# Lexical tone in word activation

Yang, Q.

# Chapter 5

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

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

**Abstract**

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

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

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

It is by now widely agreed that bilinguals' two languages interact extensively (see Kroll & Tokowicz, 2005 for a review). Words of bilinguals' two languages are constantly active in parallel, resulting in cross-language interaction at all levels of speech processing and planning. For example, when naming a picture in one language, bilinguals may experience shorter naming latency when the translation equivalent of the target picture name in their other language is present (e.g., Costa & Caramazza, 1999). Conversely, they may experience longer naming latency when a homophone of the translation equivalent is present (e.g., Hermans et al., 1998). Moreover, bilinguals may experience difficulty detecting phonemes that are present in the translation equivalent of a word, compared to those that are not (e.g., Colomé, 2001). While these observations show clear evidence that bottom-up lexical and sub-lexical (phonological) overlap creates cross-language interaction and competition, it is important to note that most previous studies drew evidence from the effects of segmental overlap in Indo-European language (e.g., Hermans et al., 1998; Colomé, 2001; Costa et al., 2003). What has remained a little-known area is how languages with great typological differences such as English and SC influence each other at the suprasegmental level.

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

One pioneering study on the role of lexical tone in non-tonal speech comprehension is Shook & Marian (2015). In this study, SC-English bilinguals were asked to listen to an English spoken word (e.g., "tree") and select its SC translation from two Mandarin words on display (e.g., *shu* with a falling pitch contour "tree" and *long* with a rising pitch contour *"*dragon*"*). Critically, the pitch contour of the spoken English target word was manipulated to either match or mismatch the lexical tone of its SC translation. With eye-tracking recordings, Shook and Marian (2015) found that, when the pitch contour of the English target matched the lexical tone of the SC translation (e.g., *tree* with a falling pitch contour), SC-English bilinguals fixated on the correct translation earlier and more frequently compared with when the pitch contour did not match (e.g., *tree* with a rising pitch contour). It is most likely that the matching tonal information was retrieved and exploited to co-activate the SC translation equivalents through top-down and/or lateral translation links. According to the authors, this finding demonstrates a clear influence of suprasegmental information across languages; the effects of native language knowledge on L2 speech processing are not limited to segments but also extend to suprasegmental information.

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

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

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

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

While these findings suggest that interaction between bilinguals' two languages can occur not only at the segmental level but also at the suprasegmental level, it is important to note that all aforementioned studies focused on the domain of speech comprehension. To our knowledge, there is no evidence that came from the production domain in the literature. Studies on second language acquisition may be able to provide some evidence on the influence of lexical tone in non-tonal speech production, as native speakers of tonal languages are found to have unique patterns of producing prosody in stress languages. For example, SC learners of English tend to avoid de-accentuation in post-focal contexts (e.g., McGory, 1997); produce intonational pitch accent in English with a most closely matching tonal contour (Ploquin, 2013); produce stressed syllables in words with pitch peak (e.g., Visceglia & Fodor, 2006). Phonological analysis on tonal substrates of English is also revealing, as researchers recently proposed that, as a result of language contact, Cantonese English has at least two contrastive lexical tones (e.g., Yiu, 2014; Wee, 2016; but see Köhnlein et al., 2019 for a different opinion). None of these studies directly investigate whether, and if so, to what extent lexical tone affects pitch processing in non-tonal speech production. To reach a more comprehensive understanding of language interaction at the suprasegmental level, more empirical data on the influence of lexical tone during spoken word production is needed. Addressing the question of whether lexical tone plays a role in speaking English words not only has important implications for our understanding of bilingual language interaction at the suprasegmental level but could also contribute to a more comprehensive view of how bilinguals' mental lexicon is organized and represented.

The goal of the current study was to fill in this gap by examining the "falling-f0 bias" hypothesis in spoken word production. As findings in speech comprehension suggested (Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2020), SC-English bilinguals represent non-tonal English words with a falling "tone", resulting in a "falling-f0 bias" in their English spoken word recognition. Following this conjecture, one may infer that the processing contrast between falling and rising tones is also evident in English spoken word production. To test this hypothesis, we employed the picture-word interference paradigm (hereafter PWI; Rosinski et al., 1975), the most widely used paradigm in studying spoken word production, and asked native SC-English bilinguals and English monolinguals to name pictures in English while ignoring simultaneously played SC distractor words. Crucially, for the same target word (e.g., *lung*), there were four types of SC distractors: 1) the target's cross-language homophone with a falling tone (CH_F; e.g., *lang* with a falling tone*, "wave"*); 2) the target's cross-language homophone with a rising tone (CH_R; e.g., *lang* with a rising tone*,* *"wolf"*); 3) an unrelated distractor with a falling tone (UN_F; e.g., *you* with a falling tone, *"right"*); 4) an unrelated distractor with a rising tone (UN_R; e.g., you with a rising tone, *"swim"*).

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

## 5.1 Methodology

### 5.1.1 Participants

Forty-eight SC-English bilinguals (39 females and 9 males; average age 24, SD = 1.2) and 48 American English monolinguals (26 females and 22 males; average age 29, SD = 1.5) participated in this study. All SC-English bilingual participants were native SC speakers who grew up in Beijing and spoke no regional dialect. All participants started learning English at an average age of 5.8 (SD = 2.3). Before participating in this experiment, they all passed the College English Test Band 6 or scored above 6 in the International English Language Testing System (IELTS). Their English proficiency level was further accessed with an adapted LEAP-Q questionnaire (Marian et al., 2007) and the multilingual naming test (MINT; Gollan et al., 2012). Using a Likert scale from one to ten, participants' self-rated frequency was 8.5 (SD = 1.4) in reading, 6.7 (SD = 1.8) in speaking, and 7.1 (SD = 1.8) in listening. The average correct response of MINT was 43% (SD = 5.1%). The English monolingual participants had no previous exposure to Mandarin or any other tone languages. All participants had no history of language disorder. This study was approved by the Ethics Committee at Leiden University Centre for Linguistics. All participants provided informed consent and were compensated for their participation.

### 5.1.2 Stimuli

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

### 5.1.3 Procedure

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

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

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

After the PWI task, participants were asked to complete a language background survey, the MINT test (Gollan et al., 2012), and a phonological similarity rating task on targets and their cross-language homophone distractors. In total, the experiment took about 30 minutes.

### 5.2 Results

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

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

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

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

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

Response times were analysed using the generalized linear mixed-effects model (GLMM) with inverse Gaussian distribution (Lo & Andrews, 2015). All the statistical analyses were run in R Studio (R Core Team, 2022) with the package *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Given that error rates were low in each condition, no further analysis on response accuracy was conducted. A maximum model including fixed effects of distractor type (CH_R, CH_F, UN_R and UN_F), participant groups (SC-English bilinguals and English monolinguals), the interaction between distractor type and group, by-subject and by-item random intercept, and by-subject and by-item random slopes for each fixed term were constructed first. Each term was then tested for exclusion. When the model failed to converge, we first increased the number of iterations and then simplified the model by removing correlation parameters in the random structures (Brauer & Curtin, 2018). The final GLMM consists of fixed effects of distractor type, the interaction between distractor type and group, and random intercepts for subject and item. As there was a significant interaction between participant group and distractor type ($p < 0.05$), pairwise comparisons between group and distractors were also computed using the *multcomp* package (Hothorn et al., 2022). Holm–Bonferroni method was implemented to correct family-wise errors (Holm, 1979).

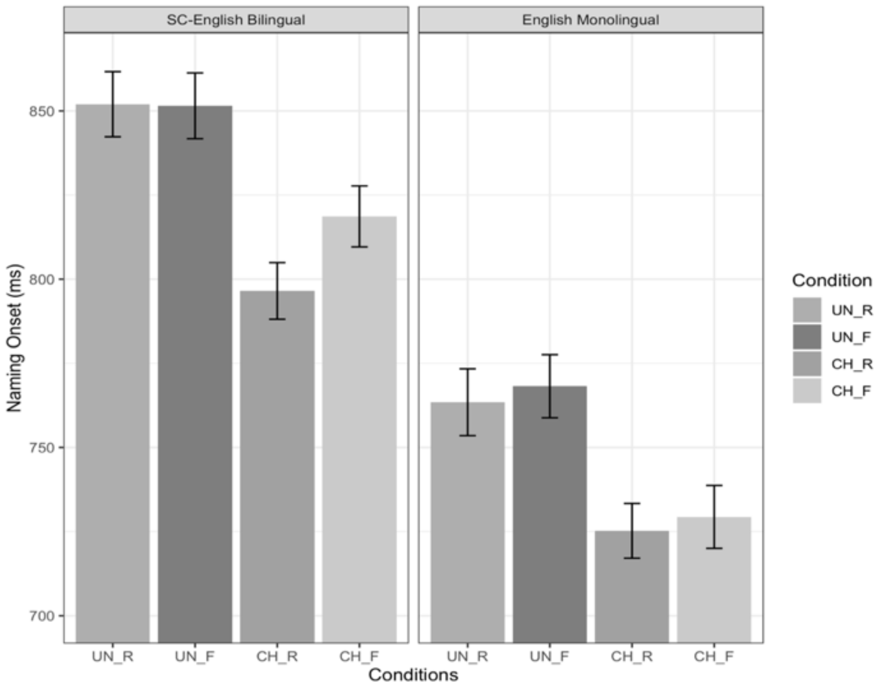


**Figure 1.** *Mean RTs and SDs of SC-English bilingual speakers' and English monolingual speakers' naming latencies in each experimental condition. In the CH_F condition, the distractor is the target's cross-language homophone with a falling lexical tone; in the CH_R condition, the distractor is the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

According to the model estimations (see Table 2), both English monolinguals and SC-English bilinguals took longer to name targets with unrelated distractor words than cross-language homophone distractors (English monolinguals: UR_R vs. CH_R, $p$ < 0.001; UR_F vs. CH_F, $p$ < 0.001; SC-English bilinguals: UR_R vs. CH_R, $p$ < 0.001; UR_F vs. CH_F $p$ < 0.01). This result suggests facilitatory effects of cross-language phonological overlap for both groups. As for differences between the rising- and falling-tone distractors, there was no significant difference between unrelated distractors (UN_R vs. UN_F: $p$ = 0.903) and cross-homophone distractors (CH_R vs. CH_F: $p$ = 0.596) for English monolinguals. Importantly, for SC-English bilinguals, there was a significant difference between cross-homophone distractors (CH_R vs. CH_F distractors: $p$ < 0.05) but no significant difference between unrelated distractors (UN_R vs. UN_F: $p$ = 0.925). This suggests that only SC-English bilinguals' naming latency, but not English monolinguals', was affected by the contrast between falling-tone and rising-tone cross-language homophones.

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

### 5.3 General Discussion

Although it is widely agreed that bilinguals' two languages interact extensively, there is limited finding on whether and how languages interact at the suprasegmental level. A full understanding of the bilingual mind should be backed up with data on whether and to what extent suprasegmental properties such as lexical tone plays a role in bilingual language processing. Our study aimed to fill in this gap by examining the effect of lexical tone on bilingual lexical access during English spoken word production. With PWI, both SC-English bilinguals and English monolinguals were asked to name pictures in English while ignoring auditory SC distractors that were either cross-language homophones with the targets' English names or unrelated. Crucially, we manipulated the pitch contour of the distractors to be either rising (SC Tone 2) or falling (SC Tone 4). This was done to examine the so-called "falling-f0 bias" in SC-English bilinguals from the largely overlooked speech production domain. According to recent comprehension studies (Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2020), SC-English bilinguals represent non-tonal English words with a falling pitch contour similar to with a falling lexical tone, and it is therefore easier for them to access English words with a falling-f0 than their rising-f0 counterparts. If lexical tone indeed "reshapes" pitch representation in English lexical representations as Ortega-Llebaria and her colleagues claimed, we expected to find a significant difference between SC-English bilinguals and English monolinguals in their naming responses to the rising and falling-tone SC distractors. Our results showed that, regardless of the pitch shape difference, both SC-English bilinguals and English monolinguals took less time to name pictures with cross-language homophones than with unrelated distractors. Consistent with previous bilingual PWI studies (e.g., Costa & Caramazza, 1999; Hermans et al., 1998; Costa et al., 2003), this finding suggests that phonologically similar words (cross-language homophones in our case) facilitate picture naming across languages. Moreover, we identified the significant pitch processing difference between SC-English bilinguals and English monolinguals: while both rising and falling-tone cross-homophones were equally facilitative to picture naming in English monolinguals, SC-English bilinguals took significantly longer to name pictures with falling-tone cross-homophone distractors than their rising-tone counterparts.

The results of our speech production study align with previous comprehension studies (Ortega-Llebaria et al., 2017; Ortega-Llebaria & Wu, 2020) in revealing a difference in pitch processing during the lexical access of a non-tonal language. However, upon closer inspection, there appears to be a contradiction in the findings. With primed lexical decision tasks, Ortega-Llebaria et al. (2017) and Ortega-Llebaria & Wu (2020) found that SC-English bilinguals were significantly faster to make lexical decisions on words with a falling pitch contour than words with a rising pitch contour. However, in this study, the "falling-f0 bias" was reversed: SC-English bilinguals took significantly more time to name pictures with falling-tone cross-language homophones than their rising-tone counterparts. If a falling-f0 word is indeed a closer lexical representation in the bilingual lexicon than the corresponding rising-f0 word, one may expect the falling-tone cross-language homophones to facilitate picture naming more, in contrast to the rising-tone ones. In the following, we offer a few possible explanations for such a contrast.

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

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

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

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

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

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

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