Lexical tone in word activation

Yang, Q.

Lexical Tone in Word Activation

Published by
LOT
Binnengasthuisstraat 9
1012 ZA Amsterdam
The Netherlands

phone: +31 20 525 2461
e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration: ‘Wave’, by Matthew Sebastian & Qing Yang.

ISBN: 978-94-6093-453-7
DOI: https://dx.medra.org/10.48273/LOT0668
NUR: 616

Lexical Tone in Word Activation

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr.ir. H. Bijl,

volgens besluit van het college voor promoties

te verdedigen op donderdag 16 mei 2024

klokke 11.15 uur

door

Qing Yang

geboren te Shangqiu, China

in 1993

Promotores: Prof. dr. Yiya Chen
Prof. dr. Niels O. Schiller (City University of Hong Kong)

Promotiecommissie: Prof. dr. Claartje Levelt
Prof. dr. Robert Hartsuiker (Ghent University)
Dr. Mirjam Broersma (Radboud University)
Dr. Maria del Carmen Parafita Couto

To My Grandpa

献给我的爷爷杨嘉运

Table of Contents

Acknowledgements

The completion of this dissertation is indebted to the wonderful people who provided support and assistance throughout this journey. First and foremost, my deepest gratitude goes to my supervisor Prof. Yiya Chen. I thank Yiya for accepting me as a PhD student and for her unwavering belief in my capabilities. Her invaluable guidance not only shaped my research but also transformed me into an independent researcher. I extend my appreciation to my co-supervisor, Prof. Niels Schiller, for his continual encouragement throughout this academic journey.

I owe my special thanks to Prof. Wen Cao, who introduced me to the fascinating world of experimental linguistics and has been my mentor since. I am also indebted to Qian Li for her encouragement and assistance, which made my study at Leiden University possible in the first place.

Heartfelt thanks to my close friends and colleagues who provided support during the highs and lows at Leiden. Tingting Zheng, you have been my go-to person whenever I feel stuck. Thank you for your understanding and genuine care. Jiang Wu, I'm incredibly lucky to have you by my side since day one of my PhD journey. Thank you for sharing helpful tips on absolutely everything and showing me a new way of living. Hang Cheng, thank you for being my productive officemate and collaborator. You are the warmest person I have ever known; without you, my days in Leiden would not be half as bright. Han Hu, thank you for being my elder sister and running buddy at Leiden. Your laughter and positive energy are always delights to my heart.

I extend my gratitude to remarkable colleagues within and outside LUCL. To my outstanding groupmates Xinyi Wen, Hana Nurul Hasanah, Priscilla Lam, Yiran Ding, Charlie Xu, Xin Li, Matthew Sung, and Tingting Zheng - working with such brilliant minds has been an honour. Xinyi, thank you for proofreading my thesis and assisting me with my teaching responsibilities. Your presence and

support have provided me with comfort and strength. Hana, collaborating with you was a pleasure, and I've learned a great deal. Menghui Shi and Dan Yuan, thank you for your research suggestions and ongoing support. Qian Li, Min Liu, Ting Zou, Yang Yang, Junru Wu, and Yifei Bi, thank you for leading the way for me. Jinlei Zhou and Chen Ran, though our time as roommates and colleagues was brief, I miss you both so much. Zhuoyi Luo, Elisabeth Kerr, Zhen Li, Jiaqi Wang, Cesko Voeten, Sarah von Grebmer Zu Wolfsthurn, Rasmus Puggaard-Rode, Andrew Wigman, Ami Okabe, Jin Wang, Ruixue Wu, Zhaole Yang and many other colleagues, thank you for the companionship and the engaging conversations we had during lunch and social events. Jos Pacilly, Ben Bonfi, Evelyn Bosma, Andreea Geambasu and Leticia Pablos Robles - your help in the eye-tracking lab is greatly appreciated, even during urgent and tricky requests. Special thanks to Andrew Wigman for being my native English speaker, and to Thom van Hugte for generously reviewing and proofreading my Nederlandse samenvatting.

I express gratitude to the anonymous reviewers of my submitted papers and the doctoral committee for this dissertation: Prof. dr. Claartje Levelt, Prof. dr. Robert Hartsuiker, Dr. Mirjam Broersma, and Dr. Maria del Carmen Parafita Couto.

Last but not least, my family - especially my mum, grandpa, and husband - deserves immeasurable thanks for their unconditional love and unwavering belief in me.

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

Chapter 1

General Introduction

In our minds, words are not organized in isolation but rather stored in interconnected networks, forming a complex web of associations (e.g., Collins & Loftus, 1975; McClelland & Rumelhart, 1985; Hutchison, 2003). Even when attempting to comprehend or produce a single word independently, we cannot simply match its meaning with its corresponding sound; instead, we involuntarily activate a list of related or similar words from our mental lexicon (Traxler, 2011).

For spoken word recognition, perhaps the most convincing evidence for parallel activation comes from the eye-tracking visual world paradigm (e.g., see Huettig et al., 2011 for a detailed review). In this paradigm (e.g., Allopenna et al., 1998), listeners are often asked to follow spoken instructions (e.g., "pick up the beaker") and identify the target (e.g., "beaker") from a few pictures displayed on the computer screen. Among the pictures, there is generally one phonological competitor which shares phonological overlaps with the target (e.g., target "beaker" – phonological competitor "beetle"), and two distractors that are unrelated to the target (e.g., target "beaker" – unrelated distractor "carriage"). During the process of identifying the correct target (e.g., "beaker"), listeners' eye movements are recorded, and they are found to fixate more on the phonological competitors (e.g., "beetle") than unrelated distractors (e.g., "carriage"). The fact that phonological competitors draw more visual attention than unrelated ones suggests that, while recognizing the target spoken word, listeners automatically activate phonologically related words and temporally consider them as potential word candidates for selection.

As for spoken word production, the most convincing evidence for parallel activation comes from the picture-word interference paradigm (Rosinski et al., 1975). In this paradigm, speakers are instructed to name pictures while ignoring the distractor word superimposed on the picture. Although phonologically related words appear to interfere with the process of spoken word recognition, they are found to facilitate the process of spoken word production (e.g., Meyer & Schriefers, 1991; Schriefers et al., 1990; Jescheniak & Schriefers, 2001). Specifically, if the distractor word is phonologically related to the target picture's

name (e.g., target "dog" – phonological distractor "fog"), speakers generally take less time to name the target picture compared with unrelated distractors (e.g., target "dog" – unrelated distractor "roof"). A generally accepted explanation of this effect is that the presence of the phonological distractor activated its corresponding sound in the speakers' mind, which overlaps with the target's sound form; speakers thus received not only top-down activation from the selected target lemma but also extra bottom-up activation from the phonological distractor during picture naming (e.g., Roelofs, 2000). Moreover, if the distractor word is categorically related to the target (e.g., target "dog" – categorically related semantic distractor "cat"), speakers take a longer time to name the target picture compared with unrelated distractors (e.g., Costa et al., 2003; La Heij, 1988; Lupker, 1984; Schriefers et al., 1990). A commonly suggested explanation for this effect is that the retrieval of the concept of the target picture not only activates the target word (e.g., "dog") but also words that are categorically related to the target (e.g., "cat"), which receive activation from both the target picture and the distractor word. Compared with unrelated distractors (e.g., "roof"), which only receive activation from the present distractor word itself, the activation level of the categorically related semantic distractor is higher and thus it is more demanding for speakers to select the intended target word (e.g., Levelt et al., 1999).

Both the processes of spoken word recognition and production are incremental, i.e., listeners and speakers do not just wait until the selection of a word to map meaning or plan for articulation. Rather, listeners and speakers activate multiple related words and their word forms for parallel processing. Besides evidence from the visual world paradigm and the picture-word interference paradigm, the co-activation of multiple related representations have also been validated by behavioural data of various tasks (e.g., lexical decision tasks, priming paradigms, blocked cyclic naming) and recent neuroimaging and electrophysiological data (see Nozari & Pinet, 2019 for a review). As one of the core principles of how our mind retrieves words, parallel activation has been

incorporated into most theories of language comprehension and production (e.g., Chen & Mirman, 2012; Dell, 1986; Levelt et al., 1999; McClelland & Elman, 1986).

During the last decades, research on parallel activation has also been extended to bilinguals, which provides a great opportunity to improve our understanding of the mental lexicon and lexical access. One critical issue that lies at the heart of the bilingual literature is whether bilingual lexical access is language-specific or language-nonspecific selection. Specifically, a central question is whether bilinguals co-activate words of the non-target language as well when comprehending or producing words in a target language. Classic paradigms introduced above, i.e., the visual world paradigm and the picture-word interference paradigm, have been applied to explore this issue. In a seminal eye-tracking study by Spivey and Marian (1999), Russian-English bilinguals were asked to follow instructions such as *Poloji marku nije krestika* "Put the stamp below the cross" and move objects on a whiteboard while their eye movements were being recorded. In critical trials, objects such as "marker" which share initial phonetic features with *marku* "stamp" were also presented. Eye movement analysis showed that the cross-language homophone "marker" attracted participants' visual attention from the target *marku* "stamp" significantly more than that of the unrelated control stimulus object (e.g., *lineika* "ruler"). This has been taken as evidence for parallel activation of words of two languages during spoken word recognition. As for spoken word production, evidence from the picture-word interference paradigm shows that phonologically and semantically related distractors of different languages manipulate the speed and accuracy of target picture naming (e.g., target "dog" – cross-language phonological distractor *muñeca* "doll" – cross-language semantic distractor *gato* "cat" - translation distractor "perro"), similar as distractors of the same language (e.g., target "dog" – phonological distractor "doll" – semantic distractor "cat"; see Hall, 2010 for a review). Such cross-language interaction provides strong evidence for the co-activation of bilinguals' two languages during spoken word production. There is

a consensus that bilinguals automatically access words of both their languages in speech comprehension and production (see Kroll et al., 2012 for a detailed review).

Although language non-specific parallel activation is widely agreed upon, it is important to note that most previous studies drew empirical evidence from Indo-European and Germanic languages such as English and Dutch. These languages are stress languages, which employ relative prominence between syllables (cued with salient pitch contours, lengthening, intensity increase, and vowel quality contrast; see Gordon & Roettger, 2017 for a review on cues of stress) to distinguish a limited number of word pairs (e.g., REcord and reCORD in English[1]). However, most of the world's languages are tonal languages, which use lexical tone (realized via pitch variation) to differentiate word meanings (Yip, 2002; Zsiga, 2012; Fromkin, 2014). For example, in Standard Chinese, a representative tonal language, the same segmental syllable *ma* means "mother" with a high-level tone (Tone 1, hereafter T1), "hemp" with a rising tone (Tone 2, hereafter T2), "horse" with a low dipping tone (Tone 3, hereafter T3), and "to scold" with a falling tone (Tone 4, hereafter T4). Therefore, to successfully recognize or produce a Standard Chinese word, it is crucial for Standard Chinese speakers to retrieve its corresponding lexical tone accurately and efficiently. Since most previous studies have focused on Western languages, the nature of lexical access in tonal languages, such as Mandarin, is not yet fully understood. For instance, the relative weighting and timing of utilizing segments versus lexical tone and the role of lexical tone in activating lexical candidates during the process of Mandarin spoken word recognition have remained controversial. Chapter 2 of this dissertation aimed to resolve the controversies with a series of eye-tracking visual world experiments.

The fact that many tonal language speakers are bilinguals further complicates the picture. As bilinguals activate words of both their languages

[1]Throughout the manuscript, we use capital letters to signal lexical stress.

during lexical access, one important issue that arises is the role of lexical tone in bilingual language co-activation. With bilinguals of two tonal systems, it is unclear whether their two tonal systems interact during lexical access, and if so, how they resolve the potential lexical conflicts. To investigate this issue, Chapter 3 of this dissertation studied the process of spoken word recognition with a unique type of bilinguals who speak two closely related dialects with mapping tones, namely Standard Chinese and Xi'an Mandarin bi-dialectals. As for bilinguals who speak both tonal and non-tonal languages, it is unclear whether lexical tone plays a role in non-tonal lexical access, especially during spoken word production. Are lexical tones activated even when bilinguals are speaking a non-tonal language? How does bilinguals' experience of speaking a tonal language affect non-tonal lexical access? With a series of picture-word interference tasks, Chapters 4 and 5 attempted to address these issues with Standard Chinese and English bilinguals.

In sum, to develop a more comprehensive account of lexical access, it is necessary to account for the role of lexical tone in both monolingual and bilingual speech comprehension and production. This dissertation aimed to fill in this gap by investigating the process of spoken word recognition and production in native speakers of Standard Chinese, bi-dialectal speakers of Standard Chinese and Xi'an Mandarin, and bilingual speakers of Standard Chinese and English. Specifically, we highlighted four issues: 1) the role of lexical tone in Mandarin spoken word recognition; 2) tonal interference in bi-dialectal spoken word recognition; 3) the activation of lexical tone in bilingual spoken word production; 4) the influence of lexical tone on bilingual mental lexicon. The rest of this chapter introduces each of the four issues and briefly explains how they are addressed using the visual world paradigm and the picture-word interference paradigm in this dissertation.

1.1 The Role of Lexical Tone in Mandarin Spoken Word Recognition

Several studies have shown activation and competition of phonologically similar words during the process of spoken word recognition in Mandarin (e.g., Lee, 2007; Liu & Samuel, 2007; Malins & Joanisse, 2010; Sereno & Lee, 2010; Zhao et al., 2011), similar to Indo-European languages. However, a few issues concerning how exactly segmental and tonal cues are taken up and processed remain open. First, it is controversial whether segmental syllables have a special status in Mandarin lexical processing. Different from syllables of Indo-European languages, Mandarin syllables are simpler in structure and limited in types and numbers (e.g., Roelofs, 2015; Verdonschot et al., 2015; Chen & Chen, 2002); as most Mandarin morphemes are monosyllabic and the writing system is based on syllable-sized characters, researchers have entertained the idea that the (segmental) syllable is a holistic processing unit in Mandarin lexical access (e.g., Zhao et al., 2011; Sereno & Lee, 2016). However, experimental data on this issue are rather controversial (e.g., see Zhao et al., 2011 and Sereno & Lee, 2016 for evidence supporting a special status of the segmental syllable; but see Malins & Joanisse, 2012 for evidence against it). Second, existing studies on Mandarin spoken word recognition differ on whether and to what extent sub-lexical components such as onset, rhyme and lexical tone affect lexical activation. In Indo-European languages, word candidates with the same onset are generally activated earlier and greater than word candidates with the same rhyme. For instance, with the visual world paradigm, Allopenna, Magnuson, and Tanenhaus (1998) found that when listening to targets (e.g., “beaker”), while both cohort competitors (e.g., “beetle”) and rhyme competitors (e.g., “speaker”) drew more of listeners’ attention than unrelated distractors (e.g., “carriage”), listeners’ eye fixations towards cohort competitors were significantly earlier than those of rhyme competitors. However, the effects of cohort and rhyme competitors are found to be less reliable in Mandarin spoken word recognition. For instance, using the visual world paradigm, Malins and Joanisse (2010) found a null effect of rhyme competitors whereas Zou

(2017) found a null effect of cohort competitors. Third, the time course of utilizing segmental and tonal cues during Mandarin spoken word recognition is not clear. Previous studies have reported a perceptual disadvantage of lexical tone compared with segmental information (Taft & Chen, 1992; Yip, 2001; Cutler & Chen, 1997; Ye & Connine, 1999, experiment 1; Hu et al. 2012; Sereno & Lee 2015; Gao et al., 2019). Such a disadvantage is often attributed to the fact that tonal information that arrives later and thus is processed later than segmental information (Cutler & Chen, 1997). However, recent evidence from eye-tracking and event-related potentials (ERPs) data show that tonal information plays an equivalent role to segmental information, and it is processed timely during Mandarin spoken word recognition (Malins & Joanisse, 2010; 2012). Questions about the time course of segmental and tonal processing and cue utilization during Mandarin lexical processing thus remain to be answered.

We addressed these issues in Chapter 2 with three eye-tracking visual world paradigm experiments. We aimed to clarify the role of segmental syllable and sub-syllabic constituents, as well as to investigate the time course of using segmental and suprasegmental tonal information during Mandarin lexical processing. In Experiments 1 and 2, native Standard Chinese (hereafter SC) speakers listened to monosyllabic SC words with the presence of a phonological competitor, which overlaps with the target in either segmental syllable, onset and tone, rhyme and tone, or just tone. Experiments 1 and 2 differ in how long listeners were allowed to preview pictures on the screen before hearing the spoken target word, as previous studies found that the length of preview time plays a crucial role in observing phonological competition effects or not (e.g., Huettig & McQueen, 2007; Huettig et al., 2011). Eye movement results of both Experiments 1 and 2 confirmed a robust competition effect of segmental syllable overlap competitors, and null effects of onset, rhyme and tone overlap distractors. Experiment 3 investigated the time course of segmental versus tonal information utilization by manipulating their point of divergence in acoustic cues. We found that both sub-syllabic information (i.e., segment vs. tone) and cue timing (i.e.,

early vs. late point of divergence) affect phonological competition effects. Regardless of the nature of the cues, the point of divergence determines the amplitude and time course of the competition effect: the earlier the point of divergence, the sooner the competition, suggesting that despite the dominant role of the segmental syllable, Mandarin listeners use both segmental and tonal information as soon as they are available to constrain lexical activation.

1.2 Tonal Interference in Bi-dialectal Spoken Word Recognition

In Chapter 3, we investigated the process of spoken word recognition in bi-dialectal speakers of Standard Chinese and Xi'an Mandarin. Both Standard Chinese and Xi'an Mandarin belong to the Mandarin Chinese family. They share similar syntactic structures, a large number of etymologically related translation equivalents, the same writing system and nearly the same segmental inventories. Moreover, the lexical tone systems of Standard Chinese and Xi'an Mandarin have a one-to-one mapping relation (Liu et al., 2020), resulting in a large number of cross-dialect homophones across the two dialects. For example, *ma* with a high-level tone means "mother" in SC, whereas it means "to scold*"* in Xi'an Mandarin. Such a unique case of Standard Chinese and Xi'an Mandarin bi-dialectals provides us with the opportunity to test whether and to what extent the mapping of tonal systems elicits cross-dialect interference while keeping the orthographic, morphological, semantic, and segmental aspects constant.

Using generalized lexical decision tasks with auditory priming, Liu (2018) manipulated five types of target and prime contrasts based on the cross-dialect phonological similarity between Standard Chinese and Xi'an Mandarin: 1) within-dialect segment and tone overlapping target and prime; 2) within-dialect segment overlapping target and prime; 3) cross-dialect segment and tone overlapping target and prime (i.e., cross-dialect homophones); 4) cross-dialect segment overlapping target and prime; 5) unrelated target and prime. Results of reaction times showed that with Standard Chinese primes, there was a significant

interference effect for the cross-dialect segment and tone overlapping targets but not for the cross-dialect segment overlapping targets, compared with unrelated targets. Liu (2018) interpreted these results as evidence for cross-dialect tonal interference during bi-dialectal spoken word recognition. Note that this effect was found in a mixed-dialectal context, i.e., bi-dialectals were exposed to words in both Standard Chinese and Xi'an Mandarin, which might have created or boosted cross-dialectal interference. Using the eye-tracking visual world paradigm, Chapter 3 sought to further understand the effects of tonal interference in a controlled mono-dialectal context. Specifically, we asked Standard Chinese and Xi'an Mandarin bi-dialectals to listen to sentences in one dialect and identify the target word among four Chinese characters shown on the screen. The characters included the target, two unrelated distractors, and a phonological competitor which share the same segmental syllable with the target within and across dialects. Among the phonological competitors, besides segmentally overlapping distractors which does not share lexical tone with the target within and across dialects (Segment Condition), there were also cross-dialect homophone competitors that share the same lexical tone with the target across dialects (Homophone Condition) and translation-induced cross-dialect homophones that share the same lexical tone with the targets' dialectal translation equivalent (Translation Condition). We hypothesized that, if both sets of lexical tones are activated, the Homophone and Translation Condition would elicit larger competition effects than Segment Condition; if only one set of lexical tones are activated, Segment Condition would elicit most competition effects, because the tonal contours of the target and competitor of the Segment Condition share most acoustic similarity. Listeners' eye movements show that distractors in the Segment Condition interfere with participants' eye fixations significantly more than Homophone and Translation Conditions, suggesting a lack of cross-dialectal interference effect. It is likely that the mono-dialectal sentence context has cancelled out the cross-dialect interference effect shown in Liu (2018). Overall, this finding marks a convergence between bi-dialectal and bilingual speech

processing. Based on these findings, a preliminary model of bi-dialectal spoken word recognition which emphasises active control of dialect activation was proposed.

1.3 The Role of Lexical Tone in Bilingual Spoken Word Production

Bilinguals not only retrieve the form of the target language but also that of the non-target language during spoken word production (See Costa, 2009 for a review). For example, with the picture-word interference paradigm, Dutch-English bilinguals were found to take longer to name pictures in their L2 English when the Dutch auditory distractor was phonologically similar to the Dutch translation of the target picture, compared with unrelated distractors (e.g., target *berg* "mountain" – phono-translation distractor *berm* "verge" – unrelated distractor *kaars* "candle"; Hermans et al., 1998). This finding indicates that the translations of the non-target language are activated at the phonological level. However, most previous studies on bilingual word production have focused on segments. It remains open whether suprasegmental information such as lexical tone is co-activated during bilingual spoken word production.

In Chapter 4, we aimed to address this issue by examining the role of lexical tone in English spoken word production with bilinguals of Standard Chinese and English. Specifically, we asked: if Standard Chinese and English bilinguals co-activate both Standard Chinese and English names during English word production, is lexical tone co-activated and utilized during the process? With four picture-word interference experiments, Standard Chinese and English bilingual speakers were instructed to name pictures in English (e.g., feather) while ignoring four types of simultaneously presented SC distractors: 1) the translation distractor, which is the translation equivalent of the English target name (e.g., *yu3mao2* "feather"); 2) the tone-sharing distractor, which shares both tone and segments with the SC translation in the first syllable (e.g., *yu3zhou4* "universe"); 3) the no-tone-sharing distractor, which shares segments only with the Standard

Chinese translation in the first syllable (e.g., *yu4mi3* "corn"); 4) the unrelated distractor, which shares no phonological overlap with target and its translation (e.g., *lei4shui3* "tear"). To further explore potential factors that may constrain the lexical tone effect, we also manipulated two additional factors that have been found to affect picture naming onset with the picture-word interference paradigm. One was distractor modality (e.g., Hantsch et al., 2009; Jonen et al., 2021); the SC distractors were presented either auditorily or visually. The other was familiarization mode (e.g., Llorens et al., 2014); bilinguals were asked to familiarize with the target pictures' English names only (i.e., English mode) or both English and Standard Chinese names (i.e., mixed mode). In Experiment 1 (with auditory distractor and English mode), translation distractors significantly facilitated bilingual English picture naming, while tone-sharing distractors significantly inhibited the process. Importantly, the tone-sharing distractors elicited significantly longer naming latency than the no-tone-sharing distractors, demonstrating the co-activation of lexical tone during English spoken word production. Overall, this study replicated previously found translation facilitation effect (e.g., Costa et al., 1999) and observed a significant interference effect of lexical tone. These findings suggest that Standard Chinese and English bilinguals not only co-activate the Standard Chinese translation equivalents but also the lexical tones of the Standard Chinese translations during English spoken word production. Results of Experiment 2 (auditory distractor and mixed mode), Experiment 3 (visual distractor and English mode), and Experiment 4 (visual distractor and mixed mode) further demonstrated that the polarity and robustness of the lexical tone effect are modulated by external factors such as distractor modality and familiarization mode.

1.4 The Effect of Lexical Tone on the Bilingual Mental Lexicon

Although it is widely agreed that words of bilinguals' two languages interact in their mental lexicon (see Kroll et al., 2012 for a review), how suprasegmental

features interact is still unclear, especially for bilinguals of two typologically different languages such as Standard Chinese and English. In Chapter 5, we further asked whether and to what extent lexical tone modulates pitch processing in non-tonal speech production with Standard Chinese and English bilinguals.

Previous studies on bilingual lexical access have identified an important distinction between Standard Chinese and English bilinguals and native English speakers in pitch processing during spoken word recognition (Ortega-Llebaria et al., 2017; 2020). With primed-lexical decision tasks, Ortega-Llebaria et al. (2020) asked Standard Chinese and English bilinguals and English monolinguals to make lexical judgments on English target words produced in either falling or rising pitch contours, while the prime and target were manipulated to fully match, fully mismatch, mismatch in segments, or mismatch in pitch. Results showed that in the full match and pitch mismatch conditions, Standard Chinese and English bilinguals experienced larger facilitation when the targets were produced with a falling pitch contour than with a rising pitch contour. Yet, such a "falling-f0 bias" was only found in SC-English bilinguals but not English monolinguals. This fact has led Ortega-Llebaria et al. (2017; 2020) to reason that words with falling pitch contours are closer English lexical representations than their rising counterparts in the mental lexicon of SC-English bilinguals; crucially, it is the long-term experience with lexical tone that reshapes their mental lexicon. This assumption is also consistent with the observation that Standard Chinese learners of English tend to produce stressed syllables with an H* pitch accent, giving the English words a falling-like pitch contour (McGory, 1997). However, no study so far has directly tested the effect of lexical tone on pitch representation and processing in bilingual spoken word production.

In Chapter 5, we adopted the picture-word interference paradigm to investigate this issue. Previous bilingual picture-naming studies have shown that cross-language homophone distractors facilitate picture naming in non-target languages (e.g., Hermans et al., 1998; Costa et al., 1999, 2003). In this study, we asked Standard Chinese and English bilinguals and native English monolinguals

to name pictures in English (e.g., *lung*) while ignoring simultaneously played SC cross-language homophones that either have a falling or a rising lexical tone (*lang4* with a falling tone, "wave"; *lang2* with a rising tone, "wolf"). We hypothesized that if lexical tone indeed influences bilinguals' pitch representation in non-tonal second languages, the effect of lexical tone (falling vs. rising) on English picture naming should differ between Standard Chinese and English bilingual and English monolingual speakers. Results showed that, compared with unrelated Standard Chinese distractors, both falling and rising cross-language homophones facilitated English word naming for both SC-English bilingual and English monolingual speakers. Most importantly, SC-English bilinguals showed significantly longer naming latencies with falling-tone in cross-language homophones than their rising-tone counterparts, whereas English monolingual speakers did not show such a pattern. As one of the first studies that investigated the influence of lexical tone in non-tonal lexical access during spoken word production, we identified a significant difference between SC-English bilinguals and English monolinguals in terms of how falling versus rising lexical tones affect English picture-word naming. This finding provides important implications for understanding pitch representation and processing in the bilingual lexicon, as well as the interaction between bilinguals' two languages at the suprasegmental level.

Chapter 2

Phonological Competition in Mandarin Spoken Word Recognition

A version of this chapter has been published as: Yang, Q., & Chen, Y. (2022). Phonological competition in Mandarin spoken word recognition. *Language, Cognition and Neuroscience*, 37(7), 820-843.

Abstract

Most of the world's languages use both segment and lexical tone to distinguish word meanings. However, the few studies on spoken word recognition in tone languages show conflicting results concerning the relative contribution of (sub-)syllabic constituents, and the time course of how segmental and tonal information is utilized. In Experiments 1 & 2, participants listened to monosyllabic Mandarin words with the presence of a phonological competitor, which overlaps in either segmental syllable, onset and tone, rhyme and tone, or just tone. Eye movement results only confirmed the segmental syllable competition effect. Experiment 3 investigated the time course of segmental vs. tonal cue utilization by manipulating their point of divergence (POD) and found that POD modulates the look trajectories of both segmental and tonal phonological competitors. While listeners can use both segmental and tonal information incrementally to constrain lexical activation, segmental syllable plays an advantageous role in Mandarin spoken word recognition.

Keywords: Mandarin spoken word recognition; Eye-tracking; Phonological competition effects; Lexical tone

The majority of the world's languages are tonal, in which pitch variation, known as lexical tone, distinguishes word meanings (Yip, 2002). For example, in Mandarin Chinese, the same segmental syllable *ma* means 'mother' with a high-level tone but 'horse' with a low (dipping) tone. Thus, it is expected that speakers of tonal languages such as Mandarin Chinese need to utilize tonal information effectively for successful and efficient spoken word recognition. Despite the importance of tone in the lexicon of the majority of the world's languages, existing models of spoken word recognition (SWR) have only begun to investigate the role of lexical tone. Understanding lexical processing in tonal languages would provide insights into the potential universal and diverse patterns of SWR across languages of the world and benefit the development of existing SWR models, which have based mainly on data from Indo-European non-tonal languages (e.g., Luce & Pisoni, 1998; McClelland & Elman, 1986; Marslen-Wilson 1987; Gaskell & Marslen-Wilson 2002; Norris, 1994; Norris, McQueen, & Cutler, 2000).

One broad consensus in current models of SWR is that the process of recognizing a word is incremental. Listeners activate several possible word candidates as the incoming speech signal unfolds. Sub-lexical phonemic features influence online lexical processing (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; McMurray, Clayards, Tanenhaus, & Aslin, 2008; Salverda, Dahan, & McQueen, 2003). There is some evidence from tonal languages, mainly limited to Mandarin Chinese (cf. Burnham et al., 2011 for Tai tones), that suggests incremental activation and competition of sub-lexical phonologically similar word candidates (e.g., Lee, 2007; Liu & Samuel, 2007; Malins & Joanisse, 2010; Sereno & Lee, 2010; Zhao et al., 2011). Despite possible similarities of SWR processes across tonal and non-tonal typologically different languages, several issues, as detailed out below, have remained outstanding and need to be clarified for SWR in tonal languages. Briefly speaking, in tonal languages, it is commonly recognized that segmental and suprasegmental tonal information both play a role during SWR (e.g., Malins & Joanisse, 2010; Malins & Joanisse, 2012a; Zhao et

al., 2011). What has remained open is how exactly segmental and tonal cues are taken up and processed during SWR. At the segmental level, the non-tonal syllable seems to play a critical role as a functional unit of processing (e.g., Sereno & Lee, 2015; Zhao, Guo, Zhou, & Shu, 2011). Relatedly, overlaps in sub-syllabic constitutes (segmental syllable onset and rhyme) have been found to exert no influence on Mandarin lexical competition (see Malins & Joanisse, 2010 for a null rhyme competition effect; see Zou, 2017 for a null onset competition effect). Given the different experimental paradigms/designs and their different levels of sensitivity to the time course of speech processing, it remains debatable whether segmental syllables are processed incrementally or holistically. The present study aimed to employ the eye-tracking technique to address the following issues by seeking answers to three specific research questions: 1) Do segmental syllables have a special status in Mandarin lexical processing? 2) What are the relative contributions of sub-syllabic segmental constituents (such as onsets and rhyme) and suprasegmental lexical tone? 3) What is the time course of segmental and suprasegmental processing and cue utilization during online lexical processing?

2.1 The Role of Segmental Syllable in Spoken Word Recognition

One issue to be resolved is the role of the non-tonal segmental syllable as a primary and holistic processing unit during SWR. Thus far, Mandarin has served as the main empirical base in the extant literature. It is well-known that Mandarin syllables differ from syllables of Indo-European languages in several aspects. First, Mandarin syllables consist of both segmental and suprasegmental information, i.e., lexical tone. The segmental syllables in Mandarin are simple in structure and have a relatively small number of syllable types. For example, they do not have consonant clusters, and only two nasal consonants (/n/ and /ŋ/) are allowed as codas. The total number of syllables is also rather limited; about 1,200 tonal syllables and 400 segmental syllables. Second, most morphemes in Mandarin are monosyllabic (i.e., segmental syllable plus tone), rendering

syllables as a unit of meaning. Last but not least, the writing system in Mandarin is based on syllable-sized characters, reinforcing the notion of the syllable as a holistic unit. These unique properties have motivated researchers to entertain the idea that Mandarin syllables may be an ideal lexical processing unit.

The evidence on the role of syllable and sub-syllabic units in SWR, however, has been mixed. Zhao et al. (2011) proposed that Mandarin SWR is "syllable-based holistic processing rather than phonemic segment-based processing." In Zhao et al. (2011), Mandarin speakers made semantic judgments on pictures while listening to an auditory distractor word. Event-related potentials (ERPs) showed that when the distractor mismatched the name of the picture in either onset, rhyme, tone, or the whole syllable (see Table 1 for sample stimuli used in the study), N400 (a negative ERP component elicited with semantic or phonological violations of expectations; Kutas & Hillyard, 1984; Praamstra & Stegeman, 1993) was elicited. Crucially, the earliest and highest amplitude was elicited by the whole syllable (i.e., segmental and tonal) violation. Sereno & Lee (2016) reached a similar conclusion with two auditory lexical decision tasks. In their study, participants' responses were only facilitated when the primes and targets had overlapping segmental syllables or syllables, and no priming effect was found for those with only partial segmental overlap (i.e., onset and tone overlaps; rhyme and tone overlaps).

Counter evidence against segmental syllables as the basic unit of processing in SWR has also been reported. With EEG recording, Malins & Joanisse (2012a) asked participants to make judgments on whether the auditory words and simultaneously presented pictures match or not. The picture names overlapped with the auditory words in either segmental syllable, onset, rhyme, tone, or unrelated (see Table 1 for sample stimuli). Results showed that all conditions modulated the phonological mapping negativity effects (PMN; an ERP component associated with pre-lexical processing; Connolly & Phillips, 1994; Newman & Connolly, 2009) and N400 effects (associated with lexical word meaning processing; Kutas & Hillyard, 1984; Praamstra & Stegeman, 1993).

Moreover, the PMN effects did not differ between the rhyme, tone, and the unrelated condition, suggesting that neither syllable nor segmental syllable in Mandarin merits any special status as a holistic processing unit. More recently, Ho et al. (2019) investigated the role of syllables in Mandarin word processing in sentence context with the cross-modal priming paradigm. In this task, prime words were embedded in the middle of a visually presented sentence, while target words were embedded in a following aural sentence. The targets and primes were mismatched in onset, tone, or syllable (see Table 1 for sample stimuli). Compared with identical sentences, all three critical mismatching conditions modulated PMN and N400 components. Crucially, the smallest amplitudes for PMN and N400 components were elicited by the whole syllable mismatching condition. This was interpreted as due to the lack of phonological competition between target and prime by Ho et al. who further suggested that Mandarin listeners process spoken words segment by segment rather than by the whole syllable.

It is clear that the above studies differed in the experimental paradigms employed, the specific behavioural and neural measurements taken, and the exact segmental conditions compared. More research on the topic is necessary to clarify the role of segmental syllable as a holistic unit of lexical processing. Experiment 1 aimed to address this issue.

Table 1. *Experimental conditions and sample stimuli of previous studies and the present study. All listed previous studies were discussed in the Introduction in the same order.*

Study	Task	Conditions	Sample Stimulus *(Pinyin)*	Shared Information	Divergent Information
Zhao, Guo, Zhou, & Shu, 2011	Picture/spoken word/picture task	Match	*bi2-bi2*	Phonemes & tone	None
		Onset Mismatch	*bi2-li2*	Rhyme and tone	Onset
		Rime Mismatch	*bi2-bo2*	Onset and tone	Rhyme
		Tone Mismatch	*bi2-bi3*	Phonemes	Tone
		Syllable Mismatch	*bi2-ge1*	None	Phonemes & tone
Sereno & Lee, 2015	Auditory priming paradigm, Experiment 1	Tone-segment Overlap	*ru4-ru4*	Phonemes & tone	None
		Segment-only Overlap	*ru4-ru3*	Phonemes	Tone
		Tone-only Overlap	*sha4-ru4*	Tone	Phonemes
		Unrelated	*qin1-ru4*	None	Phonemes & tone
	Auditory priming paradigm, Experiment 2	Tone-segment Overlap	*ru4-ru4*	Phonemes & tone	None
		Only-onset segment Overlap	*ru4-re4*	Onset & tone	Rhyme
		Only-rime Overlap	*ru4-pu4*	Rhyme & tone	Onset
		Unrelated	*ru4-qin1*	None	Phonemes & tone
Malins & Joanisse, 2012a	Picture/spoken word matching task	Segmental	*hua1-hua4*	All phonemes	Tone
		Cohort	*hua1-hui1*	Onset, glide and tone	Rhyme
		Rhyme	*hua1-gua1*	Rhyme and tone	Onset
		Tonal	*hua1-jing1*	Tone	Phonemes
		Unrelated	*hua1-lang2*	None	Phonemes & tone

Study	Task	Conditions	Sample Stimulus *(Pinyin)*	Shared Information	Divergent Information
Ho et al., 2019	Cross-modal priming paradigm	Match	*jia1-jia1*	Phonemes & tone	None
		Onset Violation	*jia1-xia1*	Rhyme and tone	Onset
		Tone Violation	*jia1-jia4*	Phonemes	Tone
		Syllable Violation	*jia1-tang2*	None	Phonemes & tone
Malins & Joanisse, 2010	Visual World Paradigm	Segmental	*chuang2-chuang1*	Phonemes	Tone
		Cohort	*chuang2-chuan2*	Onset, glide, and tone	Rhyme
		Rhyme	*chuang2-huang2*	Rhyme and tone	Onset
		Tonal	*chuang2-niu2*	Tone	Phonemes
Zou, 2017	Visual World Paradigm	Segmental	*chuang1-chuang2*	Phonemes	Tone
		Cohort	*chuang1-che1*	Onset and tone	Rhyme
		Rhyme	*chuang1-guang1*	Rhyme and tone	Onset
		Tonal	*chuang1-ji1*	Tone	Phonemes
The present study	Visual World Paradigm, Experiment 1&2	Segmental syllable	*chuang2-chuang1*	Phonemes	Tone
		Cohort	*chuang2-cha1*	Onset and tone	Rhyme
		Rhyme	*chuang2-huang2*	Rhyme and tone	Onset
		Tonal	*chuang2-ya2*	Tone	Phonemes
	Visual World Paradigm, Experiment 3	Segmental Early	*dian4-dou4*	Onset and tone	Rhyme
		Segmental Late	*xue3-xuan3*	Onset, glide, and tone	Rhyme
		Tonal Early	*ying1-ying3*	Phonemes (nasal onset)	Tone (onset and offset)
		Tonal Late	*chou3-chou2*	Phonemes (obstruent onset)	Tone (offset)

2.2 Relative Weighting of Sub-syllabic Constituents in Spoken Word Recognition

A second issue is whether, and if so, to what extent sub-syllabic segmental constituents (i.e., onset and rhyme) and lexical tone affect lexical activation. Continuous mapping models such as TRACE (McClelland & Elman, 1986) predicted that word candidates with the same onset are activated earlier and greater than word candidates with the same rhyme. Such a prediction was confirmed for English by a seminal eye-tracking study by Allopenna, Magnuson, & Tanenhaus (1998) with the visual world paradigm. In this study, participants were asked to follow instructions (e.g., *Pick up the beaker*) and move objects around on a computer screen. They looked at both the target *beake*r and its phonological competitors (i.e., the cohort competitor *beetle* and the rhyme competitor *speaker*). Moreover, participants' eye fixations towards cohort competitors were significantly earlier than those of rhyme competitors.

However, the effect of sub-syllabic units (i.e., cohort and rhyme) reported for English seems less reliable in Mandarin SWR. With the same visual world paradigm, Malins & Joanisse (2010) examined the effect of phonological similarity on Mandarin word recognition. Their results showed that given a target such as *chuang2* 'bed', both segmental syllable (*chuang1* 'window') and cohort (*chuan2* 'boat') competitors distracted fixations towards target pictures significantly, with no difference between the two conditions in terms of effect size and time course. However, in contrast to the findings of Allopenna et al. (1998), rhyme competitors (e.g., *huang2* 'yellow') did not influence participants' gaze patterns more than unrelated distractors. These findings led Malins & Joanisse (2010) to propose that sub-syllabic constitutes weigh differently in Mandarin and English SWR.

Results reported in Malins & Joanisse (2010) are not fully replicated. Zou (2017) used a similar design and investigated phonological competition effects in

Mandarin SWR. Although the goal of the study was to examine SWR by second language learners of Mandarin (with Dutch as the first language), native Mandarin listeners were also included as a control group. Zou (2017) showed that the presence of rhyme competitors distracted participants' looks to targets the most. In contrast, the cohort competitors did not, which presents an opposite pattern from Malins & Joanisse's study. Another difference between Malins & Joanisse (2010) and Zou (2017) is the role of lexical tone in SWR. Malins & Joanisse (2010) reported an early interference effect of tonal competitors. Zou (2017), however, did not observe this effect. Similar to Zou (2017), Connell (2017) examined the process of word recognition in L1 and L2 Mandarin listeners with a visual world eye-tracking experiment. Unlike Malins and Joanisse (2010), in which comparable effects of segmental syllable and cohort competition were found, Connell (2017) found significantly more target eye-fixations in the segmental syllable condition than in the cohort condition. Overall, these different results raise further questions about the role of all sub-syllabic units (i.e., onset, rhyme, tone) in Mandarin spoken word processing.

It is worth noting that discrepant results between Malins & Joanisse (2010) and Zou (2017) are likely to lie in two major differences in their methods. One is the stimuli used for different competitor conditions, and the other is the different preview times for participants to view pictures before listening to the auditory stimuli.

About the stimuli, there are two differences. One concerns the cohort competitors. Malins & Joanisse (2010) defined the cohort competitors as sharing onset, tone, and the glide or rhyme with the targets (e.g., *hua1* 'flower'- *hui1* 'grey'; *tu3* 'dirt'- *tui3* 'leg'). In Zou (2017), cohort competitors were controlled more consistently as sharing only the lexical tone and the first phoneme in the onset (e.g., *tang2* 'candy'- *tou2* 'head'). The other concerns the repeated items. In Malins & Joanisse (2010), a few items were used repeatedly, especially in the tonal condition. For example, all tonal competitors were also presented as segmental/rhyme competitors; the word *mi3* (rice) was not only used as a tonal

competitor for both target word *xin1* (heart) and *tu3* (dirt), but also a segment competitor for target word *mi4* (honey). This led to an overall unequal number of occurrences for various phonological competitors and increased familiarity with tonal competitors. Zou (2017) avoided using the same stimuli for different phonological competitors.

As for the preview time difference, Malins & Joanisse (2010) allowed for a preview time of 1500 ms while Zou (2017) presented the pictures and auditory stimuli simultaneously. Preview time has been shown to affect phonological competition effects in the visual world paradigm (Huettig & McQueen, 2007; Huettig et al., 2011). Huettig & McQueen (2007) found that when participants viewed pictures at sentence onset (with an estimation of 700 ms -1000 ms preview time), substantial online phonological competition effects were found during Dutch word recognition. However, no phonological competition effect was found when participants viewed pictures with a preview time of 200 ms. Huettig & McQueen (2007) thus proposed that a 200 ms preview may not be sufficient for participants to retrieve the names of the displayed objects and associate them with locations in their visuospatial working memory. It is worth noting that Huettig & McQueen (2007) adopted a modified version of the visual world paradigm in which no target, but three different types of competitors were presented. Also, their participants were not instructed to give any explicit response. Given that how participants approached this task is still unclear (Magnuson, 2019), it leaves open the question of the impact of preview time on phonological competition effects. Specifically, is a 200 ms preview a prerequisite for observing phonological competition using a standard visual world paradigm? More importantly, to what extent the length difference of preview time could help to account for the discrepant results in Mandarin SWR.

To summarize, the different findings on the role of sub-syllabic constituents in Mandarin SWR may have resulted from different preview times and the unequal occurrences of the same stimuli as various phonological competitors. Therefore, new experiments with stricter control of stimuli

(Experiment 1) and different preview time (Experiment 2) would illuminate resolving the conflicting results.

2.3 Segment and Lexical Tone Processing in Mandarin Spoken Word Recognition

Whether the primary processing unit in Mandarin is a segmental syllable or a sub-syllabic unit, the third issue to address is when exactly segmental information and tonal information are recognized and utilized during SWR. Existing studies on the role of lexical tone in spoken word processing have mainly focused on whether lexical tone plays a similar role as segments. Using various behavioral tasks, a perceptual disadvantage of lexical tone, compared with segmental information, has been reported in earlier studies (Cutler & Chen, 1997; Taft & Chen, 1992; Yip, 2001; Ye & Connine, 1999, experiment 1). Such a view has also been supported by a few recent studies (e.g. Hu et al. 2012 with comparison to vowels; Sereno & Lee 2015 with comparison to segmental syllables; Gao et al., 2019 with comparison to segmental syllables). This body of literature reasoned that tonal information plays a weaker role during lexical processing because such information "often arrives later than does information about the vowel that bears the tone" (Cutler & Chen, 1997) and is "less informative than segmental information" (Tong et al., 2008).

An increasing number of studies, many with experimental techniques that are more sensitive to the time course of speech processing, have provided evidence that lexical tonal information is processed timely and can play an essential role during SWR. For example, Schirmer et al. (2005) showed that mismatched tonal and segmental (rhyme) targets induce comparable ERPs in Cantonese word processing with a sentence completion task. They thus argued that tone and segment play comparable roles and are accessed with a similar time course during spoken word processing. Using the visual world paradigm, Malins and Joanisse (2010) found comparable competition effects between cohort and

segmental competitors (in terms of amplitude and time course). This was interpreted as evidence that tonal and segmental information is accessed concurrently during online SWR.

Connell, Tremblay, and Zhang (2016) tapped into this debate by examining the low-level perceptual difference between tone and segments with a gated AX-discrimination task. Their native Chinese listeners showed a delay of about 28ms to perceive tonal contrast than segmental contrasts even when they have comparable acoustic divergent points. This raises a question: if there is indeed a delay of tonal perception, why is it not reflected in time-sensitive online experiments? Connell (2017) examined this issue further by conducting an eye-tracking visual world experiment. With the acoustic and perceptual divergence points strictly controlled, Connell found that lexical tones are used no later and even more rapidly than segments in constraining word activation. One possible explanation is that lexical tones are more efficient in eliminating potential lexical candidates than vowels. If this is the case, tonal information must be used in lexical access even before the tone can be recognized. Qin (2017) looked into this issue by conducting an eye-tracking experiment with tone pairs that either has early pitch height overlapping (T1-T2) or not (T1-T4). Qin found a larger target-over-competitor activation when there was an early pitch height difference. This suggests that pitch height information can be used early to constrain word recognition. Overall, findings of Connell (2017) and Qin (2017) and previous online studies, have provided evidence that lexical tone can be used before being recognized in lexical access.

Nevertheless, no studies have systematically examined and compared how the point of divergence (hereafter POD) affects tone and segments processing in Mandarin SWR with paradigms such as the visual world paradigm that are sensitive to the time course of speech processing. Experiments are needed to address the following open questions. First, given that there is evidence for holistic processing of syllable in Mandarin (Zhao et al., 2011), whether and how does POD affect the spoken word recognition process? Second, does lexical tone

(with early/late diverging pitch contours) constrain word recognition more than vowels? Answers to these questions would lend strong evidence to the exact time course of how the two tiers of information (i.e., segmental vs. tonal) are utilized for SWR. Experiment 3 was designed to fill this knowledge gap, which also serves to replicate existing findings in Qin (2017) and Connell (2017).

To summarize, the present study consists of three experiments and aimed to clarify the role of segmental syllable and sub-syllabic constituents in Mandarin SWR, as well as to investigate the time course of when segmental and suprasegmental tonal information is utilized during lexical processing. All three experiments were conducted within the visual world paradigm (Allopenna et al., 1998; Tanenhaus et al., 1995).

2.4 Experiment 1

Experiment 1 examined the role of segmental syllable, sub-syllabic segmental constituent (onset and rhyme), and lexical tone in Mandarin SWR, as indexed by how much participants' visual attention on the target word is disrupted by the presence of a phonological competitor (with an overlapping segmental syllable, onset, rhyme or tone) when they listen to a target Mandarin word.

Given the debates in the existing literature, particularly the discrepancies between Malins & Joanisse (2010) and Zou (2017), our goal was to replicate some of the findings conceptually to resolve the discrepancies. We followed Malins & Joanisse (2010) for most of the design, but made some necessary modifications as motivated earlier:

First, we avoided using the same stimuli as different phonological competitors and kept the number of occurrences of tonal competitors the same as other competitors. This would ensure that the tonal competitor effect reported in Malins & Joanisse (2010) is introduced by lexical tone overlap and not due to the effect of familiarity. Note that following both studies, we made sure that there was an equal number of reciprocal trials in which the role of target and competitor in

critical trials was reversed. Thus, participants' chances of hearing the target and competitor in a trial remained the same. Additionally, we also made sure participants' chances of seeing the target and competitor pictures were the same by arranging competitors of one target as unrelated distractors of another. In this way, participants' chances of predicting the targets and developing strategic responses were controlled to be small.

Second, we changed a subset of the stimuli used in the cohort condition. Following Zou (2017), we defined the cohort competitor as sharing only the first onset phoneme with the target. This is because our stimuli are monosyllabic words. In SC, monosyllables either constitute a word or at least a morpheme; the syllable structure (C)V(C) (with optional onset and coda) which serves as the bearing unit of lexical tone is also relatively simple. Based on such characteristics, previous studies on Chinese lexical access often examined the role of onset and rhyme, respectively (e.g., Ho et al., 2019; Yip, 2001; Zhao et al., 2011; Zou, 2017). With an auditory priming lexical decision task, Yip (2001) observed that onset and tone overlapping between target and prime elicited an inhibitory effect whereas rhyme and tone overlapping introduced a facilitatory effect in Cantonese. The first onset phoneme (plus lexical tone) likely plays an independent and rather different role from rhyme (plus lexical tone) in Chinese. Thus, to better compare the relative contribution of sub-syllabic constituents in SC, we selected words that share the first onset phoneme and tone with targets as cohort competitors, despite that traditionally cohort words for studies in Germanic and Romance languages have been defined as sharing two or more phonemes (Marslen-Wilson, 1987).

2.4.1 Method

2.4.1.1 Participants

Twenty (mean age: 20, standard deviation: 0.8; 12 females, 8 males) native Mandarin speakers participated in the experiment. All participants were college students from Shaanxi Normal University. All of them reported normal hearing and no history of speech or language disorders. All participants identified

Standard Chinese as their first language, and none of them speak other regional Chinese dialects. This study was approved by the Ethics Committee at Leiden University Centre for Linguistics. All participants provided informed consent before participation and were paid 30 RMB in compensation for their time.

2.4.1.2 Stimuli

The stimuli consisted of 60 monosyllabic Mandarin words which are easily picturable nouns (see Table A1 in Appendix A). Among the stimuli, 12 were critical targets. For each critical target, competitors of four conditions were defined based on their phonological overlap with the target. Segmental syllable competitor shared all phonemes but differed in tone with the target; cohort competitor shared the initial consonant and tone with the target; rhyme competitor shared rhyme and tone with the target; tonal competitor shared tone alone with the target. See Table 1 for sample stimuli and their comparison with previous studies. No item was used in more than one competitor condition.

Word frequency, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010), was balanced across target words and the four competitor conditions [$F(4, 55) = 0.83$, $p > 0.5$]. All stimuli were recorded through a Sennheiser MKH416T microphone (44.1 kHz, 16 bit) at the Phonetics Lab of Leiden University, produced by a female native speaker of Standard Chinese who was born and grew up in Beijing. Each word was read four times in isolation using a randomized list. One token of each word was chosen based on its clarity. All stimuli were normalized for intensity at 70dB. The matching pictures were real object pictures selected with the assistance of three native Chinese speakers who did not participate in the experiment.

2.4.1.3 Procedure and Design

To ensure participants were familiar with all stimuli, a naming task was assigned preceding the eye-tracking recording. During the naming session, participants were shown the pictures and asked to name them with appropriate

Standard Chinese words. If the name produced was not the intended word, participants were provided with the intended name.

During the subsequent experiment, participants were tested in a sound-attenuated booth at the Psychology Lab of Shaanxi Normal University. While performing the task, participants' eye movements were recorded with SR Eyelink Portable DUO eye-tracker at a sampling rate of 500Hz. For visual stimuli display, a 24-inch DELL U2412M monitor was located behind the eye-tracker. The camera of the eye-tracker was at a distance of about 52 cm from the participants' eye, which was fixed with the help of a chin rest. The auditory stimuli were played over a Beyer DT-770 Pro dynamic headphone at a constant and comfortable hearing level.

Before the test, participants' eye gaze position was validated and calibrated with a 9-point grid. At the beginning of each trial, a central cross appeared on the screen for 500 ms. Participants were asked to look directly at the fixation for a drift check. Four pictures then appeared on the screen for 1,500 ms before an auditory word. The four pictures (300 × 300 pixels) were placed top-left, top-right, bottom-left, and bottom-right; each comprising a distinct quadrant of the display. Participants were required to click on the picture that matches the auditory word with a mouse. The next trial appeared 1,000 ms after the click. The target picture's position was counterbalanced so that the target picture appeared an equal number of times in each location, and did not appear in the same location in two consecutive trials.

All the instructions were given in Standard Chinese. Participants were first asked to complete a practice block of four trials. In total, there were 360 trials for four blocks of 90 trials. The block order was counterbalanced across participants. Between each block, participants were given time to rest and proceed as they wish. Each of the syllable, cohort, rhyme, and tonal conditions has 36 trials, in which the participants listened to the targets with corresponding phonological competitors in the display. Additionally, there was a baseline condition in which no competitor but only distractors were presented along with the target. Following

the design of Malins & Joanisse (2010), half of the trials (180) were fillers, in which the role of target and competitors were reversed (i.e., the phonological competitors were played as auditory targets). This was done so that the chances of hearing the target and competitors with the same display were equal. Furthermore, to balance the overall occurrences of target and competitor items as picture displays, competitors were taken as unrelated distractors in another set of stimuli. The same target did not appear in three consecutive trials. After the test, participants were asked to fill in a language background questionnaire.

2.4.2 Data Analysis

2.4.2.1 Analysis of Behavioural Data

Reaction time and response accuracy for mouse clicks were collected for statistical analysis. Reaction times were calculated with respect to the onset of the auditory word. Trials for which the reaction time was shorter than 250 ms were excluded for both accuracy and RT analyses. Furthermore, only correct responses were considered for RT analyses. RTs were analysed using the generalized linear mixed-effects model (GLMM) to account for the skewed distribution without the need to transform raw data (Lo & Andrews, 2015). A backward algorithm was used to select the model (Barr et al., 2013). A maximum model including fixed effects of experimental conditions, by-subject and by-item random intercept, by-subject and by-item random slopes for experimental conditions was constructed first. If a model failed to converge, we first increased the number of iterations, then simplified the model by removing correlation parameters and the random structure's main effects (Brauer & Curtin, 2018). Fixed effects and the random structure were tested by comparing the likelihood ratio test with the simpler model. Response accuracy was modelled using the same approach using GLMM. All the analyses were run in the R software (R Core Team, 2021) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015).

2.4.2.2 Analysis of Eye-Tracking Data

We excluded trials for which the target was not correctly identified and trials for which the reaction time was shorter than 250 ms. The time window of 200-980 ms post auditory stimuli onset was chosen as our interest period. The lower boundary was chosen because there is a 200 ms delay to launch an eye movement (Hallett, 1986), while the upper boundary reflects when there were approximately maximum looks towards the targets. As the gaze position and duration of participants' eye fixation were recorded, looks toward targets, competitors, and distractors during the interest period were collected. The collected eye-tracking data were first down sampled to 50Hz (a 20 ms bin), following the tutorial of Porretta et al. (2018). Then, the proportions of fixations to target, competitor, and distractors at each time point were calculated by dividing the sum of fixations on the four pictures (target, competitor, and two distractors) by the number of fixations toward each picture type. The eye-fixation data in the visual world paradigm is intrinsically binary, i.e., participants are either looking at the target/competitor or not. It has been questioned that treating the eye-tracking data as a ratio variable on a linear scale averaging across conditions may cause problems such as data distortion and the violation of the assumptions of parametric statistics (Huang & Snedeker, 2020). To avoid these issues, we performed empirical logit transformation with weights for variance estimation on eye-fixation proportions following the advice of Mirman (2014) and Porretta et al. (2018).

We used generalized additive mixed modelling (GAMM; Wood, 2011; Wood, 2017) to analyse the eye-tracking data. GAMM is a type of generalized mixed-effects model that uses smooth functions to model the non-linearity between predictor(s) and the dependent variable. The smooth function (e.g., the thin plate regression spline) combines a number of pre-defined basic functions by multiplying them with individual coefficients. With cross-validation or maximum likelihood estimation, GAMM adds a penalization to the estimation of the

coefficients to avoid over-fitting and minimize errors. GAMM is well-established and has been applied to eye movement data of the visual world paradigm (e.g., Nixon et al., 2016; Nixon & Best, 2017; Porretta et al., 2018).

We used the *mgcv* package (version 1.8-23; Wood, 2011; Wood, 2017) in R (R Core Team, 2021) to implement GAMM. The model was fit by first entering all predictors of interest. Model comparison was conducted by means of $\chi 2$ tests of fREML scores, using the "compareML" function in the *itsadug* package (Van Rij et al., 2020). Model residuals were examined to check for non-normality, heteroscedasticity, and auto-correlation. The model summary of GAMM includes parametric coefficients and smooth terms. The parametric coefficients can be interpreted the same way as linear models, with the intercepts indicating the overall heights of the trajectories. The smooth terms capture the shape of the looking trajectories. To test the statistical difference between each experimental condition, we used ordered factors to model the difference smooth. The p-value in the smooth terms thus indicates the statistical difference between the trajectories in terms of shape. To control the family-wise error rate, the Holm–Bonferroni method was applied to adjust the p-values (Holm, 1979). We also plotted the difference smooths with *tidymv* (Coretta, Van Rij, & Wieling, 2021) to show when and how the look trajectories differ.

2.4.3 Results

2.4.3.1 Behavioural Data

Reaction time and response accuracy for mouse click are shown in Table 2. For reaction time, the maximum likelihood estimation of the maximum model and the simplified random slope models failed to reach convergence. The best-fit model included fixed effects of experimental condition, by-subject and by-item random intercepts (note that random-intercepts-only models may inflate Type-I error rate; Barr et al., 2013). The fixed effects of experimental conditions ($\chi 2$ (4) = 81.221, $p < 0.001$) suggested that participants' reaction time differed across conditions. Post-hoc analysis revealed that only when segmental syllable

competitors were present, participants took longer time to identify the targets (segmental syllable condition: $p < 0.001$; cohort condition: $p = 0.454$; rhyme condition: $p = 0.775$; tonal condition: $p = 0.075$). The error rate was low in each condition (all approximately under 1%). Thus, no further analyses were conducted on the response accuracy.

Table 2. *Mean Reaction time (ms) and response accuracy percentage of Experiment 1. Standard Errors are in parentheses.*

Condition	Reaction Time (SE)	Percent Accuracy (SE)
Baseline	1053 (25.4)	99.7 (2.32)
Cohort	1067 (29.4)	98.9 (8.65)
Rhyme	1056 (26.9)	99.8 (2.05)
Segmental syllable	1116 (31.2)	99.4 (3.41)
Tonal	1088 (32.6)	99.7 (2.76)

2.4.3.2 Eye Movement Data

Looks to target

The final model of target fixations includes a fixed effect of condition, a smooth term of time, a smooth over time for each level of condition, and a non-linear random effect of subject by condition. The final model explains 98.4% of the deviance. The summary of model fit is provided in Table 3. The upper half of this table presents the parametric coefficients of the model. The first row presents the intercept of the baseline condition. The following rows indicate the changes in the intercept for the other four experimental conditions. As shown in Table 3, no condition was found to be significantly different from the baseline condition in the intercept.

The second half of Table 3 describes the thin plate regression spline smooths for different levels of conditions over time. The first smooth presents the trajectory of the (empirical logit transformed) proportions of eye fixations over time for the baseline condition. The next four smooths evaluate the curves'

difference with respect to the baseline condition. The model summary indicates that there was a significant difference between the segmental syllable and the baseline conditions ($p < 0.005$).

The smooths for all levels of conditions are visualized in Figure 1A. Figure 1B plots the difference between the two smooths comparing the segment and baseline condition.

Table 3. *GAMM analysis of fixation proportions to targets in Experiment 1 with 1500 ms preview time.*

	Estimate	Std. Error	t value	*p*-value
Intercept	0.467	0.126	3.716	<0.001
Cohort–Baseline	-0.017	0.179	-0.097	0.923
Rhyme–Baseline	0.007	0.181	0.039	0.969
Segmental syllable–Baseline	-0.161	0.179	-0.896	0.370
Tone–Baseline	0.014	0.179	0.079	0.937
	edf	Ref.df	F	*p*-value
s(Time)	8.658	8.739	175.435	<0.001
s(Time): Cohort–Baseline	1.001	1.001	0.043	0.836
s(Time): Rhyme–Baseline	1.001	1.001	1.027	0.311
s(Time): Segmental syllable–Baseline	4.037	4.389	4.197	0.002
s(Time): Tone–Baseline	1.001	1.001	0.615	0.433

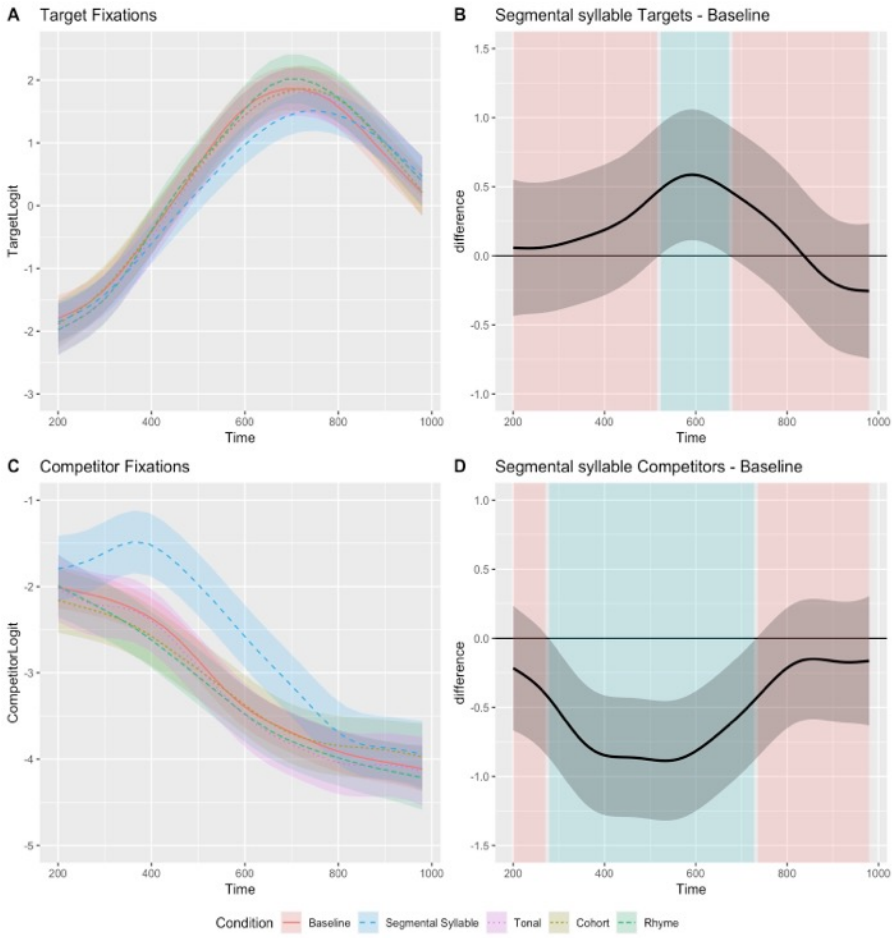

Figure 1. *Estimated smooths for all conditions and smooth differences in Experiment 1. A. Smooths of target fixations for all conditions; B. Difference between the two smooths comparing the segmental syllable and baseline condition of target fixations model; C. Smooths of competitor fixations for all conditions; D. Difference between the two smooths comparing the segmental syllable and baseline condition of competitor fixations model. The pointwise 95%-confidence intervals are shown by shaded bands. The green background in B and D indicate that the shaded confidence band is significantly different from zero.*

Looks to competitors

Same as target fixations, the final model of competitor fixations includes a fixed effect of condition, a smooth term of time, a smooth over time for each level of condition and a non-linear random effect of subject by condition. The final model explains 97.2% of the deviance. The summary of model fit is provided in Table 4.

Table 4. *GAMM analysis of fixation proportions to competitors in Experiment 1 with 1500 ms preview time.*

	Estimate	Std. Error	t value	*p*-value
Intercept	-3.179	0.096	-33.283	<0.001
Cohort–- Baseline	-0.009	0.140	-0.063	0.950
Rhyme- Baseline	-0.054	0.127	-0.422	0.673
Segmental syllable–- Baseline	0.536	0.169	3.170	0.002
Tone–- Baseline	-0.030	0.135	-0.224	0.823
	edf	Ref.df	F	*p*-value
s(Time)	7.675	8.050	32.905	<0.001
s(Time): Cohort–- Baseline	1.000	1.000	0.738	0.390
s(Time): Rhyme- Baseline	1.001	1.001	0.114	0.736
s(Time): Segmental syllable–- Baseline	5.396	5.811	5.230	<0.001
s(Time): Tone–- Baseline	1.000	1.000	0.075	0.785

The parametric coefficients of the model indicate that only the segmental syllable condition was significantly different from the baseline condition in intercept ($p < 0.005$). In the segmental syllable condition, the empirical logit of eye-fixation proportions towards competitors was higher than that of the baseline condition by 0.536.

The smooth terms of the GAMMs (as shown in Table 4) indicate that there was a significant difference between the segmental syllable and the baseline conditions over time ($p < 0.001$). The smooths for all levels of conditions are

visualized in Figure 1C. Figure 1D plots the smooth difference between the segmental syllable and baseline condition.

2.4.4 Discussion

Results of Experiment 1 showed a significant segmental syllable competitor effect, confirming findings reported in Malins and Joanisse (2010) and Zou (2017). Different from findings in Malins and Joanisse (2010) but confirming Zou (2017), there were no cohort and tonal competition effects. Note that the different cohort effects are likely due to the different definitions of the cohort (see further discussion below). Furthermore, different from Zou (2017), no rhyme competition effect was observed, confirming the lack of rhyme competition effect reported in Malins & Joanisse (2010). The different findings in the rhyme condition may be in part due to the different preview times. The possible effects of preview time on spoken word processing were addressed in Experiment 2.

To summarize, our study confirmed that segmental syllable competitors exhibit a larger competition effect over cohort, rhyme, and tonal competitors. The results thus lend further support that segmental syllable has an overall advantage over sub-syllabic segmental constituents and lexical tone during SWR. The effects of sub-syllabic units seem much more variable and seem to be subject to the influence of factors such as preview time.

2.5 Experiment 2

Experiment 2 aimed to investigate the impact of preview time on the phonological interference effects during lexical processing. In this experiment, we changed the preview time to 200 ms (from the 1500 ms in Experiment 1) while keeping everything else the same across the two experiments. If the amount of preview time given to participants is indeed a critical factor for some of the inconsistent findings, we should observe different results from Experiment 1, in similar ways as some of the results of Malins & Joanisse (2010) differ from that of Zou (2017).

We opted for a 200 ms preview instead of no preview for two main reasons. First, preview time allows listeners to perform object recognition, visual search, and other non-lexical processes before the onset of the spoken word; without it, listeners must attend to visual properties simultaneously, which has been found to add noise to the phonological competition effects (Apfelbaum, Klein-Packard & McMurray, 2021). Second, as mentioned earlier, 200 ms has been found to be insufficient for observing phonological competition with a non-standard visual world paradigm (Huettig & McQueen, 2007). Whether such a short preview time would delay or even cancel phonological competition effects with a standard visual word paradigm has been questioned since and is worthy of further investigation (Magnuson, 2019).

2.5.1 Methods

2.5.1.1 Participants

Twenty-three (mean age: 19, standard deviation: 1.8; 14 females, nine males) new native Mandarin speakers participated in the experiment. As in Experiment 1, all participants were college students from Shaanxi Normal University, with normal hearing and no history of speech or language disorders. All participants speak Standard Chinese and no other Chinese varieties. This study was approved by the Ethics Committee at Leiden University Centre for Linguistics. All participants provided informed consent before participation and were paid 30 RMB in compensation for their time.

2.5.1.2 Stimuli

The same stimuli of Experiment 1 were used.

2.5.1.3 Procedure and Design

The same procedure of Experiment 1 was used, except that the amount of time given to participants for viewing the pictures before the auditory stimuli was shortened from 1,500 ms (in Experiment 1) to 200 ms.

2.5.2 Results

2.5.2.1 Behavioural Data

Reaction time and response accuracy for mouse click are shown in Table 5. The best-fit reaction time model included fixed effects of experimental condition, by-subject, and by-item random intercepts. There was significant effect for fixed effects of experimental condition ($\chi2(4) = 37.521$, $p < 0.001$). Post-hoc analysis revealed that only in the segmental syllable condition, reaction time was significantly different from the baseline condition (segmental syllable condition: $p < 0.005$; cohort condition: $p = 0.175$; rhyme condition: $p = 0.464$; tonal condition: $p = 0.445$). For the best-fit accuracy model, the fixed factor of condition did not improve model fit, which suggested that participants' response accuracy did not differ across conditions ($\chi2(4) = 4.6957$, $p = 0.32$).

Table 5. *Mean Reaction time (ms) and mean percent response accuracy of Experiment 2. Standard Error are in parentheses.*

Condition	Reaction Time (SE)	Percent Accuracy (SE)
Baseline	975 (41.9)	96.9 (2.12)
Cohort	976 (39.9)	97.6 (1.29)
Rhyme	988 (46.3)	99 (0.81)
Segmental syllable	1054 (43.8)	98.2 (1.05)
Tonal	995 (46.5)	98.0 (1.25)

2.5.2.2 Eye Movement Data

Looks to target

The model of target fixations includes the main effect of condition, a smooth term of time, a smooth over time for each level of condition and a non-

linear random effect of subject by condition. The final model explains 98.5% of the deviance. The summary of model fit is provided in Table 6[2].

The parametric coefficients of GAMM analysis indicate that no condition was significantly different from the baseline condition in intercept. The smooth terms indicate that there was a significant difference in target fixations between the syllable and baseline conditions over time ($p < 0.001$). The smooths for all levels of conditions are visualized in Figure 2A. Figure 2B plots the smooth difference between the segmental syllable and the baseline condition. [3]

Table 6. *GAMM analysis of fixation proportions to targets in Experiment 2.*

	Estimate	Std. Error	t value	*p*-value
Intercept	0.116	0.206	0.565	0.572
Cohort – Baseline	-0.076	0.300	-0.254	0.800
Rhyme – Baseline	0.007	0.273	0.027	0.979
Segmental syllable – Baseline	-0.212	0.274	-0.776	0.438
Tone – Baseline	-0.001	0.280	-0.005	0.996
	edf	Ref.df	F	*p*-value
s(Time)	8.609	8.686	121.067	<0.001
s(Time):Cohort – Baseline	1.000	1.000	1.910	0.167
s(Time):Rhyme – Baseline	1.000	1.000	0.988	0.320
s(Time): Segmental syllable – Baseline	5.445	5.863	5.050	<0.001
s(Time):Tone – Baseline	1.000	1.000	0.484	0.486

[2] While Table 6 shows a significant difference between the segmental syllable and baseline condition, the plot of the difference smooth in Figure 2B did not show any difference over time. This discrepancy was most likely due to the use of different R packages ("mgcv" for the model summary; Wood, 2011; "tidymv" for visual inspection; Coretta, 2020). Given that model summary using ordered factors as significance testing are generally more reliable than visual inspections in GAMM (Soskuthy, 2021), we referred to the model summary as the final results of significance testing.

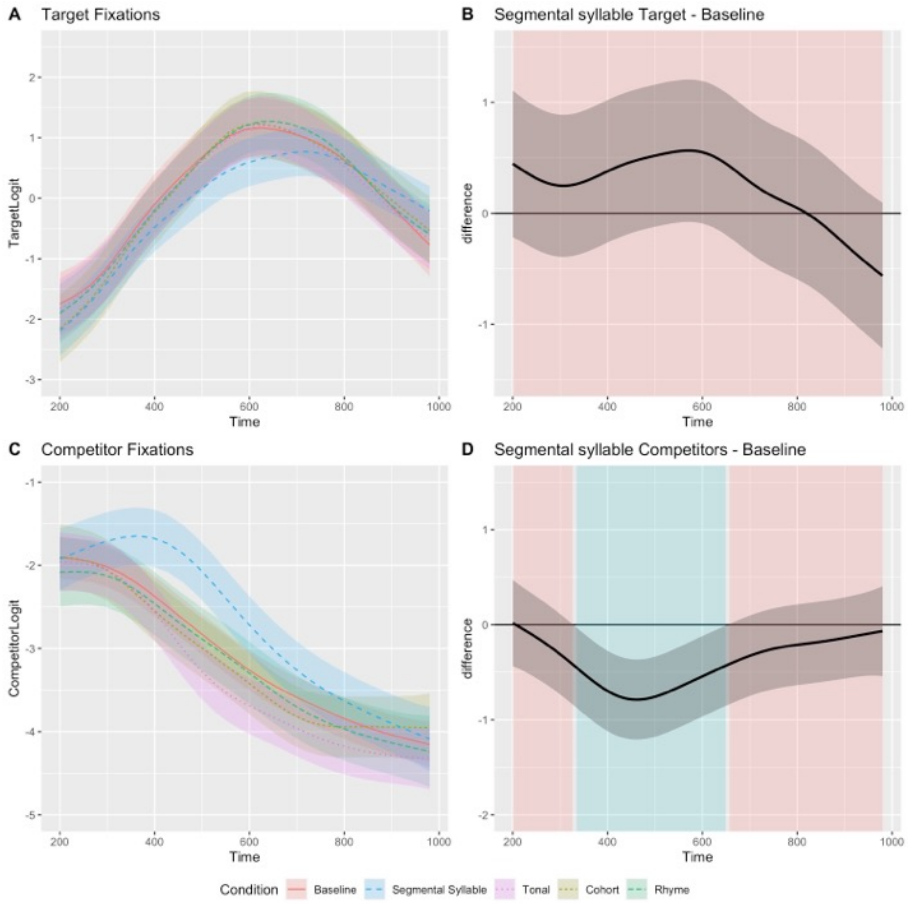

Figure 2. *Estimated smooths for all conditions and smooth differences in Experiment 2. A. Smooths of target fixations for all conditions; B. Difference between the two smooths comparing the segment and baseline condition of target fixations model; C. Smooths of competitor fixations for all conditions; D. Difference between the two smooths comparing the segmental syllable and baseline condition of competitor fixations model. The pointwise 95%-confidence intervals are shown by shaded bands. The green background in B. and D. indicates that the shaded confidence band is significantly different from zero.*

Looks to competitors

The final model of competitor fixations includes a fixed effect of condition, a smooth term of time, a smooth over time for each level of condition and a non-linear random effect of subject by condition. The final model explains 97.1% of the deviance. The summary of model fit is provided in Table 7.

Table 7. *GAMM analysis of fixation proportions to competitors in Experiment 2.*

	Estimate	Std. Error	t value	*p*-value
Intercept	-3.144	0.085	-37.067	<0.001
Cohort – Baseline	0.089	0.098	0.907	0.365
Rhyme – Baseline	-0.129	0.122	-1.058	0.290
Segmental syllable – Baseline	0.408	0.142	2.880	0.004
Tone – Baseline	-0.125	0.110	-1.132	0.258
	edf	Ref.df	F	*p*-value
s(Time)	7.317	7.720	26.253	<0.001
s(Time):Cohort – Baseline	1.000	1.000	0.161	0.688
s(Time):Rhyme – Baseline	1.000	1.000	0.025	0.874
s(Time):Segmental syllable – Baseline	5.125	5.525	4.616	<0.001
s(Time):Tone – Baseline	1.001	1.001	0.153	0.696

The parametric coefficients of the model indicate that only the segmental syllable condition was significantly different from the baseline condition in intercept ($p < 0.005$). In the segmental syllable condition, the empirical logit of eye-fixation proportions towards competitors was higher than that of the baseline condition by 0.408.

The smooth terms of the GAMMs (as shown in Table 7) indicate that there was a significant difference between the segmental syllable and the baseline conditions over time ($p < 0.001$). The estimated smooths for all levels of conditions are visualized in Figure 2C. Figure 2D plots the estimated smooth difference between the segmental syllable and baseline condition.

We examined further the effect of preview time on Mandarin phonological competition effects. Similar general additive modelling procedures described above were applied to the combined data of Experiment 1 and Experiment 2. The interaction of preview time and experimental condition was added to the model and tested for exclusion.

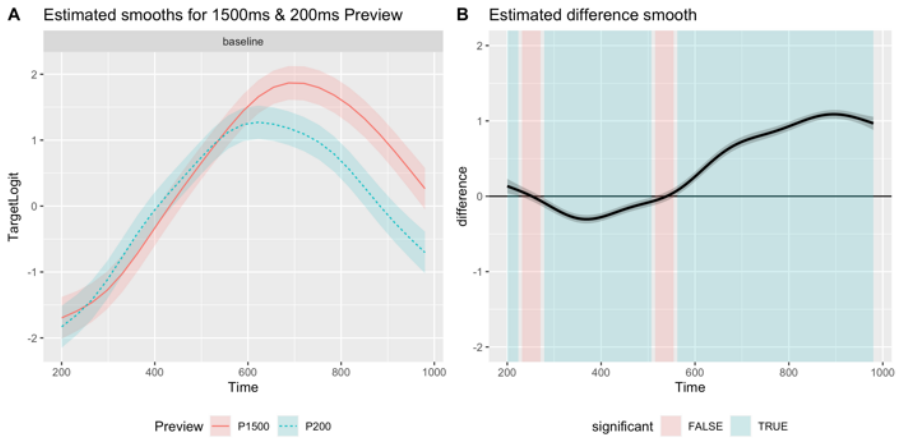

Figure 3. *Estimated smooths and smooth difference of target fixations between Experiment 1 & 2 (1,500 ms & 200 ms). A. Smooths of target fixations with 1,500 ms and 200 ms preview; B. Difference between the two smooths comparing target fixations with 1,500 ms and 200 ms preview. The pointwise 95%-confidence intervals are shown by shaded bands. The green background in B indicates that the shaded confidence band is significantly different from zero.*

Compared with models including the factor of condition, adding preview time significantly improved model fit of target fixations ($p < 0.001$). The interaction between preview and condition did not significantly improve model fit. Coefficients of parameter estimations (see Table 8) showed significant differences in intercept and smooth terms for different preview times (all $p < 0.001$). Figure 3A shows estimated smooths of target fixations of both preview times in baseline condition. Figure 3B shows the estimated smooth difference

between the two preview times. As can be seen from Figure 3A, target eye-fixations reach the peak around 700 ms post stimuli onset in Experiment 1; around 600 ms in Experiment 2. Moreover, the target fixation peak in Experiment 1 has higher empirical logit transformed proportion than in Experiment 2. The estimated difference smooth in Figure 3B shows a consistent pattern. Overall, with a short preview time (200 ms), participants' target fixation proportions reached the peak earlier with a relatively lower proportion compared with a long preview (1,500 ms).

Table 8. *GAMM analysis of fixation proportions to targets in Experiment 1 vs. Experiment 2.*

	Estimate	Std. Error	t value	*p*-value
Intercept	0.493	0.130	3.798	<0.001
Segmental Syllable	-0.347	0.029	-12.022	0.969
Cohort	-0.408	0.029	-13.860	0.969
Rhyme	-0.462	0.029	-15.789	0.314
Tonal	-0.316	0.027	-11.567	0.969
Preview P200-1500	-0.292	0.029	-10.165	<0.001
	edf	Ref.df	F	*p*-value
s(Time):Intercept	8.636	8.712	68.094	0.496
s(Time):Segmental Syllable	3.522	3.818	2.988	0.813
s(Time):Cohort	5.340	5.747	7.028	0.969
s(Time):Rhyme	4.289	4.654	3.455	<0.001
s(Time):Tonal	4.958	5.358	5.729	0.969
s(Time):Preview P200-1500	8.042	8.552	211.766	<0.001

As for the model of competitor fixations, the interaction of preview time and condition also significantly improved model fit ($p < 0.001$). Same as modelling target fixations, five ordered factors each presenting the difference between two preview times of one condition was created. Table 9 shows the estimations of parametric coefficients and smooth terms of the final model. The results show that while there was no significant difference between Experiment 1

and Experiment 2 in the baseline condition (intercept: $p = 0.910$; smooth term: $p = 0.720$), there were significant differences in the cohort condition (intercept: $p < 0.05$; smooth term: $p < 0.001$), the segmental syllable condition (smooth term: $p < 0.001$), the rhyme condition (intercept: $p < 0.05$; smooth term: $p < 0.001$), and the tonal condition (intercept: $p < 0.05$; smooth term: $p < 0.001$). Figure 4 shows the estimated smooth differences between two preview times for each experiment condition. Compared with Experiment 1, the segmental syllable competitors in Experiment 2 have more competitor fixations around 440-600 ms post stimuli onset, but fewer competitor fixations before 380 ms and after 800 ms post stimuli onset (see Figure 4B); the cohort condition has more competitor fixations around before 300 ms and after 640 ms, but fewer fixations during around 320-560 ms (see Figure 4C); the rhyme condition has more competitor fixations around 220-340 ms and 480-560 ms (see Figure 4D); the tonal condition has more competitor fixations before around 380 ms, 580-600 ms, but less during around 400-520 ms (see Figure 4E). As for the baseline condition, there is no significant difference between Experiments 1 and 2 (see Figure 4A). Overall, while the preview time difference (1,500 ms vs. 200 ms) did not affect the fixation in the baseline condition (in which no phonological competitors were presented), it did affect fixations towards different types of phonological competitors at different time intervals along the time course of recognizing the targets.

Table 9. *GAMM analysis of fixation proportions to competitors in Experiment 1 vs. Experiment2.*

	Estimate	Std. Error	t value	*p*-value
Intercept	-3.130	0.074	-42.544	<0.001
Baseline	-0.004	0.034	-0.113	0.910
Segmental syllable	0.055	0.040	1.383	0.500
Cohort	0.104	0.037	2.806	0.020
Rhyme	-0.100	0.031	-3.212	0.007
Tonal	0.121	0.036	3.348	0.005
	edf	Ref.df	F	*p*-value
s(Time)	6.403	7.101	42.484	<0.001
s(Time):Baseline	2.206	2.605	1.167	0.720
s(Time):Segmental syllable	6.976	7.705	15.890	<0.001
s(Time):Cohort	7.668	8.234	5.769	<0.001
s(Time):Rhyme	4.480	5.312	15.325	<0.001
s(Time):Tonal	7.802	8.372	17.400	<0.001

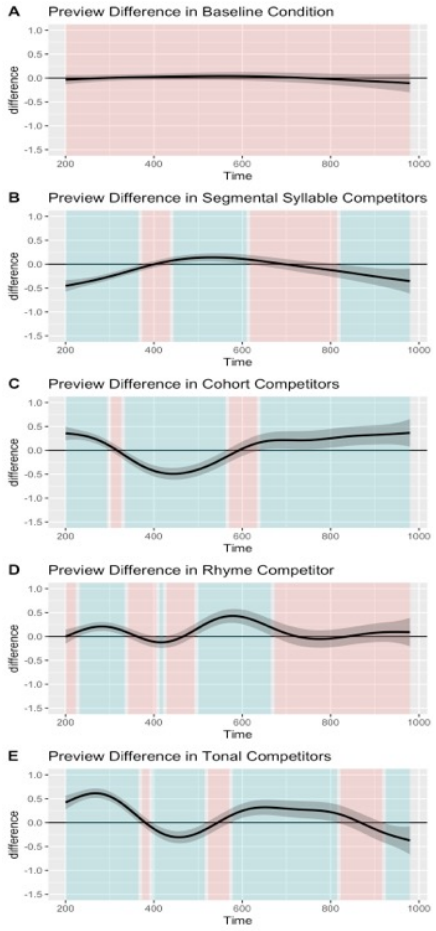

Figure 4. *Estimated smooths difference of competitor fixations between Experiment 1 & 2 (1,500 ms & 200 ms). A. Smooths difference between preview times in the baseline condition; B. Smooths difference between preview times for competitor fixations in the segmental syllable condition; C. Smooths difference of cohort competitor fixations between preview times; D. Smooths difference of rhyme competitor fixations between preview times; E. Smooths difference of tonal competitor fixations between preview times. The pointwise 95%-confidence intervals are shown by shaded bands. The green background indicates that the shaded confidence band is significantly different from zero.*

2.5.3 Discussion

As in Experiment 1, we found a robust competition effect when the segmental syllable competitors were present but null effects for the cohort, rhyme, and tonal competitors in Experiment 2. Nevertheless, with a close look at the effect of preview time on each experiment condition, we found that different preview times did affect participants' visual attention to targets and phonological competitors differentially. With a short preview time, target eye-fixations reached the peak sooner. Moreover, the peak of target fixation proportions with a short preview was lower than that with a long preview. These indicated that participants completed the visual search faster. It is likely that the short preview time created an overall faster rhythm of the task. Furthermore, under different preview times, there were also slightly different look trajectories for phonological competitors. Participants seemed to pay more attention to different phonological competitors at different time points over the time course of the target recognition. Such differences in competitor fixations between the two preview times need to be further verified.

Overall, regardless of having a short or a long preview time (i.e., 200 ms vs. 1,500 ms), the segmental syllable competitor exhibits a significant phonological competition effect. Unlike in Huettig and McQueen (2007), which found reduced phonological competition with a 200 ms preview, our results indicate that the length of preview does not affect the general phonological competition patterns in Mandarin SWR. Possible explanations for such discrepancy and its implications are discussed in the general discussion.

It is necessary to note that both Experiment 1 and 2 have a small size with each having around 20 participants. Brysbaert (2019) recommended at least 50 participants using repeated measures and warned that studies underpower are more likely to miss genuine effects or increase false-positive results in the long run. We recognize the size limitation of our experiments and hereby remind the readers to interpret the results with caution. Given that Experiment 1 and 2

consistently generate the same eye-tracking pattern, however, we also feel that our findings on the segmental syllable condition are unlikely to be false outcomes.

2.6 Experiment 3

Experiment 3 took a closer look at the time course of how listeners utilize tonal and segmental information during online spoken word processing. We manipulated the timing of the point of divergence (POD: early vs. late) for acoustic cues in two information tiers (segmental vs. tonal) and set up five conditions accordingly. To bring the reader's attention to the divergent information, we named the conditions of Experiment 3 by the component that diverged; unlike Experiment 1 and 2, in which the conditions are named by the shared component. The five conditions are the early segmental (diverging) condition, which has word pairs with early diverging segmental information; the early tonal (diverging) condition, which has word pairs with early diverging tonal information; the late segmental (diverging) condition, which has word pairs with late diverging segmental information; the late tonal (diverging) condition, which has word pairs with late diverging tonal information; the baseline condition, which has unrelated word pairs. Participants' gaze patterns across conditions would effectively inform us when and how Mandarin listeners use tonal and segmental information during SWR. We hypothesized that if both lexical tone and segment are utilized during online lexical processing, phonological competition effects (indexed by more eye fixations towards targets and fewer eye fixations towards competitors compared with the baseline condition) should be observed for both tonal and segmental diverging word pairs. In case the utilization of segmental and tonal cues is time-locked to the presence of the cues, significant differences between the early and late diverging word pairs' competition effects should be observed. Specifically, the late conditions should show larger cumulative competition effects than the early ones regardless of the information tier.

Unlike in Experiment 1 and 2, we used Chinese characters as visual displays in Experiment 3. Huettig & McQueen (2007) have reported a stronger phonological competition effect in Dutch when using printed words than pictures as a visual display. They suggested that the version of printed words visual world paradigm is "more sensitive to phonological manipulations than the version using pictures". Experiment 3 tapped into how subtle phonetic cues are used during auditory word recognition. Suppose the use of Chinese characters serves the same function as alphabetic scripts in the visual world paradigm. In that case, it can help to zoom into subtle phonological competition effects that otherwise may not be found. Another benefit of using printed words is that it makes our experiment feasible. This is because we adopted a between-subject design (i.e., participants of Experiment 1 and 2 also participated in Experiment 3). To avoid using the same stimuli, we had only a limited number of picturable nouns available as stimuli, making the design practically infeasible.

2.6.1 Methods

2.6.1.1 Participants

Thirty-seven native Mandarin speakers (mean age: 19, standard deviation: 1.5; 21 females, 16 males) who participated in Experiment 1 or 2 also participated in Experiment 3. The order of participating in Experiment 1 and 3 (or Experiment 2 and 3) was counterbalanced.

2.6.1.2 Stimuli

In a total of 96 Mandarin monosyllable words, two groups of stimuli were used in Experiment 3 (see Table A2 in Appendix A). One group consists of 24 tonal pairs, of which one word differs from the other only in the lexical tone; the other group consists of segmental pairs, of which one word differs from the other only in the segment. Based on the POD, both groups were further classified as with early POD or late POD. The early tonal POD word pairs either had a nasal onset or no onset, so the entire syllable carries tonal information from the beginning of the syllable. Their lexical tones contrast with each other from the

beginning of their tonal pitch contours (e.g., Tone1, high-level tone vs. Tone3, low rising-falling dipping tone; Tone 4, high falling tone vs. Tone 3, low rising-falling dipping tone). The late tonal POD pairs have obstruent onsets that do not carry tonal information. Lexical tones either both start low (e.g., Tone 2, rising tone vs. Tone 3, rising-falling dipping tone) or both start high (Tone 1, high-level tone vs. Tone 4, high-falling tone), so their tonal divergence point occurs late. For word pairs diverging in segmental information, the early POD word pairs share the same onset which contains only one segment and diverges in rime (e.g., *pa2 - ping2*), while the late POD word pairs share not only the one onset segment but also the following glide (e.g., *xue3 – xuan3*), which has been analysed either as part of the onset or part of the rhyme (for further discussion on the treatment of glides, see e.g., Chen & Gussenhoven, 2015). Table 1 provides sample stimuli in Pinyin[3], an alphabetic writing system of Standard Chinese.

As discussed earlier, we used printed words instead of real object pictures as a visual display in this experiment. Word frequency, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010), was balanced across conditions [$F(3, 92)=1.871$, $p>0.1$]. Also, the number of components and strokes of the characters were controlled across conditions [Strokes: $F(3,92)=0.538$, $p > 0.5$; Component: $F(3,92)= 1.564$, $p>0.1$]. All stimuli were recorded in the Phonetics Lab of Leiden University through a Sennheiser MKH416T microphone (44.1 kHz, 16 bit). The speaker is a male native Standard Chinese speaker who was born and grew up in Beijing. Each word was read four times in isolation using a randomized list. One token of each word was chosen based on its clarity and normalized for intensity at 70dB.

[3] Note that the Pinyin system is designed for spelling out the Standard Chinese syllables, not for phonetic transcription or phonological analysis as the international phonetic alphabet.

2.6.1.3 Procedure and Design

The procedure in Experiment 3 was the same as that in Experiment 1. Each of the early segmental, early tonal, late segmental, late tonal, and baseline conditions have 24 trials. Another 72 trials were included as fillers in which no phonological-related items were presented. In total, there were 192 trials distributed in four blocks. As in Experiments 1 and 2, the order of blocks was counterbalanced between participants.

2.6.2 Results

2.6.2.1 Behavioural Data

Reaction time and response accuracy for mouse click are shown in Table 10. The same criteria used for data analysis (as described in Experiment 1) were adopted. To enable better comparison with baseline, we used experimental conditions with five levels of early segmental, late segmental, early tonal, late tonal, and baseline conditions as fixed effects to model fixation proportions to target. The best-fit reaction time model included fixed effects of experimental conditions [$\chi2(4) = 41.802$, $p < 0.001$], by-subject and by-item random intercepts, and by-subject random slop for testing conditions. Post-hoc analysis showed that participants' reaction times in all critical conditions were significantly longer than baseline (early segmental: $p < 0.05$; late segmental, early tonal, late tonal: $p < 0.001$). Furthermore, the reaction time in the early segmental condition was significantly shorter than in the late segmental condition ($p < 0.001$). So did the early tonal condition compared with the late tonal condition ($p < 0.05$). There was no significant improvement in model fit for the best-fit accuracy model after adding fixed effects of experiment conditions [$\chi2(4) = 7.627$, $p = 0.106$], suggesting that response accuracy did not differ across experimental conditions.

Table 10. *Mean Reaction time (ms) and mean percent response accuracy of Experiment 3. Standard Errors are in parentheses.*

Information	Timing	Reaction (SE)	Percent Accuracy (SE)
Segmental	Early	1103 (26)	97.6 (1)
	Late	1256 (38.2)	96.5 (0.8)
Tonal	Early	1160 (25.9)	97.5 (1.5)
	Late	1202 (31.7)	97.8 (1.46)
Baseline		1086 (26)	99.2 (0.8)

2.6.2.2 Eye movement Data

Looks to target

Generalized additive modelling (Wood, 2011; 2017) was also employed to model participants' eye fixations[4]. The same modelling procedure as in Experiment 1 and 2 was applied. The resulting model of target fixations includes a fixed effect for condition, a smooth over time for each level of condition (the baseline, early segmental, late segmental, early tonal, and late tonal conditions), and a non-linear random smooth of subject by condition. This final model explains 97.8% of the deviance. Pairwise comparisons between each level were conducted with ordered factors of different reference levels. The estimates for the parametric and smooth terms are summarized in Table 11[5]. The estimated smooths for all conditions are visualized in Figure 5A.

[4] For the ease of comparison to the existing findings (i.e., Malins & Joanisse, 2010; Zou, 2017), We have also analyzed the eye-tracking data with the growth curve analysis (GCA; Mirman, 2014). The results converge for most analyses except for Experiment 3, in which the results of GAMM are more conservative. Given the discussion in Huang & Snedeker (2020), we report our results based on the GAMM analysis.

[5] While Table 11 shows no significant difference between conditions, the plots of estimated smooth in Figure 6 did show some significant difference over time. The results shown in Table 11 are more conservative because the *p*-values were corrected with the Holm-Bonferroni method (Holm, 1979) to avoid family-wise errors.

Smooth term differences between each level are plotted in Figure 6. As shown in Figure 6, compared with the baseline condition, there are fewer target fixations in early segmental, late segmental, early tonal, and late tonal conditions about 250 ms after the auditory stimuli onset. However, according to the estimated parameters of GAMMs (see Table 11), all the differences against baseline were not statistically different.

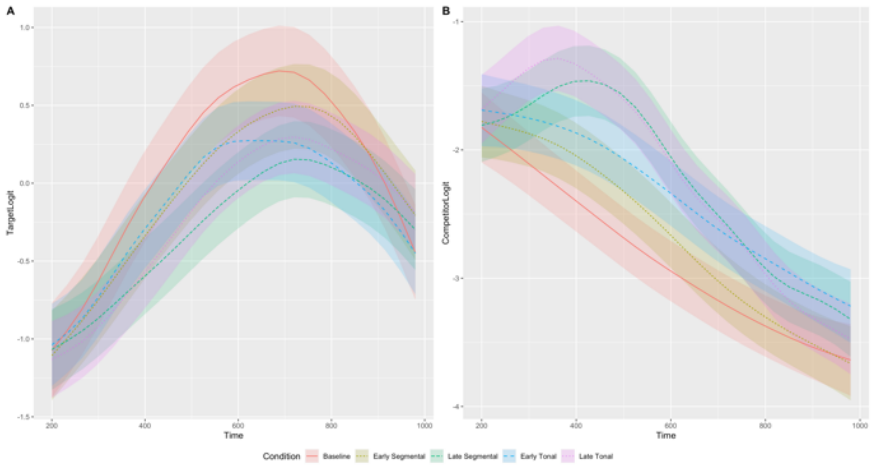

Figure 5. *Estimated smooths for all conditions in Experiment 3. A. Smooths of target fixations for all conditions; B. Smooths of competitor fixations for all conditions. The pointwise 95%-confidence intervals are shown by shaded bands.*

Table 11. *GAMM analysis of targets' fixation proportions in Experiment 3.*

	Estimate	Std. Error	t value	*p*-value
Intercept	-0.052	0.107	-0.487	0.627
Early Segmental – Baseline	0.056	0.141	0.395	1.000
Late Segmental – Baseline	-0.206	0.143	-1.446	1.000
Early Tonal – Baseline	-0.078	0.143	-0.542	1.000
Late Tonal – Baseline	-0.125	0.138	-0.907	1.000
Early Tonal – Early Segmental	0.097	0.127	0.766	1.000
Early Tonal – Late Segmental	-0.081	0.132	-0.619	1.000
Early Tonal – Late Tonal	0.001	0.126	0.008	1.000
Early Segmental – Late Segmental	-0.178	0.135	-1.325	1.000
Early Segmental – Late Tonal	-0.096	0.129	-0.746	1.000
Late Tonal – Late Segment	-0.047	0.118	-0.400	1.000
	edf	Ref.df	F	*p*-value
s(Time)	8.108	8.265	38.905	<0.001
s(Time):Early Segmental – Baseline	1.571	1.660	1.409	1.000
s(Time):Late Segmental – Baseline	3.995	4.335	3.865	0.053
s(Time):Early Tonal – Baseline	2.825	3.062	0.627	1.000
s(Time):Late Tonal – Baseline	3.617	3.927	3.168	0.268
s(Time):Early Tonal – Early Segmental	1.000	1.000	1.857	1.000
s(Time):Early Tonal – Late Segmental	3.048	3.307	2.207	1.000
s(Time):Early Tonal – Late Tonal	2.410	2.596	1.524	1.000
s(Time):Early Segmental – Late Segmental	3.047	3.307	2.145	1.000
s(Time):Early Segmental – Late Tonal	2.409	2.595	0.615	1.000
s(Time):Late Tonal – Late Segment	1.453	1.520	0.186	1.000

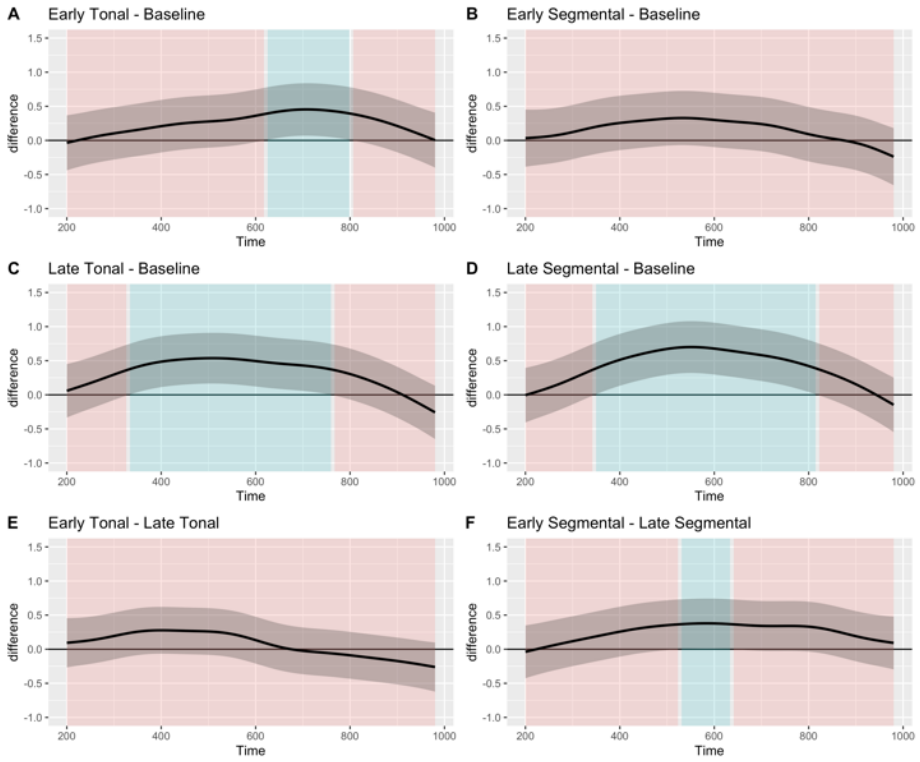

Figure 6. *Estimated smooths difference between experimental conditions for target fixations in Experiment 3. A. Smooths difference between the early tonal condition and the baseline condition; B. Smooths difference between the early segmental condition and the baseline condition; C. Smooths difference between the late tonal condition and the baseline condition; D. Smooths difference between the late segmental condition and the baseline condition; E. Smooths difference between the early tonal condition and the late tonal condition; F. Smooths difference between early segmental condition and the late segmental condition. The pointwise 95%-confidence intervals are shown by shaded bands. The green background indicates that the shaded confidence band is significantly different from zero.*

As for the effect of POD (early vs. late), although proportions of target fixations in both early conditions were slightly higher than that of late conditions

(Figure 6), the differences were not statistically different. The estimated parameters of GAMMs (see Table 11) indicate no significant differences between the information tiers (tonal vs. segmental) either.

Overall, target fixation proportions show two trends: 1) target pictures were less frequently looked at in the early segmental, late segmental, early tonal, and late tonal conditions than in baseline; 2) the late POD conditions generally has a larger effect on target fixations than the early conditions. Although these trends are observable with visual inspection, our GAMM analyses did not yield any statistical significance.

Looks to competitors

Same as models of target fixations, the final model of competitor fixations includes a fixed effect for condition, a smooth over time for each level of condition, and a non-linear random smooth of subject by condition. Pairwise comparisons between each level were conducted with ordered factors. The final model explains 96% of the deviance. The estimated smooths for all levels of conditions are visualized in Figure 5B. The estimates for the parametric and smooth terms are summarized in Table 12.

Table 12. GAMM analysis of competitors' fixation proportions in Experiment 3.

	Estimate	Std. Error	t value	*p*-value
Intercept	-2.789	0.095	-29.305	<0.001
Early Segmental – Baseline	0.126	0.133	0.948	1.000
Late Segmental – Baseline	0.592	0.146	4.053	0.001
Early Tonal – Baseline	0.425	0.136	3.127	0.020
Late Tonal – Baseline	0.613	0.135	4.546	<0.001
Early Tonal – Early Segmental	-0.269	0.117	-2.291	0.198
Early Tonal – Late Segmental	0.162	0.146	1.114	1.000
Early Tonal – Late Tonal	0.182	0.134	1.358	1.000
Early Segmental – Late Segmental	0.431	0.142	3.029	0.025
Early Segmental – Late Tonal	0.451	0.131	3.451	0.007
Late Tonal – Late Segment	0.057	0.142	0.397	1.000
	edf	Ref.df	F	*p*-value
s(Time)	4.622	4.987	15.420	<0.001
s(Time):Early Segmental – Baseline	2.840	3.073	1.257	0.312
s(Time):Late Segmental – Baseline	6.804	7.188	7.339	<0.001
s(Time):Early Tonal – Baseline	2.848	3.080	1.663	0.248
s(Time):Late Tonal – Baseline	6.782	7.160	8.172	<0.001
s(Time):Early Tonal – Early Segmental	1.001	1.001	1.243	0.312
s(Time):Early Tonal – Late Segmental	6.359	6.768	4.281	<0.001
s(Time):Early Tonal – Late Tonal	6.220	6.628	4.769	<0.001
s(Time):Early Segmental – Late Segmental	6.359	6.768	4.522	<0.001
s(Time):Early Segmental – Late Tonal	6.220	6.628	4.690	<0.001
s(Time):Late Tonal – Late Segment	6.220	6.628	4.769	0.312

As we can see from Figure 7, all experimental competitors attract more fixations than the baseline condition after 250 ms post auditory stimuli onset. As GAMMs parameters indicate (see Table 10), model fits of the late segmental, early tonal and late tonal conditions were significantly different from the baseline condition in intercept ($p < 0.001$; $p < 0.05$; $p < 0.001$). As for differences in the estimated smooth terms, only late segmental and late tonal conditions were significantly different from the baseline condition ($p < 0.001$; $p < 0.001$). The

early segmental condition did not significantly differ from the baseline in either intercept or smooth term.

As for the effect of POD (point of divergence in segmental and tonal information), model fits of the early and late segmental conditions significantly differed in intercept ($p < 0.01$) and smooth term ($p < 0.001$), while the early and late tonal conditions significantly differed in smooth term ($p < 0.001$). As Figure 7E and 7F show, participants looked more frequently at the late segmental and tonal competitors than the early competitors.

As for the differences between information tiers (segmental vs. tonal information), participants' competitor fixations in the early segmental condition seem to be overall less frequent than that in the early tonal condition, but the difference was not statistically significant. For word pairs of late POD, segmental competitors had a slightly higher proportion of eye fixations than the tonal competitors at the late time window. The difference was not statistically significant either.

Overall, competitors' eye fixations confirmed the general trends observed with target fixations. First, both tonal (early and late) and segmental (late) competitors attract participants' visual attention; Second, POD affects the proportion of eye-fixations on competitors regardless of the information tier: the later the information diverges, the more frequent eye-fixations on the competitors.

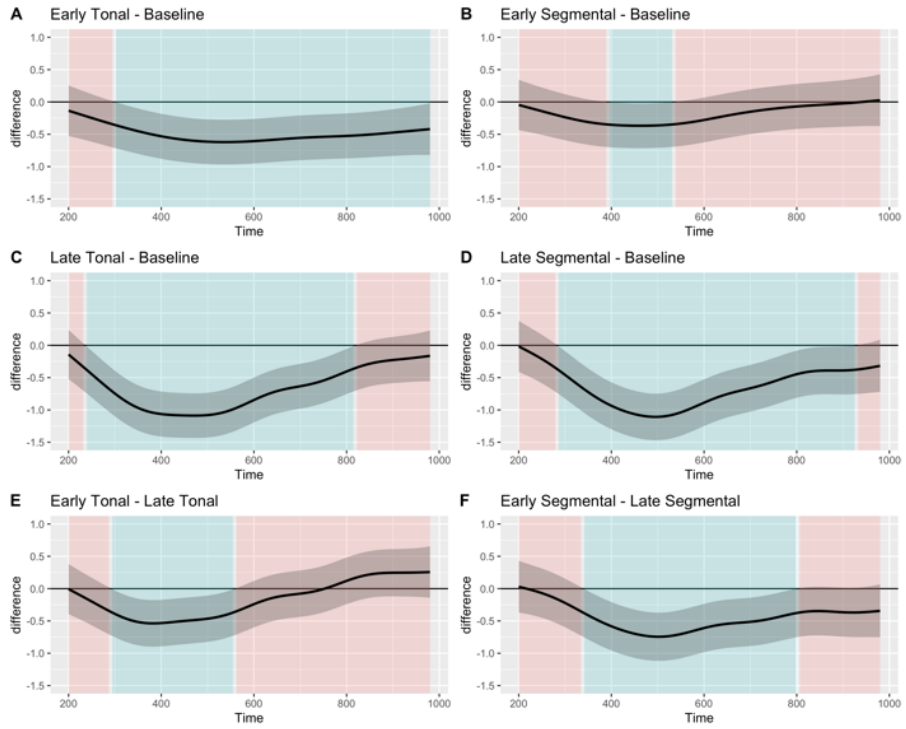

Figure 7. *Estimated smooths difference between experimental conditions for competitor fixations in Experiment 3. A. Smooths difference between the early tonal condition and the baseline condition; B. Smooths difference between the early segmental condition and the baseline condition; C. Smooths difference between the late tonal condition and the baseline condition; D. Smooths difference between the late segmental condition and the baseline condition; E. Smooths difference between the early tonal condition and the late tonal condition; F. Smooths difference between early segmental condition and the late segmental condition. The pointwise 95%-confidence intervals are shown by shaded bands. The green background indicates that the shaded confidence band is significantly different from zero.*

2.6.3 Discussion

Through careful control of the phonological overlap and timing of acoustic information (i.e., the Point of Divergence; POD), we confirmed that phonologically similar words, be it segmental or tonal, drew participants' visual attention more than unrelated distractors. This indicates that similar words with divergent phonemes or lexical tones are co-activated and compete for recognition in Mandarin lexical access. Moreover, the size and the time course of the phonological competition effects were modulated by the timing of the point of divergence in the acoustic signal. The later the information disambiguates, the larger the competition effects. These findings lend solid support to the view that Mandarin listeners use both tonal and segmental information incrementally during Mandarin SWR.

2.7 General Discussion

The present study examined the role of segmental syllables, sub-syllabic constituents, and lexical tone in Mandarin SWR and the time course of how segmental phoneme and suprasegmental lexical tone are utilized during lexical processing. Our findings suggest that segmental syllable plays a dominant role in Mandarin lexical processing while the effects of onset, rhyme, and lexical tone are more subtle and variable. Moreover, when all else is controlled, both segmental information and suprasegmental information can be used to constrain word competition as soon as their respective acoustic cues are present. Results of the three experiments have implications for both models of word recognition in tonal languages and methodological issues of using the visual world paradigm for SWR.

Experiments 1 and 2 examined the relative contribution of segmental syllable, onset, rhyme, and lexical tone. Specifically, we investigated when and to what extent participants' gazes are distracted by the presence of phonological competitors in recognizing Mandarin monosyllabic words. While Experiment 1

allowed participants to preview the pictures for 1,500 ms before listening to the target, Experiment 2 only allowed a short preview of 200 ms. Both experiments consistently found that only competitors with the same segmental syllable significantly distract participants' visual attention towards the target word. Cohort competitors (with onset and lexical tone overlaps), rhyme competitors (with rhyme and lexical tone overlaps), or tonal competitors (with lexical tone overlap) do not introduce more fixations than unrelated distractors.

Results of Experiments 1 and 2 thus replicate Malins and Joanisse's (2010) findings of the segmental (syllable) competitors and rhyme competitors, but not of the cohort and tonal competitors. Note that we have followed Zou (2017) and defined the cohort competitors as having only onset and lexical tone overlap with the targets. Our results confirmed Zou's (2017) finding of the null cohort effect. The robust cohort competition effect reported in Malins and Joanisse (2010) is likely due to the more extended overlap beyond a single onset phoneme (e.g., *hua1* 'flower' - *hui1* 'grey'; *tu3* 'dirt' - *tui3* 'leg'). Furthermore, by assigning tonal competitors an equal number of occurrences as other types of competitors and thereby avoiding a potential familiarity effect within the experiment, our results confirm the lack of tonal competition effect as in Zou (2017). This suggests that lexical tone alone does not have an impact on Mandarin SWR.

With a short preview time of 200 ms, Experiment 2 replicated the results of Experiment 1. This finding differs from that reported in Huettig and McQueen (2007), which showed a lack of phonological competitor activation with a short preview (200 ms). Huettig and McQueen (2007) proposed that 200 ms may be insufficient for Dutch participants to pre-activate the object names and consequently bias phonologically guided eye-fixations. With a series of eye-tracking experiments, Apfelbaum et al. (2021) have argued that phonological competition is not contingent upon pre-naming or pre-activating names during the preview. Instead, the preview allows participants some time to recognize visual objects so that their eye movements can better reflect lexical processing. Note that one particular design of Huettig and McQueen (2007) is that three types of

competitors, namely the visual, semantic, and phonological competitors, were all presented in one display. It is possible that when the preview is short, the visual search is delayed such that listeners may fixate first and primarily on the visual and semantic competitors in display and may not manage to fixate on the phonological competitor. Our results are consistent with Zou (2017) and Apelbaum et al. (2021), both of which found evidence of phonological competition even without preview. These studies suggest that the length differences in preview time (in our case a difference of 200 ms vs. 1,500 ms) do not influence the general pattern of phonological competition in Mandarin lexical access. Although the length of preview time is not a determining factor in phonological competition, it does influence how participants distribute their visual attention. We found that, with a shorter preview, participants located the target picture faster with fewer fixations. Moreover, there were different fixation patterns for phonological competitors when the preview was short. For example, there were slightly more frequent fixations on the rhyme competitors at a later processing stage than in Experiment 1. Future studies are still needed to fully understand how the length of preview time might affect looks to different phonological competitors.

Experiment 3 zoomed further into the time course of SWR and in particular, listeners' sensitivity to the acoustic details of segmental and tonal information. Word pairs (target and competitor) with divergent segmental information or tonal information were contrasted. With all else controlled, we were interested in whether the lexical co-activation and competition effect is modulated by the timing of the POD (i.e., the point of information divergence; early vs. late) along both the segmental and tonal dimensions. Results show that, while both early and late tonal competitors significantly attracted participants' visual attention, the late tonal competitors (which share the same segment and the onsets of tonal pitch contours with the target) exhibited a significantly larger effect than the early tonal competitors (which share the same segment with the target); segmental competitors only exhibited a significant effect when the

segmental information diverges late (which share the onset, glide, and tone with the target) but not early (which share the first onset phoneme and tone with the target). As for the relative weighting between the role of tone and segments in lexical access, no statistically significant difference was found between either early or late tonal and segmental conditions. Overall, we found that the competition effects were less persistent and weaker when the information diverges early in both conditions. The results of tonal conditions are consistent with the previous findings of Qin (2017), which confirms that lexical tone can be used early to constrain word activation before it is recognized. The results of segment condition provide further evidence against the view of holistic processing in Mandarin lexical access. Together with previous findings, our results show that both tonal and segment phonemic cues are incrementally processed as soon as they arrive.

In sum, the results of Experiments 1 & 2 indicate an advantageous role of segmental syllable over onset, rhyme, and lexical tone in activating word candidates. While Experiment 3 shows that, both tonal and segmental information can be used incrementally to constrain word candidates' activation during the process of Mandarin SWR.

How to model such effects? Previous studies have proposed several accounts of SWR in tonal languages (Gao et al., 2019; Malins & Joanisse, 2012b; Ye & Connine, 1999; Yue, 2016; Zhao et al., 2011; Tong et al., 2014; Shuai & Malins, 2017). The classic TRACE model (McClelland & Elman, 1986) posits a three-layer (word, phoneme, feature) architecture and bi-directional interconnections between layers. Existing models of SWR in tonal languages such as Mandarin typically add the "toneme" (Ye & Connine, 1999; Malins & Joanisse, 2012b; Zhao et al., 2011) or "tone" node (Gao et al., 2019; Yue, 2016) as the representation of lexical tone.

One disagreement among these existing models is whether an extra level of (tonal) syllable (Zhao et al., 2011) or segmental syllable (Yue, 2016; Gao et al., 2019) is necessary. The syllable node in Zhao et al. (2011) incorporates syllabic

morpheme (which includes both segmental syllable and tone) as a phonological representation to store morphemic syllables. By "hiding" phonemes and tonemes, the Reverse Accessing Model (RAM) in Gao et al. (2019) treats atonal syllable (i.e., segmental syllable) as "the earliest and smallest unit of phonological information immediately available for mental operations. In line with the proposal of RAM (Gao et al. 2019), our results also argue for the inclusion of segmental syllables at the sub-lexical level to account for its advantageous role in Mandarin lexical access. However, we remain sceptical about "hiding" phonemes and tonemes. The RAM proposes that tones and segmental phonemes are "hidden", i.e., can only be accessed when the information at the (atonal) syllable level is insufficient for the task at hand. This assumption well-explained the findings of the speeded discrimination tasks in Gao et al. (2019). For instance, it was easier for participants to make identical/different judgments on (segmental) syllables than phonemes or tones, because the latter would require re-activation of the phonemic and tonal information as a mental replay. Nevertheless, this assumption was not made for explaining the findings of the visual world paradigm. If only segmental syllable information is accessible when listeners were presented with spoken words, only words with the same or similar segmental syllable would be co-activated and compete for recognition. However, despite the robust competition effect of segmental syllables, effects of sub-syllabic components have also been found (e.g., late segmental competition effect in our Experiment 3; Malins & Joanisse, 2010; Zou, 2017). These findings of visual world paradigm seem to indicate that all information is maintained and can be used to aid SWR, which agrees more with the assumption of the TRACE model.

Another disagreement in the current models of tone-word recognition is whether the segment and tone processing are integrated (e.g., the TTRACE model; Tong et al., 2014) or separated (e.g., the TRACE-T model; Shuai & Malins, 2017). Zou et al. (2017) showed that native Mandarin listeners found it difficult to attend only to one of these two tiers of information, suggesting that at a certain level of processing, segmental and tonal information are integrated. Furthermore, it is also

relatively easier for them to attend only to segments (compared to only to tone), suggesting an asymmetrical relationship between the processing of these two tiers and, therefore the need for separate processing at other levels. There are also data from neural processing to shed light on this issue. Choi et al. (2017) examined the pre-attentive and phonological perceptual integration of vowels and tones in Cantonese using the oddball paradigm; the mismatch negativity (MMN) suggests the integration of vowel and tone processing at the phonological level. With the violation paradigm, Zou et al. (2020) reported different ERP effects for the rhyme and tone violation conditions, indicating different roles of tone and vowel at different stages of speech processing. Our study was not designed to explicitly test the integration or separation of segment ad tone processing. However, in Experiment 3, we do see substantial time course differences between the tonal and segmental diverging conditions. For example, the tonal condition had a significant early competition effect while the early segmental condition did not. Also, considering previous findings of tonal and segmental processing differences in terms of timing (e.g., Ye & Connie, 1999), speed (e.g., Connell, 2017), and relative weighting (e.g., Zou et al., 2020), it is more prudent for us to posit that at some levels of processing, tone and segments are processed independently, rather than integrated and holistically throughout the SWR process.

Based on our findings and data reported in the literature, we suggest a revised TRACE model for Mandarin SWR with a four-layer structure: syllable (i.e., segmental syllable and tone), segmental syllable, phonemes, and toneme, as well as their respective features. The extra level of segmental syllable accounts for the overall larger and more stable phonological competition effects of segmental syllable over a combination of sub-lexical phonological components (e.g., onset plus tone; rhyme plus tone) during Mandarin SWR. Moreover, with independent representations of phonemes and tonemes, both phonemic and tonal information can be used to resolve phonological competition when the context introduces enhanced sensitivity to the phonological information.

Having an extra unit of segmental syllable also echoes findings during online speech production. With classic paradigms such as implicit priming (Meyer, 1991), previous studies on Mandarin word production found effects of the atonal syllable (i.e., segmental syllable) but not of the initial onsets, which clearly differed from Indo-European languages (e.g., Chen, Lin, & Ferrand, 2003; Chen & Chen, 2013; Chen, O'Seaghdha & Chen, 2016; Wang, Wong, & Chen, 2018). O'Seaghdha (2010) therefore proposed that, whereas the proximate phonological encoding units in Indo-European languages are phonemic segments, it is the atonal syllable that is proximate in Mandarin. Roelofs (2015) adopted the proximate unit principle to the WEAVER++ Model (W. J. M. Levelt et al., 1999). Computational simulation results successfully explained the divergent findings between Mandarin and English, confirming cross-language differences in terms of the phonological planning units. Also adopting the proximate unit principle, Alderete et al. (2019) proposed a two-stage model for tone word production that not only incorporated the primary role of atonal syllables but also an early selection process of lexical tone, similar to the model structure we proposed. Nevertheless, to further explore the relation between tonal word production and recognition, future studies are needed.

Nonetheless, the findings of this study have to be seen in light of limitations. First, according to recent statistical advice (Brysbaert & Stevens, 2018; Brysbaert, 2019), our sample sizes are relatively small, which might reduce power and increase the margin of error. Second, due to the difficulty of finding sufficient items, we followed the design of Malins & Joanisse (2010) in using the targets repeatedly without dividing them into counter-balancing lists. How this practice may affect the data is still unclear, but it should be noted in interpreting the results and be avoided in future studies.

In summary, this study found that Mandarin listeners are sensitive to the unfolding segmental information and suprasegmental information and utilize both to constrain word recognition as soon as possible. Unlike in English or other West-Germanic languages, segmental syllable (syllable without specifying lexical

tone) plays a more advantageous role in Mandarin lexical access. Our results provide further data to adjudicate current and future models of tonal word recognition and shed new insights into the universal and diverse patterns of spoken word recognition across languages.

Chapter 3

Do bi-dialectal listeners activate both dialects during spoken word recognition?

A version of this chapter is under review: Yang, Q., & Chen, Y. (Under Revision). Do bi-dialectal listeners activate both dialects during spoken word recognition?. *Language and Speech*.

Abstract

Bilinguals are known to activate their two languages in parallel during spoken word recognition. What has remained debated is whether and, if so, to what extent speakers of two closely related dialects (i.e., bi-dialectals) also co-activate both dialects when listening to one. This study tested bi-dialectal speakers of Xi'an Mandarin and Standard Chinese. Both Standard Chinese and Xi'an Mandarin belong to the Mandarin Chinese family, sharing the same writing system and utilize lexical tones to differentiate words meanings. Using the visual world paradigm, we asked Standard Chinese - Xi'an Mandarin bi-dialectals to listen to sentences produced in either of the two varieties and identify the target word among four Chinese characters shown on screen. The characters included the target, two unrelated distractors, and a phonological competitor. The phonological competitor is either a cross-dialect homophone to the target or a cross-dialect translation-induced homophone. In addition, we also included a within-dialect condition, which contains competitors that share the same segmental syllable as the target but have different lexical tones. Listeners' eye movements showed that cross-dialect competitors (both as cross-dialect homophones and translation-induced homophones) did not influence participants' eye fixations more than the within-dialect segmentally overlapping competitors. These results suggest a lack of co-activation across dialects, which indicates a divergence between bilingual and bi-dialectal speech

processing. A bi-dialectal spoken word comprehension model is proposed to account for the results.

Keywords: Bi-dialectal; Spoken word recognition; Lexical tone; Language co-activation

Bilinguals differ from monolinguals in many aspects. One significant distinction is that bilinguals activate both their languages even when their task is to use only one (e.g., Marian & Spivey, 2003a, 2003b; Spivey & Marian, 1999). How about bi-dialectal speakers? This group of speakers is often ignored in research on speech processing. Bi-dialectals produce and comprehend both dialects in their daily lives, similar to bilinguals who are confronted with two languages. However, unlike bilinguals, the two varieties of bi-dialectals are typically similar and likely to be mutually intelligible. One question that has remained open is: do bi-dialectals activate their two dialects similarly to how bilinguals activate their two languages? In this study, we addressed this question by investigating whether bi-dialectals of Standard Chinese and Xi'an Mandarin experience cross-dialect interference during spoken word recognition, similar to what has been reported for bilinguals.

3.1 Language Co-activation in Bilingual Word Recognition

During spoken word recognition, multiple word candidates are co-activated and compete for selection. For bilingual speakers, word candidates from both languages are co-activated even when listening to just one. For example, in a seminal eye-tracking study by Spivey and Marian (1999), Russian-English bilinguals were asked to follow instructions such as *Poloji marku nije krestika* "Put the stamp below the cross" and move objects on a whiteboard while their eye movements were being recorded. In critical trials, objects such as "marker", which share initial phonetic features with *marku* "stamp", were also presented. Eye movement analysis showed that an interlingual near homophone "marker" attracted participants' visual attention from the target *marku* "stamp" significantly more than that of the unrelated control stimulus object (e.g., *lineika* "ruler"). Such an interference effect has been taken as evidence for the co-activation and interaction of bilinguals' two languages. Using the same eye-tracking task (i.e., the visual world paradigm; Allopenna et al., 1998), bilingual co-activation has

since been repeatedly found in different languages (e.g., Weber & Cutler, 2004 for co-activation of Dutch and English; Blumenfeld & Marian, 2007 for German and English; Shook & Marian, 2012 for American Sign Language and English).

Follow-up studies further explored potential factors that may remove or constrain language co-activation. With auditory lexical decision tasks, Lagrou and her colleagues (Lagrou et al., 2013a; 2013b) tested whether language co-activation is restricted by sentence context and semantic constraint. They found that highly predictive sentence context reduced cross-language interference compared with low-constraining context when tested in both L2 and their native language. Non-linguistic context such as task environment has also been found to play a role. For example, Marian and Spivey (2003) had monolingual experimenters in a bilingual study in which the bilingual participants were unaware of the bilingual nature of the study. In this way, they tried to create a monolingual lab environment. As a result, they did not replicate the significant interlingual interference effect observed during Russian spoken word recognition in Spivey and Marian (1999). Instead, they found a reversed interference effect from Russian to English spoken word recognition. It is, however, important to note that although factors such as semantic constraints of sentence context and task environment were found to inhibit language co-activation, they do not eliminate cross-language interference in bilingual spoken word recognition.

While most studies on language co-activation focus on interlingual homophones (e.g., marker – *marku* "stamp" in Russian), an increasing number of bilingual studies have also found evidence for "covert co-activation", i.e., the co-activation of translation equivalents (e.g., Thierry & Wu, 2007; Shook & Marian, 2017). Thierry and Wu (2007) asked Chinese-English bilinguals to judge whether a pair of English words were related in meaning. Unknown to the participants, in half of the trials, the Chinese translation equivalents of the English word pairs shared the first Chinese syllable (e.g., *you2chai1*"post"- *you2jian4* "mail"). This hidden repetition significantly modulated the N400 component (an ERP component associated with word processing; Kutas & Federmeier, 2011), similar

to the effect of the Chinese word pairs processed by Chinese monolinguals. This finding suggests activation of the native language's phonology even without any bottom-up input. Using the visual world paradigm (Allopenna et al., 1998), Shook and Marian (2017) replicated the covert co-activation effect with Spanish-English bilinguals. They found that when asked to listen to English words such as duck, Spanish-English bilinguals looked more to competitors such as a shovel compared with control pictures because the target and competitor overlap phonologically in Spanish (*duck* "pato"- *shovel* "pala"). These findings suggest that bilinguals not only co-activate both languages but also spread phonological competition across languages through translation links.

While most bilingual studies have focused on the segmental properties of the sound systems, a few studies have also examined whether co-activation can be observed in the suprasegmental domain of spoken words. For example, Wang, Wang and Malins (2017) investigated the role of Standard Chinese lexical tone in language co-activation. Unlike English or other Indo-European languages, Standard Chinese is a lexical tone language, in which lexical tone (realized via pitch variation) differentiates word meanings just as consonants and vowels. Using the visual world paradigm, Wang et al. (2017) found that when listening to an English word (e.g., rain), Chinese-English bilinguals looked more toward feather, whose Chinese translation equivalent (*yu* with a dipping tone) is a homophone with the target rain (*yu* with a dipping tone). What is interesting is that listeners did not look more toward fish, of which the Chinese translation equivalent (*yu* with a rising tone) has identical segments but a different tone. Such a contrast in the presence vs. absence of lexical tonal sharing between target and competitor not only provides further evidence for the non-selective access of bilinguals' two languages but also argues for a significant role that lexical tone plays in constraining cross-language activation.

In sum, the existing literature has provided quite convincing evidence that during spoken word recognition, bilinguals experience cross-language lexical competition even with highly predictive sentence context and under a

monolingual environment. Moreover, the phonological overlap between lexical items within/across languages plays a key role in automatic co-activation. What is particularly relevant for this project is that lexical tonal information is crucial when a tone language is involved.

3.2 Two Views of Bi-dialectalism

While bilinguals have been extensively studied with a general consensus on bilingual co-activation, only a few studies have examined bi-dialectal speech processing. There are two dominant views of bi-dialectalism: the independent view and the co-dependent view (as discussed in Melinger, 2018). According to the independent view (Hazen, 2001), dialects are independently represented and maintained in the same way as languages. Bi-dialectals are therefore predicted to be able to switch between dialects and would experience cross-dialect interference, in exactly the same way as bilinguals. The co-dependent view (Labov, 1998), however, argues that dialects are not independent but co-exist. Under this view, dialects are not expected to be co-activated or inhibited like languages. As a result, dialect processing should resemble that of monolingual processing.

To date, there have been few studies on bi-dialectal speech processing. Results from bi-dialectal spoken word production show mixed findings. Using the classic picture-word interference paradigm (Rosinski et al., 1975), Melinger (2018) investigated whether simultaneously processing a dialectal translation equivalent facilitates or inhibits picture naming in Scottish bi-dialectals. The predictions of this study are based on previous findings that within-language semantically related distractors should interfere with picture naming (Schriefers et al., 1990) while the presence of language translation equivalents should facilitate naming (e.g., Costa et al., 1999; Costa & Caramazza, 1999). Melinger (2018) found robust interference effects with Scottish bi-dialectals, which is similar to a within-language semantic interference effect and different from a dialectal translation equivalent facilitation effect, leading her to conclude that

these findings have "identified a clear point of processing departure between languages and dialects." The dialectal translation equivalent interference effect was recently replicated with American and British English (Melinger, 2021), which further validates the processing divergence between bilinguals and bi-dialectals.

The findings of Kirk et. al. (2018), however, lend support to the independent view. Previous studies have identified two indicators of cross-language interference in bilingual speech production. One is the language switch cost; bilinguals take longer to produce words or sentences after having had a trial to speak in a different language, compared with speaking in the same language in two consecutive trials (e.g., Meuter & Allport, 1999). The other is the cognate facilitation effect; bilinguals name cognates (i.e., etymologically related translation equivalents which overlap phonologically or orthographically) faster than non-cognate words (e.g., Costa, Caramazza, & Sebastian-Galles, 2000). Using a dialect switch task, Kirk et. al., (2018) observed both switch cost and cognate facilitation effect with German bi-dialectal speakers and Scottish bi-dialectal speakers. Kirk et. al., (2018) therefore concluded that bi-dialectals are similar to bilinguals in terms of the architecture of the lexicon and the control mechanism.

Note that the findings reported in Melinger (2018) and Kirk et al. (2018) all concern speech production. In the speech comprehension domain, listeners with exposure to more than one dialect have shown benefits or costs in their processing of dialectal variations (e.g., Sumner & Samuel, 2009; Clopper, 2014; Clopper & Walker, 2017). For instance, with a cross-modal lexical decision task, Clopper & Walker (2017) found that multi-dialectal listeners were less affected by phonetic dialect variation (i.e., the phonetic-acoustic similar vowels of the prime and target) in lexical judgment, compared with mono-dialectal listeners. They suggested that multi-dialectal listeners have relatively weaker vowel category boundaries, resulting in reduced activation of related lexical representations. While studies along this line of research have demonstrated the

significant role of linguistic experience in perceiving and representing dialectal variations (see Clopper, 2021 for a review on the perception of dialect variation), they do not directly tap into the question of whether bi-dialectal listeners experience activation and competition across dialects, similar to bilinguals.

Liu (2018) investigated whether bi-dialectal lexical access is non-selective for bilinguals. Participants were bi-dialectal speakers of Standard Chinese and Xi'an Mandarin both of which belong to the Mandarin dialect family within the Sinitic language family. They share similar syntactic structures, a large number of etymologically related translation equivalents, the same writing system, and largely overlapping segmental inventories. Moreover, the lexical tone systems of Standard Chinese and Xi'an Mandarin have a one-to-one mapping relation (Liu et al., 2020), resulting in a large number of homophones across Standard Chinese and Xi'an Mandarin. For example, *ma* with a high-level tone means "mother" in Standard Chinese, whereas it means "to scold*"* in Xi'an Mandarin. In a generalized lexical decision task with auditory priming, Liu (2018) manipulated five contrasts based on cross-dialect phonological similarity between the prime (e.g., Standard Chinese *bang* with a level tone meaning "help") and the first syllable of the target: 1) within-dialect segment and tone overlapping (i.e., identical; e.g., Standard Chinese *bang* with level tone meaning "help"); 2) within-dialect segment overlapping (e.g., Standard Chinese *bang* with a falling tone meaning "baseball"); 3) cross-dialect segment and tone overlapping (i.e., interdialectal homophone; e.g., Xi'an Mandarin *bang* with a level tone meaning "baseball"); 4) cross-dialect segment overlapping (e.g., Xi'an Mandarin *bang* with a falling tone meaning "help"); 5) unrelated (e.g., Standard Chinese *wan* with a rising tone meaning "finish"). The results showed that with Standard Chinese primes, there was a subtle facilitatory priming trend for identical and within-dialect segment overlapping targets. Furthermore, a significant interference effect for cross-dialect homophones was observed but not for cross-dialect segment overlapping targets, compared with the unrelated targets. Liu (2018) interpreted these results as evidence for non-selective access to the lexical

representations of both Standard Chinese and Xi'an Mandarin. In the identical condition, the co-activation of the Xi'an Mandarin words reduced the facilitation effect; in the cross-dialect homophone condition, the co-activation of the Standard Chinese words interfered with the recognition of Xi'an Mandarin targets. Moreover, the null result in the cross-dialect segment overlapping condition, in comparison with the cross-dialect homophone condition, was taken as due to the role of lexical tone in constraining non-selective lexical access of bi-dialectal spoken word recognition.

As a pioneer of bi-dialectal speech comprehension in a tonal language, Liu (2018)'s findings, however, remain to be further clarified, due to the following observations. First, it remains unclear whether bi-dialectal listeners co-activate words from both dialects when listening in one dialect. With a generalized lexical decision task, Liu (2018) presented either Standard Chinese or Xi'an Mandarin monosyllabic words as primes, followed by mixed Standard Chinese and Xi'an Mandarin disyllabic target words. According to Liu (2018), bi-dialectal listeners may have mistaken Xi'an Mandarin primes (e.g., *bang* with a high-level tone; "baseball") as their interdialectal homophone counterparts in Standard Chinese (e.g., *bang* with a high-level tone; "help"). It is thus important to investigate further whether bi-dialectal listeners experience cross-dialect interference when listening to their native dialect Xi'an Mandarin. Moreover, mixed contexts have been questioned for forming artificial dual-language environments and biasing bilinguals towards parallel activation (Grosjean, 1998; Thierry & Wu, 2007). Stronger evidence of bi-dialectal co-activation would come from spoken word recognition in a mono-dialectal sentence context.

To further understand whether and to what extent bi-dialects are analogous to bilinguals, we aimed to examine dialect non-selectivity in a mono-dialectal sentence context. Moreover, Liu (2018) drew evidence only from reaction time data, leaving the time course of possible cross-dialect competition effects unknown. To uncover such a time course, we used the eye-tracking technique and visual world paradigm (Allopenna et al., 1998). Third, Liu (2018)

mainly focused on how the phonological similarity of segments and lexical tone affect lexical competition and has thus left unaddressed whether dialectal translation equivalents are co-activated across languages. To further understand the degree of non-selectivity in bi-dialectal lexical access, we investigated not only the co-activation of inter-dialectal homophones but also dialectal translation equivalents.

To address the above remaining issues, we conducted a follow-up study of Liu (2018) with the following changes. First, we added a short mono-dialectal phrase *wo3 yao4 shuo1…* "I will say…" before each of the individual Standard Chinese or Xi'an Mandarin words to avoid dialect membership ambiguity. Second, we used a different task (i.e., visual world paradigm and eye-tracking) to tap into the time course of the dialect interference effect. Third, we added dialectal translation equivalents (i.e., translation-induced homophone condition), in addition to cross-dialect homophones, in order to gather more and hopefully, converging evidence on whether bi-dialectals co-activate both dialects during spoken word recognition.

3.3 Method

3.3.1 Participants

Thirty-four native Xi'an Mandarin speakers (mean age: 20, standard deviation: 2.1; 23 females, 11 males) who grew up in the urban area of Xi'an participated in the experiment.[6] All participants were college students from Shaanxi Normal University. All of them reported no history of speech or language disorders and normal hearing. All participants are proficient speakers of Xi'an Mandarin and Standard Chinese, and none speak other regional Chinese dialects. Their language background and proficiency were checked through a survey

[6] One participant's data were excluded from analysis for not completing the task.

adapted from the LEAP-Q questionnaire (Marian et al., 2007). This study was approved by the Ethics Committee at Leiden University Centre for Linguistics. All participants provided informed consent before participation and were paid 40 RMB in compensation for their time.

3.3.2 Design

The experiment includes two visual world paradigm tasks: the Standard Chinese task in which participants listened to Standard Chinese sentences only, and the Xi'an Mandarin task with Xi'an Mandarin sentences only. The instructions for the tasks were given orally in either Standard Chinese or Xi'an Mandarin according to the task. All participants performed both tasks and the order of the two tasks was counterbalanced. Between the two tasks, participants were asked to take a short break.

In each dialect task, participants listened to an auditory sentence that contains a target word and were instructed to select the corresponding target from four Chinese characters on the computer screen. The four Chinese characters included the correct target word, a phonological competitor word, and two unrelated distractor words. Based on the phonological relationship between the target and competitor, there were four experimental conditions: 1) the cross-dialect homophone condition (hereafter Homophone Condition), in which the target and competitors share segments within a dialect, while also sharing lexical tone across the two dialects; 2) the cross-dialect translation-induced homophone condition (hereafter Translation Condition), in which target and competitors share segments within a dialect, while the translation equivalents of the target also share lexical tone with the competitor; 3) the within- and cross-dialect segmentally overlapping condition (hereafter Segment Condition), in which target and competitors share only segments within a dialect or across dialects; 4) the baseline condition, in which target and competitors have no phonological overlap within a dialect or across dialects. See Table 1 for the within- and cross-dialect phonological overlap in critical conditions.

Table 1. *Experimental conditions with sample stimuli in Standard Chinese. The Pinyin system is the standard transcription for spelling out the Chinese syllables. SC is short for Standard Chinese. XM is short for Xi'an Mandarin. Phonological overlaps were indicated in* ***bold****.*

Experiment Condition		Target	Competitor
Homophone Condition	Character	借	姐
	Gloss	borrow	sister
	Pinyin	***jie4***	***jie3***
	SC pitch contour	**high-falling**	dipping
	XM pitch contour	level	**high-falling**
Translation Condition	Character	菜	猜
	Gloss	vegetable	guess
	Pinyin	***cai4***	***cai1***
	SC pitch contour	high-falling	**level**
	XM pitch contour	**level**	dipping
Segment Condition	Character	纸	直
	Gloss	paper	straight
	Pinyin	***zhi3***	***zhi2***
	SC pitch contour	dipping	rising
	XM pitch contour	high-falling	rising
Baseline Condition	Character	醋	猴
	Gloss	vinegar	monkey
	Pinyin	*cu4*	*hou2*
	SC pitch contour	high-falling	rising
	XM pitch contour	level	rising

Our choice of stimuli was based on the cross-dialect segmental and lexical tone properties described in Liu et al. (2022). As we can see from *Figure 1* (Liu et al., 2022, p.2808), Standard Chinese Tone 4 (T4) and Xi'an Mandarin T3, Standard Chinese T1 and Xi'an Mandarin T4 share identical pitch contours (for detailed mapping relation between Standard Chinese and Xi'an Mandarin

tones, see Liu et al., 2022). So, in the Homophone condition, the Standard Chinese targets and competitors are T4 and T3 monosyllabic words; the Xi'an Mandarin targets and competitors are T4 and T1 monosyllabic words. In the Translation condition, the Standard Chinese targets and competitors are T4 and T1 monosyllabic words; the Xi'an Mandarin targets and competitors are T4 and T3 monosyllabic words. As for the Segment condition, both the Standard Chinese and Xi'an Mandarin tasks include T3-T2, T4-T2, and T1-T2 monosyllabic word pairs. The stimulus pairs in all critical conditions share the same segmental syllables. Note that word pairs in the Segment conditions generally share more tonal similarity than that in the Homophone and Translation conditions. For instance, Standard Chinese word pairs of T2 (rising tone) and T3 (dipping tone), which were included in the Segment condition only, share more acoustic-phonetic similarity in their pitch contours and may elicit more lexical competition than the other tonal pairs (Shen et al., 2013; Qin et al., 2019). Therefore, we hypothesized that, if only one dialect is accessed during the task, we should find a relatively larger competition effect (indexed by fewer eye fixations towards the target and more eye fixations towards competitors) in the Segment condition than in the Homophone and Translation conditions. However, if both Standard Chinese and Xi'an Mandarin are activated, word pairs in the Homophone and Translation conditions would become homophones and should elicit a larger or similar competition effect than those in the Segment condition.

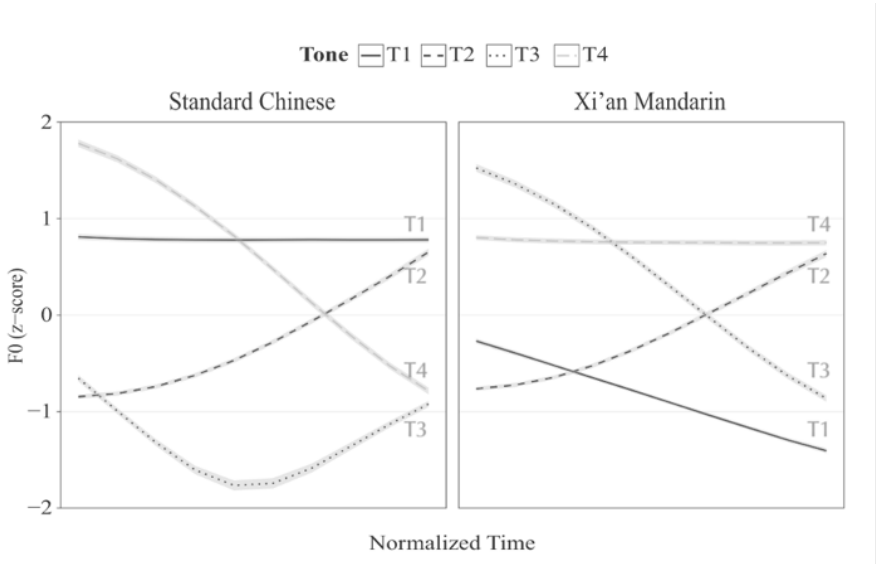

Figure 1. *Mean F0 (Z-score) contours of the four tones in Standard Chinese and Xi'an Mandarin. The grey areas indicate the 95% confidence interval of the corresponding mean. This figure is reprinted from Liu et al. (2020, p.2808).*

In both Standard Chinese and Xi'an Mandarin tasks, participants were asked to complete a practice block of four trials before performing the task. In each task, there were 72 critical trials (12 pairs of target & critical competitor × 3 critical conditions × 2 repetitions). In addition, there were 36 baseline trials, in which the competitors had no phonological or semantic overlap with the target (12 pairs of target & unrelated competitor × 3 critical conditions). The same number of filler trials were also added, in which the role of the target and critical/unrelated competitors was reversed. By doing so, participants' chances of hearing the target or competitor in the same display were kept equal. In this way, they were discouraged from developing strategic responses (following the practice of Malins & Joanisse, 2010). In total, each task included 216 trials (72 critical trials + 36 baseline trials + 108 filler trials), which were divided into four blocks of 54 trials. The order of the four blocks was counterbalanced. Participants were encouraged to take a short break between blocks.

3.3.3 Stimuli

The Standard Chinese stimuli consisted of 72 Standard Chinese monosyllabic words or morphemes (see Appendix B). The Homophone, Translation, and Segment conditions each have 12 pairs of target and competitor words. No item was used in more than one condition. Word frequency, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010), was balanced across target words and the three competitor conditions [$F(2, 69) = 0.432$, $p = 0.095$]. As Chinese characters were used as the visual display in the task, the number of components and strokes of the characters were also balanced across conditions [Strokes: $F(2, 69) = 0.044$, $p = 0.957$; Component: $F(2, 69) = 0.793$, $p = 0.457$). A group of 20 Xi'an Mandarin-Standard Chinese bi-dialectals, who did not participate in the eye-tracking experiment, judged the familiarity of the words on a scale from 1 to 10 (M = 7.094; SE = 0.517). The familiarity score of each condition was balanced [$F(2, 69) = 0.129$, $p = 0.88$].

The Xi'an Mandarin stimuli also consisted of 72 monosyllabic words or morphemes (see Appendix B). Homophone, Translation, and Segment conditions each have 12 pairs of words which all overlap in segments and differ in lexical tone. Word frequency, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010), was balanced across target words and the three competitor conditions [$F(2, 69) = 0.215$, $p = 0.807$]. The number of components and strokes of the characters was also controlled across conditions [Strokes: $F(2, 69) = 1.339$, $p = 0.269$; Component: $F(2, 69) = 0.231$, $p = 0.795$]. The same group of Xi'an Mandarin-Standard Chinese bi-dialectals who judged the word familiarity of the Standard Chinese stimuli also judged the Xi'an Mandarin stimuli on a scale from 1 to 10 (M = 7.078; SE = 0.564). The familiarity of each condition was also balanced [$F(2, 69) = 0.325$, $p = 0.724$].

All auditory stimuli were recorded in 2019 through a Sennheiser MKH416T microphone (44.1 kHz, 16 bit) and a Scarlett 2i2 sound card at a

sound-proof booth of Shaanxi Normal University. The Standard Chinese stimuli were produced by a male native speaker (age 22) of Standard Chinese who was born and grew up in Beijing. The Xi'an Mandarin stimuli were produced by a male native speaker of Xi'an Mandarin who was born and grew up in the city of Xi'an (age 20). Each word was read four times in isolation using a randomized list. One token of each word was chosen based on its clarity. The Standard Chinese and Xi'an Mandarin carrier phrase *wo3 yao4 shuo1…* "I will say…" were also recorded by the same respective speakers. The carrier phrase is sufficient for listeners to disambiguate which dialect is being spoken based on the tonal features of the first syllable. Using the software Praat (Boersma & Weenink, 2020), the Standard Chinese and Xi'an Mandarin carrier phrase were normalized to have the same duration of 1,000 ms and the same intensity of 70dB; the target stimuli were also normalized for intensity at 70dB; the normalized carrier phrase was then concatenated with each target word. No listener questioned the naturalness of the stimuli.

3.3.4 Procedure

Participants were tested in a sound-attenuated booth at the Psychology Lab of Shaanxi Normal University. While performing the task, participants' eye movements were recorded with an SR EyeLink Portable DUO eye-tracker at a sampling rate of 500Hz. For visual stimuli display, a 24-inch DELL U2412M monitor was located at a distance of about 52cm from the participant's eyes which were fixed with the help of a chin rest. The auditory stimuli were played over a Beyer DT-770 Pro dynamic headphone at a constant and comfortable hearing level.

Before the test, participants' eye gaze position was validated and calibrated with a 9-point grid. At the beginning of each trial, a central cross appeared on the screen for 500 ms. Participants were asked to look directly at the fixation for a drift check. After the central cross, four Standard Chinese characters

appeared on the screen. Meanwhile, the carrier phrase (which is 1,000 ms in duration) and the target were played. Participants were required to click on the corresponding character with a mouse. The next trial appeared 1,000 ms after the click or 2,000 ms post stimuli onset.

3.3.5 Data Analysis

3.3.5.1 Analysis of Behaviour Data

Reaction time and response accuracy for mouse clicks were collected for statistical analysis. Reaction times (hereafter RT) were calculated with respect to the onset of the auditory word. Trials for which the reaction time is shorter than 250 ms were excluded for both accuracy and RT analyses. Furthermore, only correct responses were considered for RT analyses. RT was analysed using the generalized linear mixed-effects model (GLMM) to account for the skewed distribution without the need to transform raw data (Lo & Andrews, 2015). A backward algorithm was used to select the model (Barr et al., 2013). RTs of the Standard Chinese and Xi'an Mandarin tasks were modelled separately. A maximum model including fixed effects of experimental conditions (i.e., Homophone, Translation, Segment conditions and the baseline), by-subject and by-item random intercepts, as well as by-subject and by-item random slopes for experimental conditions was constructed first. If a model failed to converge, we first increased the number of iterations, then simplified the model by removing correlation parameters and main effects in the random structure (Brauer & Curtin, 2018). Fixed effects and the random structure were tested by comparing the likelihood ratio test with a simpler model. All the analyses were run in the R software (R Core Team, 2020) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015).

3.3.5.2 Analysis of Eye-Tracking Data

We excluded trials for which the target was not correctly identified as well as trials for which the reaction time was shorter than 250 ms. Given the well-

recognized 200 ms delay for programming a saccade, the time window of 200-1,000 ms post auditory stimulus onset was chosen as our interest period. As the gaze position and duration of participants' eye fixation were recorded, looks toward target, competitor, and distractors during the interest period were collected. The collected eye-tracking data were first resampled to 50Hz. Then, the proportions of looks to target, competitor, and distractors at each time point were calculated by dividing the number of fixations toward each picture type by the sum of fixations on the four Chinese characters (target, competitor, and two distractors). Eye-movement data of the Standard Chinese and Xi'an Mandarin tasks were analysed separately.

Growth curve analysis (Mirman, 2014), a type of curvilinear regression, was used to model non-linear changes in the proportions of participants' eye fixations over time. This method has been widely accepted to analyse eye-tracking data of the visual world paradigm (e.g., Malins & Joanisse, 2010; Wang et al., 2017; Ito et al., 2018; Qin et al., 2019; Shook & Marian, 2019). With growth curve analysis, the orthogonal polynomials can capture subtle differences in the slope and curvature of the fixation lines: the linear term reflects the overall angle of a curve; the quadratic term reflects the shape of (i.e., the rise and fall) a curve with a single inflection point; and the cubic and quartic terms reflect the steepness of a curve with two or three inflection points (see Mirman et al., 2008; Mirman, 2017 for a detailed explanation regarding the significance of the polynomial terms in modelling the visual world paradigm data). Growth curve analysis is particularly useful for capturing temporal dynamics in eye-tracking data collected over time. It reveals how eye movements evolve over time and detects trends or patterns in the data. There are other ways analysing eye-tracking data such as generalized additive mixed-effect modelling (Wood, 2017) and divergent point analysis (Stone et al., 2021). Generalized additive mixed-effect modelling can handle multiple continuous and categorical predictors and detect specific intervals of difference in the general trajectory of eye-tracking data. Divergent point analysis is useful for examining attentional shifts or transitions in eye-tracking data,

allowing the identification of specific points in time when attention diverges or converges. The effectiveness of different statistical methods depends on the specific research question and data characteristics. In our research, we aimed to compare the general trends of eye movement changes between conditions rather than exploring specific intervals or time points of difference. Therefore, growth curve analysis was chosen over generalized additive mixed-effect modelling and divergent point analysis.

In this study, all analyses were carried out in the R software (R Core Team, 2020) using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015). Proportions of eye fixations to targets and competitors across experimental conditions were analysed using a fourth-order (quartic) orthogonal polynomial. Fixed effects of the experimental condition (i.e., Homophone, Translation, Segment conditions and the baseline) were tested on all time terms. The experimental condition was dummy-coded with the baseline condition as the reference level, so that the effect of each critical condition was tested relative to the unrelated baseline. Pairwise comparison between each critical condition was tested with a contrast matrix using the *multcomp* package (Hothorn et al., 2022). All analyses included participant as the random intercept and the orthogonal time polynomials as random slopes for the participant. The random intercept of items and random slopes of items were not included because the models with them did not converge. Each parameter's effect on the model fit was evaluated using model comparisons indexed by -2 times the change in log-likelihood distributed as $\chi 2$.

3.4 Results

3.4.1 Results of Standard Chinese Spoken Word Recognition

3.4.1.1 Behavioural Data

Reaction time and response accuracy for mouse click are shown in Table 2. For reaction time, the maximal likelihood estimation of the maximal model and the simplified random slope models failed to reach convergence. The final model

includes fixed effects of experimental conditions (i.e., Homophone, Translation, Segment conditions and the baseline), by-subject random intercepts, by-subject random slope for experimental conditions, and by-item random intercepts. The fixed effects of experimental conditions [$\chi 2$ (3) = 24.522, $p < 0.001$] suggested that participants' reaction time differed across conditions. Post-hoc analysis revealed that the reaction time of all critical conditions was significantly longer than the baseline condition (Homophone: $p < 0.001$; Translation: $p < 0.001$; Segment: $p < 0.001$), but there was no significant difference among the critical conditions (Homophone vs. Translation, $p = 0.270$; Homophone vs. Segment, $p = 0.670$; Translation vs. Segment, $p = 0.542$). This suggests that, while all competitors in the Homophone, Translation, and Segment conditions delayed the recognition of the target words in comparison to the baseline condition, the effect size across the three conditions was not significantly different. The error rate was low in each condition (all approximately under 1.5 %), thus no further analyses were conducted on the response accuracy.

Table 2. *Mean Reaction time (ms) and mean percent response accuracy in Standard Chinese. Standard deviations are in parentheses.*

Condition	Reaction Time (SD)	Percent Accuracy (SD)
Baseline	1100 (319)	99.9 (3.07)
Homophone Condition	1257 (613)	98.4 (12.7)
Translation Condition	1172 (319)	99.3 (8.34)
Segment Condition	1222 (388)	99.0 (10.1)

3.4.1.2 Eye Movement Data

Looks to target

Average fixations toward targets of each experiment condition are presented in *Figure 2* (a). As we can see, looks to the targets in the Homophone, Translation and Segment conditions all have overall fewer target fixations than the baseline condition over the interested time window. The Segment condition

has the least target fixations around 400-600 ms post stimuli onset. According to the estimated parameters of the growth curve analysis (as shown in Table 3), the time course of the target fixations in the Homophone (Intercept term: $p < 0.001$), Translation (Intercept term: $p < 0.001$; Linear term: $p < 0.05$; Quadratic term: $p < 0.01$; Quartic term: $p < 0.05$) and Segment (Intercept term: $p < 0.001$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.05$; Quartic term: $p < 0.01$) conditions were all significantly different from the baseline condition. Moreover, the target fixations in the Homophone and Segment conditions were significantly different from each other (Intercept term: $p < 0.001$; Quadratic term: $p < 0.01$; Quartic term: $p < 0.01$), so did Translation and Segment conditions (Intercept term: $p < 0.001$; Linear term: $p < 0.05$; Cubic term: $p < 0.05$). These results suggest that target fixations were distracted more in the Segment condition than that in the Homophone and Translation conditions.

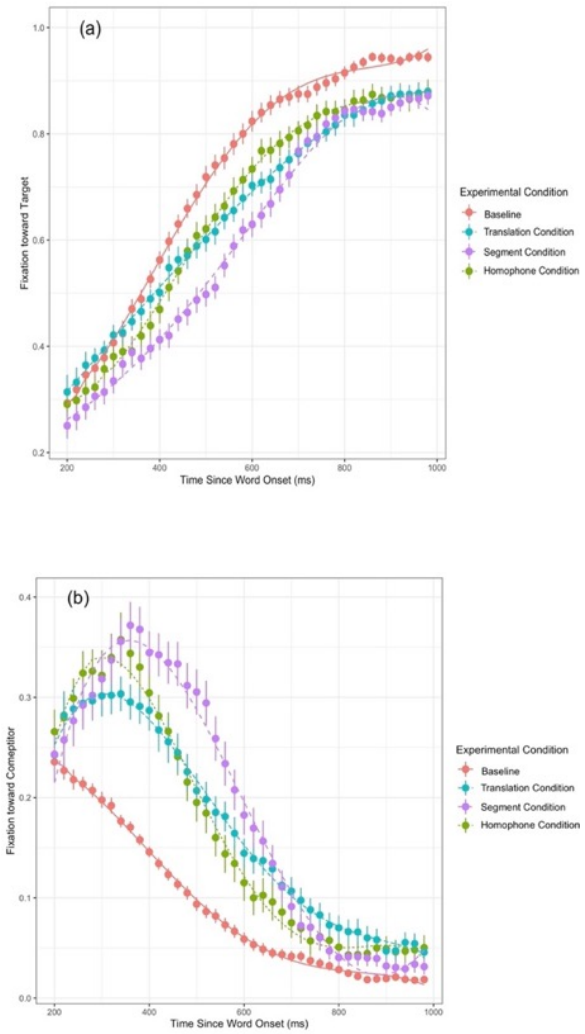

Figure 2. *Time course of eye fixations toward the target (a) and competitors (b) of each experimental condition plotted against baseline in the Standard Chinese task. The points with range represent mean proportions of fixations across participants and items with standard error. The lines represent the growth curve analysis model fits. Note that to make the different patterns of target and competitor fixations clearer, the scales of the y-axis in plot (a) and plot (b) are different.*

Looks to competitors

Average fixations toward competitors of each experiment condition are presented in *Figure 2 (b)*. As we can see, in the Homophone, Translation and Segment conditions, there were more competitor eye fixations than in the baseline condition. According to estimated parameters of the growth curve analysis (as shown in Table 4), the time course of the target fixations in the Homophone condition (Intercept term: $p < 0.001$; Linear term: $p < 0.05$; Cubic term: $p < 0.001$; Quartic term: $p < 0.05$), the Translation condition (Intercept term: $p < 0.001$; Linear term: $p < 0.05$; Quadratic term: $p < 0.05$; Cubic term: $p < 0.01$) and the Segment condition (Intercept term: $p < 0.001$; Linear term: $p < 0.001$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.001$) all significantly differed from the baseline condition. Moreover, among the three conditions, the Segment condition has the most competitor fixations around 400-600 ms post stimulus onset. According to estimated parameters of the growth curve analysis (as shown in Table 4), the Homophone and Segment conditions were significantly different (Intercept term: $p < 0.05$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.05$; Quartic term: $p < 0.05$), so did the Translation and Segment conditions (Linear term: $p < 0.05$; Quadratic term: $p < 0.01$; Cubic term: $p < 0.001$). This suggests that the competitors in the Homophone and Translation conditions were less disruptive than that in the Segment condition.

Table 3. *Growth curve analysis of looks to target in the Standard Chinese task.*

	Parameter estimates							
	Homophone: Baseline				Homophone: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	-0.068	0.014	-4.974	<0.001	0.054	0.014	3.977	<0.001
Linear	-0.046	0.085	-0.540	0.589	-0.080	0.084	-0.950	0.342
Quadratic	0.085	0.066	1.275	0.202	-0.175	0.065	-2.682	0.007
Cubic	-0.064	0.051	-1.253	0.210	0.060	0.049	1.210	0.226
Quartic	-0.012	0.039	-0.308	0.758	0.107	0.038	2.840	0.005
	Translation: Baseline				Translation: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	-0.071	0.014	-5.195	<0.001	0.051	0.014	3.753	<0.001
Linear	-0.172	0.085	-2.023	0.043	-0.206	0.084	-2.446	0.014
Quadratic	0.214	0.066	3.232	0.001	-0.045	0.065	-0.696	0.487
Cubic	0.002	0.051	0.049	0.961	0.126	0.049	2.546	0.011
Quartic	-0.092	0.039	-2.326	0.020	0.028	0.038	0.733	0.464
	Segment: Baseline				Translation: Homophone			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	-0.123	0.014	-8.992	<0.001	0.003	0.014	0.222	0.825
Linear	0.034	0.084	0.405	0.685	0.126	0.085	1.483	0.138
Quadratic	0.260	0.065	3.974	<0.001	-0.130	0.066	-1.957	0.050
Cubic	-0.123	0.050	-2.494	0.013	-0.066	0.051	-1.303	0.193
Quartic	-0.120	0.038	-3.157	0.002	0.080	0.039	2.021	0.043

Table 4. *Growth curve analysis of looks to competitors in the Standard Chinese task.*

	Parameter estimates							
	Homophone: Baseline				Homophone: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	0.074	0.011	6.850	<0.001	-0.026	0.011	-2.410	0.016
Linear	-0.239	0.074	-3.243	0.001	0.065	0.073	0.898	0.369
Quadratic	-0.045	0.051	-0.886	0.376	0.212	0.050	4.240	<0.001
Cubic	0.183	0.043	4.258	<0.001	-0.105	0.042	-2.512	0.012
Quartic	-0.093	0.037	-2.491	0.013	-0.087	0.036	-2.415	0.016
	Translation: Baseline				Translation: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	0.079	0.011	7.292	<0.001	-0.021	0.011	-1.963	0.050
Linear	-0.149	0.074	-2.021	0.043	0.155	0.073	2.131	0.033
Quadratic	-0.124	0.051	-2.435	0.015	0.133	0.050	2.661	0.008
Cubic	0.124	0.043	2.882	0.004	-0.164	0.042	-3.926	<0.001
Quartic	-0.031	0.037	-0.828	0.408	-0.025	0.036	-0.689	0.491
	Segment: Baseline				Translation: Homophone			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	0.099	0.011	9.333	<0.001	-0.005	0.011	-0.442	0.658
Linear	-0.304	0.073	-4.170	<0.001	-0.090	0.074	-1.222	0.222
Quadratic	-0.257	0.050	-5.140	<0.001	0.079	0.051	1.550	0.121
Cubic	0.288	0.042	6.884	<0.001	0.059	0.043	1.377	0.169
Quartic	-0.006	0.036	-0.171	0.864	-0.062	0.037	-1.665	0.096

3.4.1.3 Preliminary Discussion

While the reaction time data indicated no difference between the cross-dialect conditions (the Homophone and Translation conditions) and the within-dialect condition (the Segment condition), the analysis of eye fixations on targets and competitors consistently showed that the competitors of Homophone and Translation conditions introduced smaller interference effects than that of the

Segment condition. This suggests that when listening to Standard Chinese, Standard Chinese and Xi'an Mandarin bi-dialectal participants did not experience competition or interference from cross-dialect homophones and translation equivalents of Xi'an Mandarin.

3.4.2 Results of Xi'an Mandarin Spoken Word Recognition

3.4.2.1 Behavioural Data

Reaction time and response accuracy for mouse click are shown in Table 5. For reaction time, the maximum likelihood estimation of the maximum model and the random slope models failed to reach convergence. The final model included fixed effects of experimental conditions, by-subject random intercepts, by-subject random slope for experimental conditions, and by-item random intercepts. The fixed effects of experimental conditions ($\chi 2$ (3) = 20.429, $p < 0.001$) suggested that participants' reaction time differed across conditions. Post-hoc analysis revealed that the reaction time of all critical conditions was significantly different from that of the baseline condition (Homophone: $p < 0.001$; Translation: $p < 0.001$; Segment: $p < 0.001$) but showed no significant difference from each other (Homophone vs. Translation, $p = 0.843$; Homophone vs. Segment, $p = 0.843$; Translation vs. Segment, $p = 0.843$). The error rate was low in each condition (all approximately under 1.5 %), thus no further analyses were conducted on the response accuracy. These results suggest that while all competitors in the Homophone, Translation, and Segment conditions delayed the recognition of the Xi'an Mandarin target words more than in the baseline condition, the size of the interference effect across the three critical conditions was not statistically significantly different.

Table 5. *Mean Reaction time (ms) and mean percent response accuracy in Xi'an Mandarin. Standard deviations are in parentheses.*

Condition	Reaction Time (SD)	Percent Accuracy (SD)
Baseline	1165 (418)	99.9 (2.77)
Homophone Condition	1291 (409)	98.0 (14.0)
Translation Condition	1233 (366)	99.2 (8.78)
Segment Condition	1307 (414)	99.1 (9.52)

3.4.2.2 Eye Movement Data

Looks to target

Average fixations toward targets of each experiment condition are presented in *Figure 3 (a)*. As we can see, there are fewer eye fixations towards targets in the Homophone, Translation and Segment conditions than in the baseline condition over the interested time window. Among these, the Segment condition has the least target fixation around 400-700 ms post stimuli onset. This pattern was also confirmed by the estimated parameters of the growth curve analysis (as shown in Table 6), the time course of the target fixations in the Homophone condition (Intercept term: $p < 0.001$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.05$; Quartic term: $p < 0.001$), the Translation condition (Intercept term: $p < 0.001$; Quadratic term: $p < 0.001$) and the Segment condition (Intercept term: $p < 0.001$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.05$; Quartic term: $p < 0.001$) were all significantly different from the baseline condition. Moreover, both the Homophone and Translation conditions were significantly different from the Segment condition (Homophone: Quadratic term: $p < 0.05$; Translation: Intercept term: $p < 0.05$; Quadratic term: $p < 0.01$; Cubic term: $p < 0.01$). These results indicate that the target fixations in the Homophone, Translation and Segment conditions were all significantly less than that of the baseline condition. Furthermore, the Homophone and Translation conditions exhibited smaller interference effects than the Segment condition.

Looks to competitors

Average fixations toward competitors of each experiment condition are presented in *Figure 3 (b)*. As we can see, there are more competitor fixations in the Homophone, Translation and Segment conditions than the baseline condition over the interested time window. Among them, the Segment condition has the most competitor fixations around 250-799 ms post stimuli onset. According to estimated parameters of the growth curve analysis (as shown in Table 7), the time course of the target fixations in the Homophone (Intercept term: $p < 0.001$; Linear term: $p < 0.05$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.01$), Translation (Intercept term: $p < 0.001$; Linear term: $p < 0.01$; Cubic term: $p < 0.01$) and Segment (Intercept term: $p < 0.001$; Linear term: $p < 0.001$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.001$) conditions were all significantly different from the baseline condition. Moreover, competitor fixations in the Homophone and Segment conditions were significantly different from each other (Intercept term: $p < 0.05$; Linear term: $p < 0.05$; Quadratic term: $p < 0.01$; Cubic term: $p < 0.01$), so did Translation and Segment conditions (Intercept term: $p < 0.001$; Linear term: $p < 0.05$; Quadratic term: $p < 0.001$; Cubic term: $p < 0.01$; Quartic term: $p < 0.05$). These findings suggest that competitors in the Homophone and Translation conditions were less disruptive than that of the Segment conditions in recognizing Xi'an Mandarin target words.

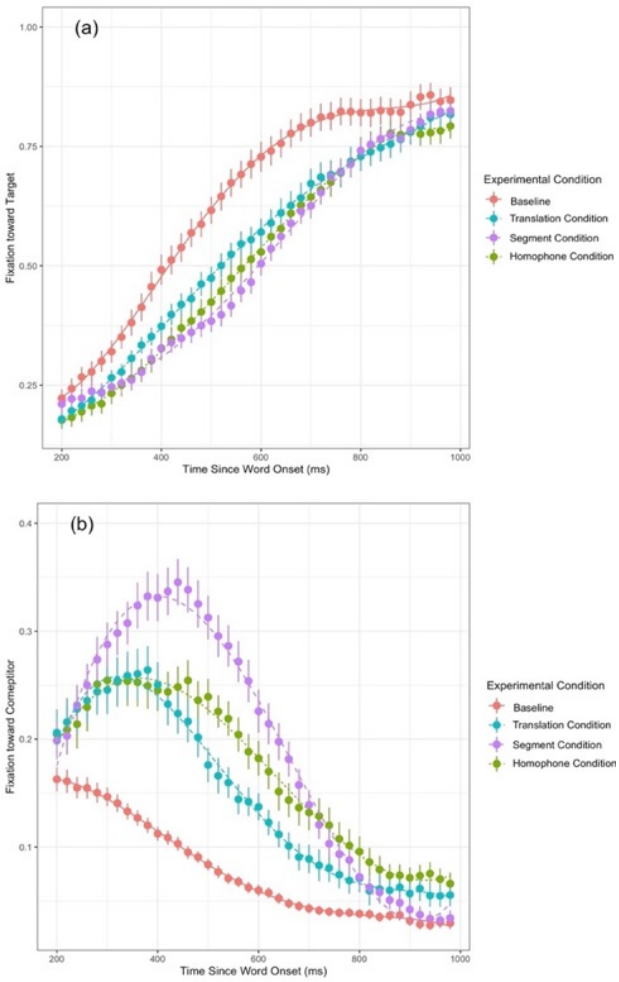

Figure 3. *Time course of eye fixations toward the target (a) and competitors (b) of each experimental condition plotted against baseline in the Xi'an Mandarin task. The points with range represent mean proportions of target fixations across participants and items with standard error. The lines represent the growth curve analysis model fits. Note that to make the different patterns of target and competitor fixations clearer, the scales of the y-axis in plot (a) and plot (b) are different.*

Table 6. *Growth curve analysis of looks to targets in the Xi'an Mandarin task.*

	Parameter estimates							
	Homophone: Baseline				Homophone: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	-0.125	0.014	-8.628	<0.001	0.006	0.014	0.416	0.678
Linear	0.074	0.059	1.267	0.205	-0.005	0.058	-0.084	0.933
Quadratic	0.299	0.058	5.133	<0.001	-0.143	0.057	-2.490	0.013
Cubic	-0.086	0.042	-2.065	0.039	0.018	0.041	0.455	0.649
Quartic	-0.098	0.034	-2.839	0.005	0.019	0.033	0.587	0.557
	Translation: Baseline				Translation: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	-0.099	0.014	-6.878	<0.001	0.031	0.014	2.177	0.029
Linear	0.015	0.059	0.261	0.794	-0.064	0.058	-1.105	0.269
Quadratic	0.244	0.058	4.194	<0.001	-0.197	0.057	-3.443	0.001
Cubic	0.010	0.042	0.249	0.803	0.115	0.041	2.838	0.005
Quartic	-0.057	0.034	-1.656	0.098	0.060	0.033	1.824	0.068
	Segment: Baseline				Translation: Homophone			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	-0.131	0.014	-9.097	<0.001	-0.025	0.014	-1.751	0.080
Linear	0.079	0.058	1.370	0.171	0.059	0.059	1.006	0.314
Quadratic	0.442	0.057	7.701	<0.001	0.055	0.058	0.939	0.348
Cubic	-0.105	0.041	-2.582	0.010	-0.097	0.042	-2.314	0.021
Quartic	-0.117	0.033	-3.557	<0.001	-0.041	0.034	-1.183	0.237

3.4.2.3 Preliminary Discussion

Similar patterns were found in the Xi'an Mandarin and the Standard Chinese experiments. Participants' reaction times in the Homophone, Translation and Segment conditions were delayed to the same extent compared to that in the baseline condition. Looks towards targets and competitors showed that the competitors in the Homophone, Translation and Segment conditions all significantly distracted participants' visual attention from targets. Among these,

Segment competitors distracted participants' looks the most. Overall, like in the Standard Chinese task, the reaction time, target and competitor fixations consistently demonstrated within-dialect interference but not cross-dialect interference. This suggests that when listening to Xi'an Mandarin only, it is unlikely that participants have accessed cross-dialect homophones and translation equivalents of Standard Chinese.

Table 7. *Growth curve analysis of looks to competitors in the Xi'an Mandarin task.*

	Parameter estimates							
	Homophone: Baseline				Homophone: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	0.092	0.009	9.752	<0.001	-0.023	0.009	-2.459	0.014
Linear	-0.132	0.054	-2.432	0.015	0.167	0.054	3.106	0.002
Quadratic	-0.177	0.041	-4.357	<0.001	0.161	0.04	4.059	<0.001
Cubic	0.117	0.038	3.071	0.002	-0.121	0.037	-3.285	0.001
Quartic	0.016	0.031	0.498	0.618	-0.015	0.03	-0.489	0.625
	Translation: Baseline				Translation: Segment			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	0.069	0.009	7.294	<0.001	-0.046	0.009	-4.944	<0.001
Linear	-0.181	0.054	-3.325	0.001	0.118	0.054	2.201	0.028
Quadratic	-0.075	0.041	-1.833	0.067	0.264	0.04	6.645	<0.001
Cubic	0.128	0.038	3.364	0.001	-0.11	0.037	-2.984	0.003
Quartic	-0.041	0.031	-1.303	0.193	-0.071	0.03	-2.366	0.018
	Segment: Baseline				Translation: Homophone			
	Est.	*SE*	*t*	*p*	*Est.*	*SE*	*t*	*p*
Intercept	0.115	0.009	12.318	<0.001	-0.025	0.014	-1.751	0.08
Linear	-0.299	0.054	-5.57	<0.001	0.059	0.059	1.006	0.314
Quadratic	-0.338	0.04	-8.523	<0.001	0.055	0.058	0.939	0.348
Cubic	0.238	0.037	6.442	<0.001	-0.097	0.042	-2.314	0.021
Quartic	0.03	0.03	1.008	0.313	-0.041	0.034	-1.183	0.237

3.5 General Discussion

To investigate whether bi-dialectal listeners co-activate both their dialects during spoken word recognition, we examined spoken word recognition in Standard Chinese-Xi'an Mandarin bi-dialectal listeners. Using the eye-tracking technique and the visual world paradigm, Standard Chinese-Xi'an Mandarin bi-

dialectal listeners were instructed to identify the target word they heard among a display of Chinese characters, which includes the target, a phonological competitor, and two unrelated distractors. All competitors share segments with the target within- and cross-dialect. Moreover, we manipulated three target-competitor conditions. In the cross-dialect homophone condition (i.e., the Homophone Condition), the target and competitor also share lexical tone across two dialects; in the translation-induced homophone condition (i.e., the Translation Condition), the translation equivalents of the target and competitor also share lexical tone across two dialects. The hypothesis is that, if lexical representations of both dialects are co-activated, Homophone and Translation competitors as cross-dialect (translation) homophones should yield a larger interference effect than the Segment competitors, which overlap with the target only in segments within the dialect (as well as across dialect). Analysis of eye fixations showed that, regardless of whether participants were listening to the target words in Standard Chinese or Xi'an Mandarin, there was larger competition (indexed by how much eye fixations towards targets are distracted by the competitors) in the Segment condition than the Homophone and Translation conditions. Overall, these findings suggest that, during spoken word recognition, bi-dialectal listeners do not experience similar interference effects from the other dialect as bilinguals do with the other language.

The lack of cross-dialect interaction seems to lend support to the co-dependent view of bi-dialecticism (Labov, 1998), which holds that bi-dialectals do not maintain two independent systems as bilinguals. However, before jumping to the conclusion, we should also take Liu (2018)'s findings into account. In Liu (2018), cross-dialect homophone primes (comparable to the Homophone competitors in our study) were found to introduce significant inhibition while within-dialect segmentally overlapping primes (comparable to the Segment condition in our study) did not, showing clear evidence for cross-dialect interference. There are two major design differences between the present study and Liu (2018). The first is the presence of sentence context. In Liu (2018), the

bi-dialectals listened to isolated words in a mixed-dialect setting, whereas in the current study, participants listened to words embedded in a short mono-dialect sentence (e.g., “I will say…”). Listening to the target words in a mono-dialect context (which was unambiguously clear due to the embedding sentence) might have constrained dialect co-activation and reduced any interference effect in the current study. Second, bi-dialectals were aware of the bi-dialectal nature of the task from the very beginning of our experiment. This might have influenced their processing mode as suggested by the findings in Wu et al. (2018).

Wu and her colleagues reported evidence from an auditory lexical decision task that bi-dialectals may inhibit cross-dialect interference as soon as they come across a bi-dialectal situation. Specifically, they found that when bi-dialectals of Standard Chinese and Jinan Mandarin were not aware that they would be tested in both dialects, cross-dialect tonal similarity significantly modulated the reaction time of recognizing Standard Chinese or Jinan Mandarin words. However, as soon as the participants became aware of the bi-dialectal situation (i.e., after switching the dialect in test), the effect was largely reduced, suggesting proactive inhibition of cross-dialect lexical competition. At the very beginning of our experiment, Standard Chinese-Xi’an Mandarin bi-dialectals were informed that they would perform two tasks, one in Standard Chinese, and the other in Xi’an Mandarin. Understanding that they were in a bi-dialectal situation at the beginning might have led Standard Chinese-Xi’an Mandarin bi-dialectals to attentionally control and inhibit lexical interference from the other dialect. Consequently, the cross-dialect interference (as shown in Liu, 2018) is likely to have been annulled by the sentence context and the awareness of the bi-dialectal context in our study.

If our interpretation of the existing results is on the right track, bi-dialectal and bilingual lexical access are then different. Previous studies have repeatedly shown that cross-language lexical competition cannot be eliminated even by a high semantic constraining sentence (e.g., Lagrou et al., 2013a, 2013b), let alone a short preceding sentence with no semantic constraints and stays

invariant during the task (e.g., "I will say…"). Moreover, according to the language mode theory (Grosjean, 1998, 2001), when bilinguals are using two of their languages (e.g., aware of the bilingual nature of the task), they are more likely to be in a "bilingual mode" and activate elements from both languages. However, in the findings of our and Wu et al. (2018)'s studies, the awareness of the bi-dialectal situation seems to only help bi-dialectals achieve more effective dialect selectivity. Simply put, even with a monolingual sentence context and/or adjacent language blocks, there were robust cross-language lexical competition effects in bilingual lexical access (e.g., Spivey & Marian, 1999; Wang et al., 2017), whereas in this bi-dialectal study, no trace of dialectal interference effect was found despite the very similar experimental set-up as the bilingual studies.

Given that the target and cross-dialect competitors (Homophone and Translation competitors) are only identical across two dialects when taking the overlapping lexical tones into account, one may speculate that the reason why cross-dialect competitors are not more disruptive than within-dialect competitors is that the role of lexical tone in constraining lexical access, compared to segments, is negligible. However, this is unlikely to be the case. First, the most recent study we are aware of that has argued for a lower priority of tone, compared to consonants and vowels, in Mandarin lexical access is Wiener and Turnbull (2016). The study, however, used a word reconstruction task and tested the tonal mutability in constraining lexical selection, which involves a very different process of lexical access from the task used in our study. A number of studies, with more comparable tasks as our study, have already shown that lexical tone plays a significant role in native monolingual tone word recognition (e.g., Schirmer et al., 2005; Malins & Joanisse, 2010, 2012; Yang & Chen, 2022). Moreover, as discussed earlier, lexical tone has been found to be critical in bilingual/bi-dialectal lexical access with English-Mandarin bilinguals (Wang et al., 2017) and Mandarin bi-dialectals (Liu, 2018; Wu, 2018).

We propose that our results lie with a different dialect control mechanism from bilingual processing. As bilingual lexicon is generally believed to "be

integrated across languages and is accessed in a non-selective way" (Dijkstra & Heuven, 2002, p.182), bilingual language comprehension models (e.g., the BIA model; BIA+ model, Dijkstra & Heuven, 2002; the BLNCS, Shook & Marian, 2013) have proposed various control mechanisms (e.g., language node; task scheme) to inhibit the non-target language and avoid catastrophic cross-language interferences. It is possible that bi-dialectals, who also switch and mix dialects often in their daily conversation, need control mechanisms to avoid cross-dialect intrusion as well. Given that languages differ considerably at all levels (e.g., syntax, lexicon, orthography, phonology, and phonetics) while dialects are generally more similar (e.g., sharing extremely similar syntax and lexicon, one writing system, and largely overlapping segmental and tonal inventories), bi-dialectals might need and have developed a stronger or more efficient control strategy, compared to bilinguals. Furthermore, bi-dialectals may be more sensitive to factors such as sentence context and tasks, and they could make better use of proactive control to suppress the intrusion of the other dialect from the beginning of a sentence or a task.

Given that the current views of bi-dialectalism (i.e., the independent and co-dependent view) are oversimplified to explain our results, we hereby propose a bi-dialectal spoken word recognition model (see *Figure 4*), drawing inspiration from bilingual comprehension and recognition models such as BLINCS (Shook & Marian, 2013), BIA (e.g., Grainger & Dijkstra, 1992; Dijkstra et al., 1998), and BIA+ (Dijkstra & Heuven, 2002). Similar to BLINCS, our bi-dialectal model has multiple levels of lexical representations: phonological, phono-lexical, ortho-lexical, and semantic representations. Between levels, the representations interact bidirectionally. Within levels, dialect-specific and dialect-shared representations are stored in the same space, allowing communication and competition between dialects. Moreover, this bi-dialectal model has an additional task scheme, following BIA+. Note that the task scheme in BIA+ cannot modulate word activation and only makes adaptations to the decision criteria. In this bi-dialectal model, the task scheme functions more like the language node level in the BIA

model, in which the entire lexicon can be suppressed top-down. We further conjecture that the differences between bi-dialectal and bi-lingual lexical access may be a continuum in terms of their co-activation in part due to the degree of similarities between the two linguistic systems (be they dialects or languages). We would like to emphasise that this bi-dialectal model is only a preliminary attempt to account for current findings of bi-dialectal lexical access in comparison with what has been documented in the bilingual literature. Given that the finding of this study was based on one experiment with a null result of the homophone and translation equivalent effects, future studies are needed to test the hypothesis of different language vs. dialect control strategies further and to validate and develop the model.

Besides the proposed model, the null result of this study may also be explained by an alternative account of bilingual lexical access without appealing to parallel activation (e.g., Costa et al., 2017; Hartsuiker, personal communication, June 1, 2023). According to Costa et al. (2017), bilinguals carry over the structure of their native language to the non-native language during learning. Consequently, the non-native lexicon would keep traces of the connections existing in the native lexicon. For instance, with Standard Chinese and English bilinguals, Standard Chinese words *huo3che1* and *huo3tui3* are strongly connected via the overlapping first syllable; their English translation equivalents “train” and “ham”, which are not related in English, are connected during the acquisition of English by Chinese learners. Based on this assumption, Costa et al. (2017) proposed that the cross-language translation effect observed in Standard Chinese and English bilinguals (Thierry & Wu, 2004) may not be due to parallel access of the Standard Chinese lexicon but rather via the “learned” connection within the English lexicon. Costa et al. (2017) further suggested that increasing proficiency in the second language may reduce activation of the non-target language. This is because as lexical activation increasingly restricts to one language, the “learned” connections weaken over time. As the bi-dialectal speakers in our study have excellent proficiency in both varieties, the “learned” connections in each dialect may have

been largely reduced over time, making any possible cross-dialectal effect difficult to observe. However, more evidence is still needed to further explore the no-activation view in bilingual speech processing, as well as its application to bi-dialectal speakers.

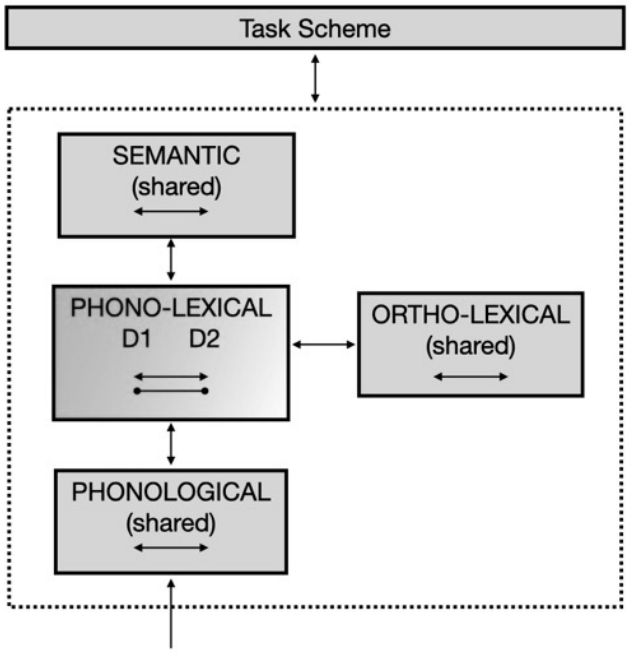

Figure 4. *The bi-dialectal spoken word recognition model. This model has phonological, phono-lexical, ortho-lexical, and semantic representations. Between levels, the representations interact bidirectionally. Within levels, dialect-specific and dialect-shared representations are stored in the same space, allowing for communication and competition between dialects. Outside the dialect system, the task scheme inserts proactive control on dialect activation based on task demands.*

To conclude, we did not find evidence that bi-dialectals experience cross-dialect interference when listening to one dialect only. Our finding marks a sharp contrast between bi-dialectal and bilingual spoken word comprehension. To

account for the lack of bi-dialectal co-activation, we proposed a preliminary bi-dialectal lexical access model, emphasizing the role of top-down control (as influenced by sentence context and task demand) in dialect interaction during processing. To further understand the extent to which bi-dialectals differ from bilinguals, as well as the locus of their differences, more work on bi-dialectal language processing, in comparison to bi-lingual language processing, is urgently needed.

Chapter 4

The Role of Lexical Tone in Bilingual Spoken Word Production

Abstract

During spoken word production, bilinguals not only retrieve the form of the target language but also that of the non-target language. However, most previous studies on bilingual word production have focused on segments. It remains open whether suprasegmental information is involved in the process as well. We aimed to address this gap by examining the role of lexical tone in English spoken word production with bilinguals of Standard Chinese (hereafter SC) and English. In four online picture-word interference (PWI) experiments, we asked SC-English bilingual speakers to name pictures in English (e.g., feather) while ignoring four types of simultaneously presented SC distractors: 1) a translation distractor, i.e., the translation equivalent of the English target name (e.g., *yu3mao2 "*feather"); 2) a tone-sharing distractor, which shares both tone and segments with the SC translation in the first syllable (e.g., *yu3zhou4* "universe"); 3) a no-tone-sharing distractor, which shares segments but not tone with the SC translation in the first syllable (e.g., *yu4mi3* "corn"); 4) an unrelated distractor, which shares no phonological overlap with the target or its translation (e.g., *lei4shui3* "tear"). We also manipulated two procedural factors, namely distractor modality (i.e., whether distractors were presented auditorily or visually) and familiarization mode (i.e., whether participants previewed the target picture names in English or in both English and SC). Not only did this study replicate the translation facilitation effect (e.g., Costa et al., 1999) but also observed significant differences between the effects of tone-

sharing and no-tone-sharing distractors. Moreover, the polarity and robustness of such effects are subject to the interaction of distractor modality and familiarization mode. Overall, our findings suggest that SC-English bilinguals co-activate the lexical tone of SC translations during English picture naming.

Keywords: lexical tone; spoken word production; language co-activation

One of the most important findings in the bilingual literature is that bilinguals co-activate both of their languages even when they are only using one. There is substantial evidence of co-activation not only for spoken word comprehension (e.g., Dijkstra & Van Heuven, 2002; Thierry & Wu, 2007; Kroll & De Groot, 2009) but also for spoken word production (see Costa, 2009 for a review). It is important to note that while the literature on spoken comprehension shows that not only segments but also suprasegmental features such as lexical tone are co-activated (e.g., Wang et al., 2017), studies on bilingual spoken word production have mainly focused on segments (e.g., Colomé, 2001; Costa & Caramazza, 1999; Hermans et al., 1998). The goal of this study is to fill in this knowledge gap and investigate whether suprasegmental information of the non-target language is also co-activated during the process of spoken word production. More specifically, we will address this question by examining the co-activation of lexical tone in English and Mandarin bilingual spoken word production within the picture-word interference paradigm (hereafter PWI; Rosinski et al., 1975).

PWI is one of the most widely used paradigms for examining the process of spoken word production. In this paradigm, participants are asked to name a picture while ignoring the presence of a written or spoken distractor word. It generally takes participants more time to name the target picture if the name of the target picture and the distractor word are semantically related (e.g., target *dog* – distractor *cat*), and less time if they are phonologically related (e.g., target *dog* – distractor *doll*), compared with an unrelated condition. These effects are known as the *semantic interference* effect and *phonological facilitation* effect, respectively. The rationale behind the effects is that the retrieval of the target picture's concept not only activates the target word (e.g., *dog*) but also words that are semantically related to the target (e.g., *cat*). Thus, a semantic distractor (e.g., *cat*) would receive activation from both the target picture and the distractor word. Compared with unrelated distractors (e.g., *table*), which only receive activation from the distractor word itself, the activation level of the semantic distractor is thus higher and interferes more with target selection (but see Mahon et al., 2007

for a different account). A phonological distractor (e.g., *doll*), on the other hand, facilitates picture naming because their shared phonological properties aid the retrieval of the targets' phonological form.

Interestingly, in a renowned bilingual study by Costa et al. (1999), *semantic interference* and *phonological facilitation* effects were also found across languages. When Spanish and Catalan bilinguals were asked to name pictures in Spanish, Catalan distractors elicited *semantic interference* and *phonological facilitation* just like Spanish distractors, suggesting that lexical activation is not limited to the target language. In that study, Costa and his colleagues also found that the Catalan translations of the targets significantly facilitated Spanish picture naming. This effect, commonly known as *translation facilitation* or between-language identity effect, has been taken as a strong indicator of bilingual language co-activation (e.g., Costa & Caramazza, 1999; Hermans, 2004). Moreover, a translation-mediated phonological effect has been found in bilingual PWI studies. For instance, in Hermans et al. (1998), Dutch-English bilinguals took longer to name pictures in their L2 English when the auditory Dutch distractor (e.g., *berm* "verge") is phonologically similar to the Dutch translation of the picture (e.g., *berg* "mountain") compared with unrelated distractors (e.g., *kaars* "candle"). Such an effect, known as the *phono-translation interference* effect, has been replicated with bilinguals of high proficiency in both their languages (Costa et al., 2003), bilinguals of typologically distant languages (French and Tunisian Arabic; Boukadi et al., 2015), and in bilinguals' native language (Klaus et al., 2018). It indicates that not only the lexical representations, but also the sub-lexical phonological representations of the translation equivalents are co-activated during speech production. The *phono-translation interference* effect also serves as evidence supporting cascaded theories of lexical access, which argues that phonological and lexical representations are linked through cascaded and interactive processing (see Schiller & Alario, 2023 for a review on cascaded models of speech production).

In addition to the *phono-translation interference* effect in PWI, evidence indicates that bilinguals co-activate the phonological representations of both languages in various speech planning and production tasks. For example, with a phoneme monitoring task, Colomé (2001) found that when Catalan-Spanish bilinguals were asked to detect certain phonemes in the Catalan name of a picture (e.g., *taula*, "table"), they took longer to reject phonemes (e.g., /m/) that are in the target pictures' Spanish translation (e.g., *mesa*) than those that are not (e.g., /f/). Similarly, Macizo (2015) asked Spanish-English bilinguals to name the colour of a pictured object in English (e.g., *brown suitcase*) and found that they took longer to respond when the names of the colour and picture were phonologically related in Spanish (e.g., *maleta marrón*, "brown suitcase") compared with unrelated counterparts (e.g., *maleta rosa*, "pink suitcase"). Consistently, evidence from bi-dialectal processing supports the same view: using an auditory lexical decision experiment, Wu et al., (2015) observed that cross-dialect tonal similarity between SC and Jinan Mandarin manipulates SC-Jinan Mandarin bi-dialectals' lexical processing of etymologically-related translation equivalents.

Despite the substantial evidence that bilinguals have access to the phonology of both languages during spoken word production (e.g., Hermans et al., 1998; Costa et al., 2000; Colomé, 2001; Roelofs, 2003; Macizo, 2016; Spalek et al., 2014; Klaus et al., 2018a), it is important to note that the existing studies mainly drew evidence from segmental information. Few studies have looked into the co-activation of suprasegmental information in bilingual word production. The only study to our knowledge is Martínez García (2018), which compared the co-activation of stress-sharing and no-stress-sharing cognates in English-Spanish bilinguals. In this study, participants were asked to name a printed Spanish word (e.g., *materia* "subject") while ignoring an English cognate competitor that was also in the display (e.g., *material* "material"). Critically, the competitor either shares the same stress pattern with the target (e.g., target *maTEria* "subject" – competitor *maTErial*) or not (e.g., target *liTEra* "bunk bed" – competitor *LIteral*). By comparing the naming latencies, Martínez García (2018) found that bilinguals

took less time to name Spanish targets with stress-sharing cognate competitors than with no-stress-sharing cognate competitors. Martínez García (2018, p. 20) interpreted this finding as evidence that "English stress modulates cross-language activation during bilingual spoken word production". It is important to note that, in this study, the bilinguals' two languages, namely English and Spanish, both have lexical stress, and only near-identical cognates were examined. Given that cognates might share "one single memory token" in bilinguals' minds, the observed stress effect may not be due to cross-language co-activation but result from the shared stress representation (or one stress assigning rule) in an integrated bilingual lexicon (Roelofs, 2003; 2006). Overall, it is still unclear whether suprasegmental information is co-activated during spoken word production, especially for bilinguals whose two languages have different systems of suprasegmental contrasts.

In this study, we aimed to fill this knowledge gap by investigating the activation of suprasegmental information with bilinguals of SC and English. SC is a representative tonal language, which uses pitch variation to differentiate morpheme or word meanings, just as consonants and vowels. For example, *ma* means "mother" when it is produced with a level pitch contour, but "horse" with a low-dipping pitch contour. Unlike tonal languages, stress languages such as English employ relative prominence between syllables to distinguish words (e.g., REcord and reCORD), which is cued not only with salient pitch contours but also salient lengthening, intensity increase, and vowel quality contrast (see Gordon & Roettger, 2017 for a review on cues of stress). Furthermore, unlike tonal contrast, which is abundant in Mandarin, the number of stress minimal pairs in English is limited (Giegerich, 1992). In short, the way suprasegmental information is utilized differs significantly in the lexicons of SC and English, which offers a unique case for investigating the interplay of phonological representations and language co-activation at the suprasegmental level.

We employed the PWI paradigm and asked SC-English bilinguals to name pictures in English while ignoring simultaneously presented SC distractors.

Critically, we manipulated the phonological overlap between the target's SC translation and the distractors. As shown in Table 1, for the same target (e.g., feather), there are four types of distractors: 1) translation distractor, which is the target's SC translation (e.g., *yu3mao2* "feather"); 2) tone-sharing (phono-translation) distractor, which shares both segments and lexical tone with the target's SC translation in the first syllable (e.g., *yu3zhou4*, "universe"); 3) no-tone-sharing (phono-translation) distractor, which shares only segments with the target's SC translation in the first syllable (e.g., *yu4mi3*, "corn"); 4) unrelated distractor, which has neither segmental nor tonal overlap with the target's SC translation (e.g., *lei4shui3*, "tear"). All the target and distractor pairs are not semantically related within and across languages. If lexical tone plays an important role in cross-language activation, we expect to find a significant difference between naming latencies in the tone-sharing and no-tone-sharing conditions.

Table 1. *A set of sample stimuli. The SC syllables are spelt out in Pinyin, an alphabetic writing system of SC. The numbers in Pinyin here indicate the lexical tone.*

	Target	Distractors			
		Translation	*Tone-sharing*	*No-tone-sharing*	*Unrelated*
English	feather	feather	universe	corn	tear
Pinyin	yu3mao2	yu3mao2	yu3zhou4	yu4mi3	lei4shui3
Character	羽毛	羽毛	宇宙	玉米	泪水

Moreover, we manipulated two procedural factors in the PWI. One is the modality of the distractors, namely whether participants listened to or viewed distractor words during picture naming (auditory vs. visual distractors). The other is the familiarization mode, i.e., whether participants were given English names only (i.e., the English mode) or both English and SC names (i.e., the mixed mode)

of the pictures before naming pictures in English.[7] These manipulations were designed to clarify the following specific issues in our understanding of language co-activation during bilingual word production.

The first issue concerns the different inhibition or facilitation effects with different distractor modalities. Despite that both *translation facilitation* and *phono-translation interference* effects have been used as indicators for language co-activation, it has been controversial how to reconcile the fact that the two effects are opposite (see Costa, 2005 and Hall, 2011 for reviews on this issue). Specifically, why would the phono-translation distractor, which is phonologically related to the targets' translation, inhibit target picture naming, while the translation distractor itself facilitates picture naming? Based on these observations, Costa (1999; 2003) proposed a language-specific lexical selection theory: words of the non-target language are excluded from lexical selection so that co-activated translations cause no lexical interference but semantic facilitation to target naming. By contrast, Hermans and his colleagues (Hermans et al., 1998; Hermans, 2004) argued for a language-nonspecific lexical selection account: there are two mechanisms underlying the translation and phono-translation effects, namely semantic facilitation at the conceptual level and cross-language competition at the lexical level; for translation distractors, semantic facilitation overrides lexical competition, leading to a net priming effect; for phono-translation distractors, the relatively weak semantic facilitation cannot overrule lexical competition, resulting in an interference effect. Despite the two opposing views, it is important to note that previous studies have mainly observed the translation facilitation effect with visual distractors (e.g., Costa et al., 1999; Costa & Caramazza, 1999; Hermans, 2004) and the phono-translation interference effect with auditory distractors (e.g., Hermans et al., 1998; Costa et al., 2003). Given that within-

[7] Typically, in PWI, there is a familiarization session before performing the naming task which allows participants to preview the target pictures and become familiarized with the intended target names.

language semantic distractors were typically found to be facilitative in the visual modality but inhibitive in the auditory modality (e.g., Hantsch et al., 2009; Jonen et al., 2021), the "translation and phono-translation paradox" (Hall, 2011) may simply be an artefact of distractor modality. To test this possibility and therefore gain a better understanding on the effect of lexical tone activation, in this study, we examined the effects of translation and phono-translation distractors in both visual and auditory domains. If indeed the contrast is caused by modality difference, we expect to replicate the translation facilitation effect only with visual distractors and the phono-translation interference effect only with auditory distractors. However, if the two effects are indeed opposite despite the modality difference, the central controversy may lie in whether language co-activation necessarily causes competition in language selection (i.e., the view of language-specific lexical selection vs. the view of language-nonspecific lexical selection; Hermans et al., 1998; Costa et al., 2003).

The second related issue concerns the conflicting views of language-specific and non-specific lexical selection, which we aimed to shed light upon by examining the consequences of increasing the activation level of the non-target language. If co-activation leads to cross-language competition as indicated by the language non-specific lexical selection view (Hermans et al., 1999), higher activation of the non-target language may interfere with target selection and prolong the naming latency. To this end, we further manipulated the SC activation level by adjusting the familiarization mode. Specifically, we expect to introduce a higher SC activation level by providing both English and SC target names in the familiarization session (mixed mode) than English names only (English mode). If we find a larger interference effect in the mixed mode over the English mode, this suggests the involvement of cross-language competition, lending support to the language non-specific lexical selection view (Hermans et al., 1999).

In sum, we aimed to answer whether SC lexical tone is co-activated during English picture naming by investigating the translation and phono-translation effects 1) across different distractor modalities and 2) with

familiarization modes. By doing so, we hope to gain a more comprehensive understanding of the interaction between bilinguals' two languages. More specifically, we conducted four PWI experiments with four types of distractors (i.e., translation, tone-sharing distractor, no-tone-sharing, and unrelated distractors). In Experiment 1, participants were familiarized with targets' English names only and then performed picture naming with the presence of auditory distractors. In Experiment 2, participants were familiarized with the targets' English and SC names and then performed naming tasks with auditory distractors. Experiments 3 and 4 are replications of Experiments 1 and 2, respectively, using visual distractors.

4.1 Experiment 1 (Auditory Distractor and English Mode)

4.1.1 Method

4.1.1.1 Participants

Forty-one SC-English bilinguals (30 females and 11 males; average age 24) participated in this experiment. All participants are native SC speakers who grew up in Northern China. They speak no local dialect and have no history of language disorder. All participants passed College English Test Band 6 or scored above 6 in International English Language Testing System (IELTS). We also assessed participants' English proficiency level with an adapted LEAP-Q questionnaire (Marian et al., 2007) and the multilingual naming test (MINT; Gollan et al., 2012) which has been validated by several studies (e.g., Sheng et al., 2014 with Chinese-English bilinguals). All participants learned English at an average age of 8.0 (SD = 3.0). Using a Likert scale from 1 to 10, the mean self-rated proficiency by the participants was 8.0 (SD = 1.6) in reading, 6.7 (SD = 1.7) in speaking and 7.0 (SD = 1.7) in listening. The average correct response of MINT was 58% (SD = 11%). This study was approved by the Ethics Committee of the

Faculty of Humanities at Leiden University. All participants provided informed consent and were compensated 30 RMB for their participation.

4.1.1.2 Stimuli

There are 24 sets of critical stimuli (see Appendix C). Each set consists of an English target word, an SC translation distractor, an SC tone-sharing distractor, an SC no-tone-sharing distractor, and an SC unrelated distractor. There are also 12 sets of filler words which are not phonologically or semantically related. All words are common disyllabic nouns. Word frequency of SC and English, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010) and SUBTLEX-US (Brysbaert & New, 2009), are balanced across conditions [SC: $F(3, 92) = 1.97$, $p = 0.13$; English: $F(3, 92) = 1.76$, $p = 0.16$]. Word length in English was also controlled [$F(3, 92) = 0.753$, $p = 0.52$]. The target pictures, which are black and white line drawings, were selected from the IPNP database (Bates et al., 2003) and the BOSStimuli database (Brodeur et al., 2012). Twenty-seven native Mandarin speakers who did not participate in the PWI experiments validated the choices of target stimuli in terms of picture naming agreement, translation agreement, and picture imageability of distractors. All spoken stimuli were recorded by a female native SC speaker (age 22) who was born and grew up in Beijing. The recording was done at the Leiden University Centre for Linguistics Phonetics Lab through a Sennheiser MKH416T microphone (44.1 kHz, 16 bit). All stimuli were normalized for duration of 1,000 ms and intensity at 70 dB using Praat (Boersma & Weenink, 2022).

4.1.1.3 Procedure

Participants performed the experiment online using Gorilla (www.gorilla.sc). All participants were required to wear headphones and sit in a quiet room. Participants were only allowed to join the experiment if using laptops. Prior to the experiment, a headphone task based on the dichotic pitch (Milne et al., 2020), as well as a microphone check and an auto-play check were

run to screen participants' equipment. All the instructions were given in English**.** Before the picture naming task, there was a familiarization session. During this session, participants were shown 36 target pictures (24 critical and 12 filler targets) with their matching English names printed underneath. Afterwards, participants were asked to type in the picture names in English. If participants did not respond accurately, the intended name would appear again. In the PWI task, a fixation was displayed in the centre of the screen for 500 ms, followed by a picture and simultaneously played English spoken distractors (Stimulus Onset Asynchrony = 0 ms). Participants were asked to name the picture as quickly and accurately as possible while ignoring the auditory distractor. The picture remained visible for 2,000 ms. Response times were measured from picture onset until naming onset using Chronset (Roux et al., 2017). If participants did not respond within 2,000 ms, the trial ended, and the experiment proceeded automatically. Between each trial, there was a blank screen of 1,000 ms. Before starting the task, participants were asked to complete four practice trials, with an option to practice more rounds. In total, there were 96 (24 targets × 4 conditions) critical trials and 48 (12 targets × 4 conditions) filler trials. All the trials were equally distributed into four blocks in a Latin Square design so that participants only saw each target picture once in every block. Between each block, participants were encouraged to take a short break without changing the equipment set-up. After the PWI task, participants were asked to complete the MINT test and a language background survey. In total, the experiment took about 30 minutes.

4.1.2 Data Analysis

Response times (hereafter RT) were analysed using the generalized linear mixed-effects model (GLMM) with inverse Gaussian distribution (Lo & Andrews, 2015). Incorrect responses (e.g., responses in SC), blank responses and unrecognizable responses were excluded from data analysis. For each experiment, a maximum model including fixed effects of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), random

slope for distractor type by subject and item, and random intercepts for subject and item were constructed first. If a model failed to converge, we increased the number of iterations and then simplified the model by removing correlation parameters in the random structures (Brauer & Curtin, 2018). All the analyses were run in R Studio (R Core Team, 2022) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Pairwise comparisons were computed using the *multcomp* package (Hothorn et al., 2022). Holm–Bonferroni method was implemented to correct family-wise errors (Holm, 1979).

4.1.3 Results

Incorrect trials (~2.9%), trials with no response (~2.9%), and unrecognizable responses (~0.2%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Table 2 and Figure 1 summarise the mean RT for each condition. Compared with unrelated distractors, participants took about 29 ms longer to name pictures with tone-sharing distractors (i.e., tone-sharing condition), about 10 ms longer with no-tone-sharing distractors (i.e., no-tone-sharing condition), and about 28 ms less with translation distractors (i.e., translation condition). Moreover, response time in the tone-sharing condition was about 19 ms longer than in the no-tone-sharing condition. The final GLMM consists of the fixed effects of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), random slope for distractor type by subject and item, and random intercepts for subject and item. According to the GLMM estimations (see Table 3), the interference effect in the tone-sharing condition ($p < 0.01$) and the facilitation effect in the translation condition ($p < 0.05$) are both statistically significant. Crucially, naming latency in the tone-sharing condition is significantly longer than in the no-tone-sharing condition ($p < 0.05$).

Table 2. *Mean naming latency and standard deviations of Experiment 1 (Auditory Modality and English Mode), Experiment 2 (Auditory Modality and Mixed Mode), Experiment 3 (Visual Modality and English Mode), and Experiment 4 (Visual Modality and Mixed Mode). The mixed effect shows the naming latency difference between the Mixed Mode and the English Mode (Mixed – English).*

Modality	Auditory Modality					Visual Modality				
Familiarization mode	English		Mixed		Mixed Effect	English		Mixed		Mixed Effect
Conditions	Mean	SD	Mean	SD		Mean	SD	Mean	SD	
Tone-sharing	1001	367	1064	348	64	1039	375	1002	290	-37
No-tone-sharing	982	360	1058	342	76	1033	319	1012	294	-21
Translation	944	329	1059	342	115	908	296	899	248	-9
Unrelated	972	351	1043	329	71	1044	348	1035	323	-8
Tone – Unrelated	29		21		-7	-5		-34		-29
No-tone – Unrelated	10		15		5	-11		-23		-12
Tone – No-tone	19		6		-12	6		-11		-17
Translation – Unrelated	-28		16		44	-136		-137		0

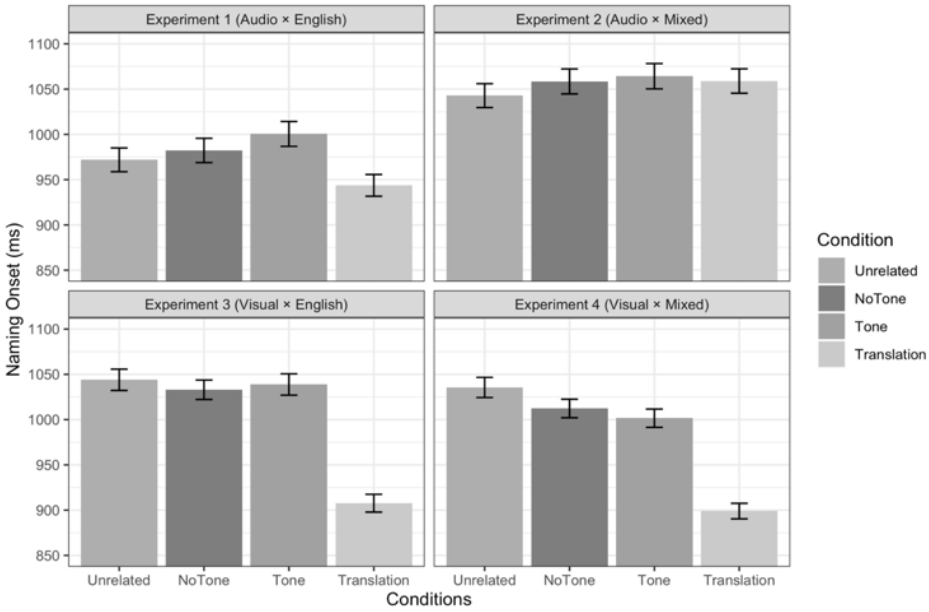

Figure 1. *Mean naming latencies of all conditions in Experiment 1, Experiment 2, Experiment 3, and Experiment 4.*

Table 3. *GLMM analysis of naming latency in Experiment 1.*

	Estimate	SE	t	*p*-value
(Intercept)	1160.440	19.200	60.440	<0.001
No-tone Sharing – Unrelated	14.540	13.070	1.112	0.266
Tone Sharing – Unrelated	54.760	15.950	3.433	0.002
Translation – Unrelated	-41.040	14.730	-2.786	0.016
No-tone Sharing – Tone Sharing	40.230	16.370	2.458	0.028

4.1.4 Discussion

In this experiment, we found that, compared with unrelated auditory distractors, target picture naming was significantly speeded up by SC translation distractors but slowed down by tone-sharing distractors. Such findings replicated previously found *translation facilitation* (e.g., Costa et al., 1999, 2003; Hermans,

2004) and *phono-translation interference* effect (e.g., Costa et al., 2003; Hermans et al., 1998) with auditory distractors. Importantly, there was a significant naming latency difference between the tone-sharing and no-tone-sharing conditions, indicating that lexical tone is co-activated and plays a significant role during the process of English picture naming.

4.2 Experiment 2 (Auditory Distractor and Mixed Mode)

4.2.1 Method

4.2.1.1 Participants

Forty-two SC-English bilinguals (31 females and 11 males; average age 24) participated in this experiment. All participants learned English at an average age of 8.6 (SD = 3.1). The mean self-rated frequency by participants was 8.1 (SD = 1.4) in reading, 6.8 (SD = 1.4) in speaking and 7.2 (SD = 1.7) in listening. The average correct response of MINT was 61% (SD = 10%).

4.2.1.2 Stimuli, Procedure & Data analysis

During the familiarization session, participants were provided with both English and SC names (i.e., mixed mode). By doing so, we expect to introduce a higher level of SC activation. If we find larger interference effects across distractors in Experiment 2 than Experiment 1, this suggests the involvement of cross-language competition in bilingual picture naming. Besides this, the same stimuli, procedure and analysis as in Experiment 1 were used in this experiment.

4.2.2 Results

Incorrect trials (~3.1%), trials with no response (~1.6%), and unrecognizable responses (~0.1%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Table 2 and Figure 1 summarize the mean RT for each condition. Compared with the unrelated condition, participants took about 21 ms longer to

name pictures in the tone-sharing condition and about 15 ms longer in the no-tone-sharing and translation condition. The final GLMM consists of the fixed effects of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractor), random slope for distractor type by subject and item, and random intercepts for subject and item. According to model estimations (see Table 4), although the tone-sharing, no-tone-sharing and translation condition all exhibited inhibitory trends towards picture naming, none of them significantly differed from the unrelated condition (tone-sharing condition: $p = 0.227$; no-tone-sharing condition: $p = 0.547$; translation condition: $p = 0.547$).

Table 4. *GLMM analysis of naming latency in Experiment 2.*

	Estimate	SE	t	*p*-value
(Intercept)	1211.350	21.880	55.363	<0.001
No-tone Sharing – Unrelated	10.150	15.640	0.649	0.547
Tone Sharing – Unrelated	33.090	17.370	1.905	0.227
Translation – Unrelated	23.150	19.350	1.196	0.547
No-tone Sharing – Tone Sharing	22.940	17.200	1.334	0.547

As shown in Table 2, naming latencies in Experiment 2 are longer than in Experiment 1 across conditions. To further investigate the effect of familiarizing both English and SC names, a joint analysis of Experiment 1 and 2 was conducted. The final joint GLMM includes the fixed effect of distractor type (i.e., translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), familiarization mode (English Mode in Experiment 1 vs. Mixed Mode in Experiment 2), the interaction between distractor type and familiarization mode, random intercepts for subjects and items, by-subject random slope for distractor type, by-item random slope for distractor type and familiarization mode, and their interaction. Results of the analysis showed a significant main effect of familiarization mode ($p < 0.001$) and a significant

interaction between condition and familiarization mode ($p < 0.05$). Pairwise comparisons showed that naming latency of all conditions in Experiment 2 were significantly longer than that of Experiment 1 (Translation: $p < 0.001$; Tone-sharing: $p < 0.05$; No-tone-sharing: $p < 0.05$; Unrelated: $p < 0.05$).

4.2.3 Discussion

When bilingual participants were familiarized with both English and SC names (instead of English names only as in Experiment 1), we failed to replicate the previously found *translation facilitation* (e.g., Costa et al., 1999, 2003; Hermans, 2004) and *phono-translation interference* effect (e.g., Hermans et al., 1998; Costa et al., 2003) with auditory distractors. None of the distractors (i.e., translation, tone-sharing and no-tone-sharing) in Experiment 2 had a significant impact on the picture naming latency compared with unrelated distractors. Importantly, a joint analysis showed that the naming latency of all conditions was significantly longer in Experiment 2 (mixed mode) than in Experiment 1 (English mode). This indicates that introducing SC names alongside English names increases the processing demands involved in English picture naming.

4.3 Experiment 3 (Visual Distractor and English Mode)

4.3.1 Method

4.3.1.1 Participants

Forty-three SC-English bilinguals (34 females and 9 males; average age 24) participated in this experiment. All participants learned English at an average age of 7.3 (SD = 2.8). The mean self-rated frequency by participants was 8.1 (SD = 1.6) in reading, 7.0 (SD = 1.8) in speaking and 7.5 (SD = 1.8) in listening. The average correct response of MINT was 61% (SD = 13%).

4.3.1.2 Stimuli, Procedure & Data analysis

Experiment 3 is a counterpart of Experiment 1. Instead of presenting auditory distractors, the SC distractor words were superimposed on the target pictures as Chinese characters.

4.3.2 Results

Incorrect trials (~3.3%), trials with no response (~3.4%), and unrecognizable responses (~0.1%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Compared with unrelated distractors, participants took 136 ms less to name pictures when translation distractors were present, and 5 ms and 11 ms less, respectively, with tone-sharing and no-tone-sharing distractors (see Table 2 and Figure 1). The final GLMM consists of the fixed effects of distractor type, by-subject and by-item random slope for distractor type, and random intercepts for subject and item. According to the GLMM estimation (see Table 5), only the naming latency in the translation condition was significantly different from the unrelated condition ($p < 0.001$), showing a robust translation facilitation effect. Although both tone-sharing and no-tone-sharing conditions exhibited a small facilitatory trend toward picture naming, the response times did not significantly differ from the unrelated condition (tone-sharing condition: $p = 0.631$; no-tone-sharing condition: $p = 0.451$).

Table 5. *GLMM analysis of naming latency in Experiment 3.*

	Estimate	SE	t	*p*-value	
(Intercept)	1195.080	33.051	36.159	<0.001	***
No-tone Sharing – Unrelated	-17.542	12.728	-1.378	0.451	
Tone Sharing – Unrelated	6.384	13.279	0.481	0.631	
Translation – Unrelated	-161.933	19.143	-8.459	<0.001	***
No-tone Sharing – Tone Sharing	23.926	16.638	1.438	0.451	

Given that Experiment 3 is a replication of Experiment 1 with visual distractors, a joint analysis of Experiment 3 and 1 was conducted to investigate the impact of distractor modality. The final joint GLMM includes the fixed effect of distractor type (translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), distractor modality (Auditory Modality in Experiment 1 vs. Visual Modality in Experiment 3), random intercepts for subjects and items, by condition random slope for subjects, by condition and by modality random slope for items. Results of GLMM analysis showed a significant main effect of distractor modality ($p = 0.010$) and a significant interaction between condition and modality ($p < 0.001$). Although naming latencies of tone-sharing, no-tone-sharing, and unrelated conditions in Experiment 3 were longer than Experiment 1, the differences were not statistically significant (tone-sharing: $p = 0.071$; no-tone-sharing: $p = 0.190$; unrelated: $p = 0.190$). However, the naming latency of the translation condition was significantly shorter in Experiment 3 than in Experiment 1 ($p < 0.001$).

4.3.3 Discussion

With visual distractors, we found robust *translation facilitation* effect, replicating findings from previous studies (e.g., Costa et al., 1999, 2003; Hermans, 2004) and from Experiment 1 with auditory distractors. However, in contrast to the previously found *phono-translation inhibition* effect, the tone-sharing and no-tone-sharing (phono-translation) distractors did not exhibit inhibition but rather insignificant facilitatory trends toward English picture naming.

4.4 Experiment 4 (Visual Distractor and English Mode)

4.4.1 Method

4.4.1.1 Participants

Thirty-eight SC-English bilinguals (31 females and 7 males; average age 24; SD = 1.6) participated in this experiment. All participants learned English at an average age of 7.3 (SD = 2.9). The mean self-rated proficiency by participants was 8.3 (SD = 1.6) in reading, 7.0 (SD = 2.1) in speaking, and 7.4 (SD = 2.0) in listening. The average correct response of MINT was 61% (SD = 13%).

4.4.1.2 Stimuli, Procedure & Data analysis

Experiment 4 is a counterpart of Experiment 2 with visual distractors. During the familiarization session, participants were provided with both English and SC names when familiarizing themselves with the target pictures.

4.4.2 Results

Incorrect trials (~1.8%), trials with no response (~1.7%), and unrecognizable responses (~0.1%) were excluded from the analysis. Given that error rates were low in each condition, no further analysis on accuracy was conducted. Table 2 and Figure 1 summarise the mean RTs for each condition. Compared with unrelated distractors, participants took about 137 ms less to name pictures when translation distractors were present; 34 ms less with tone-sharing distractors; 23 ms less with no-tone-sharing distractors. The final converged GLMM consists of the fixed effects of distractor type (translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors) and random intercepts for subject and item. According to the GLMM estimation (see Table 6), naming latency in the tone-sharing condition is significantly different than in the unrelated condition ($p < 0.05$); while the no-tone-sharing condition is not ($p =$

0.359). Moreover, with an estimated difference of about 117 ms, the translation condition is significantly different from the unrelated condition, showing a robust *translation facilitation* effect.

As we can see from Table 2, naming latencies in Experiment 4 were shorter than in Experiment 3 across all conditions. To further examine the effect of mixed familiarization mode in the visual modality, a joint analysis of Experiments 3 and 4 was conducted. The final joint GLMM includes the fixed effect of distractor type (translation, tone-sharing distractor, no-tone-sharing distractor, and unrelated distractors), familiarization mode (English Mode in Experiment 3 vs. Mixed Mode in Experiment 4), and random intercepts for subjects and items. Results of GLMM analysis showed a significant interaction between distractor type and familiarization mode ($p = 0.044$), and no significant main effect of familiarization mode was found ($p = 0.748$). Pairwise comparisons showed that there was no significant difference between Experiment 3 and 4 across conditions (Translation: $p = 0.198$; Tone-sharing: $p = 1.000$; No-tone-sharing: $p = 1.000$; Unrelated: $p = 0.832$).

Given that Experiment 4 is a replication of Experiment 2 with visual distractors, a joint analysis of Experiments 4 and 2 was also run to test the impact of distractor modality. The final joint GLMM includes the fixed effect of conditions, modality (auditory modality in Experiment 2 vs. visual modality in Experiment 4), random intercepts for subjects and items, by-subject random slope for distractor type, by-item random slope for distractor type and for modality. Results of the GLMM analysis showed a significant main effect of distractor modality ($p < 0.001$) and a significant interaction between condition and modality ($p < 0.001$). Pairwise comparison showed significant differences between Experiments 4 and 2 in all conditions (Translation: $p < 0.001$; Tone-sharing: $p < 0.001$; No-tone-sharing: $p < 0.001$; Unrelated: $p < 0.005$).

Table 6. *GLMM analysis of naming latency in Experiment 4.*

	Estimate	SE	t	*p*-value
(Intercept)	1133.721	22.673	50.003	<0.001
No-tone Sharing – Unrelated	-10.598	10.062	-1.053	0.359
Tone Sharing – Unrelated	-24.632	9.398	-2.621	0.026
Translation – Unrelated	-117.989	8.965	-13.161	<0.001
No-tone Sharing – Tone Sharing	-14.033	10.454	-1.342	0.359

4.4.3 Discussion

With visual distractors, the *translation facilitation* effect found in previous studies (e.g., Costa et al., 1999, 2003; Hermans, 2004) was replicated, same as in Experiments 1 and 3. The *phono-translation interference* effect (e.g., Hermans et al., 1998; Costa et al., 2003) in either tone-sharing or no-tone-sharing distractors were not found, unlike in Experiment 1. However, tone-sharing distractors introduced a significant facilitation effect to English picture naming whereas no-tone-sharing distractors did not. This indicates a crucial role of lexical tone in the process of bilingual spoken word production. Moreover, with visual distractors, familiarizing with target names in both languages or in English only, had no significant impact on the naming latency. Furthermore, with the mixed mode, auditory distractors interfered with the naming process significantly more than visual distractors.

4.5 General Discussion

To investigate the role of lexical tone in bilingual word production, four PWI experiments were conducted. SC-English bilinguals were instructed to name pictures in English while ignoring simultaneously presented SC distractors. Together with a target picture (e.g., “feather”), there were four types of SC distractors: a translation distractor (e.g., *yu3mao2* “feather”), a tone-sharing (phono-translation) distractor (e.g., *yu3zhou4* “universe”), which shares both segments and tone with the target’s translation in the first syllable, a no-tone-

sharing (phono-translation) distractor (e.g., *yu4mi3* "corn"), which shares only segments with the target's translation in the first syllable, and an unrelated distractor (e.g., *lei4shui3* "tear"). Moreover, to gain a fuller understanding of the effects of co-activation, we manipulated the four experiments along two dimensions. One was distractor modality, i.e., the SC distractors were presented auditorily in Experiments 1 and 2, but visually in Experiments 3 and 4. The other was familiarization mode. Bilinguals were familiarised with English names only in Experiments 1 and 3 (English mode) but with both English and SC names in Experiments 2 and 4 (mixed mode). In Experiment 1 (auditory distractor and English mode), compared with unrelated distractors, we found significantly shorter naming latency with translation distractors and significantly longer naming latency with tone-sharing distractors (but not with no-tone-sharing distractors), replicating the *translation facilitation* and *phono-translation interference* effects found in previous studies (e.g., Costa et al., 1999, 2003; Hermans, 2004). Moreover, there was a significant naming latency difference between the tone-sharing and no-tone-sharing distractors, demonstrating the co-activation of lexical tone during English spoken word production. In Experiment 2 (auditory distractor and mixed mode), naming latencies across conditions were significantly longer than those in Experiment 1 (auditory distractor and English mode); translation and phono-translation distractors all elicited interference towards target naming, but none of the effects was statistically significant. In Experiment 3 (visual distractor and English mode), there was a strong translation facilitation effect; the tone-sharing and no-tone-sharing distractors were also found to be facilitatory, but neither effect was statistically significant. In Experiment 4 (visual distractor and mixed mode), there was also a strong translation facilitation effect; the tone-sharing distractors significantly facilitated picture naming whereas the no-tone-sharing distractors did not, indicating an important role of lexical tone during English picture naming.

For the first time in the literature, our findings show that SC lexical tone is co-activated during English spoken word production. Moreover, we found that

the previously found *translation facilitation* and *phono-translation interference* effects are not fixed but rather dynamic, depending on procedural factors such as the distractor modality and familiarization mode: both *translation facilitation* and tone-sharing *phono-translation interference* effects were replicated using auditory distractors; visual distractors significantly strengthened *translation facilitation* effects and switched the tone-sharing *phono-translation interference* into facilitation. Moreover, mixing two languages during the familiarization session elicited more facilitation with visual distractors, but more interference with auditory distractors. Implications of these findings are discussed in the following sections.

4.5.1 The Role of Lexical Tone in Speech Production

In Experiment 1, we found a significant naming latency difference between tone-sharing and no-tone-sharing conditions. This suggests when bilinguals were naming target pictures in English, the lexical tone of the SC translations was also activated. This finding not only contributes to our understanding of language co-activation but also provides implications for tone word production models.

According to one of the most dominant models of speech production, i.e., the WEAVER++ model (Jescheniak & Levelt, 1994; Roelofs, 2000, 2015), to successfully generate a spoken word in stress languages such as English, speakers need to retrieve both phonological content and metrical frame during phonological encoding. While phonological content includes a set of ordered segments (phonemes), the metrical frame consists of syllabic information (e.g., number of syllables) and suprasegmental information (e.g., stress patterns). After retrieval, segments and the metrical frame associate and form a phonological word, i.e., a sequence of syllable(s). To account for tone word production, Roelofs (2015) proposed a level of tonal frame which functions similarly to the metrical frame. However, recent findings of tone error seem to suggest a more critical role of lexical tone. In a large Cantonese natural conversation corpus, Alderete et al.

(2019) found that tone errors are common (over 20% of the total sound errors) and tend to be influenced by adjacent tones in the same fashion as segmental errors. These findings prompt Alderete and his colleagues to propose that lexical tone is independently represented and equally selected as segments. Our finding of lexical tone co-activation in bilingual spoken word production seems to add more evidence to the view of Alderete et al. (2019). If lexical tone is represented diacritically as a tonal frame and not actively selected until phonetic spell-out, we are unlikely to observe any effect of lexical tone during bilingual word production of a non-tonal language. Together with the findings of previous studies, our results indicate that speakers of tonal languages are likely to select and encode lexical tone just as segments regardless of whether speaking in their native tonal language or the non-tonal second language.

4.5.2 The Translation Facilitation and Phono-Translation Interference Effects

As discussed in the introduction, previous studies have not reached a consensus on how to reconcile the seemingly contrasting *translation facilitation* and *phono-translation interference* effects (Costa, 2005; Hall, 2011). One possibility is that the contrast is introduced using different distractor modalities. With auditory distractors, we replicated both *translation facilitation* and tone-sharing *phono-translation interference* effects in Experiment 1. It is thus unlikely that the opposite translation and phono-translation effects are artefacts of distractor modality. Instead, distractor modality has a significant impact on the magnitude and direction of the (phono-)translation effects. With visual distractors, we found stronger effects of *translation facilitation* than with auditory distractors. Moreover, with visual distractors, the auditory *phono-translation interference* effect turned into a facilitation effect. As discussed in Hantsch et al. (2009, p. 1451), who also observed opposite effects for auditory and visual modality, respectively, this may be “due to differences of the time course with which the semantic representation of the distractor becomes available.” Given the nature of

parallel processing of visual distractors versus sequential processing of auditory distractors, semantic representations of visual distractors may become available more quickly than that of auditory distractors. As a result, facilitation at the semantic level introduced by visual (translation and phono-translation) distractors may be more likely to boost lemma activation and speed up target picture naming than auditory distractors.

As discussed earlier, there are two general views on the underlying mechanisms of *translation facilitation* and *phono-translation interference* effects. One is the language-specific selection view (Costa, 1999), which posits that co-activated translations facilitate target production at the conceptual level; the other is the language non-specific selection view (Hermans et al., 1998), which argues that co-activated translations not only introduce semantic facilitation but also interfere with the process of word selection. Our results on the mixed vs. English familiarization mode seem to agree with the latter view. As we increased the activation level of SC by exposing participants to both English and SC target names during the familiarization session, the picture naming latencies increased significantly across distractor types compared to when participants were presented with English names alone. Furthermore, the effect of *translation facilitation* was cancelled out.

However, it is important to note that the inhibitory effect of the mixed familiarization session was only found with auditory distractors (Experiment 1 vs. Experiment 2) but not with visual distractors (Experiment 3 vs. Experiment 4). There are two possible explanations for this divergence. First, with visual distractors, bilinguals viewed Chinese characters imposed on target pictures, which might have led them to encounter the ceiling level of cross-language interference; thus, further increasing SC activation by introducing SC names in the familiarization session did not elicit any significant inhibitory effect on target picture naming. Second, according to Hermans et al. (1998), increasing SC activation could result in not only more cross-language interference but also more semantic facilitation; given that visual distractors may have earlier access to

semantic representations than their auditory counterparts, it is thus possible that the increased SC activation in mixed mode boost semantic facilitation more readily with visual distractors than auditory distractors and cancelled out part of the cross-language interference effect.

4.5.3 Methodological Contributions

The present study also has a few methodological contributions to the use of PWI in bilingual word production studies. First, this study validated the significant impact of distractor modality in PWI and extended findings of distractor modality's influence on monolingual semantic effects (e.g., Hantsch et al., 2009; Jonen et al., 2021) to bilingual translation and phono-translation effects. Second, we demonstrated that asking bilinguals to become familiarized with target pictures' names in both their languages is an effective way to adjust language activation levels, especially with auditory distractors. This could be an important factor to manipulate for future use of PWI in bilingual studies. Third, successfully replicating previous lab findings with online experiments, this study showed that virtual PWI is an efficient and sound approach to studying speech production.

4.6 Conclusion

In conclusion, the data reported in this study showed, for the first time, that lexical tone is co-activated during the process of bilingual spoken word production. Moreover, we found that the effect of co-activating translations and their lexical tone is greatly impacted by experimental details such as distractor modality (i.e., whether participants see or hear distractor words) and familiarization mode (i.e., whether participants are familiarized with picture names in the target language only or both the target and non-target languages before picture naming). These findings provide new insights for understanding language co-activation at the suprasegmental level and the role of lexical tone in spoken word production.

Chapter 5

Do Standard Chinese-English Bilinguals Produce English Words with Lexical Tone in their Minds?

A version of this chapter is published as: Yang, Q., & Chen, Y. (2023). *Do Chinese-English Bilinguals Speak English Words with Lexical Tone in Mind?*.20th International Congress of Speech Sciences (ICPhS 2023), Prague, Czech Republic.

Abstract

Although it is well known that words of bilinguals' two languages interact extensively, whether and how language-specific suprasegmental features interact in bilingual lexical access remains unclear. This study investigated whether lexical tone affects pitch processing during English word production. Using the picture-word interference paradigm, we asked Chinese-English bilinguals and English monolinguals to name pictures in English while ignoring simultaneously played auditory Standard Chinese distractors. Crucially, these Standard Chinese distractors are cross-language homophones to the English target names, which have a falling or a rising lexical tone. Naming latency results showed that cross-language homophones with rising-tone facilitated picture naming more than their falling-tone counterparts for the bilinguals. This effect was not found with English monolinguals. Such a difference suggests a significant influence of lexical tone on pitch processing during spoken word production even in these bilinguals' non-tonal language, lending evidence to the interaction between bilinguals' two languages at the suprasegmental level.

Keywords: lexical tone; spoken word production, the bilingual lexicon

The functional role of pitch variation differs across languages. In lexical tone languages such as Standard Chinese (hereafter SC), pitch contour plays a crucial role in differentiating morpheme meanings, just as consonants and vowels (e.g., *ma* with a rising pitch contour means "hemp" but "scold" with a falling contour). For words in non-tonal languages such as English, pitch contour serves as a cue to distinguish a limited number of words, known as lexical stress (for cues of stress, see Gordon & Roettger, 2017 for an informative review). For both types of languages, pitch variation also serves to signal utterance-level information such as sentence mode. For example, in most varieties of English, "Mary" can be uttered with a rising pitch contour to signal a question and a falling contour to signal a statement; there is a probabilistically stable mapping between pitch contour shapes and sentence modes. In SC, however, pitch variation for sentence mode is constrained by the lexical tone pitch contours (see, e.g., Chen, 2022 for a review on intonation in tonal languages; Liu, Chen, & Schiller 2020 and references therein for question-induced pitch variation and intonation perception). Such cross-language differences between SC and English in the form and function of pitch variations offer a unique case for investigating pitch processing in the bilingual mental lexicon.

It is by now widely agreed that bilinguals' two languages interact extensively (see Kroll & Tokowicz, 2005 for a review). Words of bilinguals' two languages are constantly active in parallel, resulting in cross-language interaction at all levels of speech processing and planning. For example, when naming a picture in one language, bilinguals may experience shorter naming latency when the translation equivalent of the target picture name in their other language is present (e.g., Costa & Caramazza, 1999). Conversely, they may experience longer naming latency when a homophone of the translation equivalent is present (e.g., Hermans et al., 1998). Moreover, bilinguals may experience difficulty detecting phonemes that are present in the translation equivalent of a word, compared to those that are not (e.g., Colomé, 2001). While these observations show clear evidence that bottom-up lexical and sub-lexical (phonological) overlap creates

cross-language interaction and competition, it is important to note that most previous studies drew evidence from the effects of segmental overlap in Indo-European language (e.g., Hermans et al., 1998; Colomé, 2001; Costa et al., 2003). What has remained a little-known area is how languages with great typological differences such as English and SC influence each other at the suprasegmental level.

It is of interest to note that there has been robust evidence showing that long-term experience with a tonal language shapes a speaker's pitch processing in general. For example, compared with English listeners, SC listeners have been found to have better discriminative ability to non-native tonal contrast (e.g., Wayland & Guion, 2004), enhanced tonal sensitivity at pre-attentive and attentive processing stages (e.g., Chandrasekaran et al., 2007, 2009), and greater activity in left hemispheres during tone perception (Gandour et al., 2003, 2004; Wang et al., 2004). However, these studies mainly focused on native and non-native lexical tone processing. There is a surprising paucity of empirical research on how lexical tone affects pitch processing in the non-tonal language that bilinguals command. So far, only three studies have examined whether and if so, how lexical tone affects non-tonal speech processing.

One pioneering study on the role of lexical tone in non-tonal speech comprehension is Shook & Marian (2015). In this study, SC-English bilinguals were asked to listen to an English spoken word (e.g., "tree") and select its SC translation from two Mandarin words on display (e.g., *shu* with a falling pitch contour "tree" and *long* with a rising pitch contour *"*dragon*"*). Critically, the pitch contour of the spoken English target word was manipulated to either match or mismatch the lexical tone of its SC translation. With eye-tracking recordings, Shook and Marian (2015) found that, when the pitch contour of the English target matched the lexical tone of the SC translation (e.g., *tree* with a falling pitch contour), SC-English bilinguals fixated on the correct translation earlier and more frequently compared with when the pitch contour did not match (e.g., *tree* with a rising pitch contour). It is most likely that the matching tonal information was

retrieved and exploited to co-activate the SC translation equivalents through top-down and/or lateral translation links. According to the authors, this finding demonstrates a clear influence of suprasegmental information across languages; the effects of native language knowledge on L2 speech processing are not limited to segments but also extend to suprasegmental information.

Ortega-Llebaria et al. (2017) obtained further evidence from speech comprehension that bilinguals' access to non-tonal words is significantly influenced by having a tonal system in their native language. In their study, SC-English bilinguals, Spanish-English bilinguals, and English monolinguals' performances in an English primed-lexical decision task were compared. The prime and target in the task were manipulated to fully match (e.g., *rice* with a falling pitch contour - *rice* with a falling pitch contour), fully mismatch (e.g., *gold* with a rising pitch contour - *rice* with a falling pitch contour), mismatch in segments (e.g., *mice* with a falling pitch contour - *rice* with a falling pitch contour), or mismatch in pitch (e.g., *rice* with a rising pitch contour - *rice* with a falling pitch contour). Results showed that, among the three groups of participants, only SC-English bilinguals experienced significantly larger facilitation across conditions when the targets were produced with a falling pitch contour than that with a rising pitch contour. Ortega-Llebaria et al. (2017) thus suggested that, for SC-English bilinguals, English words with a falling pitch contour must be closer English lexical representations than those with a rising pitch contour; consequently, English words with a falling pitch contour were easier to access. Moreover, the fact that only SC-English bilinguals manifested the "falling-f0 bias" indicated that their long-term experience with a tonal language must be responsible for such an effect in their pitch processing in English.

These findings were re-examined in Ortega-Llebaria and Wu (2020) as a replication plus extension of Ortega-Llebaria et al. (2017). Besides English words, primed-lexical decision tasks with SC words and English non-words were added to further explore alternative explanations for the "falling-f0 bias" such as L1 transfer and pre-lexical pitch processing differences. Similarly, results of the

English lexical decision task showed that SC-English bilinguals responded to the falling-f0 English words significantly faster than their rising-f0 counterparts when the prime and target were identical (e.g., *rice* with a falling pitch contour - *rice* with a falling pitch contour) or mismatched in tone (e.g., *rice* with a rising pitch contour - *rice* with a falling pitch contour); whereas English monolinguals did not. These findings successfully replicated the SC-English bilingual "falling-f0 bias" in Ortega-Llebaria et al. (2017). Moreover, Ortega-Llebaria and Wu (2020) found that such a bias was not observed with SC words, ruling out the explanation of L1 transfer. As for English non-words, there was an opposite "rising-f0" bias: SC-English bilinguals responded to the rising-f0 English non-words significantly faster than their falling-f0 counterpart, ruling out the possibility that the locus of the "falling-f0" bias in English real-words was at the pre-lexical processing stage. Based on these findings, Ortega-Llebaria and Wu (2020) reached a validated conclusion: SC-English bilinguals represent non-tonal L2 words with a tonal-like falling pitch contour due to their long-term experience with SC, a typical tonal language.

Overall, these studies have provided evidence that the lexical tone of a bilingual's native language can significantly influence their pitch and lexical processing of a non-tonal language. Specifically, the bilingual lexicon may be organized in such a way that non-tonal words are represented with a falling pitch contour, similar to the falling lexical tone of words in their native tonal language.

While these findings suggest that interaction between bilinguals' two languages can occur not only at the segmental level but also at the suprasegmental level, it is important to note that all aforementioned studies focused on the domain of speech comprehension. To our knowledge, there is no evidence that came from the production domain in the literature. Studies on second language acquisition may be able to provide some evidence on the influence of lexical tone in non-tonal speech production, as native speakers of tonal languages are found to have unique patterns of producing prosody in stress languages. For example, SC learners of English tend to avoid de-accentuation in post-focal contexts (e.g.,

McGory, 1997); produce intonational pitch accent in English with a most closely matching tonal contour (Ploquin, 2013); produce stressed syllables in words with pitch peak (e.g., Visceglia & Fodor, 2006). Phonological analysis on tonal substrates of English is also revealing, as researchers recently proposed that, as a result of language contact, Cantonese English has at least two contrastive lexical tones (e.g., Yiu, 2014; Wee, 2016; but see Köhnlein et al., 2019 for a different opinion). None of these studies directly investigate whether, and if so, to what extent lexical tone affects pitch processing in non-tonal speech production. To reach a more comprehensive understanding of language interaction at the suprasegmental level, more empirical data on the influence of lexical tone during spoken word production is needed. Addressing the question of whether lexical tone plays a role in speaking English words not only has important implications for our understanding of bilingual language interaction at the suprasegmental level but could also contribute to a more comprehensive view of how bilinguals' mental lexicon is organized and represented.

The goal of the current study was to fill in this gap by examining the "falling-f0 bias" hypothesis in spoken word production. As findings in speech comprehension suggested (Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2020), SC-English bilinguals represent non-tonal English words with a falling "tone", resulting in a "falling-f0 bias" in their English spoken word recognition. Following this conjecture, one may infer that the processing contrast between falling and rising tones is also evident in English spoken word production. To test this hypothesis, we employed the picture-word interference paradigm (hereafter PWI; Rosinski et al., 1975), the most widely used paradigm in studying spoken word production, and asked native SC-English bilinguals and English monolinguals to name pictures in English while ignoring simultaneously played SC distractor words. Crucially, for the same target word (e.g., *lung*), there were four types of SC distractors: 1) the target's cross-language homophone with a falling tone (CH_F; e.g., *lang* with a falling tone*, "wave"*); 2) the target's cross-language homophone with a rising tone (CH_R; e.g., *lang* with a rising tone*,*

"wolf"); 3) an unrelated distractor with a falling tone (UN_F; e.g., *you* with a falling tone*, "right"*); 4) an unrelated distractor with a rising tone (UN_R; e.g., you with a rising tone*, "swim"*).

Previous bilingual PWI studies have found robust facilitation effects of cross-language homophones (e.g., Costa & Caramazza, 1999; Hermans et al., 1998; Costa et al., 2003; see Hall, 2011 for a detailed review on cross-language effects with PWI). We, therefore, expect to observe significant facilitation effects in cross-language homophone conditions (i.e., CH_F and CH_R), compared with unrelated conditions (i.e., UN_F and UN_R) for both SC-English bilingual and monolingual speakers. Importantly, if lexical tone indeed shapes pitch processing in English spoken word production, the influence of falling vs. rising-tone SC homophone distractors on English picture naming is expected to differ between SC-English bilinguals and native English monolinguals. Furthermore, the "falling-f0 bias" of SC-English bilinguals, if at play in English spoken word production, would lead to processing differences between the two homophone (i.e., CH_F vs. CH_R) conditions.

5.1 Methodology

5.1.1 Participants

Forty-eight SC-English bilinguals (39 females and 9 males; average age 24, SD = 1.2) and 48 American English monolinguals (26 females and 22 males; average age 29, SD = 1.5) participated in this study. All SC-English bilingual participants were native SC speakers who grew up in Beijing and spoke no regional dialect. All participants started learning English at an average age of 5.8 (SD = 2.3). Before participating in this experiment, they all passed the College English Test Band 6 or scored above 6 in the International English Language Testing System (IELTS). Their English proficiency level was further accessed with an adapted LEAP-Q questionnaire (Marian et al., 2007) and the multilingual

naming test (MINT; Gollan et al., 2012). Using a Likert scale from one to ten, participants' self-rated frequency was 8.5 (SD = 1.4) in reading, 6.7 (SD = 1.8) in speaking, and 7.1 (SD = 1.8) in listening. The average correct response of MINT was 43% (SD = 5.1%). The English monolingual participants had no previous exposure to Mandarin or any other tone languages. All participants had no history of language disorder. This study was approved by the Ethics Committee at Leiden University Centre for Linguistics. All participants provided informed consent and were compensated for their participation.

5.1.2 Stimuli

There were 24 sets of critical stimuli (see Appendix D). Each set consisted of an English target word, an SC cross-language homophone distractor with a falling tone (CH_F), an SC cross-language homophone distractor with a rising tone (CH_R), an SC unrelated distractor with a falling tone (UN_F), and an SC unrelated distractor with a rising tone (UN_R). There were also 12 sets of filler words. All English targets were picturable monosyllabic nouns. All distractors were SC monosyllabic morphemes. Lexical frequency of distractors, as computed with SUBTLEX-CH (Cai & Brysbaert, 2010), was balanced across conditions [$F(3, 92) = 1.97$, $p = 0.13$]. Homophone density, as computed with DoWLS-MAN (Neergaard et al., 2022), was also controlled [$F(3, 92) = 0.855$, $p = 0.47$]. The target pictures, which were black and white line drawings, were selected from the IPNP database (Bates et al., 2003) and the BOSStimuli database (Brodeur et al., 2012). Five native Mandarin speakers who did not participate in the PWI experiments validated the choices of the target picture. All spoken stimuli were produced by a male native SC speaker (age 22) who was born and grew up in Beijing. The recording was done at the Phonetics Lab of Leiden University Centre for Linguistics through a Sennheiser MKH416T microphone (44.1 kHz, 16 bit). All stimuli were normalized for duration of 400 ms and intensity at 70 dB in Praat (Boersma & Weenink, 2022).

5.1.3 Procedure

Participants performed the experiment online using Gorilla (www.gorilla.sc). All participants were required to wear headphones and sit in a quiet room. Participants were only allowed to join the experiment if they were using laptops. Before the experiment, a headphone check procedure based on the dichotic pitch (Milne et al., 2020), as well as a microphone check and an auto-play check were run to screen participants' equipment and environment. All instructions were given in English.

Before the picture-naming task, there was a familiarization session. During the familiarization session, participants were shown 36 target pictures (24 critical and 12 filler targets) with their corresponding English names printed underneath for 1,500 ms. Afterwards, the name disappeared, and participants were asked to type in the picture's English name. If participants did not respond accurately, the intended name would be shown again.

In the PWI task, a fixation was displayed in the centre of the screen for 500 ms, followed by a picture and a simultaneously played SC spoken distractor (SOA = 0 ms). Participants were asked to name the picture as quickly and accurately as possible while ignoring the auditory distractor. The picture remained on the screen for 2,000 ms. Response time (hereafter RT) was measured from picture onset until naming onset using Chronset (Roux et al., 2017). If participants did not respond in 2,000 ms, the present trial ended, and the experiment proceeds automatically. Between each trial, there was a blank screen of 1,000 ms. Before starting the task, participants were asked to complete four practice trials with the option to practice more. In total, there were 96 (24 × 4) critical trials and 48 (12 × 4) filler trials. All trials were equally distributed into four blocks in a Latin Square design so that participants saw each target picture once in every block. Between each block, participants were encouraged to take a short break.

After the PWI task, participants were asked to complete a language background survey, the MINT test (Gollan et al., 2012), and a phonological

similarity rating task on targets and their cross-language homophone distractors. In total, the experiment took about 30 minutes.

5.2 Results

Trials with incorrect responses (~3.2%), empty responses (~2.9%) and unrecognizable responses (~2.6%) were excluded from the data analysis. Table 1 and Figure 1 summarize the mean RT for each condition. As we can see from Table 1, English monolingual participants took longer to name pictures with unrelated distractors (UN_R and UN_F) than with cross-language homophone distractors (CH_R and CH_F). Moreover, either with unrelated distractors or cross-language distractors, there was no significant difference between naming with falling-tone vs. rising-tone distractors. As for SC-English bilinguals, the overall naming latency was longer than for English monolinguals in each condition. Naming latencies with cross-language homophone distractors (CH_R and CH_F) were shorter than with unrelated distractors (UN_R and UN_F). While there was no naming latency difference between rising-tone vs. falling-tone unrelated distractors, there was an average difference of 22 ms between the rising-tone and falling-tone cross-language homophone distractors.

Table 1. *Mean RTs and SDs of SC-English bilingual speakers' and English monolingual speakers' naming latencies in each experimental condition.*

	SC-English Bilinguals		English Monolinguals	
	Mean	SD	Mean	SD
CH_R	797	218	725	179
CH_F	819	234	729	209
UN_R	852	251	763	210
UN_F	852	253	768	206
CH_F – CH_R	22		4	
UN_F – UN_R	0		5	

Table 2. *GLMM estimations of SC-English bilingual speakers' and English monolingual speakers' naming latencies.*

Conditions	Estimate	SE	t-value	*p*-value
Intercept	930.698	20.785	44.779	<0.001
English: CH_R vs. CH_F	10.347	8.049	1.285	0.596
English: UN_R vs. UN_F	-6.199	8.236	-0.753	0.903
English: UN_F vs. CH_F	50.741	8.318	6.100	<0.001
English: UN_R vs. CH_R	34.195	8.008	4.270	<0.001
Bilingual: CH_R vs. CH_F	-21.139	7.629	-2.771	0.022
Bilingual UN_R vs. UN_F	0.700	7.390	0.095	0.925
Bilingual: UN_R vs. CH_R	50.498	7.287	6.930	<0.001
Bilingual: UN_F vs. CH_F	28.659	7.916	3.620	0.002
UN_R: Bilingual vs. English	70.456	17.098	4.121	<0.001
UN_F: Bilingual vs. English	63.556	18.407	3.453	0.003
CH_R: Bilingual vs. English	54.152	18.578	2.915	0.018
CH_F: Bilingual vs. English	85.638	18.518	4.625	<0.001

Response times were analysed using the generalized linear mixed-effects model (GLMM) with inverse Gaussian distribution (Lo & Andrews, 2015). All the statistical analyses were run in R Studio (R Core Team, 2022) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Given that error rates were low in each condition, no further analysis on response accuracy was conducted. A maximum model including fixed effects of distractor type (CH_R, CH_F, UN_R and UN_F), participant groups (SC-English bilinguals and English monolinguals), the interaction between distractor type and group, by-subject and by-item random intercept, and by-subject and by-item random slopes for each fixed term were constructed first. Each term was then tested for exclusion. When the model failed to converge, we first increased the number of iterations and then simplified the model by removing correlation parameters in the random structures (Brauer & Curtin, 2018). The final GLMM consists of fixed effects of distractor type, the interaction between distractor type and group, and random intercepts for subject and item. As there was a significant interaction between participant group

and distractor type ($p < 0.05$), pairwise comparisons between group and distractors were also computed using the *multcomp* package (Hothorn et al., 2022). Holm–Bonferroni method was implemented to correct family-wise errors (Holm, 1979).

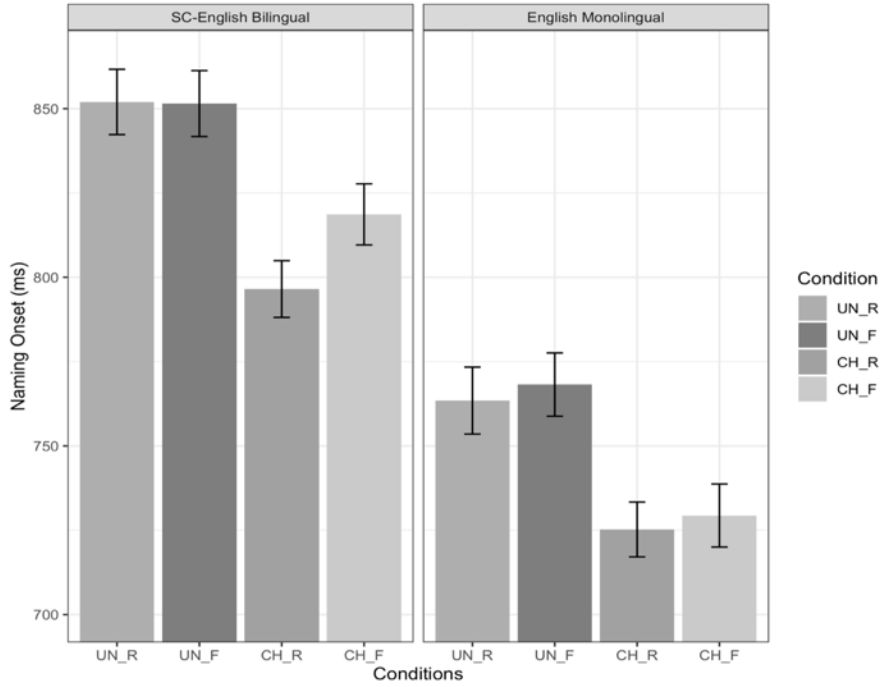

Figure 1. *Mean RTs and SDs of SC-English bilingual speakers' and English monolingual speakers' naming latencies in each experimental condition. In the CH_F condition, the distractor is the target's cross-language homophone with a falling lexical tone; in the CH_R condition, the distractor is the target's cross-language homophone with a rising lexical tone; in the UN_F condition, the distractor is an unrelated word with a falling lexical tone; in the UN_R condition, the distractor is an unrelated word with a rising lexical tone.*

According to the model estimations (see Table 2), both English monolinguals and SC-English bilinguals took longer to name targets with

unrelated distractor words than cross-language homophone distractors (English monolinguals: UR_R vs. CH_R, $p < 0.001$; UR_F vs. CH_F, $p < 0.001$; SC-English bilinguals: UR_R vs. CH_R, $p < 0.001$; UR_F vs. CH_F $p < 0.01$). This result suggests facilitatory effects of cross-language phonological overlap for both groups. As for differences between the rising- and falling-tone distractors, there was no significant difference between unrelated distractors (UN_R vs. UN_F: $p = 0.903$) and cross-homophone distractors (CH_R vs. CH_F: $p = 0.596$) for English monolinguals. Importantly, for SC-English bilinguals, there was a significant difference between cross-homophone distractors (CH_R vs. CH_F distractors: $p < 0.05$) but no significant difference between unrelated distractors (UN_R vs. UN_F: $p = 0.925$). This suggests that only SC-English bilinguals' naming latency, but not English monolinguals', was affected by the contrast between falling-tone and rising-tone cross-language homophones.

In sum, we found that for both SC-English bilinguals and English monolinguals, phonological similarity facilitated picture naming across languages. Moreover, while the naming latency of English monolinguals was not affected by the SC distractors' pitch contours, SC-English bilinguals were significantly faster when naming pictures with the falling-tone cross-language homophones than their rising-tone counterparts.

5.3 General Discussion

Although it is widely agreed that bilinguals' two languages interact extensively, there is limited finding on whether and how languages interact at the suprasegmental level. A full understanding of the bilingual mind should be backed up with data on whether and to what extent suprasegmental properties such as lexical tone plays a role in bilingual language processing. Our study aimed to fill in this gap by examining the effect of lexical tone on bilingual lexical access during English spoken word production. With PWI, both SC-English bilinguals and English monolinguals were asked to name pictures in English while ignoring

auditory SC distractors that were either cross-language homophones with the targets' English names or unrelated. Crucially, we manipulated the pitch contour of the distractors to be either rising (SC Tone 2) or falling (SC Tone 4). This was done to examine the so-called "falling-f0 bias" in SC-English bilinguals from the largely overlooked speech production domain. According to recent comprehension studies (Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2020), SC-English bilinguals represent non-tonal English words with a falling pitch contour similar to with a falling lexical tone, and it is therefore easier for them to access English words with a falling-f0 than their rising-f0 counterparts. If lexical tone indeed "reshapes" pitch representation in English lexical representations as Ortega-Llebaria and her colleagues claimed, we expected to find a significant difference between SC-English bilinguals and English monolinguals in their naming responses to the rising and falling-tone SC distractors. Our results showed that, regardless of the pitch shape difference, both SC-English bilinguals and English monolinguals took less time to name pictures with cross-language homophones than with unrelated distractors. Consistent with previous bilingual PWI studies (e.g., Costa & Caramazza, 1999; Hermans et al., 1998; Costa et al., 2003), this finding suggests that phonologically similar words (cross-language homophones in our case) facilitate picture naming across languages. Moreover, we identified the significant pitch processing difference between SC-English bilinguals and English monolinguals: while both rising and falling-tone cross-homophones were equally facilitative to picture naming in English monolinguals, SC-English bilinguals took significantly longer to name pictures with falling-tone cross-homophone distractors than their rising-tone counterparts.

The results of our speech production study align with previous comprehension studies (Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2020) in revealing a difference in pitch processing during the lexical access of a non-tonal language. However, upon closer inspection, there appears to be a contradiction in the findings. With primed lexical decision tasks, Ortega-Llebaria

et al. (2017) and Ortega-Llebaria & Wu (2020) found that SC-English bilinguals were significantly faster to make lexical decisions on words with a falling pitch contour than words with a rising pitch contour. However, in this study, the "falling-f0 bias" was reversed: SC-English bilinguals took significantly more time to name pictures with falling-tone cross-language homophones than their rising-tone counterparts. If a falling-f0 word is indeed a closer lexical representation in the bilingual lexicon than the corresponding rising-f0 word, one may expect the falling-tone cross-language homophones to facilitate picture naming more, in contrast to the rising-tone ones. In the following, we offer a few possible explanations for such a contrast.

First, the contrast might be coerced by task requirements. Though all focused on bilingual lexical access, Ortega-Llebaria et al. (2017) and Ortega-Llebaria and Wu (2020) looked into the process of spoken word recognition with a primed lexical decision task, while we examined spoken word production with PWI. It has been found that phonologically similar words generally cause inhibition in comprehension tasks (e.g., Dufour & Peereman, 2003; Magnuson et al., 2007), but facilitation in production tasks (e.g., Meyer, 1991; Meyer & Damian, 2007). According to the widely accepted interactive activation and competition framework (IAC; e.g., Chen & Mirman, 2012), both phonologically and semantically similar words join lexical competition; whether a certain word introduces facilitative or inhibitory processing depends on the task. As comprehension tasks are generally phonologically driven, phonologically similar words are often so strongly activated that they cause interference in lexical selection and therefore slow down lexical access; while production tasks are generally semantically driven, phonologically similar words are thus less activated and could potentially help speakers to overcome the ambiguity caused by automatically co-activated semantically similar words. It is thus plausible that the outcome of lexical tone on bilingual pitch processing manifests as reversed effects in production and comprehension tasks.

Alternatively, the contrast may reflect the presence vs. absence of the cross-language interference effect. While Ortega-Llebaria and her colleagues (Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2020) used English words of intonational pitch difference as the prime and target, we selected SC distractors with different lexical tones for a direct test of the tonal influence on English picture naming. Thus, besides phonologically introduced within-language activation, our study also involved the process of cross-language activation and competition. Previous studies have shown that word forms of bilinguals' two languages are co-activated during speech comprehension and production (see Kroll & Tokowicz, 2005 for a review). When the SC-English bilingual participants were naming pictures in English, they also had to resist the temptation of speaking SC. If the falling-pitch word form is indeed a closer lexical representation in the bilingual mental lexicon as Ortega-Llebaria and her colleagues proposed, it is plausible that falling-tone cross-language homophone distractors cause more cross-language interference than their rising-tone counterparts at phonological and phonetic encoding stages. The relatively larger effect of cross-language interference may cause the falling-tone cross-language homophone distractors to be less facilitative than their rising-tone counterparts.

A third possibility is that the contrast might be attributed to the robust acoustic saliency of the rising pitch contour. In the study by Ortega-Llebaria and Wu (2020), the authors not only identified a "falling-f0 bias" in lexical access among SC-English bilinguals but also revealed a "rising-f0 advantage" during non-word processing. Specifically, SC-English bilinguals were faster in detecting rising-f0 English non-words than their falling-f0 counterparts. Ortega-Llebaria and Wu reasoned that such a "rising-f0 advantage" is due to the greater acoustic saliency of the rising pitch contour than the falling contour. Moreover, the observation that only SC-English bilinguals exhibit this advantage indicates that native tonal language listeners might possess heightened sensitivity to acoustic saliency in pitch at the pre-lexical stage. In a similar vein, in the present study, the greater acoustic saliency of the rising pitch contour likely promoted swifter

responses to the cross-language homophone distractors with a rising tone. Consequently, this could have expedited word production at the pre-lexical level compared to their corresponding falling-tone distractors, which in turn facilitated the process of spoken word production.

It is worth noting that the three possibilities may not be mutually exclusive; the task requirements of PWI, the robust cross-language interference effect introduced by SC distractors, and the greater acoustic salience of the rising pitch contour may have played an interactive role, resulting in a relatively less facilitative effect of the falling-tone cross-language homophones in comparison with their rising-tone counterparts during English spoken word production. Further research is needed to investigate these possibilities.

As mentioned earlier, one of the most widely accepted assumptions in the bilingual literature is that bilinguals' two languages interact with each other extensively (see Kroll & Tokowicz, 2005 for a review). However, most evidence for this assumption came from studies on segmental processing. Only a limited number of studies examined whether language co-activation and lexical access were influenced by suprasegmental properties such as lexical tone (e.g., Shook & Marian, 2016; Wang et al., 2017; Ortega-Llebaria & Wu, 2020). While previous studies demonstrated the significant role of lexical tone in non-tonal spoken word recognition, our study offers complementary evidence regarding the role of lexical tone in non-tonal spoken word production. This contribution further enhances our understanding of the way language-specific suprasegmental features interact in bilingual lexical access.

In sum, this study found that, compared with unrelated distractors, SC cross-language homophones significantly facilitate English picture regardless of their pitch contour. SC-English bilinguals were less facilitated by SC cross-language homophones with a falling tone than with a rising tone. Such a difference in response to the falling and rising pitch contour contrast was not observed in native monolingual English speakers. Consistent with previous findings in the comprehension domain (Ortega-Llebaria et al., 2017; Ortega-

Llebaria & Wu, 2020), our findings show that, with falling-pitch SC homophones, SC-English bilinguals take longer to name pictures in English than with their rising counterparts. This indicates a significant influence of lexical tone on pitch processing during spoken word production even in bilinguals' non-tonal language. Not only does this study provide important complementary evidence for the role of lexical tone in pitch representation and processing, but it also helps develop a more comprehensive account of the bilingual mental lexicon.

Chapter 6

General Discussion

This dissertation investigates the process of spoken word recognition and spoken word production in native speakers of Standard Chinese, bi-dialectals of Standard Chinese and Xi'an Mandarin, and bilinguals of Standard Chinese and English. While most previous studies on lexical processing focused on the use of segmental information, this dissertation provides important complementary evidence with data on suprasegmental information, i.e., lexical tone, which can help us develop a more comprehensive account of (bilingual) lexical access. In the following sections, the findings of each chapter, the general implications of the findings and future directions are summarized.

6.1 Chapter-by-chapter Summary

Chapter 2 aimed to resolve three controversial issues in Mandarin spoken word recognition: 1) Do segmental syllables have a special status in Mandarin lexical processing? 2) What are the relative contributions of onset, rhyme, and lexical tone? 3) What is the time course of segmental and tonal processing during online lexical processing? To address these questions, three eye-tracking visual world paradigm experiments were conducted. Experiments 1 and 2 examined the relative contribution of the segmental syllable, onset, rhyme, and lexical tone in Mandarin lexical processing by investigating to what extent participants' visual attention is distracted by the presence of competitors during the process of recognizing the target spoken word. Critically, five types of competitors were manipulated based on their phonological overlap with the target, namely, segmental syllable competitors (with segmental syllable overlap), cohort competitors (with the onset and lexical tone overlap), rhyme competitors (with rhyme and lexical tone overlap), tonal competitors (with lexical tone overlap), and unrelated distractors (with no overlap). While Experiment 1 allowed participants to preview the pictures for 1,500 ms before listening to the target word, Experiment 2 had a shorter preview of 200 ms. Both experiments found that only segmental syllable competitors significantly distracted participants' visual

attention towards the target word more than unrelated distractors. Cohort competitors, rhyme competitors, and tonal competitors did not affect participants' visual attention more than unrelated distractors. Experiment 3 zoomed further into listeners' sensitivity to the acoustic details of segmental and tonal information. The target and competitor differed in either segmental or tonal information. Moreover, we manipulated the point of information divergence (early vs. late) between the target and competitor word pair along both segmental and tonal dimensions. Specifically, while the tonal early diverging target and competitor only share the same segments, the tonal late diverging pair share the same segment and the onset of the tonal pitch contours; while the segmental early diverging target and competitors share the onset and lexical tone, the segmental early diverging pair share the onset, glide, and lexical tone. Eye-tracking results show that, while both tonal early and late diverging competitors significantly attracted participants' visual attention, the late competitors exhibited significantly larger effects than the early competitors. Moreover, no statistically significant difference was found between tonal and segmental competitors, regardless of the point of divergence. In sum, the results of Experiments 1 and 2 indicate an advantageous role of segmental syllable over onset, rhyme, and lexical tone in activating word candidates; Experiment 3 shows that both tonal and segmental information can be used as soon as they are available to constrain word candidates' activation during the process of Mandarin spoken word recognition.

In Chapter 3, we further questioned whether and to what extent two tonal systems interact in listeners of two closely related tonal dialects. To answer these questions, we investigated the process of spoken word recognition with bi-dialectal speakers of Standard Chinese and Xi'an Mandarin. With the visual world paradigm, Standard Chinese and Xi'an Mandarin bi-dialectal speakers were asked to listen to short sentences produced in either Standard Chinese or Xi'an Mandarin (e.g., *wo3 yao4 shuo1 hua1*; "I will say flower") and identify the target word (e.g., *hua1* "flower") among four Chinese characters shown on the computer screen. The four characters included the target, two unrelated distractors, and a

phonological competitor. All phonological competitors share the same segmental syllable with the target within- and cross-dialects. Among the phonological competitors, there were cross-dialect homophone competitors that share the same lexical tone with the target across dialects (Homophone Condition), translation-induced cross-dialect homophones that share the same lexical tone with the targets' dialectal translation equivalent (Translation Condition), and competitor that does not share lexical tone with the target either within- or cross-dialects (Segment Condition). We hypothesized that, if both sets of lexical tones are activated, (translation-induced) cross-dialect homophones would elicit larger competition effects than competitors that have segmental overlap with targets only (the Segment Condition). Results of Standard Chinese and Xi'an Mandarin bi-dialectals' eye movements showed that, regardless of listening in either Standard Chinese or Xi'an Mandarin, neither the Homophone nor the Translation Condition distracted participants' eye fixations more than the Segment Condition competitors. Rather, the Segment Condition exhibited a larger phonological competition effect than the Homophone and Translation Conditions due to larger tonal similarities between the target and competitor word pairs within one dialect. Overall, this finding suggests a lack of lexical co-activation across dialects of Xi'an Mandarin and Standard Chinese. Given that previous studies on bilingualism have found consistent evidence that bilinguals co-activate both their languages during spoken word recognition with similar experimental set-ups (e.g., Spivey & Marian, 1999; Shook & Marian, 2017; Wang, Wang & Malins, 2017), our results indicate a lexical processing divergence between bilingual and bi-dialectal speech comprehension. Based on these findings, we proposed a preliminary bi-dialectal spoken word recognition model that emphasises the dialect control mechanism.

In Chapter 4, we continued to explore the role of lexical tone in bilingual spoken word production. Specifically, we asked whether Standard Chinese and English bilingual speakers co-activate lexical tone even when producing an English spoken word. With picture-word interference experiments (Rosinski et al.,

1975), Standard Chinese and English bilingual speakers were asked to name pictures in English (e.g., feather) while ignoring four types of simultaneously presented Standard Chinese distractors: 1) the translation distractor, which was the translation equivalent of the English target name (e.g., *yu3mao2* "feather"); 2) the tone-sharing distractor, which shares both tone and segments with the SC translation in the first syllable (e.g., *yu3zhou4* "universe"); 3) the no-tone-sharing distractor, which shares segments only with the SC translation in the first syllable (e.g., *yu4mi3* "corn"); 4) the unrelated distractor, which shares no phonological overlap with target and its translation (e.g., *lei4shui3* "tear"). Moreover, we manipulated the distractor modality and the familiarization mode before the naming task. Specifically, Standard Chinese distractors were presented auditorily in Experiments 1 and 2, but visually in Experiments 3 and 4. Before performing the naming task, bilinguals were familiarized with the target pictures' English names in Experiments 1 and 3, whereas both English and Standard Chinese names in Experiments 2 and 4. Results in Experiment 1 (auditory distractor and English mode) showed that translation distractors significantly facilitated target picture naming but tone-sharing distractors significantly interfered with target picture naming. Moreover, there was a significant naming latency difference between the tone-sharing and no-tone-sharing distractors, demonstrating the co-activation of lexical tone during English spoken word production. In Experiment 2 (auditory distractor and mixed mode), translation and phono-translation distractors all elicited interference towards target naming, but none of the effects was statistically significant. In Experiment 3 (visual distractor and English mode), there was a robust translation facilitation effect; the tone-sharing and no-tone-sharing distractors were also found to be facilitatory, but neither effect was statistically significant. In Experiment 4 (visual distractor and mixed mode), there was also a robust translation facilitation effect; the tone-sharing distractors significantly facilitated picture naming whereas the no-tone-sharing distractors did not, indicating an important role of lexical tone during English picture naming. Overall, replicating previously identified translation facilitation effects (e.g.,

Costa et al., 1999), this study discovers a significant difference between the tone-sharing and no-tone-sharing conditions. These findings suggest that Standard Chinese and English bilinguals not only co-activate the Standard Chinese translation equivalents but also the lexical tones of the Standard Chinese translations during English spoken word production. Moreover, the polarity and robustness of the lexical tone effect in spoken word production are modulated by procedural factors such as the distractor modality and the familiarization mode.

In Chapter 5, we investigated the influence of lexical tone on pitch representation and processing during non-tonal word production with Standard Chinese and English bilingual speakers. With the picture-word interference paradigm, we asked Standard Chinese and English bilinguals and native English monolinguals to name pictures in English (e.g., *lung*) while ignoring simultaneously played Standard Chinese cross-language homophones that either have a falling or a rising lexical tone (*lang* with a falling tone, “wave”; *lang* with a rising tone, “wolf”). We hypothesized that if lexical tone indeed affects bilinguals’ pitch representation in L2 non-tonal languages, the effect of lexical tone (falling vs. rising) on English picture naming should differ between Standard Chinese and English bilinguals and English monolingual speakers. Naming onset results show that, while both falling and rising cross-language homophones facilitated English word naming, only bilinguals of Standard Chinese and English, but not English monolingual speakers, showed significantly longer naming latencies with falling-tone cross-language homophones than their rising-tone counterparts. Such a distinction between Standard Chinese and English bilinguals and English monolinguals suggests that lexical tone plays an important role in pitch representation and processing during bilingual non-tonal spoken word production.

6.2 Theoretical Implications

Drawing empirical evidence from native Standard Chinese speakers, bi-dialectal speakers of Standard Chinese and Xi'an Mandarin, and bilingual speakers of Standard Chinese and English, our findings on spoken word recognition and production highlight the role of lexical tone during word co-activation and competition within- and across-languages. Current models of (bilingual) spoken word recognition and production should be adjusted to account for the possibilities of tonal processing. Moreover, findings on "tonal bilinguals" (Wu, 2015) of two closely related dialects should also be taken into consideration in our understanding of how the two linguistic systems of a speaker may interact in dynamically different ways.

Although previous studies have made a few attempts to modify current models to account for tonal word recognition (e.g., Ye & Connie, 1999; Yue, 2016; Gao et al., 2019; Shuai & Malins, 2017; Tong et al., 2014; Zhao et al., 2011), they disagree on whether an extra level of a syllable or segmental syllable is necessary (e.g., Zhao et al., 2011; Yue, 2016; Gao et al., 2019), and whether segment and tone processing are integrated (e.g., the TTRACE model; Tong et al., 2014) or separated (e.g., the TRACE-T model; Shuai & Malins, 2017). Given that in Chapter 2, we found an advantageous role of segmental syllables over sub-lexical constitutes such as onset and rhyme, and that both tonal and segmental information were used incrementally in Mandarin spoken word recognition, we proposed a revised TRACE model for tonal word recognition with a four-layer structure: syllable, segmental syllable, phonemes, and lexical tone. The extra level of segmental syllable accounts for the overall larger and more stable phonological competition effects of segmental syllable over a combination of sub-lexical phonological components during Mandarin spoken word recognition. Moreover, with independent representations of phonemes and tones, both phonemic and tonal information can be used to resolve phonological competition as soon as they are available.

While findings in Chapter 2 provide important implications for models of spoken word recognition in a lexical tone language, our findings on bilingual spoken word production in Chapters 4 and 5 further shed light on our understanding of the role of lexical tone in spoken word production. In general, current models of spoken word production either support the early active selection of lexical tone (e.g., Wan & Jaeger, 1998; Alderete et al., 2019) or not (e.g., Roelofs, 2015). According to the former view, the lexical tone is represented independently at the same operational level of segments and can be actively selected at an early stage of phonological encoding (Alderete et al., 2019). The latter view adapts the influential WEAVER++ model (Roelofs, 2000) to Mandarin in positing that lexical tone is represented diacritically as a metric frame, and only associated with pre-selected segments at a late stage of phonetic spell-out. In Chapter 4, we found that even during non-tonal English word production, lexical tone is co-activated and plays an important role in cross-language competition and selection, which is unlikely the case if lexical tone is only implemented during phonetic spell-out. In Chapter 5, we found further evidence that lexical tone directly modulates pitch processing in English spoken production, which in itself argues for a more influential role of lexical tone than mere labels in the metric frame. Overall, our findings strengthen the independent view of lexical tone selection in spoken word production from a bilingual lexical access perspective. Simultaneously, they validate the discrete and interactive processing of sub-lexical and lexical representations in speech production from a broader perspective on lexical access.

As previous studies either focused on bilingual or monolingual language processing, the nature of bi-dialectalism has been left largely unresearched. It has been debated whether bi-dialectal language processing should resemble that of bilinguals or monolinguals (see Melinger, 2018 for a review of the two views of bi-dialectalism). Chapter 3 focused on speakers of two closely related tonal dialects, Standard Chinese and Xi'an Mandarin, which provide important implications for our understanding of bi-dialectalism and lexical tone interaction.

Based on the finding that, unlike bilinguals, bi-dialectals of Standard Chinese and Xi'an Mandarin were able to achieve selective access to the target dialect, we proposed a spoken word recognition model for bi-dialectals. This model includes levels of phonological, phono-lexical, ortho-lexical, and semantic representations; within each level, dialect-specific and dialect-shared features are stored in the same space, allowing communication and competition between dialects. Moreover, there is a task scheme in the model that functions as a dialectal membership tag. By making use of the environment and task requirements, the task scheme can suppress the entire lexicon of the non-target dialect. This basic framework of a preliminarily verbal model is inspired by previous bilingual comprehension models such as BLINCS (Shook & Marian, 2013), BIA (e.g., Grainger & Dijkstra, 1992; Dijkstra et al., 1998), BIA+ (Dijkstra & Heuven, 2002) and further extends to account for bi-dialectal lexical processing.

6.3 Methodological Contributions

This dissertation also has a few methodological contributions to the implementation of the visual world paradigm and the picture-word interference paradigm.

Two factors in the visual world paradigm have been long questioned regarding the extent to which they affect the eye-tracking results (see reviews in Huettig et al., 2011; Apfelbaum et al., 2021). One is the length of the preview time, i.e., the time allowed for participants to view the pictures on screen before listening to the auditory stimuli. It has been proposed that a short preview time such as 200 ms is not sufficient for participants to retrieve the names of the displayed pictures and thus results in null phonological competition (e.g., Huettig & McQueen, 2007); however, there have also been studies that found effects of phonological competition even without any preview time (e.g., Zou, 2017). In Chapter 2, with a short preview time of 200 ms, we replicated results obtained with a long preview time of 1,500 ms in finding a robust segmental syllable

phonological competition effect. Besides, we found that listeners located the target picture faster with fewer eye fixations with the short preview time than with the long one, along with subtle differences in the time course of eye fixations. These findings suggest that although the length of preview time is not a determining factor for observing phonological competition, it does affect how listeners distribute their visual attention. The other factor is the use of Chinese characters as visual displays. While previous studies have validated using printed words to study spoken word recognition in Western languages (Huettig et al., 2011), it is less clear to what extent the display of Chinese characters affects Mandarin lexical processing in the visual world paradigm. In Chapter 2, we found a subtle trend of cohort competition with the display of Chinese characters, which was missing from the picture display. This discovery indicates that the incorporation of Chinese characters into the picture world paradigm not only confirms its validity but also demonstrates an increased sensitivity to phonological effects.

With the picture word interference paradigm, we also manipulated two procedural factors. One is the modality of the distractors, i.e., whether participants listened to or viewed distractor words during picture naming (auditory vs. visual distractors). The other is the familiarization mode of the target pictures, i.e., whether participants were given English names only (i.e., the English mode) or both English and SC names (i.e., the mixed mode) during the familiarization session. While both factors have been found to influence the naming process in the picture world interference paradigm (e.g., Hantsch et al., 2009; Llorens et al., 2014; Jonen et al., 2021), it has not been answered whether and if so, to what extent the two factors affect cross-language co-activation and competition in bilingual studies. For instance, it is unknown whether distractor modality might be responsible for causing opposite effects of translation and phono-translation distractors, given that previous studies mostly observed translation facilitation with visual distractors (e.g., Costa et al., 1999; Costa & Caramazza, 1999; Hermans, 2004) and phono-translation interference with auditory distractors (e.g.,

Hermans et al., 1998; Costa et al., 2003). Our findings in Chapter 4 found that, while both auditory and visual translation distractors could facilitate bilingual picture naming, only auditory phono-translation distractors, but not their visual counterparts, significantly interfered with the process. The familiarization mode was also found to have an impact on the effect of language co-activation: familiarizing target pictures' names in both languages significantly reduced the translation facilitation effect of auditory distractors, compared with familiarizing names in the target language alone. Future spoken word production studies, therefore, should take both factors into account when interpreting the direction and robustness of the cross-language effects within the (bilingual) picture-word interference paradigm. Moreover, Chapters 4 and 5 were conducted online due to the influence of COVID-19. By successfully replicating previous lab findings such as the translation facilitation effect (e.g., Costa et al., 1999), we showed that implementing the picture-word interference paradigm online is an efficient and sound approach to studying the process of spoken word production.

6.4 Limitations and Future Directions

First, further research is needed to investigate the relationship between tonal word recognition and production in Mandarin. Our findings in Chapter 2 indicate that segmental syllables play a primary role in spoken word recognition, which is consistent with previous studies on Mandarin word production (Meyer, 1991; Chen, Lin, & Ferrand, 2003; Chen & Chen, 2013; Chen, O'Seaghdha & Chen, 2016; Wang, Wong, & Chen, 2018). In our proposed revised TRACE model, we incorporated an extra level of segmental syllable and independent representation of lexical tones, which aligns with the idea in speech production that the atonal syllable is "the proximate phonological encoding" (O'Seaghdha, 2010; Roelofs, 2015). However, potential asymmetries and differential engagement of segmental syllables and lexical tone in tonal word perception and

production remain unexplored. Therefore, further studies are needed to address these questions.

Second, in Chapter 3, we proposed a model of bi-dialectal lexical access that emphasizes potential differences in the control mechanism between bi-dialectals and bilinguals. We also hypothesized that the differences between bi-dialectal and bilingual lexical access may vary along a continuum, depending on the degree of similarities between the dialects. However, further studies on other dialects and pairs of languages with typologically different prosodic systems are still needed to validate our models and to explore the extent to which bilingual and bi-dialectal lexical processing differ.

Third, although we explained our findings in spoken word recognition and production within the framework of interactive activation and competition (e.g., McClelland & Rumelhart, 1981; Chen & Mirman, 2012), it is worth noting that other theories, such as the response exclusion hypothesis (Mahon et al., 2007) may also account for our findings without referring to lexical competition. In a similar vein, although we mainly explained our bilingual findings within the framework of language co-activation, it is important to note that the learning account (e.g., Costa et al., 2017), which does not resort to language co-activation, may also explain some of our findings. Although the question of whether lexical selection depends on co-activation and competition is a compelling research topic, we did not delve into it further in this dissertation for lack of judicating evidence in our results.

Last but not least, we proposed various modifications of current models and theories of lexical processing to account for tonal word recognition and production, as well as bi-dialectal lexical access. However, our models and suggestions have remained verbal and preliminary. To validate our models, computational simulations and additional empirical evidence from various tonal languages and dialects are necessary.

Reference

Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition*, *191*, 103952. https://doi.org/10.1016/j.cognition.2019.04.021

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the Time Course of Spoken Word Recognition Using Eye Movements: Evidence for Continuous Mapping Models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Ching Lu, C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., … Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344–380. https://doi.org/10.3758/BF03196494

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660. https://doi.org/10.1080/01690960601000746

Bordag, D., Gor, K., & Opitz, A. (2022). Ontogenesis Model of the L2 Lexical Representation. *Bilingualism: Language and Cognition*, *25*(2), 185–201. https://doi.org/10.1017/S1366728921000250

Boukadi, M., Davies, R. A. I., & Wilson, M. A. (2015). Bilingual lexical selection as a dynamic process: Evidence from Arabic-French bilinguals. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *69*(4), 297–313. https://doi.org/10.1037/cep0000063

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

Brodeur, M. B., Kehayia, E., Dion-Lessard, G., Chauret, M., Montreuil, T., Dionne-Dostie, E., & Lepage, M. (2012). The bank of standardized stimuli (BOSS): Comparison between French and English norms. *Behavior Research Methods*, *44*(4), 961–970. https://doi.org/10.3758/s13428-011-0184-7

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles. *PLOS ONE*, *5*(6), e10729. https://doi.org/10.1371/journal.pone.0010729

Chen, J.-Y., Chen, T.-M., & Dell, G. S. (2002). Word-Form Encoding in Mandarin Chinese as Assessed by the Implicit Priming Task. *Journal of Memory and Language*, 46(4), 751–781. https://doi.org/10.1006/jmla.2001.2825

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417–430. https://doi.org/10.1037/a0027175

Chen, Y. (2022). Tone and Intonation. In C.-R. Huang, Y.-H. Lin, I.-H. Chen, & Y.-Y. Hsu (Eds.), *The Cambridge Handbook of Chinese Linguistics*. Cambridge University Press.

Clopper, C. G. (2014). Sound change in the individual: Effects of exposure on cross-dialect speech processing. *Laboratory Phonology*, *5*(1). https://doi.org/10.1515/lp-2014-0004

Clopper, C. G. (2021). Perception of Dialect Variation. In *The Handbook of Speech Perception* (pp. 333–364). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119184096.ch13

Clopper, C. G., & Walker, A. (2017). Effects of Lexical Competition and Dialect Exposure on Phonological Priming. *Language and Speech*, *60*(1), 85–109. https://doi.org/10.1177/0023830916643737

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Colomé, À. (2001). Lexical Activation in Bilinguals' Speech Production: Language-Specific or Language-Independent? *Journal of Memory and Language*, *45*(4), 721–736. https://doi.org/10.1006/jmla.2001.2793

Connell, K. S. (2017). The Use of Segmental And Suprasegmental Information In Lexical Access: A First- And Second-Language Chinese Investigation [University of Kansas]. https://kuscholarworks.ku.edu/handle/1808/26001

Cook, S. V., & Gor, K. (2015). Lexical access in L2. *Mental Lexicon*, *10*(2), 247–270. https://doi.org/10.1075/ml.10.2.04coo

Costa, A. (2009). Lexical Access in Bilingual Production. In J. F. Kroll & A. M. B. D. Groot, *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 308–325). Oxford University Press.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. *Bilingualism: Language and Cognition*, *2*(3), 231–244.

Costa, A., Colomé, À., Gómez, O., & Sebastián-Gallés, N. (2003). Another look at cross-language competition in bilingual speech production: Lexical and phonological factors. *Bilingualism: Language and Cognition*, *6*(3), 167–179. https://doi.org/10.1017/S1366728903001111

Costa, A., Mahon, B., Savova, V., & Caramazza, A. (2003). Level of categorisation effect: A novel effect in the picture-word interference paradigm. *Language and Cognitive Processes*, *18*(2), 205–234. https://doi.org/10.1080/01690960143000524

Costa, A., Miozzo, M., & Caramazza, A. (1999a). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection? *Journal of Memory and Language*, *41*(3), 365–397. https://doi.org/10.1006/jmla.1999.2651

Costa, A., Miozzo, M., & Caramazza, A. (1999b). Lexical Selection in Bilinguals: Do Words in the Bilingual's Two Lexicons Compete for Selection? *Journal of Memory and Language*, *41*(3), 365–397. https://doi.org/10.1006/jmla.1999.2651

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. https://doi.org/10.1075/ml.8.3.06dar

Darcy, I., & Thomas, T. (2019). When *blue* is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners.

Bilingualism: Language and Cognition, *22*(5), 1141–1159. https://doi.org/10.1017/S1366728918001050

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dijkstra, T., & Heuven, W. J. B. van. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Van Heuven, W. J. B., & Grainger, J. (1998). Simulating cross-language competition with the bilingual interactive activation model. *Psychologica Belgica*, *38*(3–4), 177–196.

Dufour, S., & Peereman, R. (2003). Inhibitory priming effects in auditory word recognition: When the target's competitors conflict with the prime word. *Cognition*, *88*(3), B33–B44. https://doi.org/10.1016/S0010-0277(03)00046-5

Gao, X., Yan, T.-T., Tang, D.-L., Huang, T., Shu, H., Nan, Y., & Zhang, Y.-X. (2019). What Makes Lexical Tone Special: A Reverse Accessing Model for Tonal Speech Perception. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02830

Giegerich, H. J. (1992). *English Phonology: An Introduction* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139166126

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of Spoken Language Dominance: A Multi-Lingual Naming Test (MINT) and Preliminary Norms for Young and Aging Spanish-English Bilinguals. *Bilingualism (Cambridge, England)*, *15*(3), 594–615. https://doi.org/10.1017/S1366728911000332

Grainger, J., & Dijkstra, T. (1992). On the Representation and Use of Language Information in Bilinguals. In *Advances in Psychology* (Vol. 83, pp. 207–220). Elsevier. https://doi.org/10.1016/S0166-4115(08)61496-X

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149. https://doi.org/10.1017/S136672899800025X

Grosjean, F. (2001). The bilingual's language modes. *One Mind, Two Languages: Bilingual Language Processing*, *122*.

Hantsch, A., Jescheniak, J. D., & Schriefers, H. (2009). Distractor modality can turn semantic interference into semantic facilitation in the picture–word interference task: Implications for theories of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(6), 1443–1453. https://doi.org/10.1037/a0017020

Hazen, K. (2001). An introductory investigation into bidialectalism. *University of Pennsylvania Working Papers in Linguistics*, *7*(3), 8.

Hermans, D. (2004). Between-language identity effects in picture-word interference tasks: A challenge for language-nonspecific or language-specific models of lexical access? *International Journal of Bilingualism*, *8*(2), 115–125. https://doi.org/10.1177/13670069040080020101

Hermans, D., Bongaerts, T., Bot, K. D., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*(3), 213–229. https://doi.org/10.1017/S1366728998000364

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Hommel, G. (1988). A stagewise rejective multiple test procedure based on a modified Bonferroni test. *Biometrika*, *75*(2), 383–386. https://doi.org/10.1093/biomet/75.2.383

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*(4), 785–813. https://doi.org/10.3758/BF03196544

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11. https://doi.org/10.1016/j.jml.2017.09.002

Jafari, M., & Ansari-Pour, N. (2019). Why, When and How to Adjust Your P Values? *Cell Journal (Yakhteh)*, *20*(4), 604–607. https://doi.org/10.22074/cellj.2019.5992

Jescheniak, J. D., & Levelt, W. J. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 824.

Jescheniak, J. D., & Schriefers, H. (2001). Priming Effects from Phonologically Related Distractors in Picture—Word Interference. *The Quarterly Journal of Experimental Psychology Section A*, *54*(2), 371–382. https://doi.org/10.1080/713755981

Jonen, M., Heim, S., Grünert, M., Neuloh, G., & Sakreida, K. (2021). Adaptation of a semantic picture-word interference paradigm for future language mapping with transcranial magnetic stimulation: A

behavioural study. *Behavioural Brain Research*, *412*, 113418. https://doi.org/10.1016/j.bbr.2021.113418

Klaus, J., Lemhöfer, K., & Schriefers, H. (2018). The second language interferes with picture naming in the first language: Evidence for L2 activation during L1 production. *Language, Cognition and Neuroscience*, *33*(7), 867–877. https://doi.org/10.1080/23273798.2018.1430837

Köhnlein, B., Dickerson, C., Leow, J., & Chávez, P. P. (2019). Lexical tone or foot structure in Hong Kong English? A response to Lian-Hee Wee. *Language*, *95*(3), e394–e405. https://doi.org/10.1353/lan.2019.0065

Kroll, J. F., Dussias, P. E., Bogulski, C. A., & Kroff, J. R. V. (2012). Chapter Seven - Juggling Two Languages in One Mind: What Bilinguals Tell Us About Language Processing and its Consequences for Cognition. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 56, pp. 229–262). Academic Press. https://doi.org/10.1016/B978-0-12-394393-4.00007-8

Kroll, J. F., Sumutka, B. M., & Schwartz, A. I. (2005). A cognitive view of the bilingual lexicon: Reading and speaking words in two languages. *International Journal of Bilingualism*, *9*(1), 27–48. https://doi.org/10.1177/13670069050090010301

Kroll, J. F., & Tokowicz, N. (2005). *Models of bilingual representation and processing: Looking back and to the future.* Oxford University Press.

Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*, *62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

La Heij, W. (1988). Components of Stroop-like interference in picture naming. *Memory & Cognition*, *16*(5), 400–410. https://doi.org/10.3758/BF03214220

Labov, W. (1998). Co-existent systems in African-American vernacular English. *African-American English: Structure, History and Use*, 110–153.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013a). Interlingual lexical competition in a spoken sentence context: Evidence from the visual world paradigm. *Psychonomic Bulletin & Review*, *20*(5), 963–972. https://doi.org/10.3758/s13423-013-0405-4

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013b). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language and Cognition*, *16*(3), 508–517. https://doi.org/10.1017/S1366728912000508

Lee, C.-Y. (2007). Does Horse Activate Mother? Processing Lexical Tone in Form Priming. *Language and Speech*, *50*(1), 101–123. https://doi.org/10.1177/00238309070500010501

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38. https://doi.org/10.1017/S0140525X99001776

Liu, M. (2018). *Tone and intonation processing: From ambiguous acoustic signal to linguistic representation*. Leiden University.

Liu, M., Chen, Y., & Schiller, N. O. (2020). Tonal mapping of Xi'an Mandarin and Standard Chinese. *The Journal of the Acoustical Society of America*, *147*(4), 2803–2816. https://doi.org/10.1121/10.0000993

Llorens, A., Trébuchon, A., Riès, S., Liégeois-Chauvel, C., & Alario, F.-X. (2014). How familiarization and repetition modulate the picture naming network. *Brain and Language*, *133*, 47–58. https://doi.org/10.1016/j.bandl.2014.03.010

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01171

Lupker, S. J. (1984). Semantic priming without association: A second look. *Journal of Verbal Learning and Verbal Behavior*, *23*(6), 709–733. https://doi.org/10.1016/S0022-5371(84)90434-1

Magnuson, J. S. (2019). Fixations in the visual world paradigm: Where, when, why? *Journal of Cultural Cognitive Science*, *3*(2), 113–139. https://doi.org/10.1007/s41809-019-00035-3

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The Dynamics of Lexical Competition During Spoken Word Recognition. *Cognitive Science*, *31*(1), 133–156. https://doi.org/10.1080/03640210709336987

Mahon, B. Z., Costa, A., Peterson, R., Vargas, K. A., & Caramazza, A. (2007). Lexical selection is not by competition: A reinterpretation of semantic interference and facilitation effects in the picture-word interference paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 503–535. https://doi.org/10.1037/0278-7393.33.3.503

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language*, *62*(4), 407–420. https://doi.org/10.1016/j.jml.2010.02.004

Malins, J. G., & Joanisse, M. F. (2012). Towards a model of tonal processing during Mandarin spoken word recognition. *Tonal Aspects of Languages-Third International Symposium*.

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics*, *24*(02). https://doi.org/10.1017/S0142716403000092

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism:*

Language and Cognition, *6*(2), 97–115. https://doi.org/10.1017/S1366728903001068

Marian Viorica, Blumenfeld Henrike K., & Kaushanskaya Margarita. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

García, M. T. M. (2020). Language bias and proficiency effects on cross-language activation: A comprehension and production comparison. *Linguistic Approaches to Bilingualism*, 10(6), 873-901.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188. https://doi.org/10.1037/0096-3445.114.2.159

McGory, J. T. (1997). *Acquisition of intonational prominence in English by Seoul Korean and Mandarin Chinese speakers* [Doctoral dissertation, The Ohio State University]. https://www.proquest.com/docview/304410387/abstract/1DBB23D441204135PQ/1

Melinger, A. (2021). Do elevators compete with lifts?: Selecting dialect alternatives. *Cognition*, *206*, 104471. https://doi.org/10.1016/j.cognition.2020.104471

Meyer, A. S. (1991). The time course of phonological encoding in language production: Phonological encoding inside a syllable. *Journal of Memory and Language*, *30*(1), 69–89. https://doi.org/10.1016/0749-596X(91)90011-8

Meyer, A. S., & Damian, M. F. (2007). Activation of distractor names in the picture-picture interference paradigm. *Memory & Cognition*, *35*(3), 494–503. https://doi.org/10.3758/BF03193289

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 1146–1160. https://doi.org/10.1037/0278-7393.17.6.1146

Michael J. Spivey & Viorica Marian. (1999). Cross Talk Between Native and Second Languages: Partial Activation of an Irrelevant Lexicon.

Psychological Science, *10*(3), 281–284. https://doi.org/10.1111/1467-9280.00151

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2020). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*. https://doi.org/10.3758/s13428-020-01514-0

Mirman, D. (2017). *Growth curve analysis and visualization using R*. Chapman and Hall/CRC.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

Neergaard, K. D., Xu, H., German, J. S., & Huang, C.-R. (2022). Database of word-level statistics for Mandarin Chinese (DoWLS-MAN). *Behavior Research Methods*, *54*(2), 987–1009. https://doi.org/10.3758/s13428-021-01620-7

Nozari, N., & Pinet, S. (2019). *A critical review of the behavioural, neuroimaging, and electrophysiological studies of co-activation of representations during word production* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/x9wq7

Ortega-Llebaria, M., Nemogá, M., & Presson, N. (2017). Long-term experience with a tonal language shapes the perception of intonation in English

words: How Chinese–English bilinguals perceive “Rose?” vs. “Rose.” *Bilingualism: Language and Cognition*, *20*(2), 367–383. https://doi.org/10.1017/S1366728915000723

Ortega-Llebaria, M., & Wu, Z. (2020). Chinese-English Speakers’ Perception of Pitch in Their Non-Tonal Language: Reinterpreting English as a Tonal-Like Language. *Language and Speech*, 0023830919894606. https://doi.org/10.1177/0023830919894606

Ploquin, M. (2013). Prosodic Transfer: From Chinese Lexical Tone to English Pitch Accent. *Advances in Language and Literary Studies*, *4*(1), 68–77.

Qin, Z. (2017). How Native Chinese Listeners and Second-Language Chinese Learners Process Tones in Word Recognition: An Eye-tracking Study. https://kuscholarworks.ku.edu/handle/1808/26474

Qin, Z., Tremblay, A., & Zhang, J. (2019). Influence of within-category tonal information in the recognition of Mandarin-Chinese words by native and non-native listeners: An eye-tracking study. *Journal of Phonetics*, *73*, 144–157. https://doi.org/10.1016/j.wocn.2019.01.002

Roelofs, A. (2000). WEAVER++ and other computational models of lemma retrieval and word-form encoding. *Aspects of Language Production*, 71–114.

Roelofs, A. (2003). Shared phonological encoding processes and representations of languages in bilingual speakers. *Language and Cognitive Processes*, *18*(2), 175–204. https://doi.org/10.1080/01690960143000515

Roelofs, A. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism : Language and Cognition*, *9*(2), 167–176. https://doi.org/10.1017/S1366728906002513

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese: Modeling of phonological encoding. *Japanese Psychological Research*, *57*(1), 22–37. https://doi.org/10.1111/jpr.12050

Rosinski, R. R., Golinkoff, R. M., & Kukish, K. S. (1975). Automatic semantic processing in a picture-word interference task. *Child Development*, 247–253.

Roux, F., Armstrong, B. C., & Carreiras, M. (2017). Chronset: An automated tool for detecting speech onset. *Behavior Research Methods*, *49*(5), 1864–1881. https://doi.org/10.3758/s13428-016-0830-1

Schiller, N., & Alario, F. X. (2023). Models of language production and the temporal organization of lexical access. *Bilingualism through the Prism of Psycholinguistics*, 17, 28-53.

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. https://doi.org/10.1162/0898929052880057

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in language production: Picture-word

interference studies. *Journal of Memory and Language*, *29*(1), 86–102. https://doi.org/10.1016/0749-596X(90)90011-N

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech*, 58(2), 131-151.

Seth Wiener & Rory Turnbull. (2016). Constraints of Tones, Vowels and Consonants on Lexical Selection in Mandarin Chinese. *Language and Speech*, *59*(1), 59–82. https://doi.org/10.1177/0023830915578000

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *The Journal of the Acoustical Society of America*, *133*(5), 3016–3029. https://doi.org/10.1121/1.4795775

Sheng, L., Lu, Y., & Gollan, T. H. (2014). Assessing language dominance in Mandarin-English bilinguals: Convergence and divergence between subjective and objective measures. *Bilingualism (Cambridge, England)*, *17*(2), 364–383. https://doi.org/10.1017/S1366728913000424

Shook, A., & Marian, V. (2013). The Bilingual Language Interaction Network for Comprehension of Speech. *Bilingualism: Language and Cognition*, *16*(02), 304–324. https://doi.org/10.1017/S1366728912000466

Shook, A., & Marian, V. (2016). The influence of native-language tones on lexical access in the second language. *The Journal of the Acoustical*

Society of America, *139*(6), 3102–3109. https://doi.org/10.1121/1.4953692

Shook, A., & Marian, V. (2017). Covert co-activation of bilinguals' non-target language: Phonological competition from translations. *Linguistic Approaches to Bilingualism*. https://doi.org/10.1075/lab.17022.sho

Shook, A., & Marian, V. (2019). Covert co-activation of bilinguals' non-target language: Phonological competition from translations. *Linguistic Approaches to Bilingualism*, *9*(2), 228–252. https://doi.org/10.1075/lab.17022.sho

Shuai, Lan, and Jeffrey G. Malins. 2017. 'Encoding Lexical Tones in jTRACE: A Simulation of Monosyllabic Spoken Word Recognition in Mandarin Chinese'. Behavior Research Methods 49(1):230–41. doi: 10.3758/s13428-015-0690-0.

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, *60*(4), 487–501. https://doi.org/10.1016/j.jml.2009.01.001

Tanenhaus, M. K., Flanigan, H. P., & Seidenberg, M. S. (1980). Orthographic and phonological activation in auditory and visual word recognition. *Memory & Cognition*, *8*(6), 513–520. https://doi.org/10.3758/BF03213770

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken

language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Thierry, G., & Wu, Y. J. (2007). *Brain potentials reveal unconscious translation* during foreign-language comprehension. 6.

Tong, Xiuhong, Catherine McBride, Chia-Ying Lee, Juan Zhang, Lan Shuai, Urs Maurer, and Kevin K. H. Chung. 2014. 'Segmental and Suprasegmental Features in Speech Perception in Cantonese-Speaking Second Graders: An ERP Study'. Psychophysiology 51(11):1158–68. doi: 10.1111/psyp.12257.

Traxler, M. J. (2011). Word Processing. In *Introduction to Psycholinguistics: Understanding Language Science*. John Wiley & Sons.

Visceglia, T., & Fodor, J. D. (2006). Fundamental frequency in Mandarin and English: Comparing first- and second-language speakers. In C. Lleó (Ed.), *Hamburg Studies on Multilingualism* (Vol. 4, pp. 27–59). John Benjamins Publishing Company. https://doi.org/10.1075/hsm.4.03vis

Wan, I.-P., & Jaeger, J. (1998). Speech errors and the representation of tone in Mandarin Chinese. *Phonology*, *15*(3), 417–461. https://doi.org/10.1017/S0952675799003668

Wang, X., Wang, J., & Malins, J. G. (2017). Do you hear ' feather ' when listening to ' rain '? Lexical tone activation during unconscious translation: Evidence from Mandarin-English bilinguals. *Cognition*, *169*, 15–24. https://doi.org/10.1016/j.cognition.2017.07.013

Weber, A., & Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, *50*(1), 1–25. https://doi.org/10.1016/S0749-596X(03)00105-0

Wee, L.-H. (2016). Tone assignment in Hong Kong English. *Language*, *92*(2), e67–e87. https://doi.org/10.1353/lan.2016.0039

Wu, J., Chen, Y., Heuven, V. J. van, & Schiller, N. O. (2018). Dynamic effect of tonal similarity in bilingual auditory lexical processing. *Language, Cognition and Neuroscience*, *0*(0), 1–19. https://doi.org/10.1080/23273798.2018.1550206

Yiu, S. (2014). *Aspects of tone in Cantonese English* [The University of Hong Kong]. http://hub.hku.hk/handle/10722/240429

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, *49*(7), 1761–1770. https://doi.org/10.1016/j.neuropsychologia.2011.02.054

Zou, T. (2017). Production and perception of tones by Dutch learners of Mandarin. LOT, Netherlands Graduate School.

Appendix A

Table A1. Stimuli used in Experiment 1 & 2 in Chapter 2.

	Target	Segmental Competitor	Cohort Competitor	Rhyme Competitor	Tonal Competitor
1	*chuang2*	*chuang1*	*cha2*	*huang2*	*ya2*
	bed	window	tea	yellow	tooth
2	*li2*	*li4*	*lun2*	*qi2*	*men2*
	pear	chestnut	tire	flag	door
3	*hu3*	*hu2*	*hai3*	*tu3*	*er3*
	tiger	pot	ocean	dirt	ear
4	*jing4*	*jing1*	*ju4*	*xing4*	*xiang4*
	mirror	whale	saw	apricot	elephant
5	*shu3*	*shu1*	*shou3*	*gu3*	*wan3*
	mouse	book	hand	drum	bowl
6	*shan4*	*shan1*	*shu4*	*dan4*	*tu4*
	fan	mountain	tree	egg	rabbit
7	*tang2*	*tang1*	*tou2*	*fang2*	*sheng2*
	sugar	soup	head	house	rope
8	*bing1*	*bing3*	*bei1*	*xing1*	*yan1*
	ice	pie	cup	star	cigarette
9	*bao4*	*bao1*	*bu4*	*pao4*	*xin4*
	leopard	bag	cloth	cannon	envelope
10	*bi3*	*bi2*	*ban3*	*yi3*	*san3*
	pen	nose	board	chair	umbrella
11	*hua1*	*hua4*	*hei1*	*gua1*	*che1*
	flower	painting	black	melon	car
12	*mao4*	*mao1*	*mo4*	*yao4*	*suan4*
	hat	cat	ink	medicine	garlic

Table A2. Stimuli used in Experiment 3 in Chapter 2.

	Tonal Pair			
	Early POD		Late POD	
1	埋	卖	丑	仇
	mai2	*mai4*	*chou3*	*chou2*
	bury	sell	ugly	hatred
2	萌	梦	读	赌
	meng2	*meng4*	*du2*	*du3*
	cute	dream	read	bet
3	鸭	哑	哭	酷
	ya1	*ya3*	*ku1*	*Ku4*
	duck	dumb	cry	cool
4	影	鹰	龟	贵
	ying3	*ying1*	*gui1*	*gui4*
	ink	wipe	turtle	expensive
5	脑	闹	金	近
	nao3	*nao4*	*jin1*	*jin4*
	brain	noisy	gold	closer
6	圆	院	抢	墙
	yuan2	*yuan4*	*qiang3*	*qiang2*
	round	yard	rob	wall
7	网	旺	证	蒸
	wang3	*wang4*	*zheng4*	*zheng1*
	nest	thriving	certificate	steam
8	玉	鱼	摆	白
	yu4	*yu2*	*bai3*	*bai2*
	cigarette	eye	arrange	white
9	盐	艳	烦	反
	yan2	*yan4*	*fan2*	*fan3*
	sault	bright	upset	opposite
10	舞	雾	猪	住
	wu3	*wu4*	*zhu1*	*zhu4*
	dance	fog	pig	live
11	夜	野	摔	帅
	ye4	*ye3*	*shuai1*	*shuai4*
	night	wild	fall	handsome
12	秒	妙	猜	菜
	miao3	*miao4*	*cai1*	*cai4*
	second	nice	guess	vegetable

	Segment Pair			
	Early POD		Late POD	
13	爬	瓶	雪	选
	pa2	*ping2*	*xue3*	*xuan3*
	climb	bottle	snow	choose
14	塔	腿	铁	舔
	ta3	*tui3*	*tie3*	*tian3*
	tower	leg	iron	lick
15	纸	找	亲	轻
	zhi3	*zhao3*	*qin1*	*qing1*
	paper	look for	kiss	light
16	电	豆	琴	晴
	dian4	*dou4*	*qin2*	*qing2*
	electricity	bean	piano	sunny
17	方	风	睡	顺
	fang1	*feng1*	*shui4*	*shun4*
	square	wind	sleep	smooth
18	热	弱	宽	筐
	re4	*ruo4*	*kuan1*	*kuang1*
	hot	weak	wide	basket
19	呆	东	喘	闯
	dai1	*dong1*	*chuan3*	*chuang3*
	stupid	east	breath	rush
20	僧	酸	球	穷
	seng1	*suan1*	*qiu2*	*qiong2*
	monk	soar	ball	poor
21	丢	灯	坏	换
	diu1	*deng1*	*huai4*	*huan4*
	lost	light	broke	replace
22	慌	喝	家	姜
	huang1	*he1*	*jia1*	*jiang1*
	old	cold	home	ginger
23	烤	苦	乖	光
	kao3	*ku3*	*guai1*	*guang1*
	grill	pain	behave	light
24	雷	楼	追	钟
	lei2	*lou2*	*zhui1*	*zhong1*
	lightening	building	chase	clock

Appendix B

Table B1. Stimuli used in the Standard Chinese task in Chapter 3.

	Homophone Condition		*Translation Condition*		*Segment Condition*	
	Target	*Competitor*	*Target*	*Competitor*	*Target*	*Competitor*
1	棍	滚	挂	瓜	火	活
	gun4	*gun3*	*gua4*	*gua1*	*huo3*	*huo2*
	stick	roll	hang	melon	fire	alive
2	害	海	怪	乖	洗	席
	hai4	*hai3*	*guai4*	*guai1*	*xi3*	*xi2*
	harm	sea	weird	good	wash	mat
3	汗	喊	醋	粗	挤	急
	han4	*han3*	*cu4*	*cu1*	*ji3*	*ji2*
	sweat	yell	vinegar	thick	squeeze	urgent
4	见	剪	兔	秃	举	桔
	jian4	*jian3*	*tu4*	*tu1*	*ju3*	*ju2*
	see	cut	rabbit	bold	lift	orange
5	旧	酒	送	松	踩	才
	jiu4	*jiu3*	*song4*	*song1*	*cai3*	*cai2*
	old	wine	send	loose	tread	talent
6	笑	小	蒜	酸	纸	直
	xiao4	*xiao3*	*suan4*	*suan1*	*zhi3*	*zhi2*
	laugh	small	garlic	sour	paper	straight
7	醉	嘴	赚	砖	怕	爬
	zui4	*zui3*	*zhuan4*	*zhuan1*	*pa4*	*pa2*
	drunk	mouth	earn	brick	afraid	climb
8	豆	抖	撞	装	厚	猴
	dou4	*dou3*	*zhuang4*	*zhuang1*	*hou4*	*hou2*
	bean	shake	hit	pack	thick	monkey
9	造	枣	变	编	县	咸
	zao4	*zao3*	*bian4*	*bian1*	*xian4*	*xian2*
	make	date	change	edit	county	salty
10	瘦	手	记	鸡	踢	提
	shou4	*shou3*	*ji4*	*ji1*	*ti1*	*ti2*
	slim	hand	note	chicken	kick	carry
11	盖	改	告	高	饭	烦
	gai4	*gai3*	*gao4*	*gao1*	*fan4*	*fan2*
	build	change	tell	high	meal	bother
12	看	砍	证	争	翘	桥
	kan4	*kan3*	*zheng4*	*zheng1*	*qiao4*	*qiao2*
	look	chop	certificate	compete	raise	bridge

Table B2. Stimuli used in the Xi'an Mandarin task in Chapter 3.

	Homophone Condition		*Translation Condition*		*Segment Condition*	
	Target	*Competitor*	*Target*	*Competitor*	*Target*	*Competitor*
1	菜	猜	课	渴	唱	尝
	cai4	*cai1*	*ke4*	*ke3*	*chang4*	*chang2*
	dish	guess	class	thirsty	sing	taste
2	夏	虾	贵	鬼	亮	凉
	xia4	*xia1*	*gui4*	*gui3*	*liang4*	*liang2*
	summer	shrimp	expensive	ghost	bright	cold
3	画	花	肚	堵	臭	愁
	hua4	*hua1*	*du4*	*du3*	*chou4*	*chou2*
	paint	flower	belly	block	smelly	worry
4	替	踢	练	脸	剩	绳
	ti4	*ti1*	*lian4*	*lian3*	*sheng4*	*sheng2*
	replace	kick	practice	face	surplus	rope
5	地	低	气	起	笔	鼻
	di4	*di1*	*qi4*	*qi3*	*bi3*	*bi2*
	earth	low	gas	up	pen	nose
6	肺	飞	报	宝	粉	坟
	fei4	*fei1*	*bao4*	*bao3*	*fen3*	*fen2*
	lung	fly	report	baby	pink	grave
7	戏	西	冻	懂	汤	糖
	xi4	*xi1*	*dong4*	*dong3*	*tang1*	*tang2*
	drama	west	freeze	understand	soup	sugar
8	裤	哭	到	岛	天	甜
	ku4	*ku1*	*dao4*	*dao3*	*tian1*	*tian2*
	pants	cry	arrive	island	sky	sweet
9	棒	帮	瞪	等	签	钱
	bang4	*bang1*	*deng4*	*deng3*	*qian1*	*qian2*
	stick	help	stare	wait	sign	money
10	病	冰	电	点	灰	回
	bing4	*bing1*	*dian4*	*dian3*	*hui1*	*hui2*
	ill	ice	electricity	dot	grey	back
11	锈	修	断	短	枪	墙
	xiu4	*xiu1*	*duan4*	*duan3*	*qiang1*	*qiang2*
	rust	repair	break	short	gun	wall
12	动	东	烫	躺	秋	球
	dong4	*dong1*	*tang4*	*tang3*	*qiu1*	*qiu2*
	move	east	hot	lie	autumn	ball

Appendix C

Table C1. Stimuli used in Chapter 3.

	Target	Distractors		
		Tone-sharing	No-tone-sharing	Unrelated
1	新娘	心脏	信封	头盔
	xin1niang2	*xin1zang4*	*xin4feng1*	*tou2kui1*
	bride	heart	envelope	helmet
2	医生	衣柜	椅子	汽车
	yi1sheng1	*yi1gui4*	*yi3zi0*	*qi4che1*
	doctor	wardrobe	chair	car
3	孔雀	恐惧	空气	希望
	kong3que4	*kong3ju4*	*kong1qi4*	*xi1wang4*
	peacock	fear	air	hope
4	企鹅	乞丐	气球	地图
	qi3e2	*qi3gai4*	*qi4qiu2*	*di4tu2*
	penguin	beggar	balloon	map
5	蜥蜴	膝盖	喜剧	城市
	xi1yi4	*xi1gai4*	*xi3ju4*	*cheng2shi4*
	lizard	knee	comedy	city
6	鲸鱼	经理	警察	香水
	jing1yu2	*jing1li3*	*jing3cha2*	*xiang1shui3*
	whale	manager	policeman	perfume
7	屋顶	乌龟	舞蹈	钢琴
	wu1ding3	*wu1gui1*	*wu3dao3*	*gang1qin2*
	roof	turtle	dance	piano
8	羽毛	雨衣	玉米	眼泪
	yu3mao2	*yu3yi1*	*yu4mi3*	*yan3lei4*
	feather	raincoat	corn	tear
9	士兵	世界	时间	经验
	shi4bing1	*shi4jie4*	*shi2jian1*	*jing1yan4*
	soldier	world	time	experience

	Target	Distractors		
		Tone-sharing	No-tone-sharing	Unrelated
10	眼镜	演员	烟花	沙漠
	yan3jing4	*yan3yuan2*	*yan1hua1*	*sha1mo4*
	glasses	actor	firework	desert
11	鸡蛋	机器	季节	比赛
	ji1dan4	*ji1qi4*	*ji4jie2*	*bi3sai4*
	egg	machine	season	match
12	橡皮	相机	香蕉	钱包
	xiang4pi2	*xiang4ji1*	*xiang1jiao1*	*qian2bao1*
	eraser	camera	banana	wallet
13	蜡烛	辣椒	垃圾	词典
	la4zhu2	*la4jiao1*	*la1ji1*	*ci2dian3*
	candle	chili	trash	dictionary
14	鼠标	薯条	数字	雕塑
	shu3biao1	*shu3tiao2*	*shu4zi4*	*diao1su4*
	mouse	fries	number	sculpture
15	礼物	理论	力量	诗歌
	li3wu4	*li3lun4*	*li4liang4*	*shi1ge1*
	present	theory	strength	poem
16	洋葱	阳光	氧气	故事
	yang2cong1	*yang2guang1*	*yang3qi4*	*gu4shi4*
	onion	sunshine	oxygen	story
17	鹦鹉	英雄	影子	邮件
	ying1wu3	*ying1xiong2*	*ying3zi0*	*you2jian4*
	parrot	hero	shadow	mail
18	贝壳	背心	悲剧	窗户
	bei4ke2	*bei4xin1*	*bei1ju4*	*chuang1hu0*
	shell	vest	tragedy	window
19	蜘蛛	芝麻	纸巾	花瓶
	zhi1zhu1	*zhi1ma0*	*zhi3jin1*	*hua1ping2*
	spider	sesame	tissue	vase

	Target	Distractors		
		Tone-sharing	No-tone-sharing	Unrelated
20	珊瑚	山坡	闪电	石头
	shan1hu2	*shan1po1*	*shan3dian4*	*shi2tou0*
	coral	hillside	lightening	rock
21	肩膀	坚果	剪刀	螃蟹
	jian1bang3	*jian1guo3*	*jian3dao1*	*pang2xie4*
	shoulder	nuts	scissor	crab
22	钥匙	药店	摇篮	收据
	yao4shi0	*yao4dian4*	*yao2lan2*	*shou1ju4*
	key	pharmacy	cradle	receipt
23	斑马	扳手	板栗	恐龙
	ban1ma3	*ban1shou3*	*ban3li4*	*kong3long2*
	zebra	wrench	chestnut	dinosaur
24	葡萄	仆人	瀑布	银行
	pu2tao0	*pu2ren2*	*pu4bu4*	*yin2hang2*
	grape	servant	waterfall	bank

Appendix D

Table D1. Stimuli used in Chapter 4.

	English Target	Standard Chinese Distractor			
		Cross-language Homophone		Unrelated	
		Rising	*Falling*	*Rising*	*Falling*
1	ball	*bao2*	*bao4*	*yang2*	*yang4*
		thin	erupt	sheep	pattern
2	bar	*ba2*	*ba4*	*ji2*	*ji4*
		pull	dad	urgent	note
3	bee	*bi2*	*bi4*	*chu2*	*chu4*
		nose	avoid	kitchen	dread
4	fan	*fan2*	*fan4*	*qing2*	*qing4*
		bother	meal	feeling	celebrate
5	hen	*hen2*	*hen4*	*xing2*	*xing4*
		mark	hate	ok	apricot
6	jar	*jia1*	*jia4*	*li2*	*li4*
		cheek	marry	pear	chestnut
7	knee	*ni2*	*ni4*	*qiang2*	*qiang4*
		mud	greasy	wall	choke
8	lake	*lei2*	*lei4*	*ti2*	*ti4*
		thunder	tired	lift	shave
9	line	*lan2*	*lan4*	*du2*	*du4*
		blue	rot	poison	ferry
10	lung	*lang2*	*lang4*	*you2*	*you4*
		wolf	wave	oil	right
11	mat	*mai2*	*mai4*	*pa2*	*pa4*
		bury	sell	climb	fear
12	meat	mi2	mi4	fu2	fu4
		maze	honey	bliss	rich
13	bat	*bai2*	*bai4*	*yan2*	*yan4*
		white	salute	salt	swallow

	English Target	Standard Chinese Distractor			
		Cross-language Homophone		Unrelated	
		Rising	*Falling*	*Rising*	*Falling*
14	pea	*pi2*	*pi4*	*tuo2*	*tuo4*
		skin	fart	camel	spit
15	pie	*pai2*	*pai4*	*di2*	*di4*
		row	assign	enemy	brother
16	pin	*pin2*	*pin4*	*hu2*	*hu4*
		poor	employ	lake	protect
17	tea	*ti2*	*ti4*	*lu2*	*lu4*
		question	replace	stove	deer
18	tie	*tai2*	*tai4*	*she2*	*she4*
		table	over	snake	shoot
19	tool	*tu2*	*tu4*	*wen2*	*wen4*
		picture	rabbit	text	ask
20	tongue	*tang2*	*tang4*	*qian2*	*qian4*
		candy	trip	money	owe
21	wine	*wan2*	*wan4*	*ren2*	*ren4*
		play	all	people	blade
22	wood	*wu2*	*wu4*	*huo2*	*huo4*
		nothing	fog	live	disaster
23	whale	*wei2*	*wei4*	*rao2*	*rao4*
		surround	flavour	forgive	circle
24	shoe	*shu2*	*shu4*	*nuo2*	*nuo4*
		redeem	tree	move	promise

Summary

It is widely acknowledged that speakers or listeners co-activate multiple-word candidates during lexical access. While this conclusion is primarily drawn from empirical evidence in stress languages such as English, it is essential to note that the majority of the world's languages are tonal languages, utilizing lexical tones to distinguish word meanings. Consequently, the nature of lexical access in tonal languages, like Standard Chinese, remains to be fully understood. For instance, the relative weighting and timing of utilizing segments versus lexical tone and the role of lexical tone in activating lexical candidates during the process of Mandarin spoken word recognition have remained controversial.

Complicating matters, many tonal language speakers are bilinguals. Given that bilinguals have been found to activate words from both of their languages during lexical access, a crucial question arises regarding the role of lexical tone in bilingual language co-activation. For instance, for bilinguals with two tonal systems, it is unclear whether both tonal systems are co-activated or interact during lexical access, and if so, how potential lexical conflicts are resolved.

Therefore, for a more comprehensive understanding of lexical access, it is necessary to consider the role of lexical tone in both native and bilingual lexical access. This dissertation aims to address this gap by delving into the process of spoken word recognition and production. It focuses on three groups of tonal speakers: native speakers of Standard Chinese, bi-dialectal speakers of both Standard Chinese and Xi'an Mandarin, and bilingual speakers of Standard Chinese and English. Four key issues were highlighted: the role of lexical tone in Mandarin spoken word recognition; tonal interference in bi-dialectal spoken word recognition; the activation of lexical tone in bilingual spoken word production; and the influence of lexical tone on the bilingual mental lexicon.

This dissertation is composed of six chapters.

Chapter 1 introduced the research questions to be explored and offered a succinct preview of each succeeding chapter.

Chapter 2 aimed to investigate the role of lexical tone in Mandarin spoken word recognition. Specifically, we examined the role of segmental syllable and sub-syllabic constituents, as well as the time course of using segmental and suprasegmental tonal information during Mandarin lexical processing. In Experiments 1 and 2, native Standard Chinese speakers listened to monosyllabic Standard Chinese words with the presence of a phonological competitor, which overlaps with the target in either segmental syllable, onset and tone, rhyme and tone, or just tone. Experiments 1 and 2 differ in how long listeners were allowed to preview pictures on the screen before hearing the spoken target word. Eye movement results of both Experiments 1 and 2 confirmed a robust competition effect of segmental syllable overlap competitors, and null effects of onset, rhyme and tone overlap distractors. Experiment 3 investigated the time course of segmental versus tonal information utilization by manipulating their point of divergence in acoustic cues. We found that both sub-syllabic information (i.e., segment vs. tone) and cue timing (i.e., early vs. late point of divergence) affect phonological competition effects. Regardless of the nature of the cues, the point of divergence determines the size and time course of the competition effect: the earlier the point of divergence, the sooner the competition, suggesting that despite the dominant role of the segmental syllable, Mandarin listeners use both segmental and tonal information as soon as they are available to constrain lexical activation.

Chapter 3 aimed to enhance understanding of tonal interference in bi-dialectal spoken word recognition. Specifically, we investigated the process of spoken word recognition in bi-dialectal speakers of Standard Chinese and Xi'an Mandarin. Using the eye-tracking visual world paradigm, we asked Standard Chinese and Xi'an Mandarin bi-dialectals to listen to sentences in one dialect and identify the target word among four Chinese characters shown on the screen. The characters included the target, two unrelated distractors, and a phonological

competitor which shared the same segmental syllable with the target within- and across dialects. Among the phonological competitors, besides segmentally overlapping distractors which do not share lexical tone with the target within and across dialects (Segment Condition), there were also cross-dialect homophone competitors that share the same lexical tone with the target across dialects (Homophone Condition) and translation-induced cross-dialect homophones that share the same lexical tone with the targets' dialectal translation equivalent (Translation Condition). We hypothesized that, if both sets of lexical tones are activated, the Homophone and Translation Condition would elicit larger competition effects than the Segment Condition; if only one set of lexical tones is activated, the Segment Condition would elicit the largest competition effects, because the tonal contours of the target and competitor of the Segment Condition share the most acoustic similarity. Listeners' eye movements show that distractors in the Segment Condition interfere with participants' eye fixations significantly more than in Homophone and Translation Conditions, suggesting a lack of cross-dialectal interference effect. This finding marks a convergence between bi-dialectal and bilingual speech processing. Based on these findings, a preliminary model of bi-dialectal spoken word recognition which emphasizes active control of dialect activation was proposed.

Chapter 3 aimed to explore the activation of lexical tone in bilingual spoken word production by examining the role of lexical tone in non-tonal spoken word production with bilinguals of Standard Chinese and English. Specifically, we asked: if Standard Chinese and English bilinguals co-activate both Standard Chinese and English names during English word production, is lexical tone co-activated and utilized during the process? With four picture-word interference experiments, Standard Chinese and English bilingual speakers were instructed to name pictures in English (e.g., *feather*) while ignoring four types of simultaneously presented Standard Chinese distractors: 1) the translation distractor, which is the translation equivalent of the English target name (e.g., *yu3mao2* "feather"); 2) the tone-sharing distractor, which shares both tone and

segments with the Standard Chinese translation in the first syllable (e.g., *yu3zhou4* "universe"); 3) the no-tone-sharing distractor, which shares segments only with the Standard Chinese translation in the first syllable (e.g., *yu4mi3* "corn"); 4) the unrelated distractor, which shares no phonological overlap with target and its translation (e.g., *lei4shui3* "tear"). To further explore potential factors that may constrain the lexical tone effect, we also manipulated two additional factors that have been found to affect picture naming onset with the picture-word interference paradigm. One was distractor modality: the Standard Chinese distractors were presented either auditorily or visually. The other was familiarization mode: bilinguals were asked to familiarize themselves with the target pictures' English names only (i.e., English mode) or both English and Standard Chinese names (i.e., mixed mode). In Experiment 1 (with auditory distractor and English mode), translation distractors significantly facilitated bilingual English picture naming, while tone-sharing distractors significantly inhibited the process. Importantly, the tone-sharing distractors elicited significantly longer naming latency than the no-tone-sharing distractors, demonstrating the co-activation of lexical tone during English spoken word production. Overall, this study replicated previously found translation facilitation effect and observed a significant interference effect of lexical tone. These findings suggest that Standard Chinese and English bilinguals not only co-activate the Standard Chinese translation equivalents but also the lexical tones of the Standard Chinese translations during English spoken word production. Results of Experiments 2, 3 and 4 further demonstrated that the polarity and robustness of the lexical tone effect are modulated by external factors such as distractor modality and familiarization mode.

Chapter 5 aimed to explore the influence of lexical tone on the bilingual mental lexicon. Specifically, we asked whether and to what extent lexical tone modulates pitch processing in non-tonal speech production with Standard Chinese and English bilinguals. Using the picture-word interference paradigm, we asked Standard Chinese and English bilinguals and native English monolinguals to name pictures in English (e.g., *lung*) while ignoring simultaneously played

Standard Chinese cross-language homophones that either have a falling or a rising lexical tone (*lang4* with a falling tone, "wave"; *lang2* with a rising tone, "wolf"). We hypothesized that if lexical tone indeed influences bilinguals' pitch representation in non-tonal second languages, the effect of lexical tone (falling vs. rising) on English picture naming should differ between Standard Chinese and English bilingual and English monolingual speakers. Results showed that, compared with unrelated Standard Chinese distractors, both falling and rising cross-language homophones facilitated English word naming for both Standard Chinese-English bilingual and English monolingual speakers. Most importantly, Standard Chinese-English bilinguals showed significantly longer naming latencies with falling-tone in cross-language homophones than their rising-tone counterparts, whereas English monolingual speakers did not show such a pattern. This finding identified a significant difference between Standard Chinese-English bilinguals and English monolinguals in terms of how falling versus rising lexical tones affect English picture-word naming, providing evidence for the interaction between bilinguals' two languages at the suprasegmental level.

Chapter 6 reviewed the research questions and main findings of each study in this dissertation. Furthermore, implications for future research were discussed in this chapter.

In summary, this dissertation has demonstrated the significant role of lexical tone during native and bilingual lexical access. In Mandarin spoken word recognition, despite the advantageous role of the segmental syllable, lexical tone is employed as soon as it becomes available to constrain word activation. Bi-dialectal listeners of two closely related Mandarin dialects are able to control tonal-induced lexical interference, suggesting dynamic interaction between tonal systems. In bilingual spoken word production, Mandarin and English bilinguals automatically activate the lexical tones of the Standard Chinese translation equivalents during English spoken word production. Moreover, the pitch processing difference during English spoken word production between Mandarin-English bilinguals and English monolinguals suggests that lexical tone may play

an important role in the mental lexicon of bilinguals. Altogether, this dissertation enhances our understanding of lexical access by providing evidence on the role of lexical tone during spoken word recognition and production within- and across languages.

Samenvatting

Het wordt algemeen erkend dat sprekers of luisteraars meerdere woordkandidaten coactiveren tijdens lexicale toegang. Hoewel deze conclusie voornamelijk wordt getrokken uit empirisch bewijs in klemtoontalen zoals het Engels, is het essentieel om op te merken dat de meerderheid van de talen in de wereld toontalen zijn, die lexicale tonen gebruiken om woordbetekenissen te onderscheiden. Bijgevolg wordt de aard van de lexicale toegang in toontalen, zoals het Standaardchinees, nog steeds niet volledig begrepen. Zo zijn bijvoorbeeld de relatieve weging en timing van het gebruiken van segmenten versus lexicale toon en de rol van lexicale toon in het activeren van lexicale kandidaten tijdens het proces van gesproken woordherkenning in het Mandarijn controversieel gebleven.

Een complicerende factor is dat veel sprekers van toontalen tweetalig zijn. Aangezien het bekend is dat tweetaligen woorden uit hun beide talen activeren tijdens lexicale toegang, rijst een cruciale vraag over de rol van lexicale toon in tweetalige taalcoactivatie. Voor tweetaligen met twee toonsystemen is het bijvoorbeeld onduidelijk of beide toonsystemen worden gecoactiveerd of interageren tijdens lexicale toegang, en zo ja, hoe potentiële lexicale conflicten worden opgelost.

Voor een beter begrip van lexicale toegang is het daarom noodzakelijk om de rol van lexicale toon in zowel moedertalige als tweetalige lexicale toegang te overwegen. Dit proefschrift beoogt deze leemte op te vullen door het proces van gesproken woordherkenning en -productie te onderzoeken. Het richt zich op drie groepen sprekers: moedertaalsprekers van het Standaardchinees, bi-dialectale sprekers van zowel Standaardchinees als Xi'an Mandarijn, en tweetalige sprekers van Standaardchinees en Engels. Vier belangrijke onderwerpen werden belicht: de rol van lexicale toon in gesproken woordherkenning van het Mandarijn; tonale interferentie in bi-dialectale gesproken woordherkenning; de activering van

lexicale toon in tweetalige gesproken woordproductie; en de invloed van lexicale toon op het tweetalige mentale lexicon.

Dit proefschrift bestaat uit zes hoofdstukken.

Hoofdstuk 1 introduceerde de onderzoeksvragen die zijn onderzocht en bood een beknopte vooruitblik op elk volgend hoofdstuk.

Hoofdstuk 2 was gericht op het onderzoeken van de rol van lexicale toon in gesproken woordherkenning in het Mandarijn. Specifiek onderzochten we de rol van segmenten en sub-syllabische constituenten, evenals het tijdsverloop van het gebruik van segmentale en suprasegmentale tonale informatie tijdens lexicale verwerking in het Mandarijn. In Experimenten 1 en 2 luisterden moedertaalsprekers van het Standaardchinees naar monosyllabische Standaardchinese woorden met de aanwezigheid van een fonologische concurrent, die overlapt met het doelwoord in ofwel alle segmenten, onset en toon, rijm en toon, of alleen toon. Experimenten 1 en 2 verschilden in de tijd die luisteraars kregen om afbeeldingen op het scherm te bekijken voordat ze het gesproken doelwoord hoorden. Eye-trackingresultaten van zowel Experiment 1 als 2 bevestigden een robuust competitie-effect van concurrenten met volledige segmentale overlap, en nul-effecten van afleiders met overlap in de onset, rijm, en toon. Experiment 3 onderzocht het tijdsverloop van het gebruik van segmentale versus tonale informatie door hun punt van divergentie in akoestische signalen te manipuleren. We ontdekten dat zowel sub-syllabische informatie (d.w.z. segment versus toon) als cue timing (d.w.z. vroeg versus laat punt van divergentie) fonologische competitie-effecten beïnvloeden. Ongeacht de aard van de cues bepaalt het punt van divergentie de grootte en het tijdsverloop van het competitie-effect: hoe eerder het punt van divergentie, hoe eerder de competitie, wat suggereert dat ondanks de dominante rol van de segmenten, Mandarijnse luisteraars zowel segmentale als tonale informatie gebruiken zodra deze beschikbaar zijn om lexicale activatie te beperken.

Hoofdstuk 3 was gericht op het vergroten van het begrip van tonale interferentie in bi-dialectale gesproken woordherkenning. Specifiek onderzochten

we het proces van gesproken woordherkenning bij bi-dialectale sprekers van het Standaardchinees en Xi'an Mandarijn. Met behulp van het eye-tracking *visual world* paradigma vroegen we bi-dialectale sprekers van het Standaardchinees en Xi'an Mandarijn om naar zinnen in één dialect te luisteren en het doelwoord te identificeren uit vier Chinese karakters die op het scherm werden getoond. De karakters bevatten het doelwit, twee ongerelateerde afleiders en een fonologische concurrent die dezelfde segmenten deelde met het doelwit binnen- en tussen dialecten. Onder de fonologische concurrenten waren er, naast segmentaal overlappende afleiders die geen lexicale toon deelden met het doelwit binnen en tussen dialecten (Segmentconditie), ook dialect-overstijgende homofone concurrenten die dezelfde lexicale toon deelden met het doelwit binnen en tussen dialecten (Homofoonconditie) en vertaal-geïnduceerde dialect-overstijgende homofonen die dezelfde lexicale toon deelden met het dialectale vertaalequivalent van het doelwit (Vertaalconditie). We stelden de hypothese dat, als beide sets van lexicale tonen geactiveerd worden, de Homofoon- en Vertaalconditie grotere competitie-effecten zouden uitlokken dan de Segmentconditie; als slechts één set van lexicale tonen geactiveerd wordt, zou de Segmentconditie de meeste competitie-effecten uitlokken, omdat de tonale contouren van het doelwit en de concurrent van de Segmentconditie akoestisch de meeste overeenkomsten vertonen. De oogbewegingen van luisteraars lieten zien dat afleiders in de Segment Conditie significant meer interfereren met de oogfixaties van deelnemers dan in de Homofoon- en Vertaalcondities, wat suggereert dat er geen dialect-overstijgend interferentie-effect is. Deze bevinding markeert een overeenkomst tussen bi-dialectale en tweetalige spraakverwerking. Op basis van deze bevindingen werd een voorlopig model van bi-dialectale spraakherkenning voorgesteld dat de actieve controle van dialectactactivatie benadrukt.

Hoofdstuk 3 had als doel de activatie van lexicale toon in tweetalige gesproken woordproductie te onderzoeken door de rol van lexicale toon in niet-tonale gesproken woordproductie te bestuderen bij tweetaligen van het Standaardchinees en Engels. De specifieke onderzoeksvraag was als volgt: als

tweetaligen van het Standaardchinees en het Engels zowel Standaardchinese als Engelse woorden coactiveren tijdens de productie van Engelse woorden, wordt lexicale toon dan ook gecoactiveerd en gebruikt tijdens dit proces? In vier beeld-woord interferentie-experimenten werden tweetalige sprekers van Standaardchinees en Engels geïnstrueerd om afbeeldingen in het Engels te benoemen (bijv. *feather*) terwijl ze vier soorten gelijktijdig gepresenteerde Standaardchinese afleiders negeerden: 1) de vertalingsafleider, die het vertaalequivalent is van het Engelse doelwoord (bijv. *yu3mao2* "veer"); 2) de toondelende afleider, die zowel toon als segmenten deelt met de Standaardchinese vertaling in de eerste lettergreep (bijv. *yu3zhou4* "universum"); 3) de niet-toondelende afleider, die alleen in de eerste lettergreep segmenten deelt met de Standaardchinese vertaling (bijv. *yu4mi3* "maïs"); 4) de niet-verwante afleider, die geen fonologische overlap heeft met het doelwit en de vertaling ervan (bijv. *lei4shui3* "traan"). Om mogelijke factoren die het lexicale tooneffect kunnen beperken verder te onderzoeken, manipuleerden we ook twee extra factoren die van invloed bleken te zijn op het begin van het benoemen van plaatjes met het beeld-woord interferentie paradigma. De eerste was de modaliteit van de afleider: de Standaardchinese afleiders werden auditief of visueel gepresenteerd. De andere was vertrouwdheidsmodus: aan tweetaligen werd gevraagd om zichzelf vertrouwd te maken met alleen de bijbehorende Engelse woorden van de doelafbeeldingen (d.w.z. de Engelse modus) of met zowel Engelse als Standaardchinese woorden (d.w.z. de gemengde modus). In Experiment 1 (met auditieve afleidingen en Engelse modus), bevorderden afleiders met vertaling significant het tweetalig Engels benoemen van afbeeldingen, terwijl afleiders met dezelfde tonen dit proces significant afremden. Belangrijk is dat de toondelende afleiders een significant langere benoemingsvertraging opwekten dan de niet-toondelende afleiders, wat de coactivatie van lexicale toon tijdens Engels gesproken woordproductie aantoont. In het algemeen repliceerde deze studie eerder gevonden vertalingbevorderende effecten en toonde een significant interferentie-effect aan van lexicale toon. Deze bevindingen suggereren dat

tweetaligen van het Standaardchinees en Engels niet alleen de Standaardchinese vertaalequivalenten coactiveren, maar ook de lexicale tonen van de Standaardchinese vertalingen tijdens de productie van Engelse gesproken woorden. De resultaten van Experimenten 2, 3 en 4 toonden verder aan dat de polariteit en robuustheid van het lexicale tooneffect worden gemoduleerd door externe factoren zoals de modaliteit van de afleider en de vertrouwdheidsmodus.

Hoofdstuk 5 onderzocht de invloed van lexicale toon op het tweetalige mentale lexicon. Specifiek onderzochten we of en in welke mate lexicale toon de toonhoogteverwerking moduleert in niet-tonale spraakproductie bij Standaardchinese en Engelse tweetaligen. Met behulp van het *picture-word* interferentieparadigma vroegen we aan tweetaligen van het Standaardchinees en Engels en aan moedertaalsprekers van het Engels om afbeeldingen in het Engels te benoemen (bijv. *lung*) terwijl ze tegelijkertijd Standaardchinese taal-overstijgende homofonen moesten negeerden die ofwel een dalende ofwel een stijgende lexicale toon hadden (*lang4* met een dalende toon, "golf"; *lang2* met een stijgende toon, "wolf"). We stelden als hypothese dat als lexicale toon inderdaad de toonhoogterepresentatie van tweetaligen in niet-tonale tweede talen beïnvloedt, het effect van lexicale toon (dalende versus stijgende toon) op het benoemen van Engelse afbeeldingen zou moeten verschillen tussen sprekers van het Standaardchinees en Engels t.o.v. eentalige moedertaalsprekers van het Engels. De resultaten toonden aan dat, in vergelijking met ongerelateerde afleiders uit het Standaardchinees, zowel dalende als stijgende taal-overstijgende homofonen het benoemen van Engelse woorden vergemakkelijkten voor zowel Standaardchinees-Engelse tweetaligen als voor eentalige moedertaalsprekers van het Engels. Het belangrijkste is dat Standaardchinees-Engelse tweetaligen significant langere benoemingslatenties vertoonden met dalende toon in taal-overstijgende homofonen dan met de tegenhangers met stijgende toon, terwijl eentalige moedertaalsprekers van het Engels geen dergelijk patroon vertoonden. Deze bevinding belichtte een significant verschil tussen Standaardchinees-Engelse tweetaligen en Engelstalige eentaligen in termen van hoe dalende versus

stijgende lexicale tonen invloed hebben op het benoemen van afbeeldingen in het Engels, wat bewijs levert voor de interactie tussen de twee talen van tweetaligen op suprasegmentaal niveau.

Hoofdstuk 6 gaf een overzicht van de onderzoeksvragen en belangrijkste bevindingen van elk onderzoek in dit proefschrift. Verder werden in dit hoofdstuk implicaties voor toekomstig onderzoek besproken.

Samenvattend heeft dit proefschrift de significante rol aangetoond van lexicale toon tijdens moedertaal- en tweetalige lexicale toegang. In gesproken woordherkenning in het Mandarijn wordt, ondanks de voordelige rol van de segmentale lettergreep, lexicale toon gebruikt zodra deze beschikbaar is om woordactivatie te beperken. Bi-dialectale luisteraars van twee nauw verwante Mandarijnse dialecten zijn in staat om tonaal-geïnduceerde lexicale interferentie te controleren, wat wijst op een dynamische interactie tussen toonsystemen. Bij tweetalige gesproken woordproductie activeren tweetaligen van het Mandarijn en het Engels automatisch de lexicale tonen van de Standaardchinese vertaalequivalenten tijdens de Engelse gesproken woordproductie. Bovendien suggereert het verschil in toonhoogteverwerking tijdens Engels gesproken woordproductie tussen Mandarijn-Engels tweetaligen en Engels eentaligen dat lexicale toon een belangrijke rol kan spelen in het mentale lexicon van tweetaligen. Al met al vergroot dit proefschrift het begrip van lexicale toegang door bewijs te leveren van de rol van lexicale toon tijdens gesproken woordherkenning en -productie binnen en tussen talen.

摘要

人们普遍认为词语提取过程中大脑会共同激活多个候选词。这一结论主要是从英语等重音语言的实证研究中得出的。但必须指出的是，世界上大多数语言都是声调语言，利用声调来区分词义。汉语等声调语言词性提取的性质仍有待充分理解。例如，在普通话口语词汇识别过程中，音段与声调的相对权重和时间以及声

调在激活候选词汇中的作用一直存在争议。

更复杂的是，许多讲声调语言的人是双语者。双语者在词语提取过程中会同时激活两种语言的词汇，但是声调在双语的共同激活中起到什么样的作用尚不清楚。例如，对于拥有两种声调系统的双语者来说，目前还不清楚这两种声调系统在词语提取过程中是否会被共同激活并相互影响；如果是，双语者又是如何解决潜在的词汇冲突的。

因此，为了更全面地理解词汇获取，有必要考虑声调在母语和二语词汇获取中的作用。本论文旨在通过深入研究口语词汇的识别和生成过程来填补这一空白。本论文重点研究了三类声调语言使用者：普通话母语者、普通话-西安普通话双方言者以及普通话-英语双语者。同时重点调研了四个关键问题：声调在普通话口语词汇识别中的作用；双方言口语词汇识别中的声调干扰；声调在双语口语词汇生成中的激活；声调对双语者心理词库的影响。

本论文一共由六章组成。

第一章介绍了要探讨的研究问题，并简明扼要地预览了后面各章的内容。

第二章旨在研究声调在普通话口语词汇识别中的作用。具体来说，我们研究了音节、声母和韵母的作用，以及在普通话词汇加工过程中加工音段和超音段信息的时间过程。在实验一和二中，以普通话为母语的人会

在听单音节普通话词同时看到四张图片。这四张图片包括了被听到的目标词，目标词的语音竞争项，以及两个无关的干扰项。竞争项与目标词在（无声调）音节、声母和声调、韵母和声调或声调上重叠。 实验 一和二的不同之处在于，在听到目标词之前，听者有多长时间可以预览屏幕上的图片。实验一和二的眼动结果都证实了（无声调）音节竞争项的强大竞争效应。实验三通过调整声学线索中音段与声调信息的分歧点，研究了音段与声调信息利用的时间过程。我们发现，不管是音段还是声调信息，分歧点的早晚决定了竞争效应的大小和时间进程：分歧点越早，词汇竞争的干扰作用越早。这表明尽管无声调的音节在词语识别起着主导作用，普通话听者会在音段和声调信息可用时立即使用它们来限制词汇激活。

第三章旨在加深对双方言口语词语识别中声调干扰的理解。具体来说，我们研究了普通话和西安普通话双方言者的口语词语识别过程。我们使用眼动跟踪视觉世界范式，让普通话和西安普通话双方言者听普通话或者西安方言句子，并从屏幕上显示的四个汉字中识别目标词。这些汉字包括目标词、两个不相关的干扰词和一个语音竞争项。在语音竞争者中，除了与目标词至共享音段不共享声调的音段重叠干扰项外，还有在共享音段和声调的跨方言同音竞争项，以及与目标词的方言翻译共享音段和声调的跨方言翻译同音竞争项。我们假设，如果两个方言的声调系统都被激活，那么跨方言同音竞争项和翻译同音竞争项将比音段重叠干扰项引发更大的竞争效应；如果只有目标方言的声调被激活，音段重叠干扰项将引发最大的竞争效应。听者的眼球运动表明，与跨方言同音竞争项和翻译同音竞争项相比，音段重叠干扰项对受试者目光关注的干扰明显更大，这表明被试的口语词语识别并没有被跨方言同音词干扰。这一发现标志着双方言和双语语音处理的不同。基于这些发现，我们提出了一个双方言口语单词识别的初步模型。

第四章旨在通过研究普通话和英语的双语者在非声调口语产出中声调的作用，探索声调在双语口语产出中的激活作用。具体而言，我们提出了这样一个问题：如果普通话和英语双语者在英语单词生成过程中同时激活了普通话和英语名称，那么在这一过程中，声调是否被共同激活和利用？在四个图片词语干扰实验中，普通话和英语双语者被要求在忽略普通话干扰词的情况下说出英语图片的名称。根据干扰词与目标词的关系，本实验共有以下四个实验条件：翻译条件，目标词的普通话翻译；声调共用条件，其第一个音节与目标项翻译的声调和音段重叠；无声调共用条件，在其第一个音节与目标项翻译的音段重叠；无关条件，与目标语及其翻译没有语音或语义上的关联。实验结果发现，相对于无关条件，翻译条件明显加快了英语图片命名的速度，而声调共用条件则明显抑制了这一过程。重要的是，声调共用条件引起的命名潜伏期明显长于无声调共用条件，这表明在英语口语词语产出过程中，汉语声调被同时激活。这些结果表明，普通话和英语双语者在英语图片命名过程中不仅共同激活了对应的普通话翻译，也共同激活了其声调。

第五章旨在探讨声调对双语心理词典的影响。具体来说，我们询问了普通话和英语双语者在非声调词语产出过程中，声调是否影响了音高加工。采用图片词语干扰范式，我们请普通话和英语的双语者以及英语为母语的单语者说出英语图片的名称，同时忽略同时播放的跨语言普通话同音词。这些同音词要么具有降调（四声），要么具有升调（二声）。 我们假设，如果声调确实会影响双语者在非声调第二语言中的音高表征，那么升调与降调对英语图片命名的作用在汉语英语双语者和英语单语者之间应该有所不同。结果显示，与无关条件相比，降调和升调汉语同音词都有助于汉语英语双语者进行英语图片命名。最重要的是，普通话英语双语者在听到降调同音词时的图片命名时间明显长于其听到升调同音词的时长，而英语单语者则没有表现出这种命名模式。这一发现体现了汉英双语者与英语

单语者在英语词语产出中音高感知的显著差异，为双语者两种语言在超音段层面上的交互作用提供了证据。

第六章回顾了本论文中每项研究的研究问题和主要结论。

总之，本论文证明了声调在词汇获取过程中的重要作用。在普通话口语词语识别中，尽管音段音节具有优势作用，但一旦声调可用，就会被及时用来限制词语激活。普通话双方言者能够调控声调引起的词汇干扰，表明两种方言系统之间存在动态互动。在双语口语词语产出过程中，普通话和英语双语者会自动激活其普通话翻译的声调。此外，普通话-英语双语者和英语单语者在英语词语产出过程中的声调处理差异表明声调可能在双语者的心理词典中扮演着重要角色。总之，本论文通过实验证据证明了声调在语言内和跨语言间词语识别与产出中的重要作用，加深了我们对词语提取的理解。

Curriculum vitae

Qing Yang (杨青) was born on 14 March 1993 in the city of Shangqiu in China. She attended Beijing Normal University Zhuhai in 2010 and received her bachelor's degree in teaching Chinese as a foreign language in 2014. She then started her Master study the same year in the phonetics lab at Beijing Language and Culture University. In 2017, she finished her studies there and obtained her Master's degree. In the fall of 2017, she started her PhD research at the Leiden University Centre for Linguistics. This dissertation is the result of her PhD research.