2024. *Proc Ling Soc Amer* 9(1). 5664. https://doi.org/10.3765/plsa.v9i1.5664.

Bidialectal acoustic realization of tones is influenced by dialect experience and homophone status

Wenqi Zeng & Christine Shea*

Abstract. While evidence shows that interlingual cognates can enhance cross-language phonetic assimilation in production, it is reasonable to assume that interlingual homophones can enhance cross-language phonological interference. Distinct from cognates, interlingual homophones do not share semantic content, which may affect the degree of co-activation observed across languages. The present study examines this hypothesis in a group of bidialectal speakers, whose lexicon consists of a large number of inter-dialectal homophones. Productions of Chengdu Mandarin tones by Chengdu Mandarin and Standard Mandarin speakers were examined in a word naming task. The results showed that bidialectal speakers' native tone productions were influenced by their experience in speaking Standard Mandarin as well as the inter-dialectal homophone status of the lexical item. Additionally, both of these influences were modulated by the structure of the inter-dialect tone categories. The findings support the similarity between bidialectal and bilingual speech processing and provide novel evidence for bilingual speech models from the level of suprasegmental processing.

Keywords. phonetic dissimilation; phonetic assimilation; tone production; bidialectalism; homophones

1. Introduction. Over the course of second language (L2) acquisition, people learn the sound system of their target language by mapping the L2 sounds to acoustically similar sound categories in the first language (L1). This leads to the creation of a shared phonological system and phonetic space where the L1 and L2 systems co-exist (Best 1995, Best & Tylor 2007, Flege 1995, Flege & Bohn 2021, Kuhl 1991, Kuhl et al. 1992). Since both systems co-exist in the same space, mutual influence occurs, and the acquisition of an L2 sound system can reshape the native phonetic categories. Most existing studies on the interaction between L1-L2 sound systems have focused on segmental features (e.g., Flege , Schirru & MacKay 2003, MacKey et al. 2001, Obsorne & Simonet 2021), and little is known about their interactions at the suprasegmental levels, such as tones. Moreover, even fewer studies examine how the use of a second tonal system affects the production of the first in bidialectal speakers. In this study we present data from speakers of Chengdu Mandarin and Standard Mandarin to address this question.

1.1. L2 INFLUENCES ON L1 PRODUCTION. According to the Revised Speech Learning Model (SLM-r, Flege & Bohn 2021), there are two ways in which an L1 category can be affected by learning an acoustically similar L2 sound, potentially leading to the formation of a new L2 category. If the perceived distance between the L2 sound and the closest L1 category is sufficient to form a new L2 category, the L1 category will drift away from the L2 category (L1-L2 dissimilation) as speakers strive to maintain cross-language phonological contrast. Alternatively, if the perceived distance is not sufficient to form a separate category for the L2 sound, speakers will

* Authors: Wenqi Zeng, University of Iowa (wenqi-zeng@uiowa.edu) & Christine Shea, University of Iowa (christine-shea@uiowa.edu).

form a composite category, in which L1 and L2 sounds will drift towards each other (L1-L2 assimilation).

Evidence for L1-L2 assimilation and dissimilation mostly came from studies investigating the acquisition of L2 stops that have voice-onset times (VOT) different from the L1 stops. For example, most Romance languages like Spanish or French are true voicing languages, in which the voiced stops are pre-voiced and the voiceless stops have a short-lag VOT (e.g., Abramson & Lisker, 1973; Flege & Eefting, 1986). On the other hand, aspiration languages, such as English, have voiced stops characterized by a short-lag VOT and voiceless stops by a long-lag VOT (e.g., Cho & Ladefoged 1999, Lisker & Abramson, 1964). In a study of Italian-English bilinguals, MacKay et al. (2001) reported an L2 effect on the production of L1 stops. Specifically, the researchers found that Italian-English bilinguals whose production of English voiced stops was more target-like had less pre-voicing (more English-like) in their Italian voiced stops. The results suggested an assimilation pattern to English voiced stops of the participants' native voiced stops.

On the other hand, Osborne and Simonet (2021) examined the productions of Portuguese stops by a group of monolingual speakers and a group of Portuguese English learners. The results showed that the Portuguese English learners' productions of the voiced stops had longer pre-voicing (more Portuguese-like) than those of their monolingual counterparts, suggesting an effect of dissimilation between Portuguese and English voiced categories (also see Flege & Eefting, 1987). L1-L2 dissimilation has also been observed in the acquisition of vowels. Flege, Schirru, and MacKay (2003) examined the production of the English [eɪ], which has more formant movements than the Italian [e], by four groups of Italian-English bilinguals who differed regarding the age of arrival in Canada and the frequency of continued Italian use after arrival. The results showed that while the groups of late arrivals produced the English vowel in an Italian-like way, the early arrival group who used Italian less frequently produced the English [eɪ] with more formant movements than the native English control group.

In summary, these results suggest that category dis/assimilation patterns is influenced by factors other than mere acoustic distances, and such factors include the age of arrival to the country where L2 is spoken, the age of L2 onset, the frequency of use for each language, and the quantity and quality of L2 input (Flege & Bohn, 2021). However, most of the existing studies have focused on the production of consonants or vowels, and little attention has been paid to how category dis/assimilation effects might play out at the suprasegmental level. Therefore, it is yet unclear whether such predictions can be applied to the acquisition of a novel tonal system and what influence it can have on the native tonal categories.

1.2. Lexical effects on bilingual acoustic realization. If the two languages share one mental lexicon of a bilingual, the acoustic realization of one language may also be influenced by higher-level processes, as there can be an interaction between the phonological/phonetic and the lexical levels' representations in bilinguals' mental lexicon (Gollan, Forster & Frost, 1997). Recent studies on cognate effects have shown that L1-L2 assimilation is enhanced in interlingual cognates (Brown & Harper 2009, Mora & Nadeu 2012). For example, Amengual (2012) showed that Spanish-English bilinguals produced Spanish [t] with longer VOT (more English-like) at the onset of interlingual cognates than in Spanish-specific words. This cognate effect on bilingual acoustic realizations may be accounted for by the co-activation of both L1 and L2 phonetic representations via the overlapping semantic representation.

Unlike cognates, which share similar phonological and semantic representations across two languages, interlingual homophones have overlapping phonological forms but different semantic representations between the two languages. The effects of interlingual homophones on bilingual acoustic realizations, however, have yet to be explored. Two possible effects can be hypothesized based on the assumption of a shared bilingual lexicon. First, like cognates, if the phonetic representation of the irrelevant language is co-activated by the shared phonological representation of homophones across the two languages, L1-L2 assimilation may be enhanced. Specifically, the acoustic realization of interlingual homophones in the L1 may assimilate to acoustic features of the L2. On the other hand, because homophones have different meanings between the two languages, bilingual speakers thus may seek to maintain or increase the acoustic distance between the acoustic realizations across the two languages, hence leading to L1-L2 dissimilation in homophones.

The current study tests the effects of using a second tonal system on the native tonal categories, as well as the effects of homophones on the interaction between two tonal systems with a group of Chengdu Mandarin and Standard Mandarin bidialectal speakers. This population is an ideal target to explore these questions. First, speech processing of bidialectal speakers is similar to that of bilingual speakers, as both populations have two sound systems in a shared phonetic space. Second, both of the Mandarin dialects are tonal; therefore, bidialectal speakers of these two dialects have two tonal systems established in their mental lexicon. Third, the segmental systems of these two dialects overlap to a great degree, resulting in a large number of inter-dialectal homophones that only differ on fine-grained acoustic features at the suprasegmental level.

2. Language background. Standard Mandarin (SM) is a standardized variety based on Beijing Mandarin, the local variety spoken in the capital of China. It has been promoted as the official variety across the country since 1956 (Tang, 2017). Chengdu Mandarin (CM) is a Mandarin dialect spoken in Chengdu, a city in Southwestern China. Due to the widespread imposition of SM across China, most CM speakers are also fluent in SM, therefore, bidialectal speakers of CM and SM.

Both CM and SM have four tone categories, referred to as T1, T2, T3, and T4. The four tone categories have cross-dialectal one-to-one phonemic correspondence; that is, if a word is in T1 in SM, in most cases it is also in T1 in CM. However, each tone category has different phonetic realizations between the two dialects (He 2015, Yip 2002). As shown in Figure 1, T1 is realized with a high level pitch contour in SM but a high rising pitch contour in CM, T2 is realized with a high rising pitch contour in SM but a low falling pitch contour in CM, T3 is realized with a dipping (falling-and-rising) pitch contour in SM but a high falling pitch contour in SM, and T4 is realized with a high falling pitch contour in SM but a dipping pitch contour in CM.

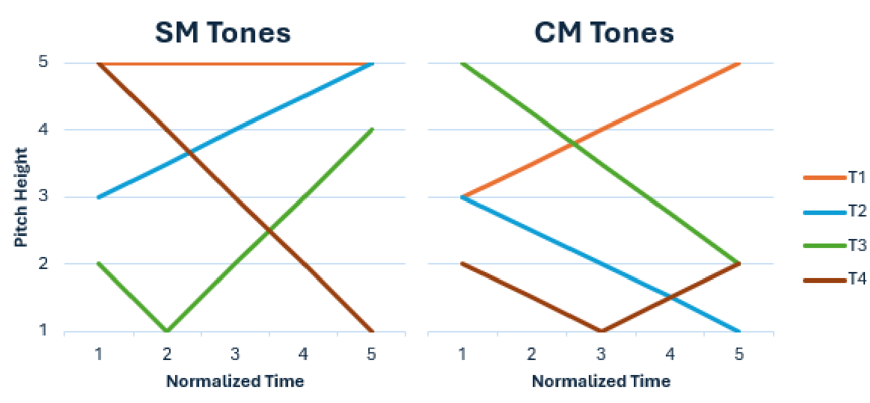

Figure 1. Relative pitch values of the four tone categories in SM (Yip, 2002) and CM (He, 2015).

Critically, it can be noted that there are pairs of tones that have similar acoustic features across the two dialects but are in different categories. Specifically, both SM-T2 and CM-T1 have a high rising pitch contour, SM-T4, CM-T2 and CM-T3 have a falling pitch contour, and both SM-T3 and CM-T4 have a dipping pitch contour. However, this similarity in pitch contour does not necessarily mean that they are identical in acoustic details. Moreover, the two dialects have a great degree of overlap in the segmental system (Cheng 1997, Tang &van Heuven 2015), resulting in a large volume of inter-dialectal homophones that may only contrast in fine-grained pitch information. Therefore, it is reasonable to suggest that using SM more frequently may encourage bidialectal CM-SM speakers to increase the phonetic distance between those similar tones to maintain cross-dialect phonological contrast, and this dissimilation effect may be modulated by inter-dialectal homophone status of the lexical item.

3. Experiment method. The current study investigated the effects of the tonal system in a second dialect on the tonal categories in the speaker's native dialect in a word-naming task. Specifically, the study addressed three research questions: (i) Are there fine-grained phonetic differences between the cross-dialectally similar tones? (ii) Are CM-SM bidialectal speakers' productions of CM tones affected by their experience in using SM? (iii) Are CM-SM bidialectal speakers' productions of CM tones affected by the homophone status of the lexical item?

3.1. PARTICIPANTS. 21 CM-SM bidialectal speakers (15 females, 6 males) and 16 SM monodialectal speakers (13 females, 3 males) participated in the study. The bidialectal group was recruited in Chengdu, and the monodialectal group was recruited in Beijing. The average age was 33.49 (*SD* = 5.53) for the bidialectal group and 26.37 (*SD* = 3.17) for the monodialectal group. All participants in the bidialectal group identified CM as their first dialect and SM as their second dialect. The range of onset age of SM for the bidialectal group was from 1 to 13, with the average being 4.78 (*SD* = 3.84). All participants in the monodialectal group identified SM as their native dialect and the only Mandarin dialect that they spoke. On a scale of 1 to 5, where 1 indicated no knowledge at all and 5 indicated native competence, the mean self-reported listening proficiency in SM was 4.76 (*SD* = 0.54) for the bidialectal group and 4.88 (*SD* = 0.32) for the monodialectal group, and the mean self-reported speaking proficiency in SM was 4.46 (*SD* = 0.95) for the bidialectal group and 4.88 (*SD* = 0.32) for the monodialectal group. For the bidialectal group, their self-reported proficiency in CM was 4.87 (*SD* = 0.33) for listening and 4.87 (*SD* = 0.33) for speaking. No participants reported any language or cognitive impairments.

3.2. WORD NAMING TASK MATERIALS. 52 monosyllabic sets were selected for the word naming task. Each set contained four stimuli with the same segmental construction but with different tone categories (T1, T2, T3, T4). The monosyllables were selected such that (i) there was no phonological or phonotactic difference between SM and CM, and (ii) each stimulus was a legal combination of tone and segments in both SM and CM. In total, 208 tonal monosyllables (52 syllables x 4 tones) were selected for the production task. Within the 52 monosyllabic sets, 32 sets had identical segmental phonetics between CM and SM and comprised the inter-dialectal homophone sets. Twenty sets were non-homophones where there were phonetic differences in either the consonants or vowels between CM and SM.

Because Mandarin dialects share one writing system (Tang, 2017), orthographic forms were chosen to present the stimuli to the participants. Due to the large number of homophones in Mandarin dialects, each tonal syllable is usually associated with multiple words, thus, multiple orthographic forms. Therefore, the most frequent orthographic form was chosen for each tonal monosyllable based on the Chinese lexical database (Sun et al., 2018), so participants would not have difficulty recognizing the words in their writing forms. The stimuli were further checked manually by the researcher who is a native CM-SM speaker to make sure each orthographic form has only one pronunciation in SM and CM. A sample of monosyllabic words used in the word naming task is provided in Table 1.

	Syllable	T1	T2	T3	T4
Homophones	[pa]	巴 “a place name”	拔 “to pull”	把 “to hold”	霸 “to dominate”
	[feɪ]	非 “not”	肥 “fat”	匪 “bandit”	费 “to cost”
Non-homophones	SM: [tɕje] CM: [tɕɛ]	阶 “step”	结 “knot”	解 “to solve”	界 “boundary”
	SM: [fɑŋ] CM: [faŋ]	方 “square”	房 “house”	访 “to visit”	放 “to put”

Table 1. Sample monosyllabic words in the word naming task.

To decrease the potential fatigue of the participants and guarantee the recording quality, the monosyllable sets were randomly divided into two lists. Half of the participants were given List 1 and the other half List 2. Therefore, each participant produced 104 tonal monosyllabic tokens.

3.3. LANGUAGE BACKGROUND SURVEY. Participants’ language background information was collected via an adapted version of the Language History Questionnaire (LHQ, Li, Sepanski & Zhao, 2006). As the LHQ was designed specifically for multilingual speakers, modifications to the questions were made to gather information on dialect use. Most relevant to the purpose of the current study, the modified questionnaire asked participants the daily proportions of using each of their language varieties in different social contexts (e.g., at home, at work, outside).

3.4. PROCEDURE. The study was conducted online via Gorilla (Anwyl-Irvine et al., 2020) and approved by the Institutional Review Board of the University of Iowa. Prior to the word naming task, a microphone check took place, requiring participants to record a sentence, listen to the playback, and check if the microphone on their devices worked properly.

After the microphone check, the participants were presented with the recording instruction, which asked them to read the written character visually presented in each trial in either CM or SM depending on the participant’s group identity. Each experimental session had four blocks, with 26 trials each. In each trial, a cross appeared in the center of the screen for 300ms, followed by a monosyllabic stimulus in its orthographic form. The visual stimulus was presented for 5s, during which the participants were instructed to read the word clearly with their normal speaking speed and repeat the word as many times as possible without overlap between any two consecutive productions. The recording started automatically at the onset of the visual stimulus display and lasted for 5s. Then, the study automatically proceeded to the next trial. The stimuli in each list were randomized across and within the four blocks.

After the word naming task, participants filled out the modified Language History Questionnaire.

4. Analysis and results.

4.1. ACOUSTIC DATA PROCESSING AND ANALYSES. The production data were processed in Praat (Boersma & Weenink, 2023) to extract the acoustic features. For each token, the voiced portion was selected, which was then equally divided into ten intervals. For each interval, the average fundamental frequency (F0) was extracted. Therefore, for each production token, there was a sequence of 10 F0 values, each indicating the pitch height of one of the ten normalized time points. The acoustic data were further processed and analyzed in R version 4.3.2 (R Core Team, 2023). To correct the data distribution for normality, the F0 values were log-transformed. The logarithms were standardized for each participant to control individual variations resulting from physiological factors such as age and gender.

Statistical analyses of F0 (standardized log F0) were carried out with growth curve analysis (Mirman, 2014) using the R package LmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). In the analysis, the curve of the standardized log F0 of a monosyllabic token was modeled with first- and second-order orthogonal polynomials of time terms, because the most complex tones in the dialects have a falling-and-rising curve. Three time terms were of interest: the intercept, and the linear and quadratic terms of time, which indicated the monosyllable's average pitch height, the direction of the pitch change, and the curvature of the pitch curve, respectively. Random effects were conditioned on both participant and segmental syllable, including random intercepts and random slopes for all terms in the fixed effects of the model, to control for individual- and item-based variability. Stepwise backward selection was used to remove overfitting random effects until each model converged.

4.2. ACOUSTIC DIFFERENCES BETWEEN SIMILAR TONES ACROSS THE TWO DIALECTS. Separate mixed-effects polynomial regression models were fitted to the four pairs of between-dialect similar tones, investigating the effects of dialect and its interaction with time terms on the standardized log F0. Dialect was a categorical variable with two levels (CM, SM), indicating which dialect the tone in each pair belonged to. This variable was dummy-coded, with SM coded as 1 and CM coded as 0. The full reports of the fixed effects of the four models are provided in Table 2.

	Standardized log F0							
	CM-T1 vs. SM-T2		**CM-T2 vs. SM-T4**		**CM-T3 vs. SM-T4**		**CM-T4 vs. SM-T3**	
Predictors	*β (SE)*	*t*	*β (SE)*	*t*	*β (SE)*	*t*	*β (SE)*	*t*
(Intercept)	0.46*** (0.04)	11.31	-0.33*** (0.04)	-8.63	0.03 (0.05)	0.59	-0.49*** (0.06)	-8.10
Time [1]	36.32*** (1.92)	18.96	-74.65*** (2.34)	-31.85	-111.05*** (3.35)	-33.17	30.15*** (2.00)	15.08
Time [2]	-17.65*** (1.67)	-10.55	2.52 (2.21)	1.14	-33.35*** (2.46)	-13.53	19.04*** (1.90)	10.01
Dialect [SM]	-0.45*** (0.06)	-8.15	0.12* (0.06)	2.00	-0.25*** (0.06)	-4.03	-0.35*** (0.08)	-4.19
Time [1] x Dialect	10.33*** (1.27)	8.16	-33.59*** (2.09)	-16.08	2.61 (2.60)	1.00	-10.46*** (1.82)	-5.76
Time [2] x Dialect	26.28*** (1.26)	20.80	-21.65*** (1.99)	-10.89	13.53*** (2.45)	5.53	30.45*** (1.84)	16.59
Observation	21856		20039		18447		17877	
Marginal/Conditional R^2	0.247 / 0.349		0.296 / 0.335		0.290 / 0.326		0.136 / 0.219	
Model structure: *LogF0Std ~ poly(Time, 2) * Group + (1|Participant) + (1+poly(Time,2)|Segments)*								

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 2. Results of the fixed effects in the polynomial models fitted to tone pairs that are acoustically similar between CM and SM.

As shown in Table 2, for all tone pairs, the model found a significant effect of dialect, and significant interactions between dialect and the first order of time term (except for the model of the two high falling tones, CM-T3 vs. SM-T4), and between dialect and the second order of time term. These results suggest that for all pairs, similar tones differed from each other regarding the average pitch height, the direction of the slope (except between CM-T3 and SM-T4), and the curvature of the slope. Model predictions are plotted in Figure 2.

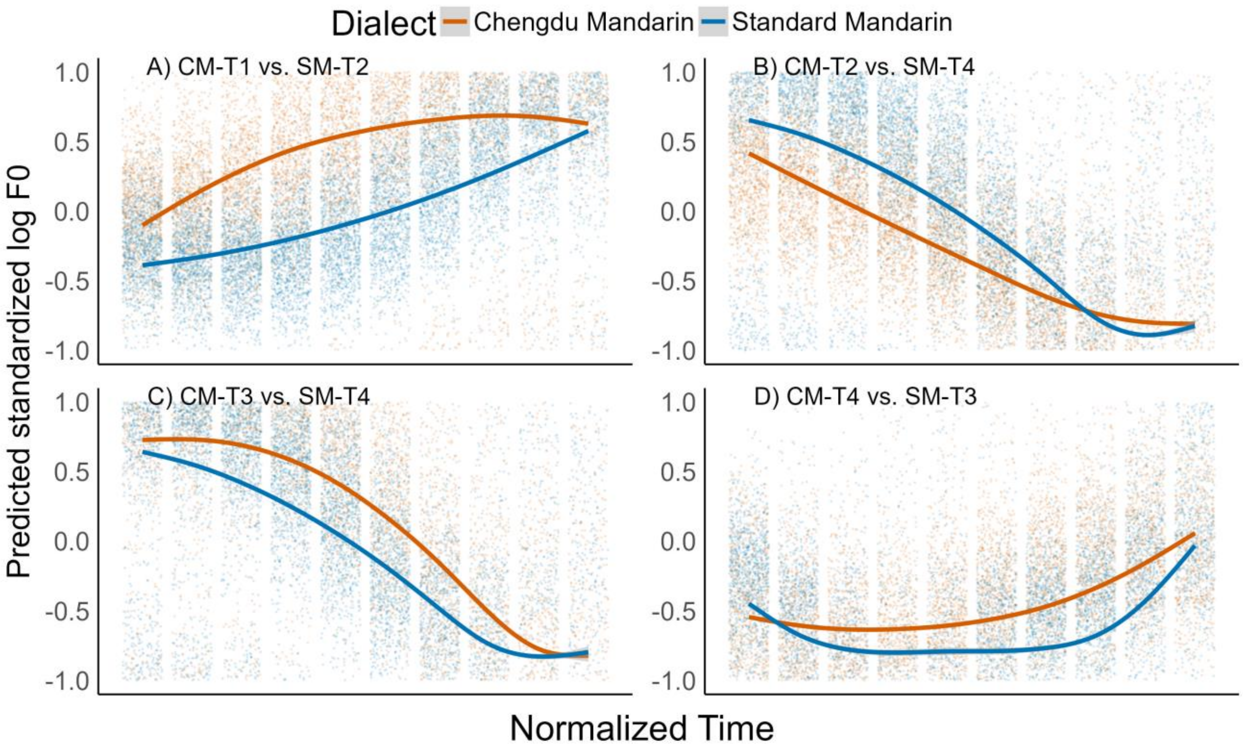

Figure 2. Standardized log F0 of the cross-dialectally similar tones. The lines indicate the dialect-level results predicted by the models reported in Section 4.2. The dots are descriptive values of individual participants.

Follow-up models were then fitted to each tone in each pair.[1] The models found that for the two rising tones, CM-T1 had a larger intercept (CM-T1: $\beta = 0.45$, $SE = 0.05$, $t = 9.90$; SM-T2: $\beta = 0.02$, $SE = 0.04$, $t = 0.45$), a smaller linear coefficient (CM-T1: $\beta = 24.91$, $SE = 2.01$, $t = 12.42$; SM-T2: $\beta = 33.46$, $SE = 1.46$, $t = 22.84$), and a negative quadratic coefficient (CM-T1: $\beta = -12.45$, $SE = 0.99$, $t = -12.58$); for SM-T2, the quadratic coefficient was positive (SM-T2: $\beta = 5.81$, $SE = 1.54$, $t = 3.76$). For the two falling tones, CM-T2 had a smaller intercept (CM-T2: $\beta = -0.36$, $SE = 0.05$, $t = -7.15$; SM-T4: $\beta = -0.22$, $SE = 0.04$, $t = -5.36$) and a larger negative linear coefficient (CM-T2: $\beta = -53.54$, $SE = 1.80$, $t = -29.82$; SM-T4: $\beta = -79.36$, $SE = 2.79$, $t = -28.45$) than SM-T4, and the models only found significant quadratic coefficient for SM-T4 (CM-T2: $\beta = 0.17$, $SE = 1.37$, $t = 0.12$; SM-T4: $\beta = -14.43$, $SE = 2.28$, $t = -6.32$). For the other falling tones, CM-T3 had a larger intercept (CM-T3: $\beta = 0.03$, $SE = 0.05$, $t = 0.50$) and a smaller negative quadratic coefficient (CM-T3: $\beta = -22.49$, $SE = 1.88$, $t = -11.95$) than SM-T4. Lastly, for the two dipping tones, The models found CM-T4 had a larger intercept (CM-T4: $\beta = -0.48$, $SE = 0.06$, $t = -7.81$; SM-T3: $\beta = -0.83$, $SE = 0.07$, $t = -12.76$), a larger linear coefficient (CM-T4: $\beta = 30.53$, SE

[1] Follow-up models were mixed-effects orthogonal polynomial models with the same as the models reported in Table 2 without the modulating effect of dialect.

= 1.38, t = 14.85; SM-T4: β = 14.42, SE = 1.94, t = 7.42), and a smaller quadratic coefficient (CM-T4: β = 12.51, SE = 1.17, t = 10.65; SM-T3: β = 35.97, SE = 1.80, t = 20.00) than SM-T3.

In summary, the results suggest that i) CM-T1 had a higher average pitch height and a flatter rising slope than SM-T2, and CM-T1 had a concave curve whereas SM-T2 had a convex curve (Figure 2A); ii) CM-T2 had a lower average pitch height and a flatter falling slope than SM-T4, and CM-T2 approximated a linear line whereas SM-T4 had a concave curve (Figure 2B); iii) CM-T3 had a higher average pitch and a narrower concave than SM-T4 (Figure 2C); and iv) CM-T4 had a higher average pitch, a steeper rising slope, and a wider convex than SM-T3 (Figure 2D).

4.3. THE EFFECTS OF SM EXPERIENCE ON FIRST DIALECT (D1) CM TONE PRODUCTIONS. Individual analyses were carried out for the data of CM-SM participants to investigate how the experience of using SM can affect CM-native speakers' CM tone productions. The daily proportion of SM use was served to quantify CM-SM bidialectal participants' SM experience. Participants' daily SM use proportions were collected via the modified LHQ (Li, Sepanski & Zhao, 2006) in three different social contexts (i.e., at home, at work, and outside), and the individual average proportion across the three contexts was used for the analysis. At the group level, the CM-SM bidialectal participants used SM 31% of the time (SD = 0.18) with a range between 0% and 68%.

Separate mixed-effects polynomial regressions were fitted to the data of CM-SM participants, predicting the effects of first- and second-orders of time and their interactions with SM use proportion on the standardized log F0 of each CM tone. The full reports of the fixed-effects of the four models are provided in Table 3. As shown in Table 3, for all CM tones, the model found significant main effects of SM use proportion and/or its interaction with time terms, suggesting that for each CM tone, SM use proportion had effects on the realization of the average pitch height, the direction of the slope, or the curvature of the slope.

	Standardized log F0							
	CM-T1		**CM-T2**		**CM-T3**		**CM-T4**	
Predictors	*β (SE)*	*t*	*β (SE)*	*t*	*β (SE)*	*t*	*β (SE)*	*t*
(Intercept)	0.59^{***} (0.07)	8.02	-0.48^{***} (0.08)	-6.05	0.09 (0.09)	0.91	-0.68^{***} (0.09)	-7.36
Time [1]	47.27^{***} (2.29)	20.67	-38.08^{***} (2.28)	-16.68	-83.61^{***} (4.36)	-19.17	16.55^{***} (1.97)	8.40
Time [2]	-4.23^{**} (1.54)	-2.74	9.57^{***} (1.89)	5.06	-17.35^{***} (2.83)	-6.14	15.41^{***} (1.89)	8.13
SM proportion	-0.46^{*} (0.20)	-2.30	0.41 (0.21)	1.93	-0.19 (0.26)	-0.73	0.67^{*} (0.26)	2.59
Time [1] x SM proportion	-73.60^{***} (3.85)	-19.10	-48.94^{***} (4.40)	-11.13	33.36^{***} (7.34)	4.55	11.11^{**} (4.27)	2.60
Time [2] x SM proportion	-26.76^{***} (3.81)	-7.02	-29.55^{***} (4.24)	-6.96	-16.66^{*} (6.97)	-2.39	-8.15 (4.28)	-1.91
Observation	10701		10106		8514		8840	
Marginal/Conditional R^2	0.160 / 0.236		0.327 / 0.391		0.269 / 0.335		0.115 / 0.208	
Model structure: *LogF0Std ~ poly(Time, 2) * SM_Prop + (1|Participant) + (1+poly(Time,2)||Segments)*								

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 3. Results of the fixed effects in the polynomial models fitted to CM tone productions as a function of time terms and SM use proportion.

To analyze the interaction effects, CM-SM bidialectal participants were divided into two groups by SM use proportion. Participants whose SM use proportion was greater than the group average were labeled as the “More SM” group, and those with SM proportion less than the group average were labelled as the “Less SM” group. The model predictions are plotted in Figure 3.

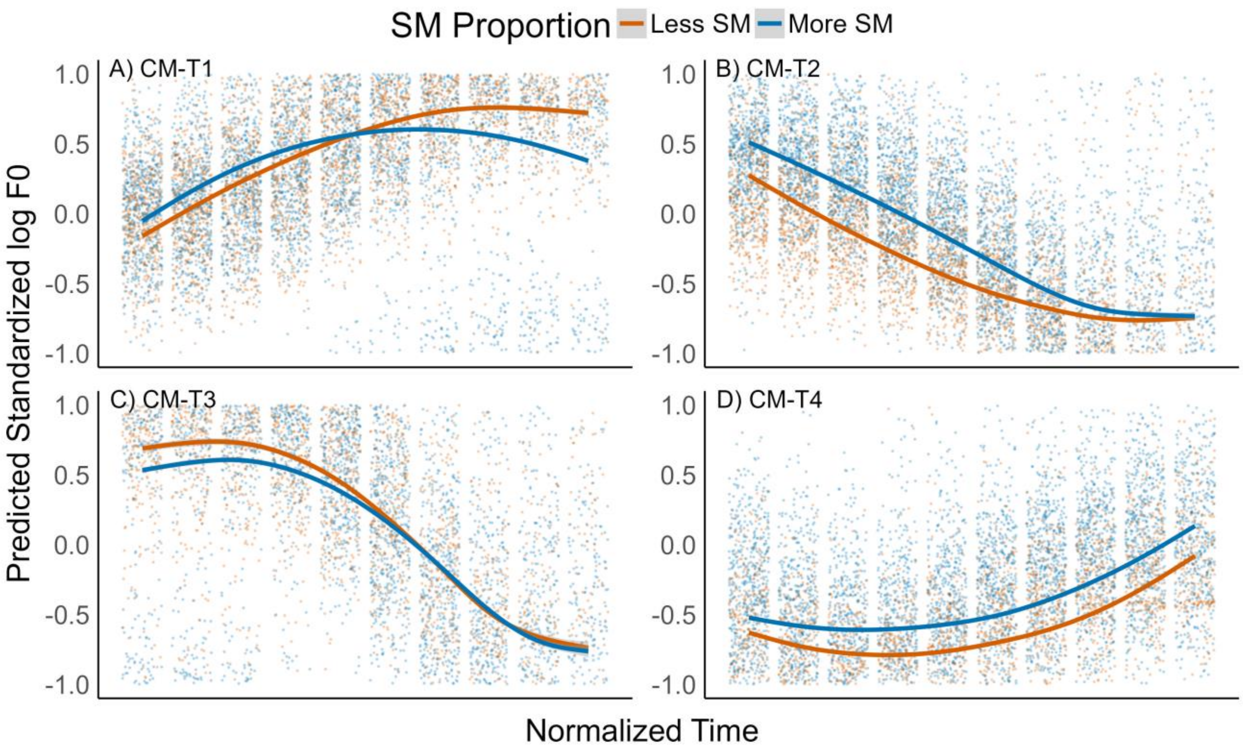

Figure 3. The effects of SM proportion of CM-SM bidialectal speakers’ CM tone productions. The lines indicate the group-level (Less SM vs. More SM) results predicted by the models reported in Section 4.3. The dots are descriptive values of individual participants.

Follow-up models were fitted to each CM tone for each group.[2] For CM-T1, the models revealed a smaller intercept (More SM: $\beta = 0.40$, $SE = 0.06$, $t = 6.41$; Less SM: $\beta = 0.51$, $SE = 0.06$, $t = 8.59$), a smaller linear coefficient (More SM: $\beta = 12.16$, $SE = 2.38$, $t = 5.12$; Less SM: $\beta = 23.66$, $SE = 0.88$, $t = 26.78$), and a smaller negative quadratic coefficient (More SM: $\beta = -12.81$, $SE = 1.09$, $t = -11.72$; Less SM: $\beta = -4.48$, $SE = 0.78$, $t = -5.74$) for the More SM group than the Less SM group. For CM-T2, the models found a smaller negative linear coefficient (More SM: $\beta = -45.52$, $SE = 1.37$, $t = -33.2$; Less SM: $\beta = -26.39$, $SE = 2.01$, $t = -13.11$) for the More SM group than the Less SM. In addition, the models found a significant quadratic coefficient only for the Less SM group (More SM: $\beta = -1.41$, $SE = 1.10$, $t = -1.28$; Less SM: $\beta = 3.32$, $SE = 1.58$, $t = 2.10$). For CM-T3, the models found a smaller negative linear coefficient (More SM: $\beta = -55.80$, $SE = 3.15$, $t = -17.74$; Less SM: $\beta = -49.65$, $SE = 3.32$, $t = -14.95$) and a smaller negative quadratic coefficient (More SM: $\beta = -18.24$, $SE = 1.68$, $t = -10.87$; Less SM: $\beta = -13.40$, $SE = 2.03$, $t = -6.59$) for the More SM group than the Less SM group. For CM-T4, the models found a larger intercept (More SM: $\beta = -0.39$, $SE = 0.07$, $t = -5.61$; Less SM: $\beta = -0.59$, $SE = 0.07$, $t = -8.04$) and a larger linear coefficient (More SM: $\beta = 15.54$, $SE = 1.33$, $t = 11.66$; Less SM: $\beta = 13.48$, $SE = 1.06$, $t = 12.72$) for the More SM group than the Less SM group.

[2] Follow-up models were mixed-effects orthogonal polynomial models with the same structure as the models reported in Table 3 without the modulating effect of SM proportion.

These results suggest that compared to the Less SM group, the More SM group produced i) CM-T1 with a lower average pitch, a flatter rising slope, and a narrower concave shape (Figure 3A), ii) CM-T2 with a steeper and more linear falling slope (Figure 3B), iii) CM-T3 with a steeper falling slope and a narrower concave shape (Figure 3C), and iv) CM-T4 with a higher average pitch and a narrower convex shape (Figure 3D). Compared to the acoustic differences in the CM and SM tone pairs reported in Section 4.3, the productions of CM-T1, -T3, and -T4 by the More SM group exhibited characteristics more similar to the corresponding CM tones, whereas their productions of CM-T2 exhibited characteristics more similar to SM-T4. In summary, the CM-T1, -T3, and -T4 produced by the CM-SM bidialectal speakers who used SM more frequently were dissimilated from the corresponding SM tones (more CM-like), whereas their CM-T2 was assimilated to SM-T4 (less CM-like), compared to the bidialectal speakers who used SM less frequently (i.e., more dominant in CM).

4.4. THE EFFECTS OF HOMOPHONE STATUS ON CM TONE PRODUCTIONS. Separate mixed-effects polynomial regression models were fitted to the CM-SM bidialectal participants' production data, predicting the standardized log F0 by the inter-dialectal homophone status of the lexical items and its interaction with the first- and second-orders of time terms. Homophone status was a categorical variable with two levels (homophone, non-homophone), and it was dummy-coded. Homophones were coded as 1 and non-homophones were coded as 0. The full reports of the fixed-effects of the models are provided in Table 4.

	Standardized log F0							
	CM-T1		**CM-T2**		**CM-T3**		**CM-T4**	
Predictors	*β (SE)*	*t*	*β (SE)*	*t*	*β (SE)*	*t*	*β (SE)*	*t*
(Intercept)	0.49^{***} (0.05)	9.13	-0.24^{***} (0.06)	-4.32	0.07 (0.06)	1.11	-0.39^{***} (0.07)	-5.53
Time [1]	22.00^{***} (3.21)	6.85	-51.96^{***} (2.84)	-18.30	-65.96^{***} (5.99)	-11.00	21.17^{***} (2.23)	9.49
Time [2]	-14.72^{***} (1.56)	-9.44	-0.89 (2.17)	-0.41	-19.27^{***} (2.91)	-6.62	13.35^{***} (1.89)	7.05
Homophone	-0.06 (0.05)	-1.26	-0.19^{***} (0.05)	-4.09	-0.07 (0.06)	-1.32	-0.14^{*} (0.06)	-2.37
Time [1] x Homophone	4.72 (4.10)	1.15	-2.63 (3.67)	-0.72	-12.48 (7.70)	-1.62	-1.07 (2.86)	-0.37
Time [2] x Homophone	3.70 (1.99)	1.86	1.74 (2.81)	0.62	-5.58 (3.79)	-1.47	-1.38 (2.43)	-0.57
Observation	10701		10106		8514		8840	
Marginal/Conditional R^2	0.117 / 0.239		0.315 / 0.389		0.262 / 0.331		0.107 / 0.228	
Model structure: *LogF0Std ~ poly(Time, 2) * Homophone + (1|Participant) + (1+poly(Time,2)|Segments)*								

Note: $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 4. Results of the fixed effects in the polynomial models fitted to CM tone productions as a function of time terms and homophone status of the lexical item.

As shown in Table 4, the model revealed a significant main effect of homophone for CM-T2 (β = -0.19, *SE* = 0.05, *t* = -4.09) and CM-T4 (β = -0.14, *SE* = 0.06, *t* = -2.37). In addition, the model found a marginal interaction effect between homophone and the second-order time term (β = 3.70, *SE* = 1.99, *t* = 1.86) for CM-T1, suggesting homophone status had an effect on the curvature of the pitch slope for CM-T1. Follow-up models found a larger negative quadratic coefficient for CM-T1 homophones (β = -8.65, *SE* = 1.00, *t* = -8.65) than non-homophones (β = -9.12, *SE* = 0.94, *t* = -9.67).[3] To exclude the possibility that the homophone effects were caused by segmental differences between the homophone and non-homophone stimuli, the same analysis was applied to the SM productions by the SM monodialectal participants. No homophone effects were found in any SM tones. The model predictions are plotted in Figure 4.

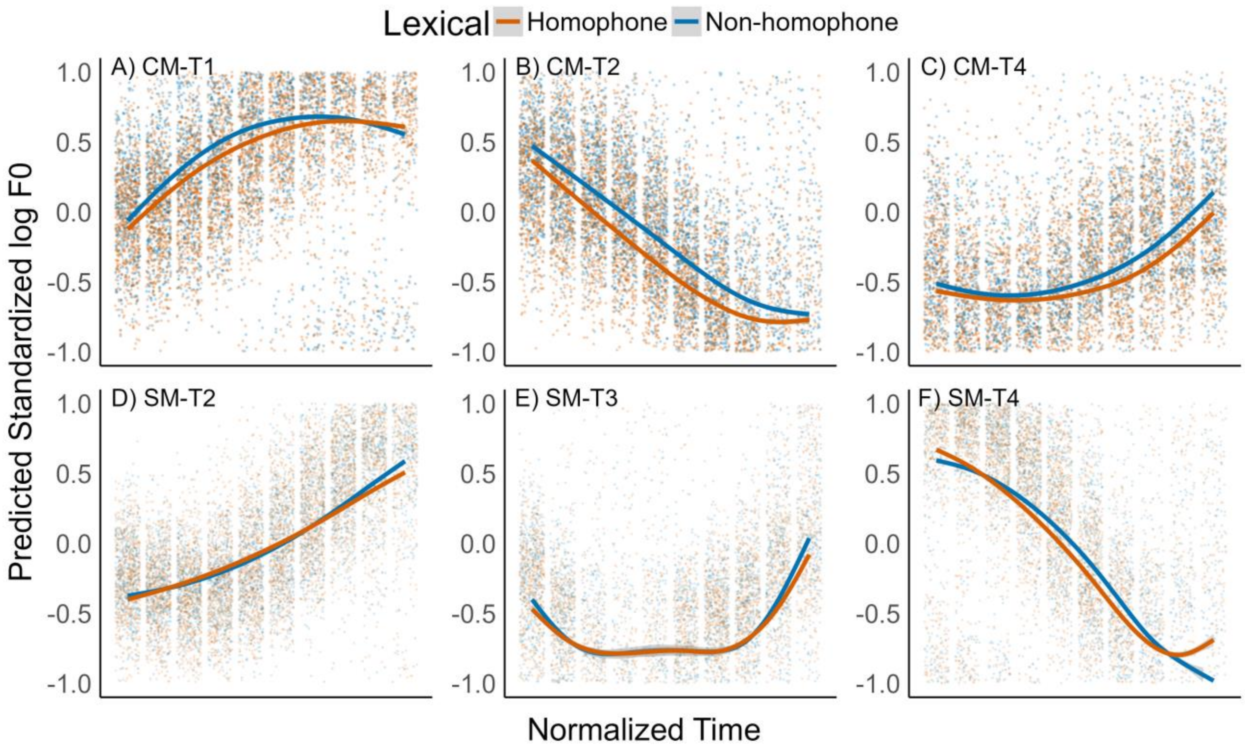

Figure 4. The effects of homophone status of CM-SM bidialectal speakers' CM tone productions. The lines indicate the group-level (homophone vs. non-homophones) results predicted by the models reported in Section 4.4. The dots are descriptive values of individual participants.

The results suggest that for CM-T1, homophones were produced with a wider concave than non-homophones (Figure 4A), and for CM-T2 and CM-T4, inter-dialectal homophones were produced with a lower average pitch than non-homophones (Figure 4B-C). No acoustic differences were found between homophones and non-homophones in SM tones (Figure 4D-F).

Compared to the acoustic differences in the CM and SM tone pairs reported in Section 4.2, the results showed that homophones in CM-T1 and CM-T4 exhibited characteristics more similar to the corresponding SM tones, whereas homophones in CM-T2 exhibited characteristics more similar to CM-T2. Compared to the acoustic differences in the productions between the More SM and Less SM reported in Section 4.3, the results showed that the homophones exhibited characteristics more similar to the productions of the Less SM group. In summary, the acoustic realizations of inter-dialectal homophones by the bidialectal group were more consistent with the productions of bidialectal speakers who were more dominant in CM. In other words, it can be

[3] Follow-up models were mixed-effects orthogonal polynomial models with the same structure as the models reported in Table 4 without the modulating effect of homophone.

summarized that the dissimilation patterns exhibited in CM-T1 and CM-T4 were eliminated in homophones, but CM-T2 showed dissimilation patterns in homophones.

5. Discussion. The current study examined the influence of using a second tonal system on the production of native tones in a word naming task with a group of bidialectal speakers of two Mandarin dialects. Three findings emerged regarding the research questions. First, the tone categories that have similar pitch contours between the two dialects are not acoustically identical. Second, for CM-SM bidialectal speakers, using SM affected their native production of CM tones. Lastly, inter-dialectal homophones showed an effect on bidialectal speakers' production of CM tones. However, the direction of the effects varied among the CM tones.

5.1. THE DISSIMILATION AND ASSIMILATION OF CM TONES. The finding that SM experience influences CM-SM bidialectal speakers' productions of CM tones is in line with the predictions of SLM-r, which claims that the bilingual's acoustic-phonetic system is malleable and responds to changes in experience across the life-span (Flege & Bohn, 2021). A closer examination of the results between the cross-dialectal tonal differences and the individual analysis on SM experience effects revealed that the productions of bidialectal speakers who used SM more frequently exhibited different dis/assimilation patterns across the four CM tone categories. Specifically, for CM-T1, -T3, and -T4, their productions showed acoustic features more similar to CM tones (more CM-like) than to SM-T2, SM-T4, and SM-T3, respectively; for CM-T2, their productions showed acoustic features more similar to SM-T4 (more SM-like). In other words, for bidialectal speakers who used SM more frequently, their native categories of CM-T1, -T3, and -T4 were dissimilated from the respective closest SM tones, but their native categories of CM-T2 were assimilated to the closest SM tone.

Two possible accounts may explain the different dis/assimilation patterns among the CM tones, or, more specifically, why the assimilation pattern was uniquely exhibited in CM-T2. First, as SLM-r posits that category dissimilation or assimilation can be used as a test for whether a new category is formed for the L2 sounds, the assimilation pattern between CM-T2 and SM-T4 might be explained as a failure in the formation of a separate category for SM-T4 by the CM-SM bidialectal speakers. This account, however, does not address the fact that we observed a dissimilation pattern between CM-T3 and SM-T4; that is, based on the account of SLM-r, a separate category should have been formed for SM-T4 in a place close to CM-T3 in the acoustic-phonetic space. One possible explanation is that at the onset of SM acquisition, CM-native learners first perceptually map the SM-T4 to both CM-T2 and CM-T3. However, the perceptual distance between SM-T4 and CM-T2 may be shorter than that between SM-T4 and CM-T3. Consequently, during the development of the interlingual phonological system, SM-T4 and CM-T2 form a composite category while the link between SM-T4 and CM-T3 eventually sunders. Hence, the observed production differences in CM-T3 may be a dissimilation from the composite category of CM-T2 and SM-T4.

The other possibility is that, given the acoustic differences between CM-T2 and SM-T4, bidialectal speakers do not map SM-T4 to CM-T2 at all. The assimilation pattern of CM-T2 observed in the current study may actually be the result of CM-dominant speakers (the Less SM group) trying to maintain within-dialect phonological contrast between CM-T2 and CM-T3, as both of them have a falling contour. This account is plausible, given the significant differences in both pitch height and pitch contour between CM-T2 and SM-T4 reported in Section 4.1. More critically, the results suggest that the CM-T2 has a linear falling contour while SM-T4 exhibits a quadratic pitch curve, and native Mandarin speakers have been shown very sensitive to pitch contours as their native tones contrast mainly by pitch change (e.g., Francis et al. 2008, Gandour

1983). It should be noted that both of the two accounts are based on some assumptions of the perceptual distance among CM-T2, CM-T3, and SM-T4, which have not been tested yet. Therefore, future studies are needed to examine how exactly bidialectal speakers perceptually map those tones, in terms of both within- and across-dialect tonal systems.

5.2. THE EFFECTS OF INTER-DIALECTAL HOMOPHONES. Regarding the second research question, the results did show effects for inter-dialectal homophone status on the production of CM tones. However, the direction of the effects varied by the CM tone categories. Specifically, homophones in CM-T1 and -T3 showed a pattern of assimilation to their respectively similar SM tones, whereas CM-T2 showed a pattern of dissimilation from SM-T4, and no homophone effects were found in CM-T4. Another possibility could be that the homophones in CM-T2, -T3, and -T4 were produced with characteristics that were in line with the productions of bidialectal speakers who were less influenced by SM experience.

With the results from the individual analysis on the effect of SM experience on CM tone productions, it seems that the direction of the homophone effects depends on the dissimilation or assimilation patterns of CM tones. For instance, in the individual analysis in Section 4.2., both CM-T1 and -T3 were found to be dissimilated from SM-T2 and -T4, but their productions in homophones showed acoustic features more similar to these respective SM tones. On the other hand, CM-T2 was found to be assimilated to SM-T4 in Section 4.2., but its productions in homophones showed acoustic features more salient in CM-T2.

Broadly speaking, the effects of homophones may be more complicated than we initially hypothesized as a binary enhancement of dissimilation or assimilation in bidialectal phonetic realizations. Instead, the results point to homophone effects that are dependent on the structure of the inter-dialectal phonetic categories. In other words, for CM tones for which a separate category of SM tone is formed, the activation of SM phonetic features was increased in homophones, whereas for CM tones that form a composite category with an SM tone, the contrast between CM and SM tones is enhanced. However, it is still not clear why no homophone effects were found in CM-T4.

6. Conclusion. The current study examined the interaction between two established tonal systems with a group of bidialectal speakers. The results show that using a second tonal dialect and inter-dialect homophone status have an effect on speakers' native tone categories, and both of these effects are modulated by the structure of the inter-dialect tonal system. The findings of the study support the similarity between bidialectal and bilingual speech processing and provide novel evidence for bilingual/bidialectal speech models that include the suprasegmental level.

References

Abramson, Arthur S. & Leigh Lisker. 1973. Voicing-timing perception in Spanish word-initial stops. *Journal of Phonetics* 1(1). 1–8. https://doi.org/10.1016/s0095-4470(19)31372-5.

Amengual, Mark. 2012. Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition* 15(3). 517–30. https://doi.org/10.1017/S1366728911000460.

Anwyl-Irvine, Alexander L., Jessica Massonnié, Adam Flitton, Natasha Kirkham & Jo K. Evershed. 2020. Gorilla in our mindst: An online behavioral experiment builder. *Behavior Research Methods* 52(1). 388–407.

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 171–206. Timonium, MD: York Press.

Best, Catherine T. & Michael D. Tyler. 2007. Nonnative and second-language speech perception. In Murray Munro & Ocke-Schwen Bohn (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, 31–52. Amsterdam: John Benjamins.

Boersma, Paul & David Weenink. 2023. Praat: doing phonetics by computer. Version 6.4.01. Retrieved from https://www.praat.org/.

Brown, Earl Kjar & David Harper. 2009. Phonological evidence of interlingual exemplar connections. *Studies in Hispanic and Lusophone Linguistics* 2(2). 257–274. https://doi.org/10.1515/shll-2009-1052.

Cheng, Chin-Chuan. 1997. Measuring relationship among dialects: DOC and related resources. *Computational Linguistics and Chinese Language Processing* 2(1). 41–72.

Cho, Taehong & Peter Ladefoged. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics* 27(2). 207–29. https://doi.org/10.1006/jpho.1999.0094.

Flege, James E. 1995. Second language speech learning: Theories, findings, and problems. In Winifred Strange (eds.), *Speech perception and linguistic experience: Issues in cross-language research*, 233–77. Timonium, MD: York Press.

Flege, James E. & Wieke Eefting. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, 15(1), 67-84. https://doi.org/10.1016/s0095-4470(19)30538-8.

Flege, James E. & Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In Ratree Wayland (eds.) *Second language speech learning*, 3–83. Cambridge University Press.

Flege, James E., Carlo Schirru & Iran R. A. MacKay. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40(4). 467–491. https://doi.org/10.1016/s0167-6393(02)00128-0.

Francis, Alexander L., Valter Ciocca, Lian Ma & Kimberly Fenn. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics* 36(2). 268–294. https://doi.org/10.1016/j.wocn.2007.06.005.

Gandour, Jack. 1983. Tone perception in Far Eastern languages. *Journal of Phonetics* 11(2). 149–175. https://doi.org/10.1016/S0095-4470(19)30813-7.

Gollan, Tamar H., Kenneth I. Forster & Ram Frost. 1997. Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23(5). 1122–1139. https://doi.org/10.1037/0278-7393.23.5.1122.

He, Wan. 2015. Experimental phonetic statistics of individual character tone of Chengdu Dialect (in Chinese). *Journal of Chengdu University (Social Sciences)* 157(1). 119–123.

Kuhl, Patricia K. 1991. Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50(2). 93–107. https://doi.org/10.3758/BF03212211.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens & Björn Lindblom. 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255(5044). 606–608. https://doi.org/10.1126/science.1736364.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. LmerTest packages: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13) 1–26. https://doi.org/10.18637/jss.v082.i13.

Li, Ping, Sara Sepanski & Xiaowei Zhao. 2006. Language history questionnaire: A web-based interface for bilingual research. *Behavior Research Methods* 38(2). 202–210. https://doi.org/10.3758/BF03192770.

Lisker, Leigh & Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measures. *Word* 20(3). 384–442. https://doi.org/10.1080/00437956.1964.11659830.

MacKay, Ian R. A., James E. Flege, Thoresten Piske & Carlo Schirru. 2001. Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America* 110(1). 516–528. https://doi.org/10.1121/1.1377287.

Mirman, Daniel. 2014. *Growth curve analysis and visualization using R.* New York, NY: Chapman & Hall/CRC.

Mora, Joan C. & Marianna Nadeu. 2012. L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism* 16(4). 484–500. https://doi.org/10.1177/1367006911429518.

Osborne, Denise M. & Miquel Simonet. 2021. Foreign-language phonetic development leads to first-language phonetic drift: Plosive consonants in native Portuguese speakers learning English as a foreign language in Brazil. *Languages* 6(3). 112–137. https://doi.org/10.3390/languages6030112.

R Core Team. 2023. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Sun, Ching Chu, Peter Hendrix, Jianqiang Ma & Rolf Harald Baayen. 2018. Chinese lexical database (CLD). *Behavior Research Methods* 50(6). 2606–2629. https://doi.org/10.3758/s13428-018-1038-3.

Tang, Chaoju. 2017. Dialects of Chinese. In Charles Boberg, John Nerbonne & Dominic Watt (eds.), *The handbook of dialectology,* 547–58. Wiley Blackwell.

Tang, Chaoju & van Heuven, Vincent J. 2015. Predicting mutual intelligibility of Chinese dialects from multiple objective linguistic distance measures. *Linguistics* 53(2). 285–311. https://doi.org/10.1515/ling-2015-0005.

Yip, Moria. 2002. *Tone*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139164559.