidence for cortical modification during learning of Mandarin lexical tone. *Journal of*

- Cognitive Neuroscience*, 15(7), 1019-1027.
doi:10.1162/089892903770007407
- Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1), 1-25.
doi:https://doi.org/10.1016/S0749-596X(03)00105-0
- Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica: International Journal of Speech Science*, 49, 25-47.
- White, C. M. (1980). *Mandarin tone and English intonation: A contrastive analysis* (M.A thesis). University of Arizona. Retrieved from <http://hdl.handle.net/10150/557400>
- Wicha, N. Y. Y., Bates, E. A., Moreno, E. M., & Kutas, M. (2003). Potato not Pope: Human brain potentials to gender expectation and agreement in Spanish spoken sentences. *Neuroscience Letters*, 346(3), 165-168.
- Wiener, S., & Ito, K. (2014). Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition and Neuroscience*, 1-13.
doi:10.1080/23273798.2014.946934
- Wong, K. F., & Xiao, Y. (2010). Diversity and difference: Identity issues of Chinese heritage language learners from dialect backgrounds. *Heritage Language Journal*, 7(2), 152-187.
- Wu, J. (2015). *Tonal bilingualism: The case of two closely related Chinese dialects* (PhD dissertation). Leiden University.
- Wu, Z. (1982). Putonghua yuju zhong de shengdiao bianhua [Tonal changes in Mandarin discourses]. *Zhongguo Yuwen*, 171(6), 439-449. (in Chinese)
- Wu, Z. (1996). A new method of intonation analysis for Standard Chinese: Frequency transposition processing of phrasal contours in a sentence. In F. G. e. al. (Ed.), *Analysis, perception and processing of spoken language*. Amsterdam: Elsevier Science B.V.
- Xu, B., & Mok, P. (2012a). *Cross-linguistic perception of intonation by Mandarin and Cantonese listeners*. Paper presented at the Speech Prosody 2012, Shanghai, China.

- Xu, B., & Mok, P. (2012b). *Intonation perception of low-pass filtered speech in Mandarin and Cantonese*. Paper presented at the The Third International Symposium on Tonal Aspect of Languages, Nanjing, China.
- Xu, B., & Mok, P. (2014). Cross-linguistic perception of Mandarin intonation. In: *Proceedings of Speech Prosody 2014*, 638-642.
- Xu, Y. (2004). Separation of functional components of tone and intonation from observed F0 patterns. In G. Fant, H. Fujisaki, J. Cao, & Y. Xu. (Eds.), *From traditional phonology to modern speech processing: Festschrift for Professor Wu Zongji's 95th birthday*. (pp. 483-505). Beijing: Foreign Language Teaching and Research Press.
- Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, 46(3-4), 220-251.
[doi:http://dx.doi.org/10.1016/j.specom.2005.02.014](http://dx.doi.org/10.1016/j.specom.2005.02.014)
- Xu, Y. (2009). Timing and coordination in tone and intonation — An articulatory-functional perspective. *Lingua: International Review of General Linguistics*, 119(6), 906-927.
[doi:http://dx.doi.org/10.1016/j.lingua.2007.09.015](http://dx.doi.org/10.1016/j.lingua.2007.09.015)
- Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33(4), 319-337.
doi:10.1016/s0167-6393(00)00063-7
- Yang, R.-X. (2011). *The phonation factor in the categorical perception of Mandarin tones*. Paper presented at the ICPHS XVII, Hong Kong, China.
- Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14(5-6), 609-630.
doi:10.1080/016909699386202
- Yip, M. (2001). Phonological priming in Cantonese spoken-word processing. *Psychologia*, 44, 223-229.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yu, K. M., & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, 136(3), 1320-1333.
doi:http://dx.doi.org/10.1121/1.4887462
- Yuan J. et al. (1989). *Hanyu fangyan gaiyao* [An introduction to Chinese dialects]. Beijing: Wenzi Gaige Chubanshe. (in Chinese)

- Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In Q. Huo, B. Ma, E.-S. Chng & H. Li (Eds.), *Chinese Spoken Language Processing 5th International Symposium* (pp. 19-30). Singapore: Springer Berlin Heidelberg.
- Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, 130(6), 4063-4069.
doi:<http://dx.doi.org/10.1121/1.3651818>
- Zhang, J. (2009). Xi'anhua yu Beijinghua danzidiao de bijiao [The comparative analysis of the monosyllabic tone between Xi'an and Beijing]. *Journal of Shaanxi Institute of Education*, 25(2), 71-75. (in Chinese)
- Zhang, J. & Shi, F. (2009). Xi'anhua danziyin shengdiao de tongji fenxi [The statistical analysis on monosyllabic tone of Xi'an dialect]. *Journal of Xianyang Normal University*, 24(5), 38-42. (in Chinese)
- Ziegler, J. C., Tan, L. H., Perry, C., & Montant, M. (2000). Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science*, 11(3), 234-248.

Summary

Spoken language processing is a task that humans continuously perform from birth. It, however, is not always easy, as speech signals are inherently noisy and ambiguous. Prosody, which structures speech, is a determinant of the form of spoken language. Ambiguity in prosody therefore can be a very important source of speech ambiguity. Yet very little is known about how ambiguity in prosodic information affects spoken language processing. Not uncommonly, the same form of a speech signal can represent different prosodic information and cause speech ambiguity. This dissertation concerns how ambiguous acoustic signals representing different prosodic information affect spoken language processing.

The most prominent prosodic feature of tonal languages such as Standard Chinese is their use of pitch (i.e., tone) to distinguish lexical meanings. However, speech ambiguity arises in Standard Chinese because the same pitch contour can cue both tone and intonation. Previous studies have shown that the dual functions of pitch in Standard Chinese cause ambiguity in speech signals and result in pitch processing difficulties at the behavioral level. However, what are the underlying neural mechanisms leading to the eventual behavioral decisions of tone and intonation processing? How do native listeners resolve the pitch processing difficulties? These issues are less well-understood and further research on tone and intonation processing is needed.

The same or similar pitch contours can also cue the same linguistic representation (e.g., tone), but different categories of that representation in two linguistic systems (e.g., different tonal categories in two tonal systems) of the same speaker. Indeed, most Standard Chinese speakers also speak a local Chinese dialect. Across some Mandarin dialects and Standard Chinese, the same or similar pitch contour(s) is often used to characterize two different tonal categories, which in turn result in different lexical meanings. For example, the high-level pitch contour with a pitch value of 55 corresponds to a different tonal category in Xi'an Mandarin (a Mandarin dialect) than in Standard Chinese. In the former, the syllable *ma55* means *scold*. In the latter, it means *mother*. For bi-dialectal speakers of such Mandarin dialects and Standard Chinese, the

question arising is whether the same or similar pitch contours from the two tonal systems are taken as representations of the same tone in pitch processing. Furthermore, what role does tone play in the activation and processing of bi-dialectal lexical representation?

This dissertation therefore set out to investigate how pitch is processed within a linguistic system (i.e., Standard Chinese) and across two linguistic systems (i.e., Standard Chinese and Xi'an Mandarin) when the same pitch contour cues different linguistic functions (i.e., tone and intonation) or different categories of the same linguistic function (i.e., tone). The dissertation is composed of six chapters.

Chapter 1 introduced the research questions to be addressed and provided a brief overview of each chapter.

Chapter 2 tapped into the neural correlates of tone and intonation processing in Standard Chinese using the event-related potential (ERP) technique. Native Standard Chinese listeners were presented with semantically neutral Standard Chinese sentences, which contrast in final tones (rising T2 or falling T4) and intonations (Question or Statement). Their behavioral and electrophysiological responses were recorded. A clear P300 effect was observed for the question-statement contrast in sentences ending with T4, but no ERP effect was found for the question-statement contrast in sentences ending with T2. These results provide ERP evidence for the interaction of tone and intonation in Standard Chinese, confirming the findings from behavioral metalinguistic data that native Standard Chinese listeners can distinguish between question and statement intonation when the intonation is associated with a final T4, but fail to do so when the intonation is associated with a final T2. Chapter 2 extends our understanding of online processing of tone and intonation 1) from the pre-attentive stage to the attentive stage; and 2) within a larger domain (i.e., multi-word utterances) than a single word utterance.

Chapter 3 further investigated the role of semantic context in resolving the pitch processing difficulties in tone and intonation processing in Standard Chinese reported in Chapter 2. Tone and intonation identification experiments were conducted in both semantically neutral and constraining contexts with the same group of native speakers of Standard Chinese. Results showed that the

overall performance of tone identification was better than that of intonation. Tone identification was seldom affected by intonation information, irrespective of semantic contexts. Intonation identification, particularly question intonation, however, was susceptible to the final lexical tone identity and was greatly affected by the semantic context. Specifically, in the semantically neutral context, questions were difficult to identify, as evidenced in the lower response accuracy and longer reaction time, regardless of the lexical tone identity. In the semantically constraining context, both intonations took significantly less time to be identified than in the semantically neutral context. Moreover, questions ending with a falling tone were more accurately identified than questions ending with a rising tone. These results suggest that top-down information provided by the semantically constraining context can play a facilitating role for listeners to disentangle intonational information from tonal information, especially in sentences with a lexical falling tone in final position. Chapter 3 provides strong evidence for the role of semantic context in resolving pitch processing difficulties in Standard Chinese, particularly from the reaction time patterns, which have not been reported in earlier studies. The results reported in Chapter 3 also resolved the puzzle of the reversed patterns of question intonation identification in sentences ending with T2 and T4 in normal natural context versus in low-pass filtered context. It was found that the stronger and more informative the linguistic context is (semantically constraining context > semantically neutral context > low-pass filtered context), the better the identification of questions ending with T4. The opposite pattern holds for questions ending with T2.

One long-neglected fact in linguistic research on Standard Chinese is that most speakers of Standard Chinese also speak a local dialect, which may share phonological features with Standard Chinese. Among these dialects, Xi'an Mandarin is particularly interesting for the seemingly simple, yet intricate mappings between its lexical tones with those in Standard Chinese. **Chapter 4** empirically compared the tonal systems of Xi'an Mandarin and Standard Chinese. Tones with similar contours from Xi'an Mandarin and Standard Chinese were paired. Both tone production and perception experiments were carried out on highly proficient bi-dialectal speakers of Xi'an Mandarin and

Standard Chinese. The two experiments together showed that there are indeed systematic mappings of tones between Xi'an Mandarin and Standard Chinese. The degree of the similarity of the mapped tone pair in tone perception was largely dependent on the acoustic phonetic similarity between the tones in tone production, with a phonological rule playing a role in the tone pair of low contour. Chapter 4 compared the two systems in tone production with a more balanced design compared to the previous acoustic study. Moreover, it provides new empirical evidence for the mapping of the two tonal systems from a perceptual point of view. It also allows for an investigation of the relationship between tone production and perception in bi-dialectal tonal language speakers.

Chapter 5 further examined the effects of cross-dialect phonological similarity in segment and tone on bi-dialectal lexical access in spoken word recognition. Balanced bi-dialectal speakers of Xi'an Mandarin and Standard Chinese took part in an auditory-auditory priming experiment with a generalized lexical decision task. The primes were monosyllabic homophones from either Xi'an Mandarin or Standard Chinese, while the targets were disyllabic Xi'an Mandarin or Standard Chinese words. Primes and the first syllable of the target words had five configurations. They either overlapped in both segment and tone within a dialect (identical) or across two dialects (interdialectal homophones), or they overlapped in segment only within a dialect or across two dialects. The baseline condition was that they overlapped neither in tone nor segment within a dialect. Results showed that Standard Chinese primes did not yield significant priming effects for within- or cross-dialect segment-only overlap targets. Standard Chinese primes did not produce significant priming effects for within-dialect identical targets either. However, they did yield significant inhibitory priming effects for cross-dialect homophone targets. This overall pattern was reversed for Xi'an Mandarin primes because these primes were not treated differently from their interdialectal homophonous primes in the current mixed dialect setting. These results suggest that cross-dialect phonological similarity in segment alone did not affect lexical access in bi-dialectal auditory word recognition, while cross-dialect phonological similarity in both segment and tone posed a threat to the recognition system of bi-dialectal listeners. It is clear that tonal information

played a significant role in constraining word activation in bi-dialectal auditory word recognition. The results reported in Chapter 5 extends our understanding of the role of segment and tone in auditory word recognition in tonal languages from the monolingual context to the bi-dialectal context, and reveals a non-selective processing mechanism in bi-dialectal lexical access during auditory word recognition, as has been demonstrated for bilingual lexical access.

Chapter 6 recapped the research questions and summarized the main findings of this dissertation. The implications for future research were also discussed in this chapter.

In summary, this dissertation has demonstrated that pitch processing in Standard Chinese is subject to both within- and cross-linguistic influences. The ambiguous acoustic signals due to dual functions of the F0 channel in signalling tone and intonation in Standard Chinese cause pitch processing difficulty at the sentential level. This pitch processing difficulty has a neural correlate and can be resolved via top-down information provided by a constraining semantic context. Acoustic ambiguities in Standard Chinese can also arise from a closely related Chinese dialect that shares tonal similarities with Standard Chinese, here Xi'an Mandarin. The cross-dialect tonal similarities affect tone processing and further interfere in lexical access during spoken word recognition in bi-dialectal tonal language speakers. Together, this dissertation revealed two of the most prominent pitch processing difficulties tonal language speakers encounter from within and across languages, advancing our current understanding of pitch processing from various aspects.

Samenvatting

Het verwerken van gesproken taal is een taak die mensen continu uitvoeren vanaf hun geboorte. Dit is echter niet altijd gemakkelijk, aangezien spraaksignalen inherent ruizig en ambigu zijn. De prosodie, die de spraak structuur geeft, is mede bepalend voor de vorm van gesproken taal. Ambiguïteit in de prosodie kan daarom een belangrijke vorm van ambiguïteit in spraak zijn. Er is echter zeer weinig bekend over wat voor invloed ambiguïteit in prosodische informatie heeft op de verwerking van gesproken taal. Niet zelden kan dezelfde vorm van een spraaksignaal verschillende prosodische informatie representeren en tot ambiguïteit in spraak leiden. Dit proefschrift onderzoekt hoe ambigue akoestische signalen die verschillende prosodische informatie representeren invloed hebben op de verwerking van gesproken taal.

Het meest prominente prosodische kenmerk van toontalen zoals het Standaardchinees is het gebruik van toon om lexicale betekenissen te onderscheiden. In het Standaardchinees ontstaat er echter spraakambiguïteit, omdat dezelfde tooncontour zowel toon als intonatie kan aangeven. Eerder onderzoek heeft aangetoond dat de dubbele functie van toon in het Standaardchinees voor ambiguïteit kan zorgen in spraaksignalen en kan leiden tot verwerkingsproblemen op het gedragsniveau. Wat zijn echter de onderliggende neurale mechanismen die leiden tot de uiteindelijke gedragsmatige beslissingen op het gebied van toon- en intonatieverwerking? Hoe lossen moedertaalsprekende luisteraars de toonverwerkingsproblemen op? Deze kwesties worden minder goed begrepen en verder onderzoek aangaande de verwerking van toon en intonatie is nodig.

Het is ook mogelijk dat dezelfde of soortgelijke tooncontouren dezelfde taalkundige representatie aangeven (bijv. toon), maar andere categorieën van die representatie in twee taalsystemen (bijv. verschillende tooncategorieën in twee toonsystemen) van dezelfde spreker. De meeste sprekers van het Standaardchinees spreken immers ook een ander plaatselijk Chinees dialect. Over sommige dialecten van het Mandarijn en het Standaardchinees worden dezelfde of gelijksoortige tooncontouren gebruikt om verschillende tonale categorieën te karakteriseren, die vervolgens tot verschillende lexicale

betekeningen leiden. De hoge vlakke tooncontour met toonwaarde 55, bijvoorbeeld, correspondeert met een andere categorie in het Xi'an-Mandarijn (een Mandarijn dialect) dan in het Standaardchinees. In het Xi'an-Mandarijn betekent de syllabe *ma55 uitschelden*. In het Standaardchinees betekent het *moeder*. Voor tweedialectische sprekers van zulke Mandarijne dialecten met het Standaardchinees ontstaat de vraag of dezelfde of gelijksoortige tooncontouren uit de twee toonsystemen als de representatie van dezelfde toon worden genomen in de verwerking van toon. Wat voor rol speelt toon daarnaast in de activatie en verwerking van tweedialectische lexicale representatie? Om deze redenen stelt dit proefschrift zich tot doel te onderzoeken hoe toon wordt verwerkt binnen een taalkundig systeem (d.w.z. het Standaardchinees) of tussen twee taalsystemen (d.w.z. het Standaardchinees en het Xi'an-Mandarijn) wanneer dezelfde tooncontour verschillende taalkundige functies aangeven (d.w.z. toon en intonatie) of verschillende categorieën van dezelfde taalkundige functie (d.w.z. toon). Het proefschrift bestaat uit zes hoofdstukken.

Hoofdstuk 1 introduceert de onderzoeksvragen en geeft een kort overzicht van ieder hoofdstuk.

Hoofdstuk 2 haakt in op de neurale correlaten van de verwerking van toon en intonatie in het Standaard-Chinees met gebruik van de *event-related-potentials*-techniek (ERP). Aan luisteraars met het Standaardchinees als moedertaal werden semantisch neutrale Standaardchinese zinnen aangeboden die contrasteerden op de laatste toon (stijgende T2 of dalende T4) en intonatie (Vraag of Stelling). Van hen werden de gedragsresponsen en de EEG-responsen opgenomen. Een duidelijk P300-effect was zichtbaar voor het contrast tussen vraag en stelling in zinnen die eindigden op T4, maar er was geen ERP-effect te zien voor het contrast tussen vraag en stelling in zinnen die eindigden op T2. Deze resultaten leveren ERP-evidentie voor de interactie tussen toon en intonatie in het Standaardchinees, en bevestigen de bevindingen uit gedragsmatige metalinguïstische data dat moedertaalsprekende Standaardchinese luisteraars in staat zijn om onderscheid te maken tussen vraag- en stellingintonatie wanneer de intonatie geassocieerd is met een finale T4, maar dit niet kunnen wanneer de intonatie geassocieerd is met een finale T2. Hoofdstuk 2 verlegt ons begrip van de online verwerking van toon en intonatie

1) van het pre-attentieve stadium naar het attentieve stadium; en 2) binnen een groter domein (d.w.z. multiwoorduitingen) dan eenwoordsuitingen.

Hoofdstuk 3 biedt verder onderzoek naar de rol van semantische context in het oplossen van de problemen met het verwerken van toon en intonatie in het Standaardchinees die genoemd worden in hoofdstuk 2. Er zijn toon- en intonatie-identificatie-experimenten uitgevoerd in zowel semantisch neutrale als beperkende contexten met dezelfde groep moedertaalsprekende luisteraars van het Standaardchinees. De resultaten laten zien dat de algehele prestatie op het gebied van toonidentificatie beter was dan die op het gebied van intonatie. Toonidentificatie werd zelden beïnvloed door intonatie-informatie, ongeacht de semantische context. Intonatie-identificatie, vooral vraagintonatie, was echter vatbaar voor de identiteit van de finale lexicale toon en werd sterk beïnvloed door de semantische context. Specifiek waren vragen moeilijk te identificeren in de semantisch neutrale context, hetgeen terug te zien is in lagere responsnauwkeurigheid en langere reactietijden, ongeacht de identiteit van de lexicale toon. In de semantisch beperkende context kostten beide intonaties significant minder tijd om geïdentificeerd te worden dan in de semantisch neutrale context. Bovendien werden vragen die eindigden in een dalende toon nauwkeuriger geïdentificeerd dan vragen die eindigden in een stijgende toon. Deze resultaten suggereren dat *top-down*-informatie, verleend door de semantisch beperkende context, een faciliterende rol kan spelen voor luisteraars om intonationale informatie te onderscheiden van tonale informatie, in het bijzonder in zinnen met een lexicale dalende toon in finale positie.

Hoofdstuk 3 biedt sterke evidentie voor de rol van semantische context in het oplossen van moeilijkheden met het verwerken van toon in het Standaardchinees, in het bijzonder uit de patronen in de reactietijden, hetgeen niet in eerder onderzoek is gerapporteerd. De resultaten in hoofdstuk 3 bieden ook een oplossing voor de puzzel betreffende de omgekeerde patronen van vraagintonatie-identificatie in zinnen die eindigen op T2 en T4 in normale natuurlijke context versus *low-pass*-gefilterde context. De bevinding was dat hoe sterker en informatiever de taalkundige context is (semantisch beperkende context > semantisch neutrale context > *low-pass*-gefilterde context), des te

beter was de identificatie van zinnen die eindigden op T4. Van het omgekeerde patroon is sprake voor zinnen die eindigen in T2.

Een lang-verwaarloosd feit in het taalwetenschappelijk onderzoek naar het Standaardchinees is dat de meeste sprekers van het Standaardchinees ook een plaatselijk dialect spreken, dat fonologische kenmerken gemeen kan hebben met het Standaardchinees. Van deze dialecten is het Xi'an-Mandarijn bijzonder interessant voor de ogenschijnlijk eenvoudige, maar ingewikkelde *mappings* tussen diens lexicale tonen en die van het Standaardchinees. **Hoofdstuk 4** geeft een empirische vergelijking van de toonsystemen van het Xi'an-Mandarijn en het Standaardchinees. Er zijn paren gemaakt van tonen met soortgelijke contouren uit het Xi'an-Mandarijn en het Standaardchinees. Zowel toonproductie- als toonperceptie-experimenten zijn uitgevoerd op zeer vaardige tweedialectische sprekers van het Xi'an-Mandarijn en het Standaardchinees. De twee experimenten samen laten zien dat er inderdaad systematische *mappings* van tonen zijn tussen het Xi'an-Mandarijn en het Standaardchinees. De mate van gelijkheid van het *gemapte* toonpaar in toonperceptie was veelal afhankelijk van de akoestisch-fonetische gelijkheid tussen de tonen in toonproductie, waarbij een toon-sandhiregel een rol speelt in het 'lage contour'-toonpaar. Hoofdstuk 4 vergelijkt de twee systemen in de productie van toon met een meer balanceerde onderzoeksopzet in vergelijking met de eerdere akoestische studie. Bovendien biedt het nieuwe empirische evidentie voor de *mapping* van de twee toonsystemen vanuit een perceptueel oogpunt. Het biedt tevens de mogelijkheid om de relatie tussen toonproductie en toonperceptie in tweedialectische toontaalsprekers te onderzoeken.

Hoofdstuk 5 onderzoekt nader wat de effecten op tweedialectische lexicale toegang in gesprokenwoordherkenning zijn van fonologische gelijkheid in segment en toon tussen dialecten. Gebalanceerde tweedialectische sprekers van het Xi'an-Mandarijn en het Standaardchinees namen deel aan een auditief-auditief-primingexperiment met een gegeneraliseerde lokale lexicalebeslissingstaak. De primes waren monosyllabische homofonen uit ofwel het Xi'an-Mandarijn ofwel het Standaardchinees, en de targets waren disyllabische woorden uit het Xi'an-Mandarijn of het Standaardchinees. De primes en de eerste syllaben van de targetwoorden hadden vijf configuraties.

Ofwel overlaptten ze in zowel segment als toon binnen een dialect (identiek) of tussen twee dialecten (interdialectale homofonen), of ze overlaptten alleen in segment binnen een dialect of tussen twee dialecten. De baselineconditie was dat ze overlaptten in toon noch segment binnen een dialect. De resultaten lieten zien dat Standaardchinese primes niet tot significante primingeffecten leidden voor targets met overlap binnen of tussen dialecten op alleen segmenten. Wel leverden ze significante inhibitieve primingeffecten op voor homofone targets tussen dialecten. Dit algehele patroon was omgedraaid voor Xi'an-Mandarijnse primes omdat deze primes niet anders werden behandeld dan hun interdialectale homofonprimes in de huidige gemixte dialectsetting. Deze resultaten suggereren dat fonologische gelijkens tussen dialecten op segment alleen geen effect had op tweedialectische gesprokenwoordherkenning, terwijl fonologische gelijkens tussen dialecten in zowel segment als toon een bedreiging vormde voor het herkenningssysteem van tweedialectische luisteraars. Het is duidelijk dat tooninformatie een significante rol speelde in het beperken van woordactivatie in tweedialectische gesprokenwoordherkenning. De resultaten in hoofdstuk 5 breiden ons begrip van de rol van segment en toon in de herkenning van gesproken woorden in toontalen van de eentalige context uit naar de tweedialectische context, en onthullen een non-selectief verwerkingsmechanisme in tweedialectische lexicale toegang tijdens de herkenning van gesproken woorden, zoals aangetoond is voor tweetalige lexicale toegang.

Hoofdstuk 6 gaat terug naar de onderzoeksvragen en vat de hoofdzakelijke bevindingen van dit proefschrift samen. Dit hoofdstuk bespreekt ook de implicaties voor vervolgonderzoek.

Samenvattend heeft dit proefschrift laten zien dat de verwerking van toon in het Standaardchinees onderhevig is aan invloeden van zowel binnen als tussen talen. De ambigue akoestische signalen door de dubbele functies van het F0-kanaal in het aangeven van toon en intonatie in het Standaardchinees leiden tot moeilijkheden met het verwerken van toon op zinsniveau. Deze moeilijkheden hebben een neurale tegenhanger en kunnen worden opgelost via *top-down*-informatie aangeleverd door een beperkende semantische context. Akoestische ambiguïteiten in het Standaardchinees kunnen ook voortkomen uit een nauw-

verwant Chinees dialect, in dit geval het Xi'an-Mandarijn, dat gelijkenissen vertoont op het gebied van toon met het Standaardchinees. De gelijkenissen in toon tussen dialecten beïnvloeden de verwerking van toon en interfereren verder met lexicale toegang tijdens het herkennen van gesproken woorden in tweedialectische toontaalsprekers. Samenvattend legt deze dissertatie twee van de meest prominente toonverwerkingsproblemen bloot die toonsprekers ervaren binnen en tussen talen, hetgeen ons begrip van de verwerking van toon vanuit verschillende invalshoeken vooruit brengt.

摘要

人类自从出生开始就在进行言语加工。由于言语信号本身含有噪音和歧义，言语加工并不总是很容易。言语信号由韵律信息承载，韵律是决定言语形式的重要因素之一。因此，韵律信息中的歧义是言语歧义的一个重要来源。然而目前学界对于韵律信息中的歧义如何影响言语加工和认识还很有限。事实上，同一形式的言语信号被用来表示不同的韵律信息，从而导致言语歧义的现象并不少见。本论文即研究表示不同韵律信息的、有歧义的声学信号如何影响言语加工。

声调语言，比如普通话，最显著的一个韵律特征是使用音高（声调）来区别词汇意义。然而，在普通话中，音高不止用于区别词汇意义，它也用来表示语调信息，同一音高曲拱有时候既包含声调信息又包含语调信息。以往的行为实验的研究结果表明普通话中音高这一既表示声调又表示语调的双功能会引发言语信号的歧义，从而导致母语者在加工音高信息时出现困难。但是，促成最终的音高加工行为的底层认知神经机制是什么？普通话母语者如何解决这种音高加工困难？这些问题学界目前还没有答案，相关研究亟待进行。

同样或相似的音高曲拱有时候还可以用来表示同一个人的不同语言系统的同一语言功能（比如声调）的不同种类（比如不同的声调调类）。这种情况在普通话使用者中很常见，大多数的普通话使用者还会说另外一种汉语方言。在一些北方官话方言和普通话中，同样或相似的音高曲拱经常用来表示两种不同的声调调类，进而导致不同的词汇意义。比如，高平调 55 在西安方言（一种北方官话方言）和普通话中的调类就不同。在西安方言中，*ma55* 这个音节表示“骂”；而在普通话中，它表示“妈”。对于官话方言和普通话的双言者来说，这些来自双声调系统的相同或相似的音高曲拱在音高加工中是不是被当做同一种声调？声调在双言者的词汇激活和加工中起什么作用？这些问题目前还悬而未决。

本论文因此致力于探讨当相同的音高曲拱用来表示同一语言系统（普通话）的不同语言功能（声调和语调），以及表示不同语言系统（普通话和西安方言）的同一语言功能（声调）的不同种类（声调调类）时，音高信息如何被加工。全文共分六章。

第一章首先介绍了各个章节的研究问题，然后简单概述了每章的主要内容。

第二章采用事件相关电位技术（ERP）研究普通话声调和语调加工的神经关联物。母语者听语义中立的普通话最小对立句，并进行行为和脑电反应。这些句子的尾字声调控制为升调 T2 或降调 T4，句子语调控制为疑问语调或陈述语调。实验结果表明当尾字声调是降调 T4 时，普通话母语者可以区分疑问和陈述语调，表现为一个 P300 的 ERP 成分；当尾字声调是升调 T2 时，普通话母语者无法区分疑问和陈述语调，表现为没有任何 ERP 成分。本章为普通话声调和语调的交互作用提供了脑电证据，用脑电技术确认了前人的行为实验的结果。相比于已有的关于声调和语调的脑电研究，本章将对声调和语调的在线加工的认识从前注意阶段扩展到了注意阶段，从单字句扩展到了多字句。

第三章进一步研究语义语境如何帮助解决由声调和语调交互作用导致的音高加工困难。本章设置了两种语义语境，一种是中立语义语境，另一种是有限制的语义语境。采用与第二章类似的设计，本章在两种语义语境中分别进行了声调和语调识别实验，被试为同一组普通话母语者。结果表明尾字声调识别的结果整体好于句子语调的识别。尾字声调的识别在两种语义语境下几乎都不会受到句子语调的影响；而语调的识别，尤其是疑问语调的识别，则会受到尾字声调的影响，且识别结果与语义语境有很大关系。具体来说，在中立语义语境下，不管尾字声调是升调 T2 还是降调 T4，疑问语调都较难识别，疑问语调的识别时间长且正确率低。在有限制的语义语境下，疑问和陈述语调的识别时间都比中立语义语境下的减少了很多，同时尾字声调为降调 T4 的疑问句的语调的识别正确率明显提高，高于相同语境下尾字声调

为升调 T2 的疑问句的语调的识别。这些结果表明由有限制的语义语境提供的自上而下的信息可以促使普通话母语者更好地从表层的音高信号中区分声调和语调信息，尤其是在尾字声调为降调 T4 的情况下。本章为语义语境可以解决普通话中由声调和语调交互作用引起的音高加工困难提供了强有力的证据。语境的作用不光体现在传统参数识别正确率上，还首次在反应时这一参数上得到了体现。在前人的研究中，尾字声调为升调 T2 和降调 T4 的疑问句的疑问语调的识别在不同的研究中表现出相反的趋势。本章的结果表明造成这种相反的趋势的一个重要原因是这些研究中句子的语义语境不同。我们发现句子的语义语境对尾字的限制越厉害（有限制的语义语境 > 中立语义语境 > 低通滤波语境），普通话母语者对尾字声调为降调 T4 的疑问句的语调的识别情况越好。尾字声调为升调 T2 的疑问句的疑问语调的识别则表现出相反的趋势。

以往关于普通话的语言学研究经常忽略的一个要点是大多数的普通话使用者同时还会说一种当地方言。这些方言有可能与普通话的音系特征非常相似，而关于这些方言与普通话的关系的研究还很少。其中一个非常典型的例子是西安方言。西安方言的声调系统看起来非常简单，却似乎与普通话的声调系统存在一种一一对应关系，这种关系还需要从实验中进行验证。**第四章**因此系统比较了西安方言和普通话的声调系统。两种语言中具有相似音高曲拱的声调组成四组声调，西安方言和普通话的流利双言使用者对这些声调对进行了产出和感知实验。声调产出和感知结果表明西安方言和普通话的声调的确存在系统性的一一对应关系。不过，两种语言的每组对应的声调在感知上的相似性并非完全一致，感知的相似性主要取决于每组声调产出上的声学相似性。对于两个低调来说，音系规则也影响了感知结果。相比于已有的比较这两种语言声调的产出研究，本章在产出实验的设计上更加科学。同时，本章首次从感知的角度研究了两种语言的声调对应情况，

明确了两个声调系统的关系，并探讨了双言者的声调感知和产出的关系。

第五章则进一步研究了跨方言的音段和声调相似性对双言者的听觉词汇通达的作用。研究采用了听觉-听觉启动实验范式，要求西安方言和普通话水平相当的双言者完成一个扩展的真假词判别实验。启动项是西安方言或者普通话中的同音单音节，目标项是西安方言或者普通话中的双音节词。启动项音节和目标项的第一个音节存在五种关系：1) 语言相同，音段和声调也相同（完全一致）；2) 语言不同，音段和声调相同（跨方言同音音节）；3) 语言相同，音段相同，声调不同（同方言声调最小对立音节）；4) 语言不同，音段相同，声调不同（跨方言声调最小对立音节）。第五种关系是基准条件，语言相同，音段和声调都不同。实验结果表明，普通话单音节启动项对同方言和跨方言的声调最小对立目标项没有启动效应，它对同方言完全一致的目标项也没有启动效应，但是却对跨方言的同音目标项有明显的抑制启动效应。西安方言单音节启动项在几种条件下表现出完全相反的结果，主要是由于在两种方言混合使用的情况下，具有不同方言属性的同音启动项没有被区别对待。这些结果表明跨方言的音段相似单独并不影响双言者的听觉词汇通达，而跨方言的音段和声调的同时相似则增加了双言者的词汇识别难度，声调信息在限制双言者的听觉词汇激活中起非常重要的作用。前人只关注音段和声调在单一语言环境下对听觉词汇加工的作用，本章将我们对音段和声调在听觉词汇加工中的作用的认知从单一语言环境扩展到了双方言环境。此外，同双语词汇通达的研究结果一致，我们发现在双言听觉词汇识别中，双言者同时激活了两种方言的词汇，表现出非选择性的词汇加工机制。

第六章回顾了所有研究问题，总结了本论文的主要发现，并提出了未来需要进一步研究的相关问题。

综上所述，本论文表明普通话的音高加工既会受到语内其他语言功能的影响，也会受到语际相同语言功能的影响。普通话的音高既表

示声调又表示语调，这一音高的双功能会引发言语信号的歧义，从而导致普通话母语者在加工音高信息时出现困难。这种音高加工困难表现出相应的底层神经关联物，并且可以通过由有限制的语义语境提供的自上而下的信息得到解决。普通话的音高歧义还可能由一种具有相似音系特征的与之紧密相关的汉语方言（比如西安方言）引发。跨方言的声调相似影响两种方言的声调感知结果，进而会对双言者的听觉词汇通达产生干扰。总之，本论文研究了声调语言使用者常遇到的来自语内和语际的两种最显著的音高加工困难情况，从多方面促进了我们对于音高加工的理解。

Curriculum vitae

Min Liu was born on 6 July 1988 in Shaanxi in the People's Republic of China. She attended Northwest University (Xi'an) in 2006, and received her bachelor's degree in Teaching Chinese as a foreign language in 2010. She then started her Master's programme the same year in the phonetics lab at Beijing Language and Culture University. In 2013, she finished her studies there and obtained her Master's degree. In the fall of 2013, she started her PhD research at the Leiden University Centre for Linguistics. This dissertation is the result of her research.