CASLAR 2021; 10(1): 31–52

Jiang Liu* and Seth Wiener

CFL learners' Mandarin syllable-tone word production: effects of task and prior phonological and lexical learning

https://doi.org/10.1515/caslar-2021-0002

Abstract: This study examined beginner-level Chinese as a Foreign Language (CFL) learners' production of newly learned words in an image naming and pinyin-reading task. Fifteen L1-English CFL learners learned 10 tonal monosyllabic minimal pairs (e.g., *shu1* and *shu3*) in a three-day sound-image word learning experiment. Ten of the 20 words were homophonous with previously learned words (e.g., participants already knew that *shu1* means 'book'), while the other 10 were not (e.g., no *shu3* words had been learned). Ten of the 20 words had frequent phonology participants were familiar with (e.g., *shi* is a high token frequency syllable), while the other 10 had infrequent phonology (e.g., *ku* is a low token frequency syllable). On the last day of learning, participants performed an image naming task followed by a pinyin-reading task. The recoded word tokens from both tasks were then played to 10 native Chinese speakers who were asked to transcribe the words in pinyin. The results showed that overall word production in the pinyin-reading task was more accurate than image naming. The pinyin-reading advantage was robust, but homophone status and syllable token frequency also interacted with task type: learners produced syllables with high token frequency but without homophones equally well in the pinyin-reading and naming tasks. These results suggest phonological encoding in long-term memory based on pinyin orthography can be affected by learners' prior phonological and lexical knowledge. Pedagogical applications and limitations of the study are discussed, as well.

Keywords: image naming; phonological and lexical knowledge; pinyin; word production

摘要: 本研究目的在于考察汉语初学者在学习完新词汇后， 在命名图片和拼音朗读任务中的发音情况。在三天的词汇学习实验中， 15名母语为英语的汉语学习者

***Corresponding author: Jiang Liu**, Department of Languages, Literatures and Cultures and Linguistics Program, University of South Carolina, 709 Humanities Office Building, 1620 College Street, Columbia, SC, 29208, USA, E-mail: jiangliu@mailbox.sc.edu
Seth Wiener, Department of Modern Languages, Carnegie Mellon University, Pittsburgh, PA, USA, E-mail: sethw1@cmu.edu

学习了10 对单音节词, 每对词由声调区分词义 (例如, 叔 *shu1* 和 鼠 *shu3*) 。20个新词汇中的10 个词汇有之前学过的同音词 (例如, 书*shu1*) , 另外10个新词汇没有之前学过的同音词 (未学过 *shu3*这个音节声调组合) 。10个新词的音节是被试熟悉的 (例如, 音节*shi*) , 10个新词的音节是被试不熟悉的 (例如, 音节*ku*) 。在第三天学习结束后, 被试完成一个图片命名任务, 之后完成一个拼音朗读任务。我们将被试所录的音档播放给10名汉语母语者, 让其把听到的词转录成拼音。结果显示拼音朗读任务中的词汇发音正确度显著高于图片命名中的词汇发音。拼音朗读任务在词汇发音准确度方面优势明显, 但目标词是否有已学过的同音词以及目标词的音节频率也与发音任务产生交互效应, 结果显示高频音节无同音词的目标词的发音在拼音朗读中的准确度与图片命名的准确度没有显著差异。这种交互效应表明由拼音编码在长期记忆中的音系表征在词汇发音过程中是会受到学习者已有音系和词汇知识的影响的。我们也对本研究的一些教学意义和一些局限性进行了讨论。

关键词: 图片命名; 拼音; 音系及词汇知识; 词汇发音

1 Introduction

Adult Chinese as a Foreign Language (CFL) learners who receive formal instruction of Mandarin Chinese (hereafter 'Chinese') in the classroom are typically exposed to pinyin from the beginning of learning. For example, commonly used CFL textbooks, such as *Integrated Chinese* (Liu and Yao 2009) and *Chinese Link* (Wu et al. 2010), all begin with familiarizing learners with the pinyin system. In a survey conducted by Shen (2010) 19 out of 23 Chinese university instructors reported that they presented pinyin along with the target phrase and its English translation for beginning learners in vocabulary instruction in a classroom setting. The same practice has been reported in Chung (2002) and Shen (2013).

Pinyin has been shown to play a facilitative role in tone learning. As lexical tones are notoriously challenging for CFL learners whose L1 is a non-tonal language (e.g., L1 English speakers; see Pelzl 2019 for an overview), several studies have explored how best to teach tone perception and production. Liu et al. (2011) examined learners' tone perception learning outcomes over one semester of classroom learning. Learners were visually presented with one of the three conditions: (1) tone contour + pinyin (the segmental and tonal components were presented separately with pinyin and visual pitch contours); (2) tone number (e.g., 1, 2, 3, 4) + pinyin (separate segmental and tonal representation but no visual pitch contour representation of tone) or (3) tone contour only. Tone identification tasks were given as pre- and posttests. The researchers found that the contour + pinyin condition had more error reduction in tone identification over the course of training than the contour only condition. The critical finding was that the participants made greater improvement in their tone judgment accuracy in the

'contour + pinyin' condition than in the 'number + pinyin' condition, which Liu et al. claimed as the effect of the audio-visual dual-modality principle. Compared with tone numbers, which were arbitrary naming of the tones, the visual input of the pitch contour of a tone offered more robust information for retaining and recalling speech information. Other studies have also pointed out the effectiveness of tone diacritics, which is similar to the visual representation of pitch contour (e.g., ā á ǎ à) in learning monosyllabic words in L2 Chinese (Chung 2002, 2003).

In a recent tone production study, Wiener et al. (2020) found that for total beginners unfamiliar with tone, the presence of tonal diacritics and pitch contour instructions only improved Tone 2 and Tone 4 production accuracy compared to instruction without the visual cue of tonal diacritics and pitch contour. Tone 1 productions improved if participants were given both tone diacritics/contours and multi-talker input. The authors' results suggest that for true beginner-level CFL learners, the effectiveness of showing tone diacritics or pitch contour may only benefit the learning of specific tones at the initial learning stage. For instance, Wiener et al. found that tonal diacritics and visual contours did not improve the production of Tone 3, which is traditionally one of the harder tones for CFL adults to acquire.

Pinyin has also been shown to play a facilitative role in word learning in adult CFL learners. Everson (1998) presented first-year university CFL learners 48 Chinese characters and asked the learners to orally produce the words. Learners were then asked to write down the English meanings of the words. There was a high correlation (around 0.96) between being able to say the word out loud and being able to define/translate it. Moreover, when participants were able to pronounce a word correctly, there was a mean probability of 90.7% that they would be able to give the correct meaning of the word. Similar findings were also reported in Jiang (2003) and Zhao (2003). More recently, Zhang et al. (2019) asked CFL learners to take a computer-based test in which they had to first select a picture that represented the meaning of a target word presented in characters only, and then answer the same item presented in both characters and pinyin. The addition of pinyin substantially increased test reliability and learners' test scores. Among those test items that were initially identified incorrectly when only characters were provided, almost all were correct the second time when pinyin was presented together with the characters.

Chang (2018) further studied how the tone format affected word learning in L2 Chinese. In her study, CFL learners were asked to learn 24 monosyllabic words. Participants were presented with three types of visual prompts while listening to the audio recordings of the target words. The first two conditions consisted of character + pinyin + English translation. The difference was that the first condition used tone diacritics + pinyin while the second condition used tone number + pinyin. The third condition only showed the character + English translation. After the learners

passed a criterion test with 90% correct, they were asked to complete a tone perception and production task. For both tone perception and production, the results showed that participants in the tone diacritics + pinyin condition outperformed participants in the tone number + pinyin condition and participants in the no-pinyin condition (character + English translation). There was no difference between participants in the tone number + pinyin condition and the no-pinyin condition. The result further strengthened the advantage of tone diacritics + pinyin in tonal word learning.

Chang (2018) also found that learners perceived and produced tones more accurately for characters with lower character frequency. The result was inconsistent with previous findings in which characters with higher frequency were perceived more accurately (e.g., Chen 2001; Chen and Wang 2001; Hao and Shu 2005). Chang (2018) pointed out that the different results may have come from the task differences. The studies by Chen (2001), Chen and Wang (2001) and Hao and Shu (2005) only collected written pinyin data by having their participants fill in the blanks with the pinyin of the vocabulary words chosen based on their frequency. In other words, these studies tested the correlation between the character frequency and pinyin knowledge the learners had already learned. In contrast, Chang (2018) studied whether words with different character frequencies were learned differently by CFL learners in terms of orthography, pronunciation, and meaning. Chang's analysis was based on the word learning data and not on the learners' existing vocabulary knowledge.

Overall, there is converging evidence that presenting pinyin together with tone diacritics can facilitate tone and word learning in L2 Chinese. Importantly, however, prior knowledge can also play a role in word learning. In this study, we expand on previous research that has mainly focused on the presentation format of pinyin (e.g., Chang 2018; Chung 2002, 2003; Liu et al. 2011) or the relation between learners' pronunciation and pinyin orthography (e.g., Bassetti 2007). We examine CFL learners' word learning within a more 'natural' environment where pinyin and character orthography are not present at all. By 'natural', we mean a direct mapping between meaning and sound without written orthographic interference. In other words, a word learning process similar to child language acquisition. This approach acknowledges Chang's (2018) finding that that the no-pinyin and tone number + pinyin conditions resulted in comparable word recognition and production accuracy. That is, even without the help of any form of phonetic transcription, learners can still reach a reasonable accuracy rate (62% in Chang's study). Because the participants in Chang's (2018) study had already learned Chinese for over one semester, it is possible that after receiving a certain amount of formal Chinese instruction or input they not only established phonological representations of syllables + tones but also use that phonological knowledge to learn new words without any pinyin orthography. If we

take a step further by not using pinyin or characters at all in word learning, then it will allow us to examine whether learners' previously learned phonological and lexical knowledge affects spoken word learning. In the current study, we explore to what extent learner's previously learned homophones (lexical knowledge) and the syllable token frequency of the target words (phonological knowledge) affect word production after meaning-to-sound association learning.

We specifically build on a recent study by Liu and Wiener (2020), which investigated how prior phonological knowledge in the form of homophones affected beginner-level CFL learners' spoken word learning and recognition of monosyllabic words. Liu and Wiener asked L1-English CFL learners to learn 10 tonal minimal pairs (e.g., 虹*hong2* 'rainbow'— 烘*hong1* 'to bake'). Ten of the 20 words were homophones of previously learned words in their classroom textbook. For example, participants had previously learned that *hong2* can mean 'red' as in 红色 *hong2se4* but not that it could also mean 'rainbow,' which served as the target word. The other half of the words had no known homophones; learners had not been exposed to a *hong1* word in their class or textbook. For three consecutive days participants acquired the 20 words through self-paced learning and then were tested using a 4-alternative-forced-choice task. The task presented a target audio word, e.g., *hong2*, and showed two distractor images along with two images that shared the same syllable but differed in tone (e.g., *hong2, hong1*). CFL learners were asked to click on the image that matched the perceived word as quickly and accurately as possible. On Day 3, participants reached an average accuracy of 90%.

Liu and Wiener found that on the first day of testing, participants were more accurate identifying words with previously learned homophones. The authors argued that new words like 虹 *hong2* 'rainbow' were learned faster than new words like 烘*hong1* 'to bake' because the CFL participants had already established a robust *hong2* phonological representation. On the second and third day of testing, no accuracy differences were found, which the authors claimed was the result of overnight sleep consolidation into the lexicon (e.g., Qin and Zhang 2019). On the second and third day, however, participants identified new words like *hong2* faster than new words like *hong1*, which reflected the homophones' more familiar phonology once the new words were integrated or configured into the L2 lexicon (e.g., Leach and Samuel 2007).

Interestingly, Liu and Wiener found no difference in accuracy or response time given the target word's syllable token frequency. For example, *shi* is a relatively frequent syllable in spoken Chinese and a syllable that CFL learners regularly encounter in class given it is the copula verb. In contrast, *ku* is a relatively infrequent syllable that CFL learners rarely encounter in class. Given that standard Chinese only makes use of roughly 400 (consonant)-vowel-(consonant) syllables like *ji, ta, san,* etc. (DeFrancis 1986; Duanmu 2007), and that native Chinese speakers are highly

sensitive to syllable token frequency information (Chen et al. 2002; Zhou and Marslen-Wilson 1995; Wiener and Ito 2015, 2016) the authors predicted that CFL learners would also show a difference in accuracy and response time, but none was found.

The current study serves as an extension of Liu and Wiener's (2020) study. We examine a subset of the participants from their study and have them perform two oral production tasks after the third day of image-to-sound learning. Given that providing pinyin with tone diacritics is highly effective for L2 Chinese tone and word learning, we examined to what extent the naming of new words through a direct meaning-to-sound mapping task is comparable to the naming/reading of the same new words written in pinyin. To do this, we asked the participants to produce the newly learned words by viewing either their picture only or their pinyin only. In this way, we examine whether learners' use of prior lexical and phonological knowledge differs given the nature of the production task.

2 Research questions

1. After learning a set of new words, do CFL learners produce newly learned words more accurately in an image-naming or pinyin-reading task?
2. Do lexical knowledge of homophone status and phonological knowledge of syllable token frequency interact with this learning across the two oral production tasks?

Given the previously reported advantage of pinyin and tone diacritics on CFL learners' behavioral performance (e.g., Chang 2018; Chung 2002, 2003; Liu et al. 2011; Wiener et al. 2020), we predict that pinyin-reading will result in more accurate productions than image naming. Given that homophones were found to boost CFL acquisition of new words (e.g., Liu and Wiener 2020), we predict that words with homophones will be produced more accurately than words without homophones, irrespective of the production task. Pinyin-reading, however, may show a larger advantage than image naming given the more transparent method of production. We also predict that words with more frequent syllables will be produced more accurately than words with less frequent syllables given that CFL learners are sensitive to syllable token frequency information in perception (Wiener and Lee 2020; Wiener et al. 2019). An interaction may occur in which syllable frequency only affects pinyin-reading given that image naming may not allow the participant to correctly identify the intended syllable.

3 Method

3.1 Participants

Fifteen beginner CFL learners (seven male; eight female; mean age = 19; SD = 0.9; age range: 18–21) who had learned Chinese for just over one semester (about 19 weeks) at a public U.S. university took part in the study. All participants were native English speakers and no participant was a heritage Chinese speaker. No participant self-reported speaking another foreign language. No participant self-reported a hearing or vision impairment, and all participants were able to discriminate two pure tones at 15 Hz or lower. All participants received class credit and a small payment for their participation. Ten native adult Chinese speakers in China with minimal exposure to CFL learners were asked to transcribe in pinyin the word tokens produced by the learners.

3.2 Stimuli

The stimuli design follow that outlined in Liu and Wiener (2020). Briefly, all the words introduced in the participants' textbook (*Integrated Chinese, level 1, part 1*; Liu and Yao 2009) were first analyzed for a total of 115 unique syllable types. From this larger set, 20 monosyllabic words or 10 minimal pairs differing in tone were identified (see Appendix for full stimuli). Homophone status (Homophone/Non-homophone) and syllable token frequency (High frequency/Low frequency) were crossed, resulting in four conditions with five items per condition. For example, the syllable-tone combination *hong2* occurred in known words such as 红色 *hong2-se4* 'red', 红衣服 *hong2-yi1fu* 'red clothes', etc. Thus, we considered *hong* a learned syllable with known homophones and identified one unlearned tonal homophone 虹*hong2* 'rainbow' as well as one unlearned tonal non-homophone 烘*hong1* 'to bake'. Syllable token frequency was calculated as the cumulative frequency of all morphemes' frequency that appeared in the learners' textbook that shared the same syllable (irrespective of tone; see Ziegler et al. 2000 and Wang et al. 2012). For example, *hong2* and *hong1* shared the same *hong* syllable with the previously learned morpheme, 红*hong2* 'red.' We categorized our stimuli as high syllable token frequency if its token count was ≥ 8 and low syllable token frequency if its token count was < 8. We note that because our participants had learned so few words at the time of testing, we were therefore unable to include all possible tonal pairs as target homophone/non-homophone words.

Prior to the experiment, all participants were screened to ensure that they did not know the 20 new words but did know the presumed 10 homophones. None of

the 20 new words were correctly identified by the participants. All participants correctly identified the 10 previously learned words' pinyin and English translation. Some participants did not identify all the morphemes in isolation initially, such as 书, but once provided with the words in which the morpheme appears in the text, e.g., 一本书, the participants identified the morpheme.

The 20 target words were recorded by a female native speaker from Beijing at 44k Hz/16 bits using Praat (Boersma and Weenink 2018). Each word was paired with a hand-drawn color image designed to establish the sound–meaning connection. Figure 1 (online for color version) shows the *hong2-hong1* minimal pair.

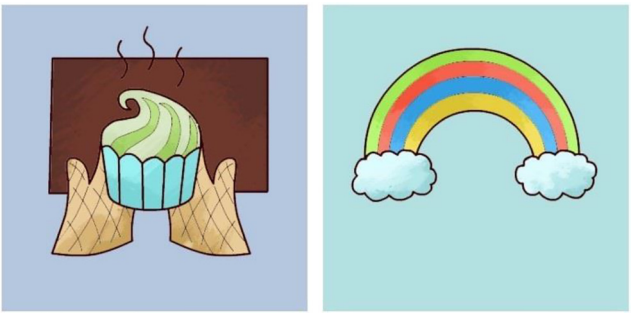

Figure 1: Images used for *hong1* 'to bake' (left, non-homophone-low frequency condition) and *hong2* 'rainbow' (right, homophone-low frequency condition).

3.3 Procedure

All aspects of the study were approved by the Institutional Review Board (IRB) of the first author's university. Before the experiment, all participants read and signed the IRB letter. For three consecutive days, participants performed the same perceptual learning-testing routine. Participants were seated in front of a computer with headphones and simultaneously presented with a spoken word and its corresponding image (20 items × 4 repetitions = 80 randomized trials). Participants were instructed to memorize the sound-image pair and to mouse-click at their own pace to advance to the next trial. After the learning trials, participants were tested using a 4-alternative-force-choice task in which four images were presented on the screen while one of the image's audio was played over headphones (20 randomized trials). Participants were asked to mouse-click on the perceived target as accurately and quickly as possible. See Liu and Wiener (2020) for additional details regarding the testing.

On the third day, after completing the testing trials, participants performed an image naming task and pinyin-reading task, which are the focus of this study. In the

image naming task, the 20 images were pseudo-randomly presented on screen. Participants were instructed to produce the audio label for the newly learned word as accurately as possible. In the pinyin-reading task, the 20 words written in pinyin along with their English translation were presented to participants. Participants were instructed to read the words aloud as accurately as possible. Because we worried that the pinyin-reading task might bias participants, it was always presented second after the image naming task. Responses from both tasks were recorded in Praat at 44 kHz/16 bits.

3.4 Word production accuracy and tone error coding

To measure the participants' production accuracy, we used intelligibility measurements (Derwing and Munro 2015). All utterances produced by learners were transcribed by 10 native Chinese speakers in pinyin. All the L1 raters were college students in China who reported rarely hearing foreigners speak Chinese. The 40 words (20 words in the image naming task; 20 words in the pinyin-reading task) produced by each participant were extracted from each participant's full recording and then normalized to an intensity of 70 dB. Each L1 rater transcribed a total of 600 word tokens (15 participants × 20 target words × 2 Task types). The tokens were split into two blocks. L1 raters completed the transcription of the two blocks on two consecutive days.

In order to quantify the participants' production accuracy, we compared each L1 rater's pinyin transcription to the correct answer in terms of the initial (syllable-initial consonant), final (rime) and tone. For word accuracy, a token received a value '1' when all three parts (i.e., initial, final and tone) were correct and a value '0' otherwise. Cronbach's α was calculated in order to verify interrater reliability among the 10 raters. For the image naming task, Cronbach's α was 0.95. For the pinyin-reading task, Cronbach's α was 0.94. Both values indicate a highly reliable interrater judgment. Any item with disagreement among L1 raters was manually checked by the first author who is a trained phonetician and experienced Chinese instructor. Tone errors were marked when the tone was incorrect, regardless of the segments. For example, a tone error was coded as '1' if the word *ji1* was transcribed either as *ji2* (segments were correct but tone was incorrect) or *ju2* (both vowel and tone were incorrect). If the word was transcribed as '*ji1*' or '*ju1*' then we coded it as '0' as the tone was correct regardless of the segment production accuracy. It is worth noting that for all the token utterances transcribed by L1 raters there was no instance in which the tone was correct but both consonants and vowels were incorrect (e.g., '*ji1*' transcribed as '*ku1*').

4 Results

4.1 Word production accuracy

To test whether task type (image naming vs. pinyin-reading), homophone status (Homophone vs. Non-homophone) and syllable token frequency (High vs. Low) affected word production accuracy, a mixed-effects logistic regression model was built with the *lme4* package in R (version 3.5.1; R Core Team 2017) with task type, homophone status, and syllable token frequency as sum coded variables (1, –1). Model fit, including the selection of predictors and their interactions, the random effect structure, and reported p-values were derived from the χ^2-test of the change in deviance between the models with and without the effect of interest. Table 1 reports the final model output and R code.

Table 1: Summary of fixed-effects coefficient estimates and significance values for outcomes on production accuracy. R code: glmer(Correct Word ~ Task + Task:Homophone + Task:Homophone:-Frequency + (1|Word) + (1|Subject), family = "binomial").

	β	SE	*z* value	*p*
(Intercept)	0.7615	0.4142	1.839	0.07
Task	–0.4322	0.1097	–3.941	<0.001
Task:Homophone	–0.3273	0.1192	–2.746	0.006
Task:Homophone:Frequency	–0.4235	0.1195	–3.544	<0.001

A main effect of task was found, indicating that the production accuracy in the pinyin-reading task (mean: 0.71, 95% Confidence Interval (CI): [0.67, 0.73]) was significantly higher than that in image naming task (mean: 0.58, 95% CI: [0.55, 0.61]). A significant two-way interaction between Task and Homophone and a significant three-way Task, Homophone and Syllable token frequency interaction were found.

To identify the locus of the two-way and three-way interactions, we subset the data into high and low syllable token frequency groups respectively. For words with high token frequency syllables, pinyin-reading had a significantly higher accuracy rate than the image naming task in the Homophone condition ($\beta = 2.35$, $SE = 0.53$, $z = 4.36$, $p < 0.001$) whereas accuracy rates in the image naming and pinyin-reading tasks did not differ in the Non-Homophone condition ($\beta = 0.68$, $SE = 0.46$, $z = 1.47$, $p = 0.14$). For words with low token frequency syllables, the pinyin-reading task had a marginal higher accuracy rate than the image naming task in the Homophone

condition, (β = 0.72, SE = 0.41, z = 1.74, p = 0.08) whereas pinyin-reading had a significantly higher accuracy rate than the image naming task in the Non-Homophone condition, (β = 1.12, SE = 0.42, z = 2.65, p < 0.01). The three-way interaction is summarized in Figure 2 (online for color version), which plots individual participant points, condition means (solid lines), distributions (violin plots), and 95% CI (white boxes).

To summarize the results of the correct syllable-tone analysis, participants were overall more accurate at reading the pinyin than image naming the target, though neither task resulted in mean accuracy above 71%. A two-way and three-way interaction indicated that this difference was primarily driven by two particular conditions. Participants were more accurate at reading the pinyin compared to image naming for new words with homophones that consisted of high token frequency syllables and for new words without homophones that consisted of low token frequency syllables.

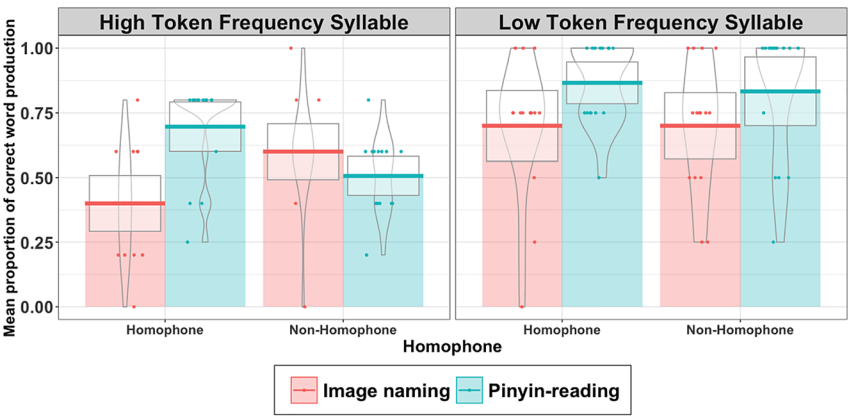

Figure 2: Mean task production accuracy results by participant (point) and condition means (solid line), distributions (violin), and 95% CI (white box).

4.2 Tone error analysis

Because the majority of errors were tonal errors, we next examined participants' tone error pattern. The tone error rate is summarized in Figure 3 (online for color version) using the same plotting scheme as Figure 2. Overall, the error rate of the pinyin-reading (mean: 0.14, 95% CI [0.12, 0.14]) was lower than that of the image naming task (mean: 0.24, 95% CI [0.22, 0.26]). To test whether task type, homophone status and syllable token frequency affected tone error rate, a mixed-effects logistic

regression model was built according to the same criterion as the previous model. Table 2 reports the final model output and R code.

A main effect of task was found, indicating that the pinyin-reading had a significantly lower tone error rate than the image naming task. A significant three-way interaction between task, homophone status and syllable token frequency was also found.

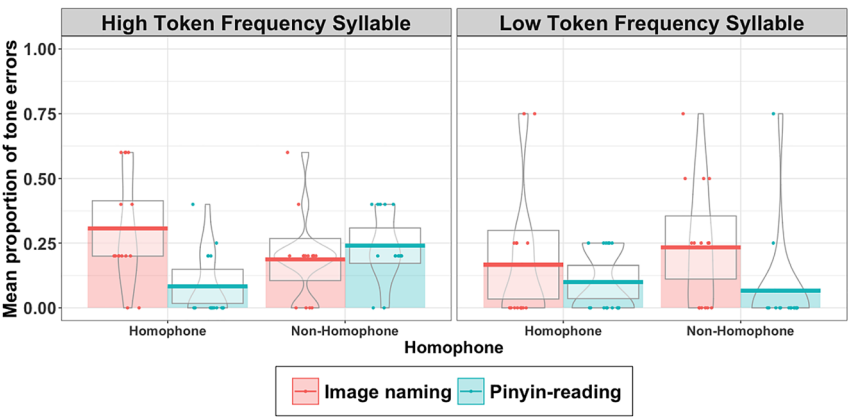

Figure 3: Mean tone error results by participant (point) and condition means (solid line), distributions (violin), and 95% CI (white box).

To identify the locus of the three-way interaction, we subset the data into high and low syllable token frequency words respectively. For words with high syllable token frequency, the pinyin-reading's tone error rate was significantly lower than the image naming task's tone error rate in the Homophone condition ($\beta = -2.57$, $SE = 0.68$, $z = -3.78$, $p < 0.001$) whereas the pinyin-reading and image naming tasks did not differ in terms of tone error rate in the Non-Homophone condition ($\beta = 0.63$, $SE = 0.55$,

Table 2: Summary of fixed-effects coefficient estimates and significance values for outcomes on tone errors. Final model ouput, R code: glmer(ToneError ~ Task + Homophone:Frequency:Task + (1|Word) + (1|Subject), family = "binomial").

	β	SE	*z*-value	*P*
(Intercept)	−2.0843	0.3810	−5.470	<0.001
Task	0.4632	0.1284	3.607	<0.001
Task:Homophone:Frequency	0.4990	0.1348	3.701	<0.001

$z = 1.14$, $p = 0.25$). For words with low syllable token frequency, the pinyin-reading and image naming tasks did not differ in terms of tone error rate in the Homophone condition ($\beta = -0.32$, $SE = 0.45$, $z = -0.72$, $p = 0.47$) whereas the pinyin-reading's tone error rate was significantly lower than the image naming task's tone error rate in the Non-Homophone condition, ($\beta = -1.47$, $SE = 0.52$, $z = -2.82$, $p < 0.01$). The result indicated that the pinyin-reading advantage over the image naming task in terms of

Table 3: Mean tone error results by tone type, syllable token frequency, homophone status, and production task. Rows in italics indicate no items were tested for this combination.

Tone	Syllable token frequency	Homophone status	Task	Mean	SD
1	High	Homophone	Image naming	0.27	0.46
1	High	Homophone	Pinyin-reading	0.00	0.00
1	High	Non-homophone	Image naming	0.06	0.26
1	High	Non-homophone	Pinyin-reading	0.06	0.26
1	Low	Homophone	Image naming	0.13	0.35
1	Low	Homophone	Pinyin-reading	0.00	0.00
1	Low	Non-homophone	Image naming	0.27	0.45
1	Low	Non-homophone	Pinyin-reading	0.06	0.25
2	High	Homophone	Image naming	0.36	0.48
2	High	Homophone	Pinyin-reading	0.20	0.41
2	*High*	*Non-homophone*	*Image naming*		
2	High	Non-homophone	Pinyin-reading	0.00	0.00
2	Low	Homophone	Image naming	0.27	0.46
2	Low	Homophone	Pinyin-reading	0.20	0.41
2	*Low*	*Non-homophone*	*Image naming*		
2	*Low*	*Non-homophone*	*Pinyin-reading*		
3	*High*	*Homophone*	*Image naming*		
3	High	Homophone	Pinyin-reading	0.00	0.00
3	High	Non-homophone	Image naming	0.22	0.42
3	High	Non-homophone	Pinyin-reading	0.38	0.49
3	Low	Homophone	Image naming	0.13	0.35
3	Low	Homophone	Pinyin-reading	0.13	0.35
3	Low	Non-homophone	Image naming	0.13	0.35
3	Low	Non-homophone	Pinyin-reading	0.07	0.26
4	High	Homophone	Image naming	0.20	0.41
4	High	Homophone	Pinyin-reading	0.00	0.00
4	*High*	*Non-homophone*	*Image naming*		
4	*High*	*Non-homophone*	*Pinyin-reading*		
4	*Low*	*Homophone*	*Image naming*		
4	Low	Homophone	Pinyin-reading	0.07	0.26
4	Low	Non-homophone	Image naming	0.27	0.46
4	*Low*	*Non-homophone*	*Pinyin-reading*		

the tone error rate was conditioned by both homophone mate status and syllable token frequency in line with the results from the word accuracy analysis in section 4.1.

Finally, as we pointed out earlier, we were unable to systematically manipulate the tones for a design with fully balanced tonal minimal pairs. Thus, the tone categories were unevenly distributed across conditions, which made any statistical comparison impossible. Given previous findings that showed CFL learners tend to struggle disproportionally across the four categories (e.g., Hao 2012; Wang et al. 2003; Wiener et al. 2020), Table 3 presents means tone error results by tone type, syllable token frequency, homophone status, and production task. Note that rows in italics with a missing mean and standard deviation indicate that no items were tested for this particular combination.

Although we refrain from making any strong conclusions across tone types, we note two general trends from Table 3. First, certain combinations in the pinyin-reading task resulted in no errors. This was not observed in the image naming task. This is noteworthy and suggests that for each tone, at least one combination was potentially easier to learn than others in terms of reading aloud the pinyin. Second, Tone 3 caused the highest error rate (0.38), followed by Tone 2 (0.36), with Tone 1 and Tone 4 having the same lower rates (0.27). These higher error rates were largely in the image naming task, though Tone 3 was driven by the pinyin-reading task. This overall pattern of errors with Tone 3 and 2 being the hardest for CFL learners is in line with previously reported studies (e.g., Hao 2012; Wang et al. 2003).

5 Discussion and pedagogical implications

In the present study, beginner-level CFL learners were exposed to new words by associating an auditory stimulus with a meaning-bearing picture. We varied whether the word had a previously learned homophone (lexical knowledge) and whether the word was composed of a frequent or infrequent syllable sound pattern (phonological knowledge). We then tested to what extent image naming and pinyin-reading of the newly learned words was comparable and whether homophone status and syllable token frequency interacted with the two production tasks. We analyzed word production in terms of whole word accuracy (both syllable and tone were correct) and tone error (tone was incorrect regardless of the segments) to explore our two research questions.

First, we found that the word production accuracy was significantly higher in the pinyin-reading task (mean: 0.71) than the image naming task (mean: 0.58). This indicates that after just over one semester's formal Chinese instruction, the CFL learners have already largely learned how to encode phonological representation

by using pinyin orthography. Moreover, this form of encoding was more reliable than the phonological representation learned through image-to-sound word learning in terms of word production accuracy. This finding is consistent with previous research demonstrating that the visual presentation of tone diacritics or pitch contour shape can facilitate tone learning (e.g., Chang 2018; Liu et al. 2011). Upon closer inspection, however, we found that participants were statistically more accurate at reading the pinyin compared to image naming only for new words with homophones that consisted of high token frequency syllables and for new words without homophones that consisted of low token frequency syllables.

Second, in the tone error analysis, we found a very similar pattern to the word production accuracy result. Overall, the pinyin-reading task had a significantly lower tone error rate (mean: 0.14) than the image naming task (mean: 0.24). Homophone status and syllable token frequency were found to interact. Pinyin-reading showed lower tone error rates than the image naming task only when the new words had homophones that consisted of high token frequency syllables and when the new words did not have homophones and consisted of low token frequency syllables. This pattern was exactly the same as that observed in the word accuracy results.

Finally, we found that neither homophone status alone nor syllable frequency alone affected the accuracy in either production task (or their comparison). This was somewhat unexpected as homophone effects are fairly robust in native Chinese speakers (Chen and Ding 2005; Chen et al. 2009; Zhou and Shu 2008; Ziegler et al. 2000), CFL learners (Liu and Wiener 2020; Wiener et al. 2018), and even naïve learners of an artificial tonal language (Wiener et al. in press). Similarly, syllable frequency effects are well documented in both native and non-native speakers (Chen et al. 2002; Wiener and Lee 2020; Wiener et al. 2019; Zhou and Marslen-Wilson 1995). Unique to our study, it was the combination of these two variables that resulted in our finding. Thus, other lexical and phonological knowledge beyond character frequency (Chang 2018), can affect CFL tone production and word learning. Our results indicate that in either meaning-to-sound or formal classroom word learning settings, learners' tone learning is not necessarily separated from word learning. It is not the case that once a tone category is perceived and produced well in one syllable or rime that it can be easily generalized to another syllable or rime for CFL learners.

To better understand these results and put our findings in a learning context, consider a word such as 塔 *ta3* 'pagoda' whose syllable frequently occurs in the textbook but almost always occurs with Tone 1 (e.g., 他/她 *ta1* 'he/she'). This item caused problems for participants in the image naming task. This is presumably because learners had so often produced *ta* with a high-level Tone 1 that producing it

with a new, unfamiliar tone caused problems, even during the pinyin-reading task. In contrast, a word such as 哭 *ku1* 'to cry' carries a relatively infrequent syllable, so learners may not have had as much experience producing the syllable and learners may have had no other *ku1* form in the lexicon to cause image naming production problems.

In terms of the tone error patterns, consider the *ren3*/*ren2* minimal pair. We found 13 of the 15 tonal errors for *ren3* were mispronounced as *ren2*. There are phonetic and lexical reasons for this pattern. Phonetically, Tone 2 and Tone 3 are the most confusing tone pair because there exists a final pitch rising in both tones if Tone 3 is pronounced in the citation or word-final position (Hao 2012; Moore and Jongman 1997; Sun 1998; Yang 2015). The higher tone error rate in this specific condition may simply come from the perceptual confusion. Lexically, *ren2* is a syllable + tone combination learners had frequently encountered in their textbook (e.g., 人 *ren2* 'people' in 中国人 *zhong1guo2-ren2* 'Chinese nationals', 美国人 *mei3guo2-ren2* 'American nationals'). Interestingly, these errors were observed even in the pinyin-reading task, suggesting that prior experience can strongly affect even straightforward monosyllabic productions.

The observed tonal bias pattern in the pinyin-reading task corroborates previous psycholinguistic research on spoken word learning in which CFL learners showed faster (and more accurate) word recognition for syllables with more probable tones. In other words, L2 learners are biased towards perceiving and producing more likely speech sounds and may relay on this bias to overcome highly variable speech (Wiener et al. 2018, 2019). As for why such prior lexical and phonological knowledge did not affect the tone production accuracy in the image naming task, it seems likely that it may have affected encoding after Day 1 or Day 2 (see Liu and Wiener 2020 for discussion). Additionally, a more sensitive measure including response time of the productions may capture different effects. Future studies may explore this approach.

Importantly, our results should be considered in light of the task. We found the average accuracy rates in our current study (mean: 0.76 in the image naming task) to be slightly higher than that of Wiener et al.'s (2020) study (mean: 0.70 in the no pinyin orthography), which were both higher than the average accuracy rate reported in Chang's (2018) study (mean: 0.62 without pinyin orthography). Two main differences between the former two studies and the latter one may contribute to the tone production accuracy difference. First, the current study only involved meaning-to-sound learning without orthography learning. Wiener et al.'s (2020) study only involved symbol-to-sound learning without meaning learning. But in Chang's (2018) study,

meaning, Chinese character, and sound learning were all included. Thus, the learning task in Chang's study was more cognitively demanding for the participants.

Second, the current study and Wiener et al.'s study designed the target words to contrast in tones whereas Chang's (2018) study did not specifically target tonal contrasts in the word learning phase. Rather, she designed the target words based on character frequency without considering tonal minimal pairs. Although the learners performed the tone perception and production tasks only after they reached 90% accuracy rate criterion in the vocabulary test, participants still did not produce tones as accurately as they did in the studies where only partial lexical representation was being learned as in the current study and Wiener et al.'s study (i.e., meaning + sound or orthography + sound). We argue the lower accuracy rate in Chang's study may come from the limited cognitive resource that can be allocated to the tone learning for beginner level CFL learners (see Barcroft 2015). It is reasonable to say that it is not 'fair' to compare learners' tone production when they are required to learn different amounts of information. To put it in another way, if the goal of instruction is to improve learners' tone production, then maybe reducing the cognitive load of the learning task may help to direct learners to allocate more attention to the tonal contrast.

Although tone learning is critical for learning L2 Chinese, we have to bear in mind that learners not only need to learn the tonal contrasts, they also need to use the tones linguistically to recognize and produce words. However, it is not until recently that researchers have started to pay attention to tone word learning (e.g., Ling and Grüter 2020; Pelzl et al. 2020). When instructors teach vocabulary, they usually focus on the character-sound-meaning association. Often, instructors (and learners) overlook the relatively large number of homophones that exist in standard Chinese. Based on the current finding that showed learners took advantage of homophone mates when producing the newly learned words with pinyin orthography, instructors may consider to use learners' previously learned homophones as an anchor point to learn new words. Such a way may not only help learners aquire new words but also enhance the long-term memory of previously learned vocabulary. This approach may build a more robust representation of the tone bearing unit, and thus, facilitate overall tone perception and production.

6 Limitations

There are two main limitations to our study. First, as previously discussed, our learners had a relatively small lexicon at the time of testing, which meant we were

unable to test all possible tone combinations as minimal pairs. Table 3 shows that several tone-homophone-syllable frequency combinations were not tested simply because our learners did not know enough words. Future studies on more advanced learners (e.g., Pelzl et al. 2020) are needed. Second, our stimuli contained an uneven distribution of nouns and verbs in different conditions. As previous research has shown that in general nouns are easier to learn than verbs (e.g., McDonough et al. 2011), one additional factor that may affect the CFL learners' word production could be the Part of Speech (POS). As the part of speech was unbalanced across conditions among the 20 target words, we were unable to test the effect of POS on the word production in the full dataset. In a post-hoc analysis suggested by a reviewer, we subset our data by selecting 4 tone minimal pairs (the eight target words marked with asterisks in the Appendix) that were balanced across conditions in terms of POS. We then included POS together with other three factors (Task, Homophone status and Syllable frequency) and found our overall results still held, and that POS neither showed a main effect nor an interaction with other factors. With this small sample size data, we cautiously conclude that POS had a null (or at least very limited) effect on the learners' word production accuracy. Future studies should include POS in the experimental design to systematically examine its possible impact on CFL learners' spoken word learning.

7 Conclusion

The current study examined CFL learners' production of newly learned words in image naming and pinyin-reading tasks after a three-day image-to-sound word learning experiment. We found that the word production in the pinyin-reading task was more accurate than production in the image naming task. We also found that lexical knowledge in the form of whether the words had previously learned homophones and phonological knowledge in the form of the syllable token frequency interacted with the task types, causing word and tone production accuracy differences. The results indicate a complex, interdependent relationship in which the CFL learner's lexical and phonological knowledge interacts to affect the production of tones.

Appendix

Experiment stimuli:

Homophone		Non-homophone	
High syllable Token frequency	**Low syllable token frequency**	**High syllable Token frequency**	**Low syllable token frequency**
ta1 塌*	shu1 叔	ta3 塔*	shu3 鼠
'to collapse'	'uncle'	'pagoda'	'mouse'
ren2 仁*	chen4 称[1]	ren3 忍*	chen2 沉
'nuts'	'scale'	'to tolerate'	'to sink'
wan3 碗	jie3 解	wan2 丸	jie1 接
'bowl'	'to untie'	'(meat) ball'	'to pick up'
hai2 骸	ku4 库*	hai3 海	ku1 哭*
'skeleton'	'warehouse'	'ocean'	'to cry'
shi4 柿	hong2 虹*	shi1 湿	hong1 烘*
'persimmon'	'rainbow'	'wetness'	'to bake'

Words marked with the asterisk * were chosen in a subset of the data, which we used for testing the possible effect of Part of Speech (POS) on the word production.

References

Barcroft, Joe. 2015. *Lexical input processing and vocabulary learning*, Vol. 43. Amsterdam: Benjamins.

Bassetti, Benedetta. 2007. Effects of Hanyu Pinyin on pronunciation in learners of Chinese as a foreign language. In Andreas Guder, Jiang Xin & Wan Yexin (eds.), *The cognition, learning and teaching of Chinese characters*. Beijing, China: Beijing Language and Culture University Press.

Boersma, Paul & David Weenink. 2018. *Praat: Doing phonetics by computer [Computer program]. Version 6.0. 43*.

Chang, Yufen. 2018. How pinyin tone formats and character orthography influence Chinese learners' tone acquisition. *Chinese as a Second Language Research* 7(2). 195–219.

Chen, Baoguo & Aihua Ding. 2005. Homophone effects in the recognition of Chinese character: The evidence of phonology influencing the graphic processing of Chinese character. *Psychological Exploration* 25(4). 35–39 (in Chinese).

1 We mistakenly used an audio file pronounced as *chen4* for the image of 'scale'. In standard Beijing Chinese, it should be pronounced as *cheng4*. Removal of this syllable had no effect on the results and therefore we considered it to be accented audio input, which occasionally happens in real L2 Chinese class. Considering the nature of the experiment, the use of '*chen4*' with a different nasal would not have affected the learners to associate the audio label to the image in terms of spoken word learning. Thus, we kept '*chen4*' in the data analysis.

Chen, Hui. 2001. An analysis of the types of error in processing phonogram characters by foreign students. *Language Teaching and Linguistics Studies* 2. 16–20.

Chen, Hui & Kuijing Wang. 2001. Experiment on identification of phonetic-semantic compound among foreign learners. *Chinese Teaching in the World* 2. 75–80.

Chen, Hsin-Chin, Jyotsna Vaid & Jei-Tun Wu. 2009. Homophone density and phonological frequency in Chinese word recognition. *Language and Cognitive Processes* 24(7–8). 967–982.

Chen, Jenn-Yeu, Train-Min Chen & Gary S. Dell. 2002. Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language* 46. 751–781.

Chung, Kevin Kien Hoa. 2002. Effective use of Hanyu Pinyin and English translations as extra stimulus prompts on learning of Chinese characters. *Educational Psychology* 22(2). 149–163.

Chung, Kevin Kien Hoa. 2003. Effects of pinyin and first language words in learning of Chinese characters as a second language. *Journal of Behavioral Education* 12(3). 207–223.

DeFrancis, John. 1986. *The Chinese language: Fact and fantasy*. Honolulu: University of Hawaii Press.

Derwing, Tracey M. & Murray J. Munro. 2015. *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. Amsterdam: Benjamins.

Duanmu, San. 2007. *The phonology of standard Chinese*. Oxford: Oxford University Press.

Everson, Michael E. 1998. Word recognition among learners of Chinese as a foreign language: Investigating the relationships between naming and knowing. *The Modern Language Journal* 82(2). 194–204.

Hao, Meiling & Hua Shu. 2005. The role of phonetic information in learning new characters for second language learners. *Language Teaching and Linguistic Studies* 4. 46–51.

Hao, Yen-Chen. 2012. Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics* 40(2). 269–279.

Jiang, Xing. 2003. Relationship between the knowledge of the sounds and meanings of Chinese characters among foreign students with different native language backgrounds [不同母语 背景的外国学生汉字知音和知义之间关系的研究]. *Language Teaching and Linguistics Studies* 24(6). 51–7 (语言教学与研究).

Leach, Laura & Arthur G. Samuel. 2007. Lexical configuration and lexical engagement: When adults learn new words. *Cognitive Psychology* 55(4). 306–353.

Ling, Wenyi & Theres Grüter. 2020. From sounds to words: The relation between phonological and lexical processing of tone in L2 Mandarin. *Second Language Research*. https://doi.org/10.1177/0267658320941546.

Liu, Jiang. & Seth Wiener. 2020. Homophones facilitate lexical development in a second language. *System* 91. 102249.

Liu, Ying, Min Wang, Charles A. Perfetti, Brian Brubaker, Sumei Wu & Brian MacWhinney. 2011. Learning a tonal language by attending to the tone: An in vivo experiment. *Language Learning* 61(4). 1119–1141.

Liu, Yuehua, Tao-Chung Yao, Nyan-Ping Bi, Liangyan Ge & Yaohua Shi. 2009. *Integrated Chinese (*中文听说读写*) Level 1 Part 1*. Boston, MA: Cheng & Tsui Company.

McDonough, Colleen, L. Song, K. Hirsh-Pasek, R. M. Golinkoff & R. Lannon. 2011. An image is worth a thousand words: Why nouns tend to dominate verbs in early word learning. *Developmental Science* 14(2). 181–189.

Moore, Corinne B. & Jongman. Allard. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *The Journal of the Acoustical Society of America* 102(3). 1864–1877.

Pelzl, Eric. 2019. What makes second language perception of Mandarin tones hard? A non-technical review of evidence from psycholinguistic research. Chinese as a Second Language. *The Journal of the Chinese Language Teachers Association, USA* 54(1). 51–78.

Pelzl, Eric, Ellen F. Lau, Taomei Guo & Robert DeKeyser. 2020. Even in the best-case scenario L2 learners have persistent difficulty perceiving and utilizing tones in Mandarin: Findings from behavioral and event-related potentials. *Studies in Second Language Acquisition*. 1–29. https://doi.org/10.1017/s027226312000039x.

Qin, Zhen & Caicai Zhang. 2019. The effect of overnight consolidation in the perceptual learning of non-native tonal contrasts. *PloS one* 14(12). e0221498.

Shen, Helen H. 2010. Imagery and verbal coding approaches in Chinese vocabulary instruction. *Language Teaching Research* 14. 485–499.

Shen, Helen H. 2013. Chinese L2 literacy development: cognitive characteristics, learning strategies, and pedagogical interventions. *Language and Linguistics Compass* 7. 371–387.

Sun, Sylvia Henel. 1998. *The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese*. Honolulu, HI: Second Language Teaching & Curriculum Center.

Wang, Wenna, Xiaojian Li, Ning Ning, & John X. Zhang. 2012. The nature of the homophone density effect: An ERP study with Chinese spoken monosyllable homophones. *Neuroscience Letters* 516. 67–71.

Wang, Yue, Jongman Allard & Joan A. Sereno. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America* 113. 1033–1043.

Wiener, Seth & Kiwako Ito. 2015. Do syllable-specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition and Neuroscience* 30(9). 1048–1060.

Wiener, Seth & Kiwako Ito. 2016. Impoverished acoustic input triggers probability-based tone processing in mono-dialectal Mandarin listeners. *Journal of Phonetics* 56. 38–51.

Wiener, Seth, Kiwako Ito & Shari R. Speer. 2018. Early L2 spoken word recognition combines input-based and knowledge-based processing. *Language and Speech* 61(4). 632–656.

Wiener, Seth, Kiwako Ito & Shari R. Speer. In press. Effects of multitalker input and instructional method on the dimension-based statistical learning of syllable-tone combinations: An eye-tracking study. *Studies in Second Language Acquisition*. 1–26. https://doi.org/10.1017/S0272263120000418.

Wiener, Seth & Chao-Yang Lee. 2020. Multi-talker speech promotes greater knowledge-based spoken Mandarin word recognition in first and second language listeners. *Frontiers in Psychology* 11. 214.

Wiener, Seth, Chao-Yang Lee & Tao Liang. 2019. Statistical regularities affect the perception of second language speech: Evidence from adult classroom learners of Mandarin Chinese. *Language Learning* 69(3). 527–558.

Wiener, Seth, Marjorie K. Chan & Kiwako Ito. 2020. Do explicit instruction and high variability phonetic training improve nonnative speakers' Mandarin tone productions? *The Modern Language Journal* 104(1). 152–168.

Wu, Su-Mei, Yueming Yu, Yanhui Zhang & Weizhong Tian. 2010. *Chinese Link (*中文天地*)*. Upper Saddle River, NJ: Pearson Prentice Hall.

Yang, Bei. 2015. *Perception and production of Mandarin tones by native speakers and L2 learners*. New York: Springer-Verlag Berlin Heidelberg.

Zhang, Dongbo, Lin Chin-His, Zhang Yining & Choi Yunjeong. 2019. Pinyin or no pinyin: does access to word pronunciation matter in the assessment of Chinese learners' vocabulary knowledge? *The Language Learning Journal* 47(3). 344–353.
Zhao, Guo. 2003. Beginning Chinese learners' character knowledge (初级阶段欧美留学生识字 量与字的构词数). *Applied Linguistics (*语言文字应用*)* 11(3). 106–112.
Zhou, Haiyan & Hua Shu. 2008. The homophone family size effect and semantic transparency effect in Chinese P–O mapping. *Psychological Science* 31. 852–855 (in Chinese).
Zhou, Xiaolin & Marslen-Wilson William. 1995. Morphological structure in the Chinese mental lexicon. *Language and Cognitive Processes* 10(6). 545–600.
Ziegler, Johannes C., Li Hai Tan, Conrad Perry & Marie Montant. 2000. Phonology matters: The phonological frequency effect in written Chinese. *Psychological Science* 11. 234–238.

Bionotes

Jiang Liu
Department of Languages, Literatures and Cultures and Linguistics Program, University of South Carolina, Columbia, SC, USA
jiangliu@mailbox.sc.edu

Jiang Liu is an Assistant Professor of Chinese and Linguistics at the University of South Carolina. He received his PhD degree in Linguistics from the University of Kansas. His primary research areas are psycholinguistics, second language acquisition, phonetics and phonology. Before joining in the Department of Languages, Literatures and Cultures at the University of South Carolina, he served as the Associate Director of Chinese Flagship Program at the University of Minnesota for three years.

Seth Wiener
Department of Modern Languages, Carnegie Mellon University, Pittsburgh, PA, USA
sethw1@cmu.edu

Seth Wiener is an Associate Professor of Second Language Acquisition and Chinese Studies in the Department of Modern Languages at Carnegie Mellon University (CMU). He received his PhD degree from The Ohio State University and carries out research on speech learning and lexical processes using psycholinguistic methods. He is the Director of the Language Acquisition, Processing, and Pedagogy (LAPP) Lab at CMU and is an Associate Editor for the journal *Applied Psycholinguistics*.