

## **Si** Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers

Seth Wiener<sup>a\*</sup> and Kiwako Ito<sup>b</sup>

<sup>a</sup>The Ohio State University, Hagerty Hall 398, 1775 College Road, Columbus, OH 43210, USA; <sup>b</sup>The Ohio State University, 1712 Neil Avenue, Columbus, OH 43210, USA

(Received 30 June 2013; accepted 2 July 2014)

An eye-tracking study investigated how the interaction between syllable frequency and syllable-specific tonal probability guides online lexical access in speakers of mutually unintelligible Chinese dialects with three disparate tonal systems. Mono-dialectal Mandarin speakers, bi-dialectal Shanghai–Mandarin speakers and bi-dialectal Cantonese–Mandarin speakers searched for target Mandarin syllable–tone combinations while their eye movements and mouse clicks were recorded. The results showed dialectal differences in online eye fixation patterns but not in offline mouse responses. For all groups, mouse clicks were fastest for infrequent syllables with most probable tones and slowest for infrequent syllables with least probable tones. In online eye movement responses, only mono-dialectal Mandarin speakers showed an interaction between syllable frequency and tonal probability. Mono-dialectal Mandarin speakers' fixations were fastest for infrequent syllables with probable tones and slowest for infrequent syllables with improbable tones. Mono-dialectal speakers also showed a greater amount of competition from the more probable segmental competitor when hearing improbable tones. Bi-dialectal speakers showed different timing in their integration of tonal probabilities. These findings suggest that highly bilingual speakers track and use Mandarin tonal probabilities, but their sensitivity to L2 tonal information may lag behind monolinguals for online word recognition.

**Keywords:** lexical access; Mandarin Chinese; lexical tone; eye-tracking; spoken word recognition

Models of spoken word recognition such as TRACE (McClelland & Elman, 1986) and Shortlist/Merge (Norris, 1994; Norris & McQueen, 2008; Norris, McQueen, & Cutler, 2000) assume that listeners map relevant speech signals to sublexical and lexical representations. Based on findings from European languages (Cooper, Cutler, & Wales, 2002; Cutler, 2005; Fear, Cutler, & Butterfield, 1995), suprasegmental features such as stress are generally believed to play a minimal role in this process. For the majority of Sino-Tibetan languages, however, tone is an integral phonetic resource for distinguishing lexical meaning.

While research on the tonal phonology of non-Western languages has contributed tremendously to the typological exploration of human linguistic systems (Clements & Goldsmith, 1984; Hombert, Ohala, & Ewan, 1979; Yip, 2002), it remains unclear how exactly tonal information is integrated during word recognition. Importantly, tones typically span multiple segments or syllables and undergo language-specific alignment and alteration rules (e.g., Gussenhoven, 2004; Hyman & Kisseberth, 1998; Odden, 1995; Xu & Wang, 2001). How tones are represented in a lexicon may vary according to language-specific factors such as the size of the tonal inventory, phonotactic complexity and the predictability of tone assignment.

The present study tests this hypothesis by examining the representation and use of tones during spoken word recognition by mono- and bi-dialectal speakers of Mandarin Chinese (Pu35 tong55 hua51<sup>1</sup>).

### Previous cross-linguistic work

The use of suprasegmental cues in lexical access is language specific. For example, English listeners do not rely heavily on stress information, presumably due to the limited set of stress minimal pairs (Cooper et al., 2002; Creel, Tanenhaus, & Aslin, 2006; Cutler, 1986). Italian, Dutch and Spanish listeners rely slightly more on stress, since it can temporarily provide more disambiguating information than segmental cues (Cutler & van Donselaar, 2001; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; Sulpizio & McQueen, 2012; Tagliapietra & Tabossi, 2005).

Speakers of tone languages, in which syllables are differentiated via a finite set of pitch cues, may benefit most from suprasegmental information. Previous studies have demonstrated that speakers of tone languages have higher sensitivities to pitch information than speakers of non-tone languages. Gandour (1983) showed that speakers of Thai, Taiwanese (Southern Min), Cantonese and Mandarin (Taiwan guo35 yu214) all rated tones differently than English speakers. Burnham and Jones (2002)

\*Corresponding author. Email: [wieners.24@osu.edu](mailto:wieners.24@osu.edu)

found that Thai speakers perceived artificial tone continua better than English speakers. In Lee, Vakoč, and Wurm (1996), Cantonese and Mandarin speakers discriminated their own tones better than tones from the opposite language, while both groups outperformed English speakers. Huang and Johnson (2010) argued that Mandarin speakers attend to tonal contours and use their knowledge of phonologically conditioned rules governing tone change, while English speakers tune to F<sub>0</sub> height for judging tonal similarities. Huang and Johnson conclude that language-specific phonological knowledge modulates tone perception.

The role and time course of tone processing in word recognition have been studied mostly with monolingual speakers (Cutler & Chen, 1997; Lee, 2007; Ye & Connine, 1999). A growing body of experimental research with the visual-world paradigm (Malins & Joanisse, 2010; Shen, Deutsch, & Rayner, 2013) and event-related potential analyses of spoken Cantonese (Schirmer, Tang, Penny, Gunter, & Chen, 2005) and Mandarin (Brown-Schmidt & Canseco-Gonzalez, 2004; Malins & Joanisse, 2012; Zhao, Guo, Zhou, & Shu, 2011) suggest that speakers of tone languages integrate tonal and segmental information in parallel.

One important factor that has been often neglected in the research on Mandarin word recognition is the fact that China's linguistic diversity exhibits fast-growing diglossia; many Mandarin speakers today use multiple dialects, often with one dialect used more dominantly than another (Li & Lee, 2004). Huang (2012) reports that the tonal sensitivities of bi-dialectal speakers largely depend on the (in)separability among the dialects, which is rooted in their historical tone correspondences (Chao, 1941; Karlgren, 1954; Norman, 1988). While Huang's findings provide an insight into the architecture of bi-dialectal lexica, it is unclear how inter-dialect tonal correspondences affect the processing of tone in bi-dialectal speakers. If the bi-dialectal representation of tone is not as fine-grained as it is in a mono-dialectal lexicon, bi-dialectal speakers may not show sensitivities to information such as tonal predictability at a similar timing as mono-dialectal speakers do.

Two previous studies on bi-dialectal Cantonese-Mandarin children suggest that tonal awareness develops early and is a result of a speaker's L1. Chen et al. (2004) and Chen, Ku, Koyama, Anderson, and Li (2008) had grade school students perform same-different judgement tasks for Mandarin and Cantonese onsets, rimes and tones. In Chen et al. (2004), second- and fourth-grade Cantonese-Mandarin bi-dialectal children showed greater sensitivity to tones than children who only spoke Mandarin. The authors suggest that Cantonese-speaking bi-dialectal children develop superior tone awareness due to exposure to Cantonese and its more complex tonal system. In a follow-up study, which tracked performance across grade levels, Chen et al. (2008) found that while tone awareness in Mandarin

increased from the first to the fourth grade, tone awareness in Cantonese improved from the first to the second grade, but then decreased from the second to the fourth grade. The authors conclude that tonal awareness is a language-specific construct that shows different developmental trajectories. Taken together, these studies suggest that bi-dialectal speakers of Mandarin and another more tonally complex dialect may exhibit a processing advantage when hearing Mandarin tones. For example, Cantonese-Mandarin bi-dialectal speakers may show greater tone sensitivity and faster word recognition than Mandarin mono-dialectals because of their experience with the more complex Cantonese tonal system. In contrast, bi-dialectal speakers of Mandarin whose dominant dialect has a simpler tone system may show a disadvantage in tone processing. That is, bi-dialectal speakers of Mandarin and a dialect with three or fewer tones may lack the sensitivity to Mandarin tones that mono-dialectal speakers exhibit and show slower word recognition.

### **Mandarin, Shanghai and Cantonese**

Languages spoken in China are traditionally divided into seven major dialect groups (Li, 1973; Norman, 1988; Yuan, 1960). The largest group is the Mandarin subset, clustered primarily in northern/northwest China, including Beijing, which the 'lingua franca' Pu35 tong55 hua51 and its phonological system are based on (Chen, 1999; Li, 2006).

Standard Mandarin syllable structure can be represented as (C)(G)V(X), with G and X, respectively, representing on-glides and off-glides/nasal codas (Duanmu, 2008). While this structure can produce as many as 2280 different syllables, only around 400 spoken syllables exist, ignoring tone (Duanmu, 2007). Each syllable carries one of four tones: high-level pitch 55, rising pitch 35, low-dipping pitch 214, and high-falling pitch 51 (Chao, 1968; Figure 1a).

Because not all 400 Mandarin syllables are produced with all four tones, there are approximately only 1300 unique syllable-tone combinations (Duanmu, 2007). On average, each syllable-tone combination corresponds to four morphemes, with some combinations carrying as many as 90 homophones (e.g., Yi51) (Yin, 1984). Due to the gaps in tonal distributions and the highly varying number of homophones, certain syllable-specific tone contours may be more predictable than others. Thus, listeners may track and exploit tonal information according to its cue validity for each word/morpheme.

Our second target dialect, Shanghai (a variety of the Wu dialect), is spoken by approximately 13 million residents of the greater Shanghai region (Norman, 1988). Here we focus on what Xu and Tao (1997) call 'New Shanghai', which has five citation tones; three realized in the high F<sub>0</sub> register – rising 35, falling 52, and short-high 44 – while two are realized in the low F<sub>0</sub> register – rising

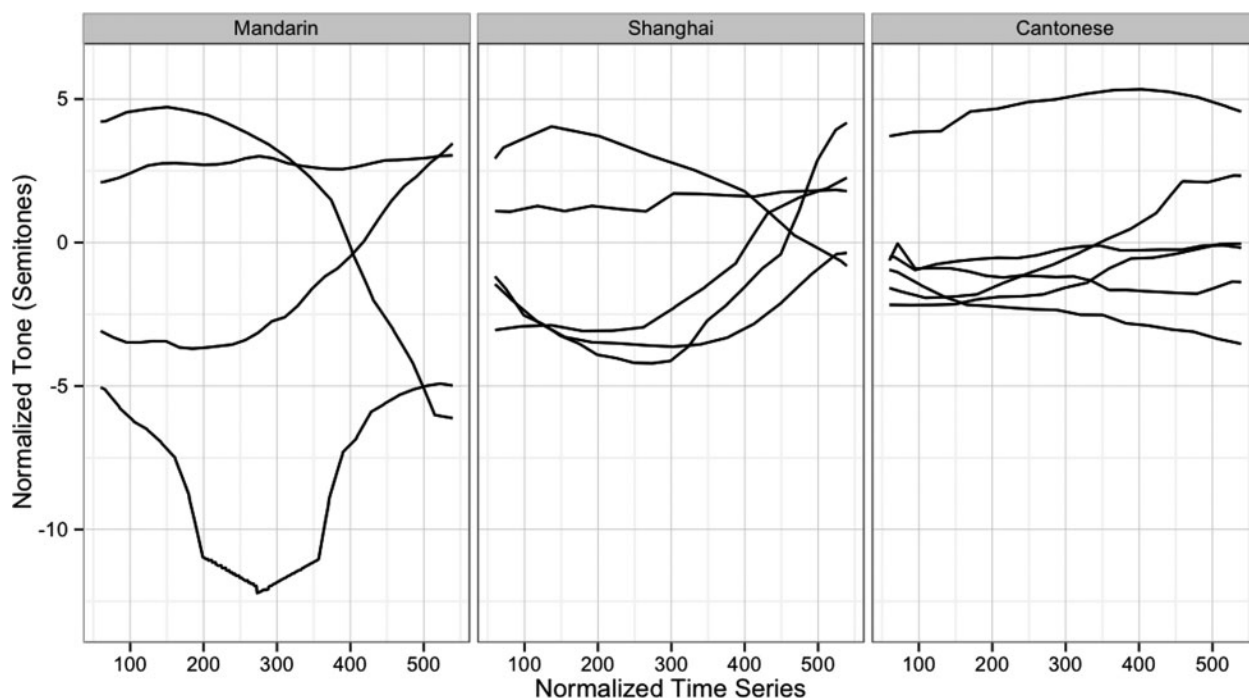


Figure 1. (a) Standard Beijing Mandarin lexical tones. (b) Standard new Shanghai lexical tones. (c) Standard Guangzhou Cantonese lexical tones. (Note: Contours in Figure 1 were produced by a native speaker of each dialect. To compensate for temporal differences across tones, productions were normalized via 550 relative time points.)

24 and short-rising 23 (Chen, 2008; Xu, Tang, You, Qian, Shi, & Shen, 1988; Yip, 1980; Zee & Maddieson, 1980; Figure 1b).

The Shanghai tonal system is particularly interesting because tones are largely predictable from the syllable onset and coda: syllables with voiceless onsets are realized only with high register tones; voiced onsets co-occur with only low-register tones; short tones co-occur with only closed syllables with a glottal coda (Chen, 2011). This co-occurrence motivates claims that Shanghai underlyingly has two tones or is a phonation register language (Duanmu, 1999; Zhu, 1999). Furthermore, Shanghai has fewer syllable–tone combinations (approximately 700) and more homophones than Mandarin (Duanmu, 2008). Given the high likelihood of specific syllable–tone combinations, the role of tone and the need to track tonal information may be similar to or less crucial in Shanghai than they are in Mandarin.

Lastly, Cantonese (a variety of the Yue dialect) is spoken throughout the southern Chinese provinces of Guangdong and Guangxi, and the municipality of Hong Kong. Cantonese exhibits considerable phonetic and tonal variation across sub-dialects (Bauer & Benedict, 1997; Norman, 1988). The present study investigates the variety spoken in the Guangzhou region of Guangdong. A traditional analysis reports nine surface tones, but of those only six are phonemic: high-level 55, mid-rising 35, mid-level 33, low-falling 21, low-rising 24, and low-level 22

(Bauer & Benedict, 1997; Figure 1c). The other three tones are conditioned by a syllable coda (what is historically known as ‘checked tones’: 55, 33, and 22).

Cantonese, unlike Mandarin, has several overlapping and similar tone contours (e.g., mid-level 33 and low-level 22; mid-rising 35 and low-rising 24). Cantonese also has a higher number of syllables (between 630 and 750: Bauer & Benedict, 1997; Zeng, 1994) and syllable–tone combinations (between 1700 and 2300: Leung, Law, & Fung, 2004; Zeng, 1994) than Mandarin. Unlike Shanghai, Cantonese syllable onsets/offsets do not foreshadow tone types (except for the aforementioned ‘checked tones’). Given the much larger number of syllable–tone combinations, and the relative similarity of several tone contours, Cantonese speakers may most benefit from tracking tonal probabilities.

### Why Mandarin, Shanghai and Cantonese?

It is estimated that over 70% of Shanghai speakers and over 60% of Cantonese speakers are able to communicate in Mandarin (Chinese National Committee on Language Affairs, 2004). When highly proficient bi-dialectal speakers engage in Mandarin spoken word recognition, their L2 tone processing may or may not be modulated by their L1 tone representations.

If Shanghai and Cantonese speakers map L2 Mandarin tonal phonology to their existent L1 phonological system (Flege, 1995; Huang, 2012), then speakers of these

Table 1.  $2 \times 2$  design of eye-tracking experiment.

	High-probability tone (PH)	Low-probability tone (PL)
High-frequency syllable (FH)	FH-PH	HF-PL
	High-frequency syllable	High-frequency syllable
	Most probable tone	Least probable tone
Low-frequency syllable (FL)	FL-PH	FL-PL
	Low-frequency syllable	Low-frequency syllable
	Most probable tone	Least probable tone

dialects may be less accurate than the mono-dialectal speakers at predicting Mandarin tonal probabilities given the lack of one-to-one mapping of Mandarin tones to tones in their respective dialects. Cantonese speakers may show higher sensitivities to Mandarin tonal probabilities, given their larger L1 tone inventory and lack of segmental cues, whereas Shanghai speakers may show lower sensitivities to Mandarin tonal probabilities given their smaller L1 tone inventory and potential over-reliance on segmental cues (e.g., initial stop distinction). Alternatively, if bi-dialectal speakers develop higher sensitivities to tone than mono-dialectal speakers due to exposure to more than one tone system (Chen et al., 2004, 2008) and they represent their L1 and L2 tonal systems separately, then both bi-dialectal groups may show similar or even higher sensitivities to Mandarin tonal probabilities as compared to the mono-dialectal group.

## Experiment

### *Participants*

Twenty mono-dialectal Mandarin speakers from the Beijing area (14 female; 6 male; mean age: 24.5 years), 20 Shanghai-dominant Mandarin bilinguals from Shanghai (11 female; 9 male; mean age: 25.8 years) and 20 Cantonese-dominant Mandarin bilinguals from Guangdong (14 female; 6 male; mean age: 23.2 years) were recruited for a small payment at the Ohio State University. All participants resided in central Ohio at the time of the study, spoke English as their second or third language and finished high school in China. Over half the participants in each group finished their undergraduate degrees in China. The mean time in the US was as follows: Mandarin – 48 months; Shanghai – 38 months and Cantonese – 36 months. All participants reported normal or corrected-to-normal vision and no hearing impairment.

### *Materials and design*

To test listeners' sensitivities to syllable-specific tonal probabilities, two factors – syllable frequency (high vs. low) and tonal probability (high vs. low probability) – were crossed to create four ( $2 \times 2$ ) conditions (Table 1).

### *Auditory stimuli*

A total of 96 Mandarin monosyllabic words were selected from SUBTLEX-CH spoken word corpus (Cai & Brybaert, 2010). Twenty-four segmental pairs that matched in segments but contrasted in tones (e.g., zhou55 'porridge' and zhou35 'axle') were recorded for the four critical conditions. Twelve of these pairs had high syllable frequencies while the other 12 had low frequencies. One member of each pair had the most probable tone while its counterpart had the least probable tone for the given syllable (see *Visual Stimuli* for the calculations of syllable frequencies and tonal probabilities). Forty-eight fillers with mid-range frequencies were recorded such that the entire syllable set played in each session reflected the natural distribution of tones and frequencies according to the corpus (see *Appendix*). All words were recorded by a 27-year-old female from Beijing (who spoke no other dialects) with the carrier phrase wo214 yao51 shuo55 ... 'I will say...' at 16 bits/44,100 Hz. One clear carrier phrase with a 100-ms pause was spliced in before each of the 96 words. Acoustic analyses of the target words revealed previously reported durational differences across tone types (Dreher & Lee, 1968; Ho, 1976; Howie, 1976; Jongman, Wang, Moore, & Sereno, 2006), but the mean duration and F0 contour of each tone did not vary across conditions (Table 2).

### *Visual stimuli*

A total of 96 (48 critical, 48 filler) slides were created. Because the target syllable–tone combinations selected from the corpus were not equally identifiable concrete objects, written forms of words (i.e., Chinese characters) were used as the visual stimuli. Although a typical visual-world paradigm uses pictures or objects (e.g., Trueswell & Tanenhaus, 2005), several studies have established the viability of the methodology using printed words both in Latin script (Dahan, Drucker, & Scarborough, 2008; McQueen & Viebahn, 2007; Salverda & Tanenhaus, 2010) and in Chinese characters (Shen et al., 2013).

Each slide contained four Chinese characters: target, segmental competitor, rhyme competitor and distractor. Twelve unique slides belonged to each of the four conditions in Table 1. The segmental competitor had the

Table 2. Mean duration and F0 of stimuli by condition and tone (with one-way repeated measures ANOVA by condition).

	Overall mean	FH-PH	FH-PL	FL-PH	FL-PL	F-score
Tone 1 (55)						
Mean duration (ms)	548	528	616	553	496	$F(3,10) = 1.52, p = .27$
Mean F0 (Hz)	304.7	280.8	293.4	330.9	308.1	$F(3,10) = 1.42, p = .30$
Tone 2 (35)						
Mean duration (ms)	563	600	505	560	587	$F(3,8) = 0.80, p = .52$
Mean F0 (Hz)	249.1	292.1	232.3	285.4	203.2	$F(3,8) = 0.64, p = .61$
Tone 3 (214)						
Mean duration (ms)	739	869	673	742	672	$F(3,6) = 3.11, p = .11$
Mean F0 (Hz)	228.1	218.1	236.0	246.2	225.1	$F(3,6) = 0.54, p = .67$
Tone 4 (51)						
Mean duration (ms)	463	409	499	419	525	$F(3,8) = 2.53, p = .13$
Mean F0 (Hz)	285.3	277.3	303.6	277.6	272.1	$F(3,8) = 2.48, p = .14$

identical syllable as the target but the opposite tonal probability. The rhyme competitor shared the rhyme and tone with the target, but had a different onset. The distractor had a distinct onset, vowel and tone from the other three characters. The spoken and printed word frequency and the numbers of strokes of the four characters on each slide were matched using three different frequency corpora (Cai & Brybaert, 2010; Liu, Shu, & Li, 2007; Wang, 1986). See Figure 2 for an annotated example. Twenty-four filler slides presented two visually similar characters while the remaining 24 filler slides presented two semantically related characters.

Syllable frequencies were calculated by taking the log of the summed tokens of all words in which the given syllable appeared as the first or only syllable (regardless of tone). The first syllable was used since it captures all members of the cohort in the lexicon (Packard, 1999, 2000; Zhou & Marslen-Wilson, 1994, 1995). Because the median common log frequency of all syllables was 4.4 (range: 1.0–6.3), syllables with log frequencies above 4.8 were classified as high frequency while those below 4.0 were classified as low

frequency. Tonal probabilities were calculated by dividing the summed token count for a given tone–syllable combination by the summed token count of that specific syllable. The most probable tone occurred in over 50% of the tokens, while the least probable tone occurred in less than 20% of the tokens with that syllable.

**Procedure**

Participants’ eyes were first calibrated with the ClearView 5-point calibration programme before hearing instructions in Mandarin. Participants completed 4 practice trials, 24 experimental trials, and 44 fillers. Trials were presented in a fixed pseudo-random order in four different lists, with each list separated into two blocks of 36 trials. At least two filler trials always separated the critical trials with a 2-second inter-trial interval. On each trial, participants fixated on the centre of the screen during the carrier phrase ‘I will say...’, then the character set and the target monosyllabic word were simultaneously presented. Participants were instructed to click on the character that

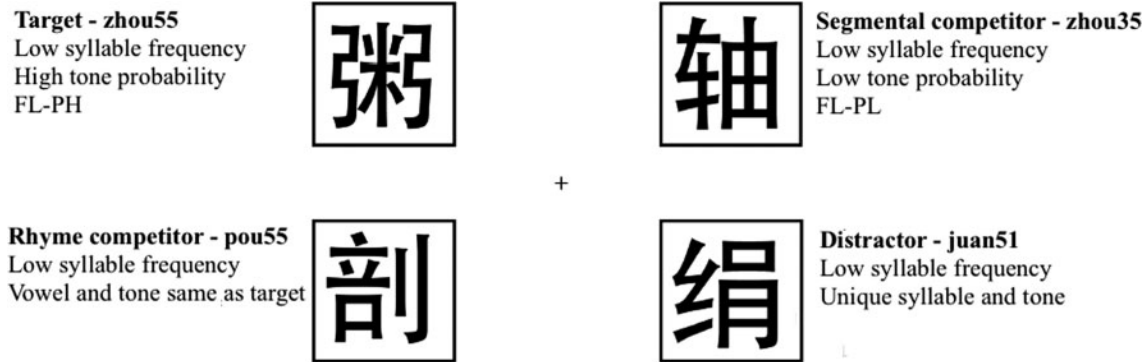


Figure 2. Example slide with a character set.  
 Target 粥 zhou55 ‘porridge’  
 Segmental Competitor 轴 zhou35 ‘axle’  
 Rhyme Competitor 剖 pou55 ‘to cut open’  
 Distractor 绢 juan51 ‘silk fabric’.

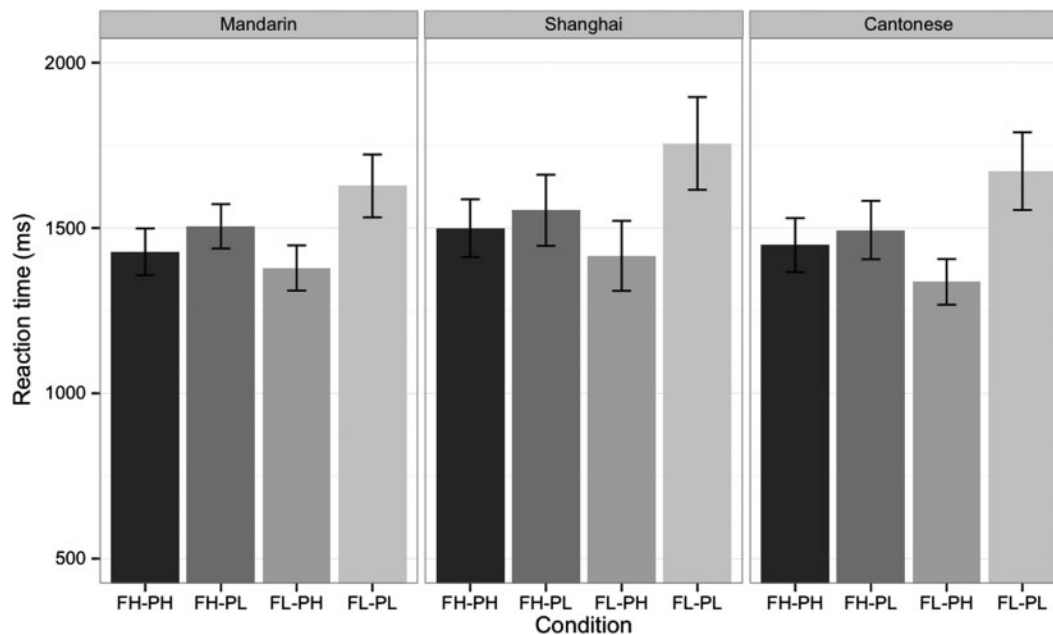


Figure 3. Mouse click response latencies (with 95% confidence intervals).

matched the perceived word. Their eyes were continuously tracked at 50 Hz using the Tobii 1750 system.

### Results

Results were analyzed using mixed-effects models (Baayen, Davidson, & Bates, 2008; Barr, 2008; Jaeger, 2008) in R's (R Development Core Team, 2013) lmer package (Bates, Maechler, & Bolker, 2011). Models included syllable frequency, tonal probability and language group as predictor factors and intercept and maximal random slope structure for both subject and item (models are provided at the top of tables) (Barr, Levy, Scheepers, & Tily, 2013). In each model, syllable frequency and tonal probability were sum coded while language was Helmert coded with two planned contrasts – mono-dialectal group versus the mean of the bi-dialectal groups, and Cantonese versus Shanghai. The  $p$ -values were obtained by likelihood ratio tests of the full model against the model without the effect in question.

Data from two target-cohort sets were excluded due to low accuracies in mouse clicking (50% and 55% respectively). The 44 remaining trial sets all had 80% or above accuracies. Across groups, participants never made incorrect clicks on rhymes or distractors; all mistakes ( $n = 79$ ) were made solely on segmental competitors (less than 5%).

### Click time analysis

The click response times showed identical patterns across the groups: while infrequent syllables with most probable tones were clicked the fastest, infrequent syllables with least probable tones were clicked the slowest (Figure 3).

A mixed-effects regression on response times did not reveal an effect of language group but found an effect of tonal probability and an interaction between syllable frequency and tonal probability (Table 3). Additionally, a marginal three-way interaction between syllable frequency, tonal probability and language group was found when the mono-dialectal speakers were compared to the bi-dialectal speakers. Further analyses within groups revealed that all three groups exhibited the two-way interaction between syllable frequency and tonal probability: Shanghai bi-dialectals ( $\beta = 449.09$ , standard error [ $SE$ ] = 188.72,  $t = 2.28$ ,  $p = .01$ ), Cantonese bi-dialectals ( $\beta = 340.62$ ,  $SE = 154.16$ ,  $t = 2.15$ ,  $p = .02$ ) and Mandarin mono-dialectals ( $\beta = 286.75$ ,  $SE = 137.95$ ,  $t = 2.03$ ,  $p = .04$ ).

### Fixation analysis

Following Barr (2008) and Barr, Gann, & Pierce (2011), grand mean fixation proportions across conditions and groups were plotted (Figure 4) and the empirical logit (elogit) (which converts fixation proportions to a continuous scale without upper/lower bounds) and the associated weight were calculated for each 200-ms time window for each trial. While approximately 200 ms is required to plan an eye movement (Matin, Shao, & Boff, 1993), the grand mean functions showed a divergence of segmental cohorts (target and competitor) from the distractors (rhyme and unrelated) at about 300 ms, indicating that participants started moving their eyes according to segmental information after 300 ms from the poststimulus onset. Thus, 200-ms analysis windows were set from 300–500 ms until the target reached its peak (900–1100 ms).

Table 3. Summary of mixed-effect regressions on mouse click response times:  $R$ : response time  $\sim$  frequency  $\times$  probability  $\times$  language + (1 + probability  $\times$  frequency|subject) + (1 + language|item).

Predictor	Est.	SE	<i>t</i>
(Intercept)	1513.66	43.96	34.43
Frequency	-48.04	71.14	-0.68
Probability	-212.72	71.17	-2.99*
Language mono-dialect vs. bi-dialect	-57.88	61.04	-0.95
Language Cantonese vs. Shanghai	-63.04	72.81	-0.87
Frequency:probability	325.18	142.33	2.28*
Frequency:language mono-dialect vs. bi-dialect	47.19	53.96	0.87
Frequency:language Cantonese vs. Shanghai	-25.19	69.12	-0.36
Probability:language mono-dialect vs. bi-dialect	68.27	53.98	1.26
Probability:language Cantonese vs. Shanghai	13.38	69.21	0.19
Frequency:probability:language mono-dialect vs. bi-dialect	-224.87	107.99	-2.08^
Frequency:probability:language Cantonese vs. Shanghai	-118.05	138.35	-0.85

Note: All fixed-effect *p*-values were obtained by likelihood ratio tests. *N* = 1223: total number of observations within the model. \**p* < .05, ^*p* < .10.

Unlike mouse click latencies, the fixation patterns clearly differentiated mono-dialectal from bi-dialectal speakers. Mono-dialectal speakers showed a clear competition between the target and segmental competitor for low-frequency syllables (Figure 5, FL-PH vs. FL-PL), whereas bi-dialectal speakers did not. In mono-dialectal participants, fixations to the target increased most quickly in the FL-PH condition and most slowly in the FL-PL condition, reflecting the inversely related looks to the segmental competitor (Figure 6). Mixed-effects models on weighted elogits for target showed the main effect of probability in the 700- to 900-ms window (which was also marginally significant in the 900- to 1100-ms window), a two-way interaction between syllable frequency and tonal probability and a three-way interaction between syllable frequency, tonal probability and language group (mono-

vs. bi-dialectal) in the 300- to 500-ms and 500- to 700-ms windows (Table 4). Further analyses within groups confirmed the two-way interaction between syllable frequency and tonal probability only in mono-dialectal speakers for the 300- to 500-ms window ( $\beta = 1.99$ ,  $SE = 0.63$ ,  $t = 3.14$ ,  $p = .001$ ) and the 500- to 700-ms window ( $\beta = 2.30$ ,  $SE = 0.57$ ,  $t = 4.03$ ,  $p = .001$ ).

Bi-dialectal speakers did not exhibit this effect of tonal probability for low-frequency syllables (300–500 ms:  $\beta = 0.93$ ,  $SE = 0.47$ ,  $t = 1.96$ ,  $p > .05$ ; 500–700 ms:  $\beta = 0.37$ ,  $SE = 0.47$ ,  $t = 0.79$ ,  $p > .05$ ) nor did they exhibit unique, dialect-specific effects; dialect-specific models showed null interaction between syllable frequency and tonal probability (300–500 ms: Shanghai:  $\beta = 0.31$ ,  $SE = 0.59$ ,  $t = 0.53$ ,  $p > .05$ ; Cantonese:  $\beta = 0.34$ ,  $SE = 0.68$ ,  $t = 0.50$ ,  $p > .05$ ; 500–700 ms: Shanghai:  $\beta = -0.70$ ,

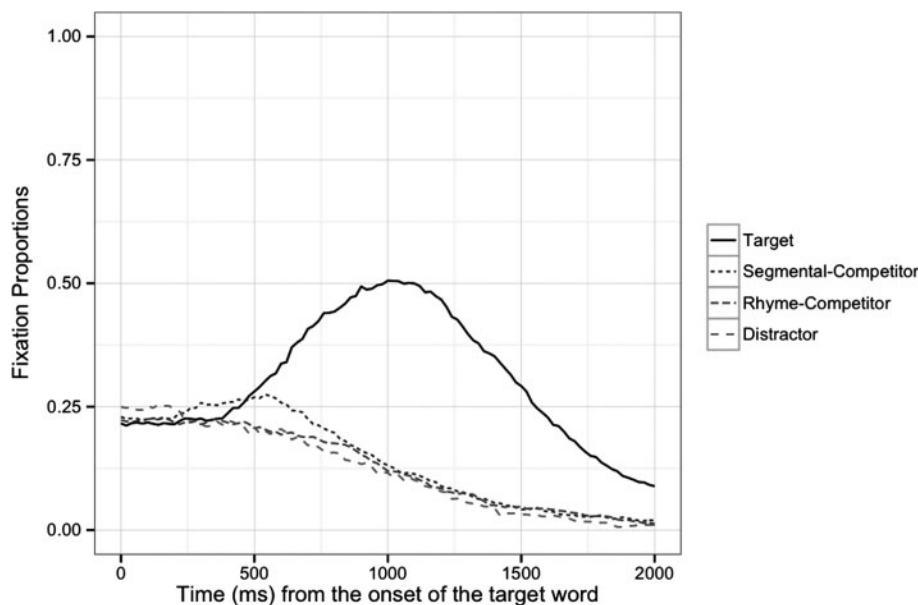


Figure 4. Grand mean fixation proportions.

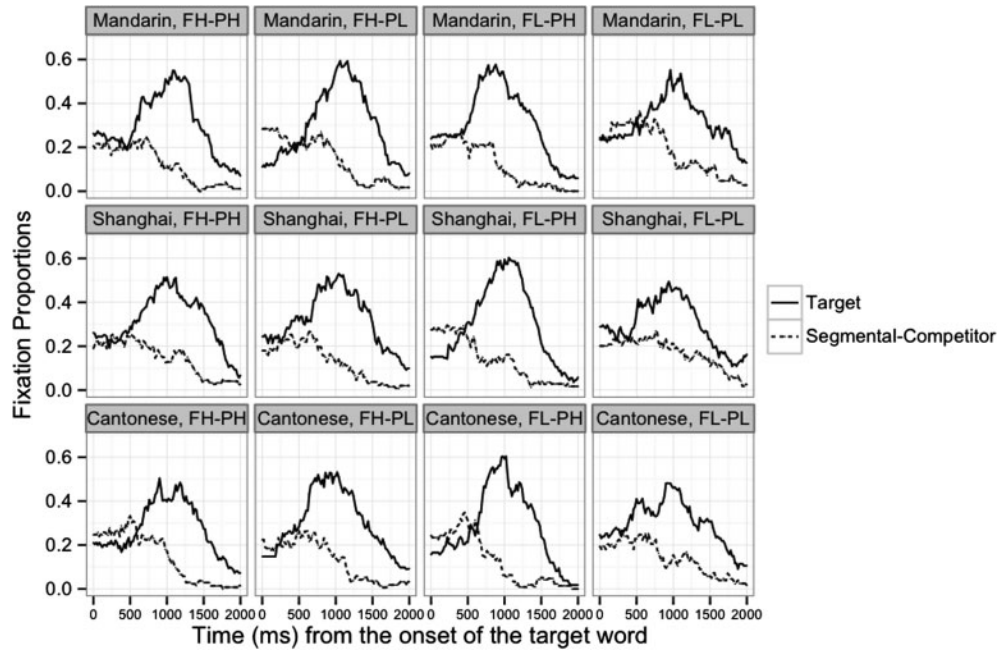


Figure 5. Language group condition fixation proportions (target vs. segmental competitor).

$SE = 0.64, t = -1.09, p > .05$ ; Cantonese:  $\beta = -1.02, SE = 0.70, t = -1.46, p > .05$ ) presumably due to less competition from the segmental competitor in FL-PL condition (Figure 6).

To test whether a stimulus of infrequent syllables with improbable tones led to a higher number of looks to the segmental competitor with more probable tones, mixed-effects models tested the weighted elogits for the segmental competitor from 300–700 ms (Table 5), where

looks to the competitor diverged from looks to distractors (Figure 4). In the 300- to 500-ms window, a main effect of probability, a two-way interaction between probability and language group (mono- vs. bi-dialectal) and a three-way interaction between probability, frequency and language group (mono- vs. bi-dialectal) were found. In the 500- to 700-ms window, a two-way interaction between probability and language group (mono- vs. bi-dialectal) and a two-way interaction between frequency and probability

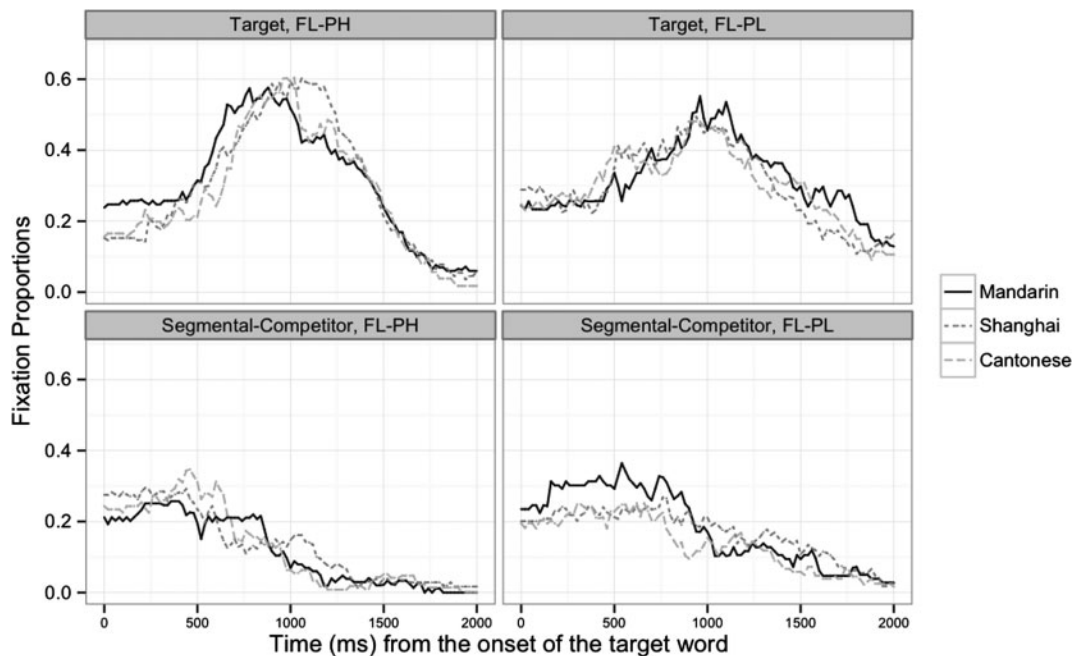


Figure 6. FL-PH and FL-PL fixation proportions by language group (target vs. segmental competitor).

Table 4. Summary of weighted mixed-effect regressions on target fixations (Helmert contrast coding):  $R: \text{e log} \sim \text{frequency} \times \text{probability} \times \text{language} + (1 + \text{probability} \times \text{frequency} | \text{subject}) + (1 + \text{language} | \text{item})$ .

Window	300–500 ms			500–700 ms		
	Est.	SE	t	Est.	SE	t
(Intercept)	-1.49	0.12	-12.09	-1.23	0.11	-10.69
Frequency	-0.38	0.24	-1.56	-0.40	0.22	-1.79
Probability	-0.36	0.24	-1.49	-0.34	0.22	-1.54
Language mono-dialect vs. bi-dialect	-0.01	0.12	0.10	-0.12	0.15	-0.84
Language Cantonese vs. Shanghai	-0.12	0.15	-0.86	-0.01	0.17	-0.03
Frequency:probability	1.28	0.48	2.64*	1.04	0.47	2.20*
Frequency:language mono-dialect vs. bi-dialect	-0.46	0.22	-2.03^	-0.13	0.28	-0.46
Frequency:language Cantonese vs. Shanghai	-0.26	0.27	-0.97	0.16	0.34	-0.47
Probability:language mono-dialect vs. bi-dialect	-0.01	0.23	-0.03	0.31	0.27	1.10
Probability:language Cantonese vs. Shanghai	-0.42	0.27	-1.54	-0.52	0.33	-1.52
Frequency:Probability:Language mono-dialect vs. bi-dialect	0.98	0.46	2.12*	-1.80	0.64	2.77*
Frequency:Probability:language Cantonese vs. Shanghai	0.25	0.55	0.45	0.18	0.76	0.24

Window	700–900 ms			900–1100 ms		
	Est.	SE	t	Est.	SE	t
(Intercept)	0.63	0.11	5.63	0.34	0.10	3.29
Frequency	0.10	0.20	0.54	-0.09	0.19	-0.49
Probability	0.60	0.21	2.86*	0.40	0.19	2.05^
Language mono-dialect vs. bi-dialect	0.22	0.17	1.26	0.24	0.16	1.47
Language Cantonese vs. Shanghai	0.05	0.21	0.27	-0.02	0.20	-0.10
Frequency:probability	-0.93	0.56	-1.64	-0.78	0.44	-1.74
Frequency:Language mono-dialect vs. bi-dialect	-0.01	0.28	-0.02	-0.05	0.30	-0.18
Frequency:language Cantonese vs. Shanghai	0.42	0.34	1.24	-0.08	0.36	-0.23
Probability:Language mono-dialect vs. bi-dialect	-0.06	0.32	-0.19	0.12	0.31	0.39
Probability:language Cantonese vs. Shanghai	0.31	0.38	0.81	0.14	0.37	0.37
Frequency:probability:language mono-dialect vs. bi-dialect	0.04	1.02	0.04	0.21	0.77	0.27
Frequency:probability:language Cantonese vs. Shanghai	-0.55	1.19	-0.46	-0.69	0.91	-0.76

Note:  $N = 996$ : total number of observations considered in the model. \* $p < .05$ , ^ $p < .10$ .

were shown. Further analyses confirmed the interaction between frequency and probability only in mono-dialectal speakers for the 300- to 500-ms window ( $\beta = -0.78$ ,  $SE =$

0.34,  $t = -2.29$ ,  $p = .02$ ). Shanghai and Cantonese groups showed no difference in the use of tonal probability in either window (i.e., no interaction between frequency and

Table 5. Summary of weighted mixed-effect regressions on segmental-competitor fixations:  $R: \text{e log} \sim \text{frequency} \times \text{probability} \times \text{language} + (1 + \text{probability} \times \text{frequency} | \text{subject}) + (1 + \text{language} | \text{item})$ .

Window	300–500 ms			500–700 ms		
	Est.	SE	t	Est.	SE	t
(Intercept)	-1.25	0.19	-6.41	-1.18	0.16	-7.20
Frequency	0.14	0.39	0.35	-0.19	0.32	-0.59
Probability	-0.35	0.16	-2.19*	-0.37	0.20	-1.80
language mono-dialect vs. bi-dialect	0.02	0.10	0.25	-0.01	0.11	-0.12
Language Cantonese vs. Shanghai	0.02	0.12	0.20	0.12	0.14	0.86
Frequency:probability	-0.48	0.32	-1.49	-0.87	0.41	-2.14*
Frequency:language mono-dialect vs. bi-dialect	-0.12	0.20	-0.60	0.02	0.23	0.09
Frequency:language Cantonese vs. Shanghai	0.10	0.24	0.42	0.12	0.28	0.44
Probability:language mono-dialect vs. bi-dialect	-0.60	0.24	-2.47*	-0.72	0.26	-2.77*
Probability:language Cantonese vs. Shanghai	0.07	0.24	0.30	0.16	0.28	0.60
Frequency:probability:language mono-dialect vs. bi-dialect	-1.04	0.49	-2.11*	-0.06	0.52	-0.12
Frequency:probability:language Cantonese vs. Shanghai	0.23	0.49	0.47	-0.01	0.56	-0.01

Note:  $N = 996$ : total number of observations considered in the model. \* $p < .05$ .

probability for either language group. 300–500 ms: Shanghai:  $\beta = -0.52$ ,  $SE = 0.80$ ,  $t = -0.65$ ,  $p > .05$ ; Cantonese:  $\beta = -0.31$ ,  $SE = 0.87$ ,  $t = -0.35$ ,  $p > .05$ ; 500–700 ms: Shanghai:  $\beta = -0.84$ ,  $SE = 0.64$ ,  $t = -1.30$ ,  $p > .05$ ; Cantonese:  $\beta = -0.96$ ,  $SE = 0.84$ ,  $t = -1.14$ ,  $p > .05$ ).

## Discussion

The current study presents two important findings about the function of lexical tone in Mandarin word recognition. First, although it is contrastive and conveys lexical semantics, tone does not seem to be represented equally across the 400 distinct syllables. In all three language groups tested, participants clicked on targets with most probable tones much faster than those with least probable tones – but crucially, only for the syllables with low frequencies. The fact that tonal probability affected click times only for infrequent syllables suggests that tonal information constrains or delimits which lexeme is accessed more for infrequent syllables than for frequent syllables. Because tonal information may not be enough to reduce the cohort size of highly frequent syllables, which may still have upwards of 20 common homophones (Packard, 1999, 2000; Yin, 1984), the role of tone and the importance of a ‘toneme’ feature (Ye & Connine, 1999) may depend on the frequency and probability of sounds co-occurring over time in speech.

Second, while proficient bi-dialectal speakers exhibited similar clicking patterns as mono-dialectal speakers, their knowledge of L2 tonal probabilities did not affect their online responses to incoming speech. Only mono-dialectal speakers showed reliably higher initial fixations to the segmental competitor for infrequent syllables with least probable tones. This group was more likely to fixate on the segmental competitor prior to the target, indicating that as soon as the critical segmental sequence was identified, a particular, highly probable tone associated with the given syllable was activated, leading to anticipatory eye movements. Thus, knowledge of tonal probabilities triggered immediate probability-based processing in the mono-dialectal group. Since both bi-dialectal groups showed an interaction between syllable frequency and tonal probability in their mouse click responses, bi-dialectal speakers must represent Mandarin tones similarly. The present findings, however, suggest that tonal information may not be accessed as swiftly during L2 processing. This finding further strengthens the claim that even among highly proficient bilingual speakers of sister languages, subtle processing differences exist (cf. Zhang, Samuel, & Liu, 2012).

Surprisingly, despite different amounts of exposure to Mandarin and the vastly disparate syllable–tone inventories, the predicted differences between the Cantonese and Shanghai groups did not emerge. The findings of the present study suggest that experience with a more or less complicated L1 tonal system may have little effect on

adult L2 tonal processing. Although the predicted dialectal differences may be observed in non-proficient populations such as elderly bi-dialectal speakers or those beginning to learn Mandarin in school (cf. Chen et al., 2004, 2008), the null group effect in the present study suggests that bi-dialectal speakers of Mandarin above a certain proficiency level may not differ in the efficacy of tone processing during L2 word recognition.

In sum, tonal information seems to be selectively represented according to how often each syllable appears in speech in both native and highly proficient non-native speakers of Mandarin. Our findings indicate that the tonal cue validity is higher for infrequent words/syllables, which are generally considered to have higher activation thresholds and thus may require more detailed phonological representations. Further research is necessary to clarify when and how learners of a tone language start tuning toward such dynamic cue validity of lexical tones and how the variation in the input in diglossia interacts with this process.

## Acknowledgements

This research was partially funded by a grant to the first author through The Center for Cognitive and Brain Sciences at The Ohio State University. This article reflects ideas and analysis that have benefited from critique, advice and support from the following individuals: Shari Speer, Marjorie Chan, Duanmu San, Chen Yiya, Kodi Weatherholtz and the audiences at the North American Conference on Chinese Linguistics hosted by the University of Michigan, Ann Arbor, and the International Conference on Multilingualism hosted by McGill University, Montreal, Canada. We are also very grateful to James Myers, the editors of this special issue, and two other anonymous reviewers for feedback that substantially improved the article.

## Note

1. Tone values follow Chao’s (1968) numerical notation of a starting and ending pitch, with 5 corresponding to the highest pitch and 1 the lowest.

## References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 413–425. doi:10.1016/j.jml.2007.12.005
- Barr, D. J. (2008). Analyzing ‘visual world’ eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59, 457–474. doi:10.1016/j.jml.2007.09.002
- Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, 137, 201–207. doi:10.1016/j.actpsy.2010.09.011
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using Eigen and S4 classes. R package version 0.999375-41. Retrieved from: <http://CRAN.R-Project.org/package=lme4>

- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin: Mouton de Gruyter.
- Brown-Schmidt, S., & Canseco-Gonzalez, E. (2004). Who do you love, your mother or your horse? An event-related brain potential analysis of tone processing in Mandarin Chinese. *Journal of Psycholinguistic Research*, 33, 103–135. doi:10.1023/B:JOPR.0000017223.98667.10
- Burnham, D., & Jones, C. (2002). Categorical perception of lexical tone by tonal and non-tonal language speakers. In C. Bow (Ed.), *Proceedings of the 9th Australian International Conference on Speech Science and Technology*, Melbourne (pp. 515–519). Canberra: Australian Speech Science and Technology Association (ASSTA).
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *Plos ONE*, 5, e10729. doi:10.1371/journal.pone.0010729
- Chao, Y. R. (1941). Distinctions within ancient Chinese. *Harvard Journal of Asiatic Studies*, 5, 303–322. doi:10.2307/2717913
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.
- Chen, P. (1999). *Modern Chinese: History and sociolinguistics*. New York, NY: Cambridge University Press.
- Chen, X., Anderson, R. C., Li, W., Hao, M., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology*, 96, 142–151. doi:10.1037/0022-0663.96.1.142
- Chen, X., Ku, Y.-M., Koyama, E., Anderson, R. C., & Li, W. (2008). Development of phonological awareness in bilingual Chinese children. *Journal of Psycholinguistic Research*, 37, 405–418. doi:10.1007/s10936-008-9085-z
- Chen, Y. (2008). Revisiting the phonetics and phonology of tone Sandhi in Shanghai Chinese. In *Proceedings of the Speech Prosody 2008*, Brazil (pp. 253–256).
- Chen, Y. (2011). How does phonology guide phonetics in segment-f0 interaction? *Journal of Phonetics*, 39, 612–625. doi:10.1016/j.wocn.2011.04.001
- Chinese National Committee on Language Affairs. (2004). Zhongguo yuyan wenzi shiyong qingkuang diaocha [Survey of spoken and written Chinese]. Retrieved from <http://www.chinalanguage.gov.cn/doc/zhongguodiaocha2004-12/shujv.doc>
- Clements, G. N., & Goldsmith, J. (1984). *Autosegmental studies in Bantu tone*. Dordrecht: Foris Press
- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45, 207–228. doi:10.1177/00238309020450030101
- Creel, S. C., Tanenhaus, M. K., & Aslin, R. N. (2006). Consequences of lexical stress on learning an artificial lexicon. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 15–32. doi:10.1037/0278-7393.32.1.15
- Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech*, 29, 201–220. doi:10.1177/00238309010440020301
- Cutler, A. (2005). Lexical stress. In D. B. Pisoni & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 264–289). Oxford: Blackwell.
- Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics*, 59, 165–179. doi:10.3758/BF03211886
- Cutler, A., & van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech*, 44, 171–195. doi:10.1177/00238309010440020301
- Dahan, D., Drucker, S. J., & Scarborough, R. A. (2008). Talker adaptation in speech perception: Adjusting the signal or the representations? *Cognition*, 108, 710–718. doi:10.1016/j.cognition.2008.06.003
- Deher, J. J., & Lee, P. C. (1968). Instrumental investigation of single and paired Mandarin tonemes. *Monumenta Serica*, 27, 343–373.
- Duanmu, S. (1999). Metrical structure and tone: Evidence from Mandarin and Shanghai. *Journal of East Asian Linguistics*, 8, 1–38. doi:10.1023/A:1008353028173
- Duanmu, S. (2007). *The phonology of standard Chinese* (2nd ed.). New York, NY: Oxford University Press.
- Duanmu, S. (2008). *Syllable structure: The limits of variation*. New York, NY: Oxford University Press.
- Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, 1893–1904. doi:10.1121/1.412063
- Flege, J. (1995). Second language speech learning: Theory, findings, and problems. In Winifred Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore, MD: York Press.
- Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, 11, 149–176.
- Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33, 353–367. doi:10.1159/000259792
- Hombert, J.-M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, 55, 37–58. doi:10.2307/412518
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge: Cambridge University Press.
- Huang, T. (2012). Cross-linguistic and inter-dialectal differences in tone perception by native speakers of three Chinese dialects and American English. *Journal of Chinese Linguistics*, 40, 155–179.
- Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica*, 67, 243–267. doi:10.1159/000327392
- Hyman, L. M., & Kisseberth, C. (1998). *Theoretical aspects of Bantu tone*. Stanford, CA: C.S.L.I.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59, 434–446. doi:10.1016/j.jml.2007.11.007
- Jongman, A., Wang, Y., Moore, C. B., & Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng (Eds.), *The handbook of East Asian psycholinguistics, Vol. 1: Chinese* (pp. 209–217). Cambridge: Cambridge University Press.
- Karlgren, B. (1954). Compendium of phonetics in ancient and archaic Chinese. *Bulletin of the Museum of Far Eastern Antiquities*, 26, 211–367.
- Lee, C. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, 50, 101–123. doi:10.1177/00238309070500010501
- Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25, 527–542. doi:10.1007/BF01758181
- Leung, M.-T., Law, S.-P., & Fung, S.-Y. (2004). Type and token frequencies of phonological units in Hong Kong Cantonese. *Behavior Research Methods, Instruments, and Computers*, 36, 500–505. doi:10.3758/BF03195596
- Li, F.-K. (1973). Language and dialects of China. *Journal of Chinese Linguistics*, 1, 1–13.

- Li, D. C. S. (2006). Chinese as a lingua franca in Greater China. *Annual Review of Applied Linguistics*, 26, 149–176. doi:10.1017/S0267190506000080
- Li, D. C. S., & Lee, S. (2004). Bilingualism in East Asia. In T. K. Bhatia & B. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 742–779). Malden, MA: Blackwell.
- Liu, Y., Shu, H., & Li, P. (2007). Word naming and psycholinguistic norms: Chinese. *Behavior Research Methods*, 39, 192–198. doi:10.3758/BF03193147
- 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*, 64, 407–420. doi:10.1016/j.jml.2010.02.004
- Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: An ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia*, 50, 2032–2043. doi:10.1016/j.neuropsychologia.2012.05.002
- Martin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Attention, Perception, & Psychophysics*, 53, 372–380. doi:10.3758/BF03206780
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86. doi:10.1016/0010-0285(86)90015-0
- McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, 60, 661–671. doi:10.1080/17470210601183890
- Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, 52, 189–234. doi:10.1016/0010-0277(94)90043-4
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115, 357–395. doi:10.1037/0033-295X.115.2.357
- Norris, D., McQueen, J. M., and Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences*, 23, 299–325. doi:10.1017/S0140525X00003241
- Odden, D. (1995). Tone: African languages. In John Goldsmith (Ed.), *The handbook of phonological theory*. Oxford: Basil Blackwell.
- Packard, J. L. (1999). Lexical access in Chinese speech comprehension and production. *Brain and Language*, 68, 89–94. doi:10.1006/brln.1999.2102
- Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge: Cambridge University Press.
- R Development Core Team (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from: <http://www.R-project.org>
- Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1108–1117. doi:10.1037/a0019901
- Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., and Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience*, 17, 1–12. doi:10.1162/0898929052880057
- Shen, J., Deutsch, D., & Rayner, K. (2013). On-line perception of Mandarin tones 2 and 3: Evidence from eye movements. *Journal of the Acoustical Society of America*, 133, 3016–3029. doi:10.1121/1.4795775
- Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, 45, 412–432. doi:10.1006/jmla.2000.2783
- Sulpizio, S., & McQueen, J. M. (2012). Italians use abstract knowledge about lexical stress during spoken-word recognition. *Journal of Memory and Language*, 66, 177–193. doi:10.1016/j.jml.2011.08.001
- Tagliapietra, L., & Tabossi, P. (2005). Lexical stress effects in Italian spoken word recognition. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings in XXVII conference of the cognitive science society* (pp. 2140–2144). Mahwah, NJ: Lawrence Erlbaum Associates.
- Trueswell, J. C., & Tanenhaus, M. K. (Eds.). (2005). *Processing world-situated language: Bridging the language-as-action and language-as-product traditions*. Cambridge, MA: MIT Press.
- Wang, H. (1986). *Modern Chinese frequency dictionary*. Beijing: Beijing Language Institute Press.
- Xu, B., Tang, Z., You, R., Qian, N., Shi, R., & Shen, Y. (1988). *Shanghai Shiqü fangyan zhi* [Urban Shanghai dialects]. Shanghai: Shanghai Jiaoyu Chubanshe.
- Xu, B., & Tao, H. (1997). *Shanghai Fangyan Cidian* [Dictionary of Shanghai dialect]. In Rong Li (Ed.), *Xiandai Hanyu Fangyan Da Cidian* [Dictionaries of modern Chinese dialects]. Nanjing: Jiangsu Jiaoyu Chubanshe.
- Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, 33, 319–337. doi:10.1016/S0167-6393(00)00063-7
- Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes*, 14, 609–630. doi:10.1080/016909699386202
- Yin, B. Y. (1984). *Hanyu yusu de dingliang yanjiu* [A quantitative research of Chinese morphemes]. *Zhongguo Yuwen* [Chinese Literature and Language], 182, 338–347.
- Yip, M. (1980). *The tonal phonology of Chinese* (Unpublished doctoral dissertation). MIT, Cambridge, MA.
- Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.
- Yuan, J. H. (1960). *Hanyu fangyan gaiyao* [The standard handbook on Chinese dialects]. Peking: Wenzhi Gaige Chubanshe.
- Zee, E., & Maddieson, I. (1980). Tones and tone Sandhi in Shanghai: Phonetic evidence and phonological analysis. *Glossa: An International Journal of Linguistics*, 14, 45–88.
- Zeng, Z. (1994). *Guangzhouhua Putonghua de duibi yu jiaoxue* [Comparison between and instruction of Cantonese and Mandarin]. Hong Kong: Joint Publishing.
- Zhang, X., Samuel, A. G., & Liu, S. (2012). The perception and representation of segmental and prosodic Mandarin contrasts in native speakers of Cantonese. *Journal of Memory and Language*, 66, 438–457. doi:10.1016/j.jml.2011.12.006
- Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: Evidence from ERP analyses. *Neuropsychologia*, 49, 1761–1770. doi:10.1016/j.neuropsychologia.2011.02.054
- Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes*, 9, 393–422. doi:10.1080/01690969408402125
- Zhou, X., & Marslen-Wilson, W. (1995). Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes*, 10, 545–600. doi:10.1080/01690969508407114
- Zhu, X. S. (1999). *Shanghai tonetics*. München: Lincom Europa.

**Appendix. Mean frequencies of each condition**

	Mean syllable Frequency (log)	Mean spoken-word Frequency (log)	Mean written-word Frequency (log)	Mean number of strokes
FH-PH Target	5.1	2.9	2.7	9.8
FH-PH Segmental-Competitor	5.1	2.9	2.7	9.8
FH-PH Rhyme	4.9	3.3	3.0	10.7
FH-PH Distractor	5.1	3.3	2.6	8.8
FH-PL Target	5.1	2.9	2.2	9.8
FH-PL Segmental-Competitor	5.1	2.9	2.2	9.8
FH-PL Rhyme	4.9	3.0	2.8	9.7
FH-PL Distractor	5.1	3.3	2.6	8.8
FL-PH Target	3.8	3.0	2.6	9.1
FL-PH Segmental-Competitor	3.8	3.0	2.6	8.8
FL-PH Rhyme	3.5	2.6	2.2	10.5
FL-PH Distractor	3.9	3.0	2.7	8.8
FL-PL Target	3.8	3.0	2.4	9.0
FL-PL Segmental-Competitor	3.8	3.0	2.4	9.3
FL-PL Rhyme	3.5	2.5	2.5	9.3
FL-PL Distractor	3.9	3.0	2.7	9.3