Do Chinese-English Bilinguals Speak English Words with Lexical Tone in Mind?

Qing Yang & Yiya Chen

Centre for Linguistics, Leiden University; Leiden Institute for Brian and Cognition, Leiden University
q.yang@hum.leidenuniv.nl; yiya.chen@hum.leidenuniv.nl

ABSTRACT

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

Keywords: lexical tone; spoken word production; the bilingual lexicon; picture-word interference paradigm

1. INTRODUCTION

The functional role of pitch variation differs across languages. In lexical tone languages such as Standard Chinese (hereafter, SC), pitch contour plays a crucial role in differentiating morpheme meanings, just as consonants and vowels (e.g., *ma* with a rising pitch contour means "hemp" but "scold" with a falling contour). For words in non-tonal languages such as English, pitch contour serves as a cue (together with cues such as duration and intensity) to distinguish a limited number of words, known as lexical stress [1]. For both types of languages, pitch variation also serves to signal utterance-level information such as sentence mode and information status [2]. For example, in most varieties of English, "Mary" can be uttered with a rising pitch contour to signal a question and a falling contour to signal a statement. There is a probabilistically stable mapping between pitch contour shape and sentence mode. In Standard Chinese, however, pitch variation for sentence mode is constrained by the pitch contours of lexical tones [3]. Such cross-language differences between SC and English in the form and function of pitch variations offer a unique case for investigating pitch processing in bilingual lexical access.

One widely accepted assumption is that bilinguals' two languages interact extensively [4]. For instance, the phonemes of bilinguals' two languages are found to be co-activated and compete for selection during speech production [5], [6]. However, most evidence for language interaction came from studies on segmental processing. Only a limited number of studies examined whether bilingual lexical access is influenced by suprasegmental properties such as lexical tone. Most of them have focused on subtle differences in tone perception between native and non-native tonal speakers [7], [8]. There is a surprising paucity of empirical research on how speaking a tonal language may affect bilinguals' pitch processing in the non-tonal language they command.

As one of the few studies that tapped into this issue, Ortega-Llebaria and her colleagues [9], [10] asked SC-English bilinguals and English monolinguals to perform a primed-lexical decision task. The prime and target were manipulated to fully match (e.g., "rice" with a falling-f0 – "rice" with a falling-f0), fully mismatch (e.g., "gold" with a rising-f0 – "rice" with a falling-f0), mismatch in segment (e.g., "mice" with a falling-f0 – "rice" with a falling-f0), or mismatch in pitch (e.g., "rice" with a falling-f0 – "rice" with a falling-f0). Results of reaction time showed that only SC-English bilinguals experienced larger facilitation across conditions when the targets were produced with a falling f0 than with a rising f0. This "falling-f0 bias" in SC-English bilinguals led the authors to propose that English words with falling f0 are closer to English lexical representations than those with rising f0 in the bilingual lexicon; therefore, English words with falling f0 were easier to access than their rising-f0 counterparts. Moreover, only SC-English bilinguals manifested the "falling-f0 bias" indicating the impact of long-term experience with lexical tone on bilinguals' pitch processing even in their non-tonal spoken word recognition.

Despite evidence from the comprehension domain on the effect of lexical tone on bilingual lexicon and spoken word recognition, to our knowledge, no study has investigated whether and, if so, how lexical tone affects non-tonal lexical access during speech production. To develop a more comprehensive account of the role of lexical tone in cross-language interaction, empirical data from non-tonal spoken word production is still needed.

In this study, we aimed to fill this knowledge gap. Following Ortega-Llebaria and her colleagues [9], [10], one may hypothesize that the "falling-f0 bias" of SC-English bilinguals also plays a role in bilingual spoken word production. Using the picture-word interference paradigm (hereafter, PWI) [11], we asked native SC-English bilinguals and English monolinguals to name pictures in English while ignoring simultaneously played SC auditory distractors. Crucially, for the same target word (e.g., *lung*), there were four types of distractors (as illustrated in Table 1): 1) the target's cross-language homophone with a falling tone (CH_F; e.g., "lang4*", wave*); 2) the target's cross-language homophone with a rising tone (CH_R; e.g., "lang2*", wolf*); 3) unrelated distractor with a falling tone (UN_F; e.g., "you4*", right*); 4) unrelated distractor with a rising tone (UN_R; e.g., "you2*", oil*).

English Target	SC Distractors			
	Cross-language Homophone		Unrelated	
	Rising (CH_R)	Falling (CH_F)	Rising (UN_R)	Falling (UN_F)
fei4 **"lung"**	***lang2*** "wolf"	***lang4*** "wave"	***you2*** "oil"	***you4*** "right"

Table 1: A set of sample stimuli.

Previous bilingual PWI studies have found robust facilitation effects of cross-language homophones (e.g., [12]). We, therefore, expected to observe significant facilitation effects in both the CH_F and CH_R conditions, compared with the UN_F and UN_R conditions for both SC-English bilingual and monolingual speakers. Secondly, if lexical tone indeed shapes pitch processing in English spoken word production, the influence of falling vs. rising-tone SC homophone distractors on English picture naming was expected to differ between SC-English bilinguals and native English monolinguals. Furthermore, the "falling-f0 bias" of SC-English bilinguals, if at play in spoken word production, would lead to processing differences between the two homophone (i.e., CH_F vs. CH_R) conditions.

2. METHOD

2.1. Participants

48 SC-English bilinguals and 48 American English monolinguals participated in this study. All SC-English bilingual participants are native SC speakers who grew up in Beijing and speak no regional dialect. All participants learned English at an average age of 5.8 (SD=2.34). Their English proficiency level was accessed with an adapted LEAP-Q questionnaire [13] and the multilingual naming test (MINT; [14]). Using a Likert scale from one to ten, participants' self-rated frequency is 8.5 (SD = 1.4) in reading, 6.7 (SD = 1.8) in speaking, and 7.1 (SD = 1.8) in listening. The average correct response of MINT is 43% (SD = 5.1%). The English monolingual participants had no previous exposure to Mandarin or any other tonal languages. None of the participants had a history of language disorder. All participants provided informed consent and were compensated for their participation.

2.2. Materials

24 sets of critical stimuli were included Each set consists of an English target word, an SC cross-language homophone distractor with a falling lexical tone (CH_F), a cross-language homophone distractor with a rising tone (CH_R), an unrelated distractor with a falling tone (UN_F), and an unrelated distractor with a rising tone (UN_R). 12 sets of fillers which have no phonological overlap were also added. All target words are English picturable monosyllabic nouns. All distractor words are SC monosyllabic morphemes. Lexical frequency of distractors, as computed with SUBTLEX-CH [16], are balanced across conditions ([$F(3, 92) = 1.97$, $p = 0.13$]). Homophone density, as computed with DoWLS-MAN [17], was also controlled ([$F(3, 92) = 0.855$, $p = 0.47$]). The target pictures, which are black and white line drawings, are selected and adapted from the IPNP database (Bates et al., 2003) and the BOSStimuli database [19]. All spoken stimuli were produced by a male native SC speaker who was born and grew up in Beijing. The recording was done at the Phonetics Lab of the Leiden University Centre for Linguistics, through a Sennheiser MKH416T microphone (44.1 kHz, 16-bit). All stimuli were normalized for a duration of 400ms and intensity at 70dB in Praat [20].

2.3. Procedure

Participants took part in the experiment online using Gorilla (www.gorilla.sc). All participants were required to wear headphones and sit in a quiet room. Participants were only allowed to join the experiment if they were using laptops. Before the experiment, a headphone check based on the dichotic pitch [21], a microphone check, and an auto-play check were run to screen participants' equipment. All the instructions were given in English. Before the naming task, there was a familiarization session, during which participants were shown all target pictures with their corresponding English names printed underneath for 1500ms. After the name disappeared, participants were asked to type in the picture's English name. If participants did not respond accurately, the intended name would be shown for 1500ms again. After the familiarization session, participants were asked to complete four practice trials with an option to practice more if they wish.

The picture naming task started after the familiarization and practice sessions. Within each trial, a fixation was first displayed in the centre of the screen for 500ms, followed by a picture and a simultaneously played auditory distractor (SOA = 0ms). Participants were asked to name the picture as quickly and as accurately as possible while ignoring the distractor. The picture remains on the screen for 2000ms. Response times were measured from picture onset until naming onset using Chronset [22]. If participants did not respond within the 2000ms interval, the next trial began automatically. Between each trial, there was a blank screen of 1000ms.

In total, there were 96 (24 x 4) critical trials and 48 (12 x 4) filler trials. All trials were equally distributed into four blocks with a Latin Square design so that participants only see each target picture once in one block. Between the blocks, participants were encouraged to take a short break. After the naming task, participants were asked to complete a language background survey (i.e., the adapted LEAP-Q questionnaire) and the language proficiency MINT test. In total, the experiment took about 30 minutes.

3. RESULTS

Trials with incorrect (3.2%) and empty responses (5.5%) were excluded from the data analysis. Table 2 summarizes the mean naming latency (ms) for each experimental condition. As we can see, English monolingual participants took about 40ms longer to name pictures with unrelated distractors (UN_R and UN_F) than with cross-language homophone distractors (CH_R and CH_F). Moreover, either with unrelated distractors or cross-language distractors, there was almost no difference (~5ms) between naming with falling-tone vs. rising-tone distractors. As for SC-English bilinguals, the overall naming latency was more than 70ms longer than English monolinguals in each condition. Naming latencies with cross-language homophone distractors (CH_R and CH_F) were shorter than that with unrelated distractors (UN_R and UN_F). While there was no naming latency difference between rising-tone vs. falling-tone unrelated distractors, there was an average difference of 22ms between the rising and falling-tone cross-language homophone distractors.

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

Table 2: Mean and SD of raw picture naming latency (ms).

Response times were analysed using the generalized linear mixed-effects model (GLMM) with inverse Gaussian distribution [23]. All the statistical analyses were run in R with the package *lme4* [24]. Given that error rates were low in each condition, no further analysis of response accuracy was conducted. A maximum model includes fixed effects of distractor type (CH_R, CH_F, UN_R and UN_F), participant groups (SC-English bilinguals and English monolinguals), the interaction between distractor type and group, by-subject and by-item random intercept, and random slopes for each fixed term were constructed first. Each term was then tested for exclusion. When the model failed to converge, we first increased the number of iterations and then simplified the model by removing correlation parameters in the random structures [25]. The final GLMM consists of fixed effects of distractor type, the interaction between distractor type and group, random intercepts for subject, and random intercepts for item. As there was a significant interaction between groups and distractor type ($p = 0.030$), pairwise comparisons between group and distractors were also computed using the *multcomp* package [26]. Holm–Bonferroni method was implemented to correct family-wise errors [27].

According to the model estimations (see Table 2), both English monolinguals and SC-English bilinguals took longer to name targets with UN_R and UN_F than CH_R and CH_F respectively (English monolinguals: $p < 0.000$; $p < 0.000$; SC-English bilinguals: $p < 0.000$; $p = 0.002$). These results confirm the facilitatory effects of cross-language phonological overlap. For English monolinguals, there was no significant difference between the distractors UN_R vs. UN_F ($p = 0.903$) and CH_R vs.

CH_F (p = 0.596). For SC-English bilinguals, there was a significant difference between CH_R and CH_F distractors (p = 0.022), but no significant difference between UN_R and UN_F distractors (p = 0.925). This suggests that only SC-English bilinguals, but not English monolinguals, responded to the rising-tone cross-language homophone distractors faster than their falling-tone counterparts.

4. DISCUSSION

Our study investigated the effect of SC lexical tone in bilingual lexical access during English spoken word production. Within the picture word inference paradigm (PWI), both SC-English bilinguals and English monolinguals were asked to name pictures in English while ignoring SC distractors that were either cross-language homophones with the target names or unrelated. Furthermore, we manipulated the lexical tone of the distractors to be either rising or falling. Our naming latency results showed that, regardless of the pitch shape difference, both SC-English bilinguals and English monolinguals took less time to name pictures when the distractors are cross-language homophones, in comparison to the unrelated distractors. Our results thus confirmed the cross-language phonological facilitation effect and showed that our participants, despite their on-line participation, behaved similarly to participants in lab experiments [12].

Furthermore, while both rising and falling-tone cross-homophones were equally facilitative for English monolinguals, SC-English bilinguals took significantly longer time to name pictures with falling-tone cross-language homophone distractors than with their rising-tone counterparts. Such a distinction indicates a significant pitch processing difference between SC-English bilinguals and English monolinguals during English spoken word production. Our results thus echoed with the bilingual comprehension studies [8-9] that lexical tone plays a significant role in non-tonal lexical access. It is, however, important to note that our findings showed an opposite pattern with regard to the direction of pitch processing advantages in bilingual speakers. Ortega-Llebaria and her colleagues found a “falling-f0 bias” [9], [10] in SC-English bilingual speakers’ lexical decision, we observed a reversed “rising-f0 bias” in SC-English picture naming. There are a few possible explanations for such a difference.

First, it might be due to different task requirements. The presence of phonologically similar words is known to elicit inhibitory effects in comprehension tasks (e.g., [28]), but facilitatory effects in production tasks (e.g., [29]). This is because, in comprehension tasks, which are phonologically driven, phonologically similar words often introduce speech ambiguity and thus slow down lexical access; while in speech production tasks, which are semantically driven, phonologically similar words often help resolve semantic competition and facilitate the processes of phonological encoding and phonetic spell-out.

Alternatively, the falling vs. rising-bias contrast may reflect cross-language interference effects. Ortega-Llebaria and her colleagues [9], [10] used English intonational falling vs. rising contrasts in the stimuli, but we selected SC cross-language homophones with lexical rising and falling tones as distractors. We know that the phonetic details of a lexical falling tone in SC and an intonational falling tone in English are different. If English words are indeed represented with an intonational falling tone, as proposed by Ortega-Llebaria and her colleagues, SC cross-language homophones with a lexical falling tone may cause more cross-language interference than their rising-tone counterparts, and therefore become less facilitative for English target naming.

A third possibility is related to the acoustic saliency of the pitch contour in the rising-tone distractors, which could have elicited faster responses, and, in turn, facilitated word production. A followup question is why such rising pitch saliency does not promote lexical decision.

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

5. CONCLUSION

Overall, the SC-English bilingual speakers in our study showed a robust facilitation effect of SC cross-language homophones. Furthermore, the rising lexical tone introduced a greater facilitation effect than a falling lexical tone. No such facilitation effects were observed in monolingual English speakers. Our findings lend novel evidence to the role of lexical tone in bilinguals’ non-tonal lexical access during spoken word production, complementary to earlier findings on the impact of lexical tone during spoken word recognition. Further research is needed for a refined and comprehensive account of an interactive bilingual mental lexicon for speech processing.

6. ACKNOWLEDGEMENT

This research was supported by the China Scholarship Council for the State Scholarship Fund (No.201707110012) and the Netherlands Organization for Scientific Research (VI.C.181.040).

6. REFERENCES

[1] M. Gordon and T. Roettger, "Acoustic correlates of word stress: A cross-linguistic survey," *Linguistics Vanguard*, vol. 3, no. 1, Dec. 2017, doi: 10.1515/lingvan-2017-0007.

[2] M. Ernestus and Y. Chen, "Message-Related Variation: Segmental Within-Speaker VariationTonal Variation," in *The Oxford Handbook of Laboratory Phonology*, A. C. Cohn, C. Fougeron, and M. K. Huffman, Eds., Oxford University Press, 2011, p. 0. doi: 10.1093/oxfordhb/9780199575039.013.0006.

[3] M. Liu, Y. Chen, and N. O. Schiller, "Context Matters for Tone and Intonation Processing in Mandarin," *Lang Speech*, vol. 65, no. 1, pp. 52–72, Mar. 2022, doi: 10.1177/0023830920986174.

[4] J. F. Kroll and F. Ma, "The bilingual lexicon," in *The handbook of psycholinguistics*, in Blackwell handbooks in linguistics. Wiley-Blackwell, 2018, pp. 294–319.

[5] À. Colomé, "Lexical Activation in Bilinguals' Speech Production: Language-Specific or Language-Independent?," *Journal of Memory and Language*, vol. 45, no. 4, pp. 721–736, Nov. 2001, doi: 10.1006/jmla.2001.2793.

[6] P. Macizo, "Phonological coactivation in the bilinguals' two languages: Evidence from the color naming task*," *Bilingualism: Language and Cognition*, vol. 19, no. 2, pp. 361–375, Mar. 2016, doi: 10.1017/S136672891500005X.

[7] J. Gandour *et al.*, "Temporal integration of speech prosody is shaped by language experience: An fMRI study," *Brain and Language*, vol. 84, no. 3, pp. 318–336, Mar. 2003, doi: 10.1016/S0093-934X(02)00505-9.

[8] B. Chandrasekaran, A. Krishnan, and J. T. Gandour, "Mismatch negativity to pitch contours is influenced by language experience," *Brain Research*, vol. 1128, pp. 148–156, Jan. 2007, doi: 10.1016/j.brainres.2006.10.064.

[9] M. Ortega-Llebaria, M. Nemogá, and N. Presson, "Long-term experience with a tonal language shapes the perception of intonation in English words: How Chinese–English bilinguals perceive 'Rose?' vs. 'Rose,'" *Bilingualism: Language and Cognition*, vol. 20, no. 2, pp. 367–383, Mar. 2017, doi: 10.1017/S1366728915000723.

[10] M. Ortega-Llebaria and Z. Wu, "Chinese-English Speakers' Perception of Pitch in Their Non-Tonal Language: Reinterpreting English as a Tonal-Like Language," *Lang Speech*, p. 0023830919894606, Jan. 2020, doi: 10.1177/0023830919894606.

[11] R. R. Rosinski, R. M. Golinkoff, and K. S. Kukish, "Automatic semantic processing in a picture-word interference task," *Child Development*, pp. 247–253, 1975.

[12] A. Costa, À. Colomé, O. Gómez, and N. Sebastián-Gallés, "Another look at cross-language competition in bilingual speech production: Lexical and phonological factors," *Bilingualism: Language and Cognition*, vol. 6, no. 3, pp. 167–179, Dec. 2003, doi: 10.1017/S1366728903001111.

[13] M. Kaushanskaya, H. K. Blumenfeld, and V. Marian, "The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later," *Bilingualism: Language and Cognition*, vol. 23, no. 5, pp. 945–950, Nov. 2020, doi: 10.1017/S1366728919000038.

[14] K. Antoniou and N. Katsos, "The effect of childhood multilingualism and bilectalism on implicature understanding," *Applied Psycholinguistics*, vol. 38, no. 4, pp. 787–833, Jul. 2017, doi: 10.1017/S014271641600045X.

[15] T. H. Gollan, G. H. Weissberger, E. Runnqvist, R. I. Montoya, and C. M. Cera, "Self-ratings of Spoken Language Dominance: A Multi-Lingual Naming Test (MINT) and Preliminary Norms for Young and Aging Spanish-English Bilinguals," *Biling (Camb Engl)*, vol. 15, no. 3, pp. 594–615, Jul. 2012, doi: 10.1017/S1366728911000332.

[16] Q. Cai and M. Brysbaert, "SUBTLEX-CH: Chinese Word and Character Frequencies Based on Film Subtitles," *PLOS ONE*, vol. 5, no. 6, p. e10729, Jun. 2010, doi: 10.1371/journal.pone.0010729.

[17] K. D. Neergaard, H. Xu, J. S. German, and C.-R. Huang, "Database of word-level statistics for Mandarin Chinese (DoWLS-MAN)," *Behav Res*, vol. 54, no. 2, pp. 987–1009, Apr. 2022, doi: 10.3758/s13428-021-01620-7.

[18] E. Bates *et al.*, "Timed picture naming in seven languages," *Psychonomic Bulletin & Review*, vol. 10, no. 2, pp. 344–380, Jun. 2003, doi: 10.3758/BF03196494.

[19] M. B. Brodeur *et al.*, "The bank of standardized stimuli (BOSS): comparison between French and English norms," *Behavior Research Methods*, vol. 44, no. 4, pp. 961–970, Dec. 2012, doi: 10.3758/s13428-011-0184-7.

[20] P. Boersma, "Praat: doing phonetics by computer [Computer program]," *http://www. praat. org/*, 2011.

[21] A. E. Milne *et al.*, "An online headphone screening test based on dichotic pitch," *Behavior Research Methods*, 2020, doi: 10.3758/s13428-020-01514-0.

[22] F. Roux, B. C. Armstrong, and M. Carreiras, "Chronset: An automated tool for detecting speech onset," *Behav Res*, vol. 49, no. 5, pp. 1864–1881, Oct. 2017, doi: 10.3758/s13428-016-0830-1.

[23] S. Lo and S. Andrews, "To transform or not to transform: using generalized linear mixed models to analyse reaction time data," *Frontiers in Psychology*, vol. 6, Aug. 2015, doi: 10.3389/fpsyg.2015.01171.

[24] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using **lme4**," *Journal of Statistical Software*, vol. 67, no. 1, 2015, doi: 10.18637/jss.v067.i01.

[25] M. Brauer and J. J. Curtin, "Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items," *Psychological Methods*, vol. 23, no. 3, pp. 389–411, 2018, doi: 10.1037/met0000159.

[26] T. Hothorn *et al.*, "Package 'multcomp,'" *Simultaneous inference in general parametric models. Project for Statistical Computing, Vienna, Austria*, 2016.

[27] S. Holm, "A Simple Sequentially Rejective Multiple Test Procedure," *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979.

[28] S. Dufour and R. Peereman, "Inhibitory priming effects in auditory word recognition: when the target's competitors conflict with the prime word," *Cognition*, vol. 88, no. 3, pp. B33–B44, Jul. 2003, doi: 10.1016/S0010-0277(03)00046-5.

[29] A. S. Meyer, "The time course of phonological encoding in language production: Phonological encoding inside a syllable," *Journal of Memory and Language*, vol. 30, no. 1, pp. 69–89, Feb. 1991, doi: 10.1016/0749-596X(91)90011-8.