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	***jie****4*	***jie****3*
	SC pitch contour	**high-falling**	dipping
	XM pitch contour	level	**high-falling**
Translation Condition	Character	菜	猜
	Gloss	vegetable	guess
	Pinyin	***cai****4*	***cai****1*
	SC pitch contour	high-falling	**level**
	XM pitch contour	**level**	dipping
Segment Condition	Character	纸	直
	Gloss	paper	straight
	Pinyin	***zhi****3*	***zhi****2*
	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.

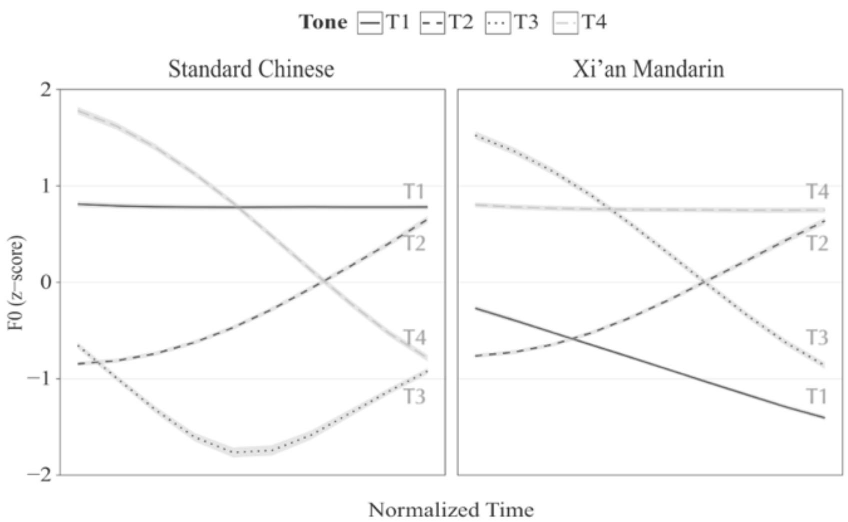

Figure 1. *Mean F0 (Z-score) contours of the four tones in Standard Chinese and Xi'an Mandarin. The grey areas indicate the 95% confidence interval of the corresponding mean. This figure is reprinted from Liu et al. (2020, p.2808).*

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 wellrecognized 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 [$\chi2$ (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.

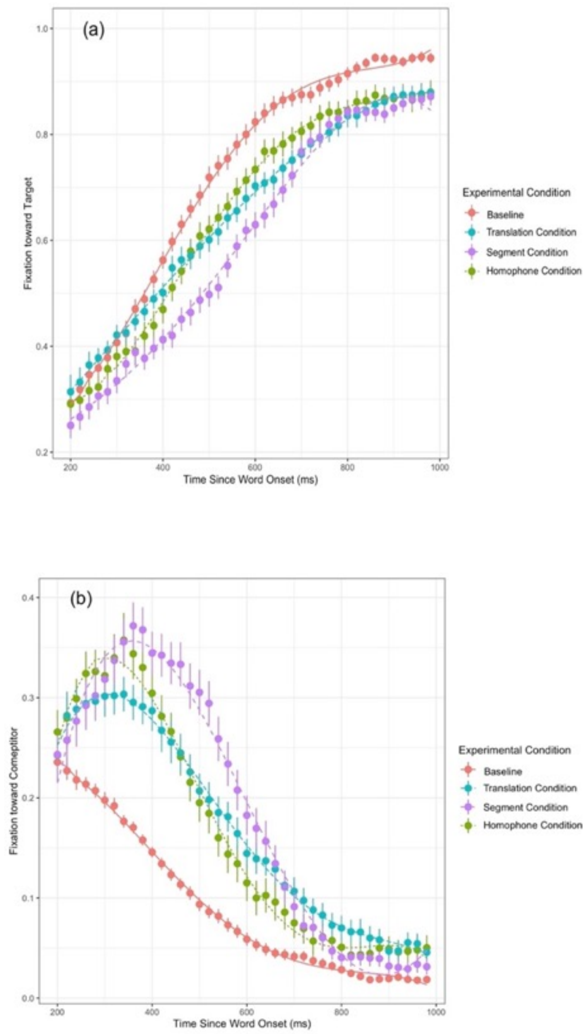

Figure 2. *Time course of eye fixations toward the target (a) and competitors (b) of each experimental condition plotted against baseline in the Standard Chinese task. The points with range represent mean proportions of fixations across participants and items with standard error. The lines represent 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 ($\chi2$ (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.

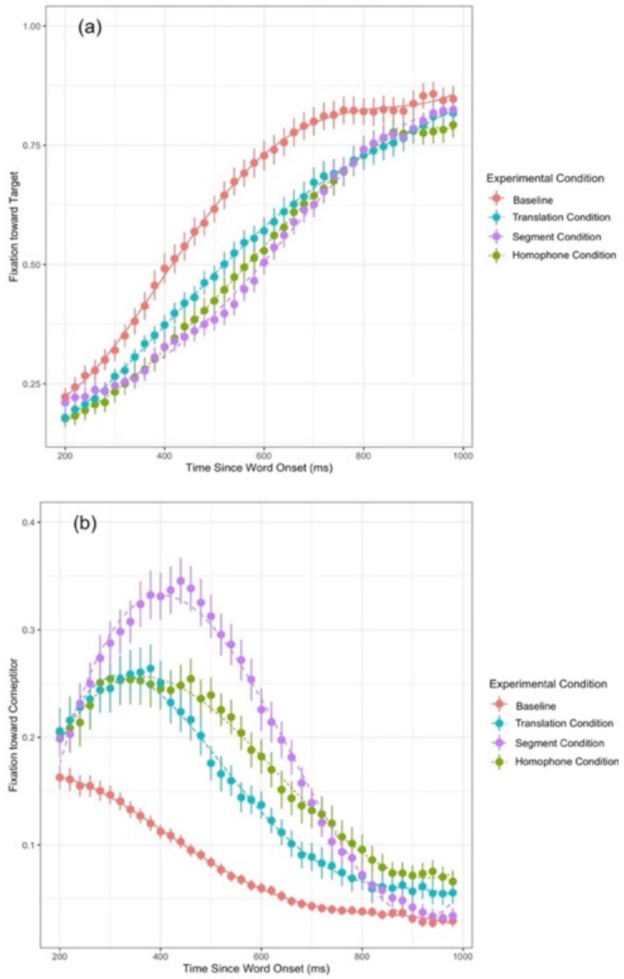

Figure 3. *Time course of eye fixations toward the target (a) and competitors (b) of each experimental condition plotted against baseline in the Xi'an Mandarin task. The points with range represent mean proportions of target fixations across participants and items with standard error. The lines represent 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.

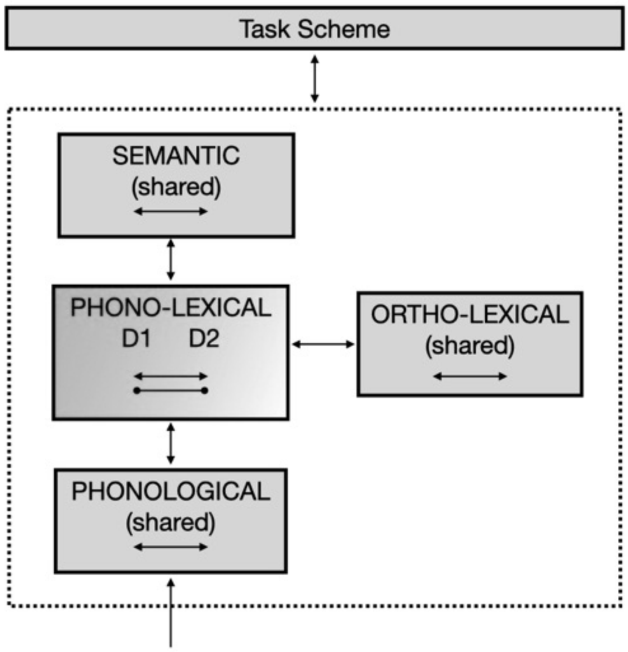

Figure 4. *The bi-dialectal spoken word recognition model. This model has phonological, phono-lexical, ortho-lexical, and semantic representations. Between levels, the representations interact bidirectionally. Within levels, dialect-specific and dialect-shared representations are stored in the same space, 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.