

ORIGINAL ARTICLE

The effect of second-language orthographic input on the phonological encoding of Mandarin words

Yen-Chen Hao^{1*}  and Chung-Lin Martin Yang²

¹Department of Modern Foreign Languages and Literatures, University of Tennessee, Knoxville, TN, U.S.A. and ²Department of Brain & Cognitive Sciences, University of Rochester, Rochester, NY, U.S.A.

*Corresponding author. Email: yhao6@utk.edu

(Received 20 October 2019; revised 9 February 2021; accepted 22 February 2021; first published online 11 June 2021)

Abstract

Previous studies have yielded mixed findings regarding the effect of familiar and novel L2 graphemes on learners' phonological encoding. The current study investigated this topic by comparing the effect of Pinyin and Chinese characters on English speakers' Mandarin word learning. Different from previous research, this study examined both segmental and tonal encoding and compared participants from different Mandarin proficiency levels. Seventeen Advanced learners, 29 Intermediate learners, and 21 Naïve English speakers participated in a word-learning experiment in which half of the participants were exposed to the Pinyin spelling of the target words while the other half to characters. After the learning phase, they did a meaning – auditory stimulus matching task. Half of the stimuli were complete matches, while in the other half the stimulus mismatched the target either in segments or tones. The results revealed that at the Advanced level, the Character group was more accurate than the Pinyin group in rejecting tonal mismatches to the target words, while the opposite tendency was observed at the Naïve level. This study suggests that novel graphemes facilitate advanced L2 learners' tonal encoding more than familiar graphemes, which is probably due to the unique nature of Chinese characters.

Keywords: L2 orthography; Mandarin word learning; different L2 proficiency levels; segmental and tonal encoding

One of the most crucial tasks in second-language (L2) acquisition is to build a lexicon, which requires forming an association between sound and meaning. The orthographic form of an L2 word, although not always required, can also be linked to the lexicon especially after schooling. Learning the orthography of an L2 is not only essential for reading and writing, but has also been found to affect learners' encoding of L2 phonological forms (e.g., Hayes-Harb et al., 2010). However, whether L2 orthographic input plays a facilitatory or negative role in phonological acquisition is still under debate. Some studies suggest that the availability of L2 orthographic forms aids learners' encoding of L2 sounds (Erdener & Burnham, 2005;

Escudero et al., 2008). For example, Escudero et al. (2008) found that Dutch speakers who were provided with written representation in addition to auditory input of English nonwords containing the vowels /*ɛ*/ and /*æ*/ during training learned to distinguish these two vowels to some degree (even though Dutch does not have the /*ɛ*/-/*æ*/ contrast), while those who received auditory input only could not discriminate this vowel contrast. In another study, Erdener and Burnham (2005) showed that Turkish and Australian English speakers made fewer errors in repeating Spanish and Irish nonwords when the spelled forms were available to them. These studies demonstrate a positive effect of L2 orthography in L2 phonological acquisition.

In contrast, other studies provide evidence for a negative effect of L2 orthographic input on L2 phonological encoding, especially when the L2 and L1 use the same alphabet but employ different grapheme-to-phoneme correspondences, or GPCs (Bassetti, 2006, 2007; Hayes-Harb et al., 2010; Young-Scholten, 2002; Young-Scholten & Langer, 2015). In one study, Hayes-Harb et al. (2010) found that when English speakers were presented with the spelled form of a new word that did not conform to their L1 GPCs, they were more likely to memorize the new word using their L1 correspondence and thus make errors. For instance, when the target word [fʌfə] was spelled as <faza>, the English speakers were more likely to memorize it incorrectly as [fuzə]. Similarly, Bassetti (2006, 2007) observed that English speakers tended to use their L1 GPCs to interpret Pinyin (Romanized transcription of Mandarin Chinese used in China). Hence, they often failed to pronounce or perceive a vowel that was not represented in Pinyin, even when it was present in the audio stimulus. In sum, these studies suggest that L2 orthographic representation may negatively interfere with phonological encoding of L2 sounds due to mismatches between L1 and L2 GPCs.

In view of the negative L1 interference when L2 uses the same alphabet but adopts different GPCs, a few studies have investigated the effect of L2 orthography that utilizes novel graphemes, defined as graphemes that are nonexistent in the learners' L1. Among them, Showalter and Hayes-Harb (2013) examined the effect of Mandarin tonal diacritics (e.g., < ā, á, ǎ, à > for Tone 1: high level, Tone 2: high rising, Tone 3: low dipping, and Tone 4: high falling, respectively) on English speakers' encoding of Mandarin tonal categories. They found that those who saw Mandarin tonal diacritics during the word-learning phase discriminated Mandarin tonal contrasts more accurately in the subsequent perception test than those who did not see the diacritics. This finding demonstrated that novel graphemes helped English speakers establish new phonological categories for novel L2 phonemes (i.e., Mandarin tones). In another study, Hayes-Harb and Cheng (2016) compared the relative benefit of familiar and novel L2 graphemes on native English speakers' phonological encoding of Mandarin nonwords. Specifically, during the word-learning phase, one group of English speakers were presented with Pinyin (familiar graphemes) and the other group with Zhuyin (phonetic transcription of Mandarin Chinese used in Taiwan utilizing graphemes nonexistent in English). The results showed that the Zhuyin group needed more time than the Pinyin group to memorize the correspondences between auditory stimuli and written forms, most likely because they had to learn new graphemes in addition to the GPC rules. In the subsequent test on the matchedness of meaning and sound of

Mandarin nonwords, the Zhuyin group achieved higher accuracy than the Pinyin group in rejecting segmental mismatches between the target words and audio stimuli, particularly when the mismatches conformed to the English GPCs. This study suggests that novel L2 graphemes are more beneficial than familiar graphemes in helping learners encode accurate phonological forms of L2 words, because they are less likely to activate L1 GPCs and thus avoid negative L1 interference. However, in another study that similarly compared the effect of Pinyin and Zhuyin input on native English speakers' ability to distinguish English and Mandarin consonants, Pytlyk (2011) did not find an advantage of Zhuyin input over Pinyin. The Zhuyin group, Pinyin group, and a third group with no orthographic input did not differ statistically in their improvement from pretest to posttest. If anything, the Pinyin group descriptively made a larger improvement than the other two groups. Pytlyk hypothesized that the Zhuyin group did not show the expected advantage probably because of their heavier cognitive load of having to learn new graphemes and new GPCs simultaneously. In addition to Pytlyk (2011), several other studies have also shown little difference in the learning outcome when exposing learners to familiar or novel L2 graphemes during the learning phase (e.g., Mathieu, 2016; Showalter, 2018; Showalter & Hayes-Harb, 2015; Simon et al., 2010). These studies posit that factors other than L2 grapheme familiarity may contribute to the effect of L2 orthography on phonological encoding, including transparency of the learners' L1 orthography, difficulty of the L2 phonological contrasts, complexity and foreignness of the novel L2 graphemes, among other factors.

One factor that has been less well studied is whether the effect of L2 orthography on L2 phonological encoding differs for learners at different L2 proficiency levels. Showalter (2020) investigated this question by comparing the phono-lexical acquisition of Russian nonwords by English speakers with no Russian experience (naïve group) and learners of Russian from elementary and advanced-level Russian classes (beginner and experienced groups, respectively). All participants learned Russian nonwords whose initial phoneme was represented by three types of Cyrillic graphemes: ones that do not exist in English (novel graphemes), exist in English and comply with the English GPCs (familiar congruent), and exist in English but have different GPCs (familiar incongruent). Showalter (2020) found that the naïve English speakers and both learner groups performed at ceiling in matching audio stimulus with its pictured meaning when the target words contained novel graphemes or familiar congruent graphemes. A significant between-group difference was found only when the target words contained familiar incongruent graphemes, with the experienced learners performing significantly more accurately than the naïve group. This study suggests that grapheme familiarity per se does not affect L2 phono-lexical acquisition, regardless of learners' L2 proficiency level. However, L2 graphemes that exist in the learners' L1 but with different GPCs have a detrimental effect for participants naïve to the L2 but not for experienced learners. In another study that compared learners with different amounts of L2 experience, Hayes-Harb and Hacking (2015) investigated whether written stress marks, which are novel graphemes for native English speakers, facilitated their phonological encoding of Russian stress contrasts. They found that for both experienced and inexperienced learners of Russian, whether the stress marks were available did not change their performance: the inexperienced learners performed at chance,

while the experienced learners were significantly more accurate than the inexperienced learners. These two studies suggest that there is little difference in the influence of novel L2 graphemes on experienced and inexperienced learners' phonological encoding.

Similar to Showalter (2020), the current study compared the effect of familiar and novel graphemes on the phonological encoding of Mandarin words by native English speakers with different amounts of L2 experience. However, our study crucially differs from Showalter (2020) and most previous studies in the following aspects. First, the novel L2 graphemes that we examined are Chinese characters, which are not considered a type of sound-based orthography. Thus, the findings should provide a unique insight into the existing literature, which has primarily focused on sound-based orthographies. Second, the majority of previous orthography studies targeted at consonantal or vowel contrasts, with a few investigating prosodic contrasts (Hayes-Harb & Hacking, 2015; Showalter & Hayes-Harb, 2013). The present study examined both segmental and tonal encoding, which should provide a more comprehensive picture of the influence of orthography on different types of sound contrasts. Finally, all the target stimuli in this study were real words in Mandarin Chinese. Hence, the findings should be more comparable and applicable to natural L2 word acquisition.

English orthography, Chinese characters, and Pinyin

English and Chinese orthographies differ extensively in the association between the written form, sound, and meaning, which affects how native speakers encode phonological information of a word in memory. Classified as a phonological orthography, the English writing system uses Roman alphabet to represent phonemes. However, there are many irregularities in the GPCs, especially for vowels. As a result, English is typically considered an opaque orthography (Van den Bosch et al., 1994). The graphemes do not provide any semantic information, which has to be obtained by converting graphemes into phonemes and then assembled into words. Therefore, English speakers have often been found to rely more heavily on phonological than orthographic information in word learning, identification, and reading (Chikamatsu, 1996; Hayes-Harb et al., 2010; Turnage & McGinnies, 1973).

Different from English, the Chinese writing system is generally regarded as a meaning-based rather than sound-based orthography. Its graphemes, Chinese characters, often contain a semantic component (radical) that cues the meaning. For example, most characters whose meanings have something to do with water contain the water radical “氵”, and most characters whose meanings have something to do with metal contain the gold radical “金”. In addition to semantic information, more than 74% of the most commonly used Chinese characters also contain a phonetic component that may serve as a pronunciation cue (Chen, 1999; DeFrancis, 1984). For instance, the pronunciations of the characters 箱 <*xiāng*> “box”, 厢 <*xiāng*> “wing-room,” and 湘 <*xiāng*> “short form for Hunan Province” are the same as their phonetic component “相”. However, characters differ from

graphemes in a phonological orthography in two crucial ways. First, characters cannot be decomposed into different phonemes such as consonants, vowel, and tones. Instead, each character corresponds to a whole syllable. Second, the number of characters far exceeds the number of possible syllables in Mandarin, and thus there are often many-to-one relations between characters and syllables. For example, 力 “power”, 利 “profit”, and 粒 “measure word for small round objects”, which do not share any phonetic component, are pronounced identically. Probably due to the abundance of homophones, native Chinese speakers have been found to rely more on visual (written forms) than audio information in word recognition (Leck, Weekes & Chen, 1995; Turnage & McGinnies, 1973), even when they recognize words in another language that uses a phonological orthography (Haynes & Carr, 1990; Koda, 1990; Wang & Koda, 2005; Wang, Koda & Perfetti, 2003). Furthermore, the literacy rate among Chinese speakers is relatively low compared to speakers of alphabetic languages due to the common dissociations between Chinese characters and pronunciation (Chen, 1996; Coulmas, 1983; DeFrancis, 1984).

Pinyin is the official Romanized transcription of Mandarin Chinese used in China. It was developed by Chinese linguists in the 1950s and has been used ever since. Pinyin utilizes the Roman alphabet to represent segments (consonants and vowels) of each syllable, and there is generally a one-to-one correspondence between graphemes and phonemes. Pinyin and English share some GPCs. However, since some Mandarin phonemes do not exist in English and vice versa, discrepancies also exist (cf. Bassetti, 2006, 2007). For example, the letter <x> stands for [ç] in Pinyin but typically stands for [ks] in English; the letter <z> stands for [ts] in Pinyin but for [z] in English. Another difference between Pinyin and English orthography is that the four lexical tones in Mandarin are marked by diacritics above the main vowel in Pinyin, namely <wēi> for Tone 1: high-level tone; <wéi> for Tone 2: rising tone; <wěi> for Tone 3: low-dipping tone; <wèi> for Tone 4: falling tone. In China, Pinyin is commonly used to type Chinese characters with a Western keyboard, aid Chinese children in learning the pronunciation of characters, and annotate sounds in dictionaries. However, it does not replace Chinese characters in written communications. As for L2 instruction, Pinyin is widely used in Mandarin instruction for English-speaking beginning learners due to its shared graphemes with English orthography as well as its regular GPCs. However, as L2 learners gain experience in Mandarin and receive more training in reading and writing, the use of characters in language instruction usually becomes more dominant.

Research questions and hypotheses

The research questions that this study addresses are (1) Do English speakers benefit more from Chinese character or Pinyin input in their phonological encoding of new Mandarin words? (2) Does the relative benefit of these two types of orthographic input differ for English speakers at different Mandarin proficiency levels? and (3) Is there a differential effect on segmental and tonal encoding between these two types of orthographic input?

It is hypothesized that Pinyin should facilitate L2 phonological encoding more than characters for English speakers with little or no knowledge of Mandarin. This is

because English speakers do not need to learn new graphemes given the similarity of Pinyin to the Roman alphabet used in English orthography, and thus they can focus on phonological encoding of Mandarin words. As for more advanced learners, they may find Chinese characters more helpful than Pinyin for their new word learning, since they have become accustomed to using characters in their language study. Furthermore, the semantic information carried by characters may also facilitate their word learning. With regard to segmental and tonal encoding, previous studies have suggested that Pinyin often negatively interferes with English speakers' phonological encoding because of discrepancies in the GPCs between Pinyin and English (Bassetti, 2006, 2007; Hayes-Harb & Cheng, 2016). On the other hand, the tonal diacritics in Pinyin have been found to facilitate English speakers' learning of Mandarin tonal categories (Showalter & Hayes-Harb, 2013). However, we have no basis to predict the effect of Chinese characters on Mandarin segmental and tonal encoding, because segments and tones are not marked in characters and there is not yet any literature on this topic. The findings of this study shall shed some light on this question.

Methods

Participants

Three groups of native English speakers differing in their Mandarin proficiency levels participated in this study. The Naïve group was composed of 21 university students who had no knowledge of Mandarin (2 M and 19 F; mean age = 21.24, $SD = 0.68$). The Intermediate (Inter) group consisted of 29 L2 learners of Mandarin recruited from second- and third-year Chinese classes (13 M and 16 F; mean age = 19.62, $SD = 1.27$). The Advanced (Adv) group consisted of 17 learners recruited from fourth-year Chinese classes (11 M and 6 F; mean age = 20.56, $SD = 1.34$). The participants at each level were randomly assigned to the Pinyin input (PY) group and Character input (CH) group. Specifically, the first participant from each level who signed up for the experiment was assigned to the PY group, the second participant assigned to the CH group, the third assigned to the PY group, the fourth assigned to the CH group, and so on. Due to a few cases of no-shows, the numbers of participants in the PY and CH groups were close but not identical (Naïve: 11 vs. 10; Inter: 16 vs. 13; Adv: 9 vs. 8 in the PY and CH groups, respectively). Their average length of Mandarin studying in years and study-abroad duration in months are summarized in Table 1.

To check whether the PY and CH groups at each proficiency level had comparable Mandarin learning experience, independent t tests were conducted to compare their length of Mandarin studying and study-abroad duration. The results are reported in Table 1. One may notice that the CH group at the Adv level had studied Mandarin and studied abroad for a longer period of time than the PY group. However, none of the difference between the PY and CH groups was significant ($ps > .05$). Independent t tests were also conducted to compare the Mandarin experience between the Inter and Adv learners. Regarding the length of studying, no significant difference was found between the Inter and Adv learners ($t(44) = .71, p = .48$). As for the study-abroad duration, the Adv learners had spent a significantly longer period of time in Chinese-speaking countries than the Inter learners ($t(44) = 3.68$,

Table 1. The average length (*M*), standard deviation (*SD*), and range of Mandarin studying (in years) and study-abroad duration (in months) of the Inter and Adv learners

Level	Input group	Length of studying (years)	Comparison of PY and CH	Study-abroad duration (months)	Comparison of PY and CH
Intermediate	PY	<i>M</i> = 3.94 <i>SD</i> = 1.82 Range = 1–7	$t(27) = .19$ $p = .86$	<i>M</i> = 1.38 <i>SD</i> = 2.82 Range = 0–12	$t(27) = .57$ $p = .57$
	CH	<i>M</i> = 4.08 <i>SD</i> = 2.10 Range = 1–8		<i>M</i> = 0.89 <i>SD</i> = 1.06 Range = 0–3	
Advanced	PY	<i>M</i> = 3.72 <i>SD</i> = 1.40 Range = 2–7	$t(15) = 1.53$ $p = .15$	<i>M</i> = 4.11 <i>SD</i> = 3.37 Range = 1–12	$t(15) = .62$ $p = .55$
	CH	<i>M</i> = 5.25 <i>SD</i> = 2.38 Range = 3–10		<i>M</i> = 5.44 <i>SD</i> = 4.91 Range = 1–16	

$p = .001$). This indicates the two learner groups differed mainly in the amount of immersion experience they had received but not in the years of formal education. In fact, the PY group at the Inter level had slightly longer average length of Mandarin studying than the PY group at the Adv level. This could be because the Inter learners were recruited from both second- and third-year Chinese classes and thus exhibited larger variation in their Mandarin experience.

Target words

Sixteen disyllabic Mandarin words were selected from the vocabulary list of the highest Hanyu Shuiping Kaoshi (HSK) levels, level 5 and 6. The HSK is a standardized Mandarin proficiency test most commonly used in China, administered by its Ministry of Education. The vocabulary list of level 5 and 6 includes low-frequency words that may be known only by very advanced L2 learners. Another selection criterion was that the individual characters used to form these target words are vocabulary words in the textbooks of Elementary Chinese classes (*Integrated Chinese Level 1 part 1 & 2*, 3rd edition, Liu et al., 2009). The purpose of these selection criteria was to ensure that the meanings of the target words were unknown to the participants, yet the characters were known to the two learner groups. Hence, the L2 learners in the CH group should be able to focus on learning the phonological form and lexical meaning of the target words rather than having to learn new graphemes at the same time.

After selecting the target words, one segmentally mismatched variant and one tonally mismatched variant were created for each target word. Table 2 lists two target words along with their segmental and tonal mismatches as an example. Half of the variations were made in the first syllable, while the other half in the second syllable. The segmental mismatches included both consonantal and vowel divergence. The divergence was created by replacing the target segment with a phonetically similar sound. For example, the vowel in the second syllable of the Mandarin word

Table 2. The English meaning, Characters (CH), Pinyin (PY), and IPA of two target words and their segmental and tonal mismatches. The mismatches are in bold

Target words		Segmental mismatch	Tonal mismatch
English	“theory”		
CH	学说		
PY	xuéshuō	xuésh ōu	xuéshu ò
IPA	[ɕyɛT2 ɕuoT1]	[ɕyɛT2 ɕ ou T1]	[ɕyɛT2 ɕuo T4]
English	“day dream”		
CH	空想		
PY	kōngxiǎng	kōng sh ǎng	kōngxi áng
IPA	[kʰoŋT1 ɕiǎŋT3]	[kʰoŋT1 ɕ ɑ ŋT3]	[kʰoŋT1 ɕiǎŋ T2]

“theory” [ɕyɛT2 ɕuoT1] was replaced by [ou]; the consonant in the second syllable of the Mandarin word “day dream” [kʰoŋT1 ɕiǎŋT3] was replaced by [ɕ]. The tonal mismatches focused on the contrast between Tone 1 (high level) and Tone 4 (high falling) and that between Tone 2 (rising) and Tone 3 (dipping), because previous studies have shown that these two pairs are more confusable than others for English speakers who have no knowledge of Mandarin (So & Best, 2010, 2014) as well as for L2 learners of Mandarin (Hao, 2012; Kirilloff, 1969; Shen, 1989; Wang et al., 1999). For instance, the tonal mismatch for the Mandarin word “theory” [ɕyɛT2 ɕuoT1] was created by replacing Tone 1 in the second syllable with Tone 4, as in [ɕyɛT2 ɕuoT4]. The full list of target words and their phonological mismatches is provided in the Appendix.

One female native Mandarin speaker from northern China produced the 16 target words, 16 segmental mismatches, and 16 tonal mismatches. These recordings were used as the stimuli of this experiment.

Procedure

The experiment consisted of a target word assessment, word-learning phase, a criterion test, and a perception test. A naming test was also administered, but the result will not be reported here. All the tasks except for the target word assessment were programmed and administered by the software *E-Prime 2.0 Professional*. Each participant completed the experiment individually in a sound-attenuated room.

Target word assessment

The participants listened to the recording of the 16 target words and tried to write down their English meaning. If they did not know the meaning, they were instructed to write down “x”. This task was self-paced and they could listen to each word as many times as they needed. The purpose of this task was to confirm that all the targets were unknown to the participants.

Word-learning phase

The participants listened to the audio recording of each target word through headphones (*Sennheiser PC350*) and saw its English translation on the computer screen. For each target word, the PY group saw its Pinyin with tonal diacritics below the English translation, while the CH group saw its Chinese characters. They were instructed to repeat each audio stimulus and memorize the association between the sound and meaning (English translation). They were told that the orthographic input below the English translation was to assist their learning, and they would not be tested on the orthographic forms after the learning phase. Each target word appeared four times in a random order. The learning phase was self-paced and no feedback was given.

Criterion test

The purpose of the criterion test was to ensure that the participants had memorized the association between the sound and meaning of the target words before starting the perception test. In each trial, the participants heard an audio stimulus and saw an English meaning on the computer screen. They had to judge whether the sound matched the meaning. No Pinyin or character input was provided. Each of the 16 target words appeared twice on the screen, once paired with an audio stimulus that matched its meaning and the other time with an audio stimulus of a different target word (which had different segmental and tonal composition), resulting in 32 trials. The participants had to achieve an accuracy of 90% or higher to advance to the perception test. If the participants failed to meet the requirement, they had to repeat the learning phase and criterion test until they reached the threshold.

Perception test

In the perception test, the participants were presented with 128 pairs of audio stimuli and English meanings and had to judge whether they matched or not. The participants were told that the procedure for this test was the same as that for the criterion test, but the mismatch between the audio stimulus and English meaning was more subtle than that in the criterion test. In the mismatched trials, the audio stimulus differed from the target word only by one segment or by one tone, rather than corresponded to a different word. In this test, 64 pairs were complete matches (16 target words \times 4 repetitions), 32 pairs were segmental mismatches (16 segmental mismatches \times 2 repetitions), and 32 pairs were tonal mismatches (16 tonal mismatches \times 2 repetitions). They were presented in a random order.

Analysis

To check whether the PY and CH groups differed in the time needed to pass the criterion test, independent *t* tests were conducted within each proficiency level to compare the number of times the participants had to go through the learning phase. Aside from the time needed for learning, individual participants' accuracy rates in the perception test were converted to *d*-prime (MacMillan & Creelman, 2004) in order to assess their sensitivity to phonological mismatches, which is a measure of their phonological encoding accuracy. To further compare their ability to encode

Table 3. The average number and range of the learning phase the participants needed to go through to pass the criterion test

Level	Adv	Inter	Naïve
Input			
PY group	$M = 1.11$ (range 1-2)	$M = 1.13$ (range 1-2)	$M = 2.7$ (range 2-4)
CH group	$M = 1.13$ (range 1-2)	$M = 1.08$ (range 1-2)	$M = 2.7$ (range 2-4)

segmental and tonal contrasts, accuracy rates for segmental and tonal mismatches were separately computed and converted to d-prime as well. The target words correctly identified by each participant in the target word assessment were removed from his/her data, so that all the target words analyzed were unknown to the learners prior to the experiment. This ensured that the learners' accuracy in the perception test reflected their learning outcome from the learning phase rather than their prior knowledge of the target words.

Results

The number of times the participants needed to go through the learning phase to pass the criterion test are summarized in Table 3. It can be seen that the Naïve participants in both PY (Pinyin) and CH (Character) groups generally had to go through the learning phase three times to reach 90% accuracy in matching the sound and meaning of the target words. The Inter (Intermediate) and Adv (Advanced) learners, on the other hand, needed to go through a little over once on average. Independent *t* tests revealed that the difference between the PY and CH groups was not statistically significant at any level (Naïve: $t(19) = 0$, $p = 1$; Inter: $t(27) = .41$, $p = .69$; Adv: $t(15) = .08$, $p = .94$). In other words, the PY and CH group did not significantly differ in the number of learning cycles they needed in order to move on to the next phase.

The accuracy rates of the PY groups at the three proficiency levels in the perception test are plotted in Figure 1, while those of the CH groups are plotted in Figure 2. The accuracy rates of the Naïve group are represented by white bars, those of the Inter group by light gray bars, and those of the Adv group by dark gray bars. From left to right are the accuracy rates in the matched, segmental-mismatched, and tonal-mismatched trials.

As can be seen from Figures 1 and 2, one clear similarity between the PY and CH groups was that the participants were generally least accurate in the tonal-mismatched trials in which the audio stimulus differed from the target word only by the tone of one syllable. One noticeable dissimilarity between the PY and CH groups, on the other hand, was the greater accuracy difference in the tonal-mismatched condition between the three proficiency levels in the CH groups than in the PY groups.

The PY and CH groups' overall d-prime scores as well as those for segmental-mismatched and tonal-mismatched items are plotted in Figure 3. The PY groups' values are represented by light gray bars, and the CH groups' values by dark gray

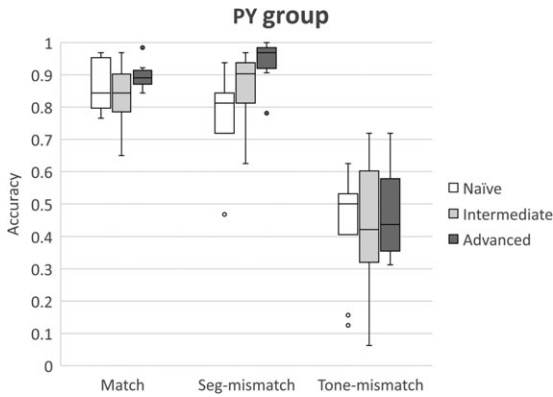


Figure 1. Accuracy rates of the PY groups at the three proficiency levels in the perception test.

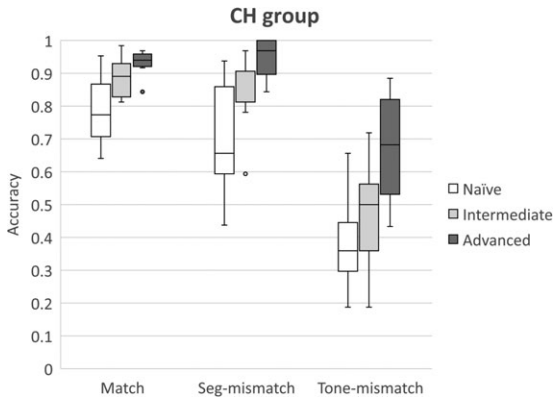


Figure 2. Accuracy rates of the CH groups at the three proficiency levels in the perception test.

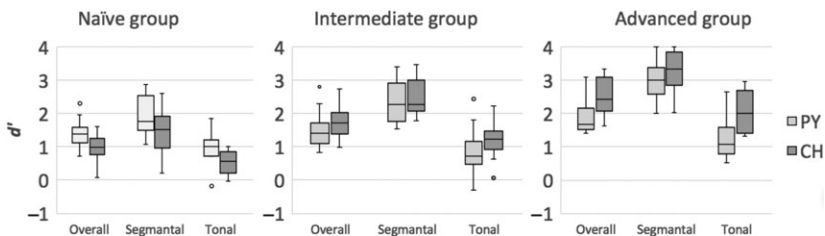


Figure 3. D-prime scores of the PY and CH groups at the three proficiency levels in the perception test.

bars. From left to right are the Naïve, Inter, and Adv level. A visual inspection of the figure revealed that at the Naïve level the PY group was generally more sensitive to phonological mismatches than the CH group, while at the Inter and Adv levels the CH groups showed higher sensitivity than the PY groups. These participants'

d-prime values were submitted to a repeated measures ANOVA with orthographic input (PY, CH) and proficiency level (Naïve, Inter, Adv) as between-subject factors and mismatch type (overall, segmental mismatch, and tonal mismatch) as within-subject factor. The main effects of proficiency level ($F(2,61) = 21.40$, $p < .001$, *partial* $\eta^2 = .41$) and mismatch type were found to be significant ($F(2,122) = 418.26$, $p < .001$, *partial* $\eta^2 = .87$), while the orthographic input was not ($F(1,61) = 0.95$, $p = .34$, *partial* $\eta^2 = .015$). *Post hoc* analysis (Bonferroni adjustment) indicated that Adv group had significantly higher sensitivity than the Inter group, who in turn had higher sensitivity than the Naïve group ($ps < .001$; Cohen's ds between 0.62 and 0.74). Regarding the mismatch type, the participants were most sensitive to segmental mismatches and least sensitive to tonal mismatches ($ps < .001$; Cohen's ds between 0.80 and 1.03). Given the medium to large effect sizes of these comparisons, these main effects were highly significant.

The interactions between proficiency level and orthographic input ($F(2,61) = 4.81$, $p = .011$, *partial* $\eta^2 = .14$), between proficiency level and mismatch type ($F(4,122) = 8.59$, $p < .001$, *partial* $\eta^2 = .22$), and between mismatch type and orthography ($F(2,122) = 3.22$, $p = .044$, *partial* $\eta^2 = .05$) all reached statistical significance, although the latter only had a medium effect size. Given these significant interactions, we conducted *post hoc* analysis (Bonferroni adjustment) to investigate the PY and CH groups' performance at different proficiency levels and to different stimuli types. The results showed that at the Adv level, the CH group had higher overall d-prime scores than the PY group ($p = .014$; Cohen's $d = 1.06$), which was driven by their higher sensitivity to tonal mismatches than the PY group ($p = .005$; Cohen's $d = 1.21$), as the two groups did not differ on segmental mismatches ($p = .32$). The CH group at the Intermediate level generally had higher d-prime scores than the PY group, although the differences did not reach statistical significance (ps between .096 and .525). At the Naïve level, on the other hand, the PY group exhibited higher d-prime scores than the CH group in all conditions, with the difference in the overall d-prime scores reaching significance with a relatively large effect size ($p = .045$; Cohen's $d = 1.01$). Comparing within the PY groups, the Adv learners were significantly more sensitive to segmental mismatches than the Inter and Naïve participants ($ps \leq .036$; Cohen's $ds \geq 1.09$), while no difference was observed in the other two conditions. Within the CH groups, in all three conditions, the Adv learners' d-prime scores were significantly higher than those of the Inter learners, which were significantly higher than those of the Naïve participants ($ps \leq .021$; Cohen's $ds \geq 1.47$).

To summarize, statistical tests revealed that the effect of orthographic input is different at different proficiency levels. At the Adv level, the CH group generally outperformed the PY group, particularly in detecting tonal mismatches to the target words. The CH group at the Inter level also had higher d-prime scores than the PY group in all conditions, although not significantly. At the Naïve level, in contrast, the PY group had significantly higher overall d-prime scores than the CH group. They also exhibited higher sensitivity to both segmental and tonal mismatches than the CH group, although the difference did not reach statistical significance. When the comparison was made across proficiency levels within each orthography group, for both the PY and CH groups, the Adv learners were more sensitive to segmental mismatches than the less experienced groups. However, proficiency did not affect participants' accuracy in rejecting tonal-mismatched items within the PY groups. As

for the CH groups, proficiency was a significant predictor for their sensitivity to tonal mismatches.

Discussion

The main purpose of this study was to assess the effect of two types of orthographic input, Pinyin and Chinese characters, on English speakers' phonological encoding of new Mandarin words. Pinyin utilizes the Roman alphabet that is familiar to English speakers, while characters are novel graphemes that do not exist in English. Our data showed that orthographic input did not influence the time English speakers needed to form association between the phonological form and meaning of new Mandarin words, as the PY and CH groups did not differ in their number of learning phase repetitions at any level. However, orthographic input was found to facilitate English speakers' phonological encoding to a different degree depending on their Mandarin proficiency level and the type of sound contrast. Specifically, at the Adv level, the Character (CH) group was more accurate than the Pinyin (PY) group overall and performed significantly better than the PY group in rejecting tonal mismatches. The Inter learners showed a similar tendency as the Adv learners, although the difference between the CH and PY groups was much smaller and not significant. The Naïve participants patterned differently from the two learner groups: the PY group had significantly higher overall *d*-prime scores than the CH group, and they also obtained higher *d*-prime scores than the CH group in segmental- and tonal-mismatched conditions, although the difference did not reach statistical significance. The finding on the Naïve participants is inconsistent with that in Hayes-Harb and Cheng (2016), which revealed greater facilitation of novel graphemes than familiar graphemes on English speakers' phonological encoding of Mandarin nonwords. We hypothesize that such a discrepancy is probably due to the fact that the novel graphemes in our study, the Chinese characters, do not represent phonemes. Furthermore, Hayes-Harb and Cheng (2016) provided a training session for their participants to learn the GPCs of the novel graphemes (Zhuyin) before the actual word-learning task, whereas our participants did not receive any training in characters prior to the word-learning phase. Hence for our participants who were used to phoneme-denoting orthography and totally unfamiliar with the nature of Chinese characters, they most likely found characters to be of no use or even distracting for their phonological encoding. As for Pinyin, although it differs from English in some GPCs, it is a sound-based orthography and does not require participants to learn new graphemes. This is probably why it seemed more beneficial than characters for English speakers' Mandarin phonological encoding when they had no knowledge of the target language.

On the other hand, the current study also found that for English speakers who had studied Mandarin, especially the Adv learners, novel L2 graphemes had a greater facilitatory effect on the phonological encoding of Mandarin sounds than the familiar graphemes, as evidenced by the CH group's higher sensitivity to phonological mismatches than the PY group at the Adv level. This is similar to the finding in Hayes-Harb and Cheng (2016), which attributed the advantage of novel graphemes over familiar graphemes to the absence of negative L1 interference. Specifically in their study, when the GPCs in Pinyin were incongruent with those

in English, English speakers had the tendency to comply with their L1 GPCs and encoded the L2 phonological forms incorrectly. However, L1 interference does not seem to be the cause for the difference between the CH and PY groups in our study, because the CH group was not more sensitive to segmental mismatches than the PY group. The CH group exhibited significantly higher sensitivity to tones, which are marked by tonal diacritics rather than by the Roman alphabet. We hypothesize that the Adv CH group performed better in their phonological encoding than the PY group probably because the presentation of characters along with the English meaning of new words is closer to their accustomed way of vocabulary learning. As learners' Mandarin proficiency increases, the use of characters in language instruction as well as literacy practices outside the classroom becomes much more prevalent than that of Pinyin. In addition, the individual characters of the target words in our study were carefully chosen, so that they were known to the learners. As a result, the learners in the CH group were not tasked with the learning of new graphemes and thus could focus on sound encoding just like the PY group. Finally, the semantic information that many characters carry may have facilitated the CH group's retention of the target words, which led to more accurate phonological encoding. For instance, the target word 知觉 <zhījué> "consciousness" is composed of the character 知, which means "to know", and 觉, which means "feeling". Hence, seeing the characters may have helped the CH group's memorization of the target words, an advantage that the PY group did not have. Certainly, these explanations are tentative until they are verified in a separate study. Specifically, Adv learners' familiarity with the characters of the target words should be assessed, as well as their accustomed way of vocabulary learning.

The most substantial difference between the CH and PY groups was that the Adv CH group was significantly more sensitive to tonal mismatches than the Adv PY group. This is intriguing because, as described in the Introduction, tones are marked in Pinyin but not in characters, and yet all three proficiency groups who received Pinyin input achieved similarly low accuracy rates (0.42–0.47) on tonal-mismatched items, revealing that even Adv L2 learners who were familiar with the Pinyin tonal diacritics did not benefit from them. This is reminiscent of the finding in Hayes-Harb and Hacking (2015) that the presence of written stress marks on Russian non-words did not improve English speakers' encoding of Russian stress contrasts, whether they were experienced L2 learners or had no knowledge of Russian. Taken together, these findings suggest that English speakers probably pay little attention to L2 prosodic diacritics regardless of their experience with the L2. We propose that this may be because tonal diacritics in Pinyin are relatively unfamiliar graphemes to English speakers compared to the Roman alphabet used to mark segments. Hence, it is likely that the participants in the PY groups attended less to tones because their attention was drawn to the more familiar Roman alphabet. This proposition, however, is not compatible with the results in Showalter and Hayes-Harb (2013), which showed that tonal diacritics in Pinyin helped English speakers encode Mandarin tonal contrasts to some extent. We hypothesize that such a discrepancy may result from differences in the stimuli. Specifically, the target stimuli in Showalter and Hayes-Harb (2013) were 2 monosyllables each combined with the 4 Mandarin tones, while our stimuli were 16 different disyllabic words with tonal variation on both the first and second syllable. Thus, the participants in Showalter

and Hayes-Harb (2013) were able to focus on tonal contrasts, whereas our participants had to encode consonantal, vowel, and tonal information of 32 syllables. This may be why our participants seemed to pay little attention to tones even though they were marked in Pinyin. Our results also suggest that when English speakers have to encode multiple types of phonological information, they tend to prioritize segments over tones.

Regarding the character input, previous research suggests that reading Chinese characters activates phonological information for native speakers just like reading in English, despite the fact that characters are not sound-based. The difference is that pronunciation of English words is obtained by assembling a sequence of phonemes represented by letters, while pronunciation of Chinese characters is obtained by retrieving the whole syllable, including both segments and tone (Perfetti & Liu, 2005; Perfetti et al., 1992). In other words, the character input is a holistic representation of a syllable. This may be why the Adv CH group attended more to tones than the PY group, because tonal information is not separate from segmental information in characters. A similar hypothesis was proposed by Liu (2013) when trying to account for the observation that Mandarin speakers recalled words much faster than Vietnamese speakers when being primed by a prompt written in their L1 orthography. Liu suggested that Vietnamese speakers probably have separate entries for segments and tones in their mental representation, because their orthography, similar to Pinyin, marks segments with Roman alphabet and tones with diacritics. Hence, when seeing a prompt written in their L1 orthography, they had to convert the sequence of letters into phonemes and combine them with tones to derive the pronunciation and meaning of the word. In contrast, due to the non-compositional nature of Chinese characters, Mandarin speakers have developed an integrated entry for each character with its pronunciation and meaning. Therefore, Mandarin speakers could retrieve the pronunciation and meaning of the prompt significantly faster than Vietnamese speakers (Liu, 2013: pp. 9–10). Our findings suggest that the Adv L2 learners of Mandarin behaved similarly to native Mandarin speakers. They benefited from the integral representation of sounds in characters and encoded tones of new words more accurately than the group who saw Pinyin. As for the participants with less Mandarin experience, such as the Naïve and Inter participants, the tone encoding accuracy rates of both the CH and PY groups were relatively low (mean accuracy rates range from 0.38 to 0.47). This indicates that it probably requires extensive L2 experience for learners to process characters as a phonological representation of syllables and attend to both segmental and tonal information in their phonological encoding.

It should be noted that we cannot rule out a potential confound that the Adv CH group encoded Mandarin tones more accurately than the PY group because they were simply more proficient in Mandarin. Even though the assignment of participants to the two orthography groups was random, a measure of their proficiency should have been administered to ensure that the two groups were comparable. A counterevidence to this possibility, however, is that the CH and PY groups did not differ in their sensitivity to segmental mismatches. Nevertheless, a measure of their proficiency would provide a more robust confirmation that the two groups did not differ in their proficiency. Furthermore, the relatively low number of participants at the Adv level also renders any conclusion tentative. A larger sample of advanced

learners would be needed in future research to confirm the findings and hypotheses of this study.

With regard to the effect of orthographic input on learners at different L2 proficiency levels, in both PY and CH groups, the Adv and Inter learners differed significantly in their sensitivity to segmental mismatches, showing that the Adv learners outperformed the Inter learners in their ability to encode segments of newly learned Mandarin words. As for tonal mismatches, the Adv learners in the CH group were significantly more sensitive than the Inter learners, while the Adv learners in the PY group did not differ from the Inter learners and not even from the Naïve participants. All three PY groups had relatively low *d*-prime scores on tonal-mismatched items, suggesting that increasing L2 experience did not improve English speakers' sensitivity to tonal contrasts. Even though we cannot definitively exclude the possibility that this is due to imprecise categorization of participants' L2 proficiency, a more plausible explanation seems to be that English speakers have more difficulty encoding tonal than segmental information in newly learned L2 words. This is a clear tendency in our data that all groups were more accurate in rejecting segmental mismatches to the target words than rejecting tonal mismatches. Even for the Adv CH learners who had the highest sensitivity to tonal mismatches among all groups, their accuracy in the tonal-mismatched trials (0.675) was still much lower than that in the segmental-mismatched trials (0.95). Similar findings have been reported in previous research, which showed that English speakers are more accurate in perceiving Mandarin vowels than tones (Gottfried & Suiter, 1997; Hao, 2018). These researchers attributed this pattern to native English speakers' L1 influence, because vowels but not tones are phonologically contrastive in English. Similarly in the current study, the English-speaking participants were more proficient in encoding segmental than tonal information when learning new words, because they do not usually have to encode the latter in their L1. The effect of orthographic input seems to be, as revealed by our data, that Pinyin hindered rather than facilitated L2 learners' improvement on tone encoding. This is probably due to its separate markings of segments and tones as discussed above.

Finally, this study overall found more similarities than differences in the performance between the PY and CH groups. These two orthography groups at all three proficiency levels did not differ in the time needed to memorize the target words. Aside from the overall sensitivity at the Naïve and Adv levels and the tonal-mismatched condition at the Adv level, the orthographic input was not found to be a significant factor in any other comparison. This suggests that L2 orthography might not play a crucial role in English speakers' encoding of Mandarin sounds. Such a claim has been proposed in former studies, which showed that English speakers who were provided with orthographic representation and audio input of target words did not memorize the phonological forms better than those who were only provided with audio input (Escudero, 2015; Hayes-Harb et al., 2010; Mathieu, 2016; Showalter, 2018; Showalter & Hayes-Harb, 2015; Simon et al., 2010). Some of these researchers proposed that this could be due to the fact that English uses an opaque orthography which has many irregularities in the GPCs. As a result, English speakers have learned not to allocate too much attentional resources to orthographic forms in phonological encoding, since they may not reflect the actual pronunciation. This proposition can account for the general observation that English speakers rely

more heavily on audio than on written input in L1 speech processing (Chikamatsu, 1996; Mori, 1998; Turnage & McGinnies, 1973). It is also supported by L2 production studies such as Erdener and Burnham (2005), which showed that L2 production of Turkish speakers was more heavily influenced by orthographic forms than that of Australian English speakers, the former having a transparent L1 orthography while the latter having an opaque orthography. To determine the role of orthography in L2 word learning, future studies could assess whether participants indeed pay attention to orthographic forms during word learning. For example, probe questions about orthographic forms of the target words can be added to the learning phase (e.g., “Was the word you just saw written as *xuéshuō* or *xuéshōu*?” for the PY group, and “Was the word you just saw written as 学说 or 学悦?” for the CH group). A more sensitive measure would be to implement eye-tracking during the word-learning phase, which would reveal whether and for how long the participants look at orthographic forms. Finally, more future research comparing transparent and opaque L1 orthography groups is also needed to clarify the influence of learners’ L1 orthographic depth on their use of L2 orthography.

Conclusion

This study has demonstrated that the effect of L2 orthography on phonological encoding of L2 words varies depending on the learners’ L2 experience and the types of phonological contrasts. An L2 orthography using graphemes similar to the learners’ L1 (i.e., Pinyin) appears to be more helpful than an orthography utilizing novel graphemes for the early beginners, like our Naïve participants. For the more experienced L2 learners, however, novel graphemes (i.e., Chinese characters) are more beneficial than familiar graphemes, especially in their tonal encoding. This may be because the more advanced L2 learners have become accustomed to seeing characters when learning new Mandarin words, and because characters represent syllables in a non-compositional manner. The current study offers a unique perspective on the effect of orthography on L2 phonological encoding due to the distinctive nature of Chinese characters. The comparison of segmental and tonal encoding by English speakers with differing L2 proficiency also complements existing literature on the relationship between orthography and L2 acquisition. One teaching implication derived from this study is that for learners at the advanced level, using Pinyin does not help them memorize the tones of new words as well as using characters. However, it should be noted that the characters used to compose the target words of this study were carefully chosen so that they were known to the learners. When learners encounter new words with unknown characters or when an auditory input is unavailable, which are possible scenarios in real-life vocabulary acquisition, the relative benefit of Pinyin and characters on L2 phonological encoding needs to be further investigated.

Acknowledgments. This study was funded by the University of Tennessee Professional Development Award to the first author. The authors would like to thank Yufen Chang, Hang Zhang, and Pei-Shan Yu for their help with data collection.

References

- Bassetti, B.** (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language & Literacy*, *9*(1), 95–114.
- Bassetti, B.** (2007). Effects of hanyu pinyin on pronunciation in learners of Chinese as a foreign language. In A. Guder, X. Jiang & Y. Wan (Eds.), *The cognition, learning and teaching of Chinese characters* (pp. 156–179). Beijing Language and Culture University.
- Chen, P.** (1996). Toward a phonographic writing system of Chinese: A case study in writing reform. *International Journal of the Sociology of Language*, *122*(1), 1–46.
- Chen, P.** (1999). *Modern Chinese: history and sociolinguistics*. Cambridge University Press.
- Chikamatsu, N.** (1996). The effects of L1 orthography on L2 word recognition: A study of American and Chinese learners of Japanese. *Studies in Second Language Acquisition*, *18*(4), 403–432.
- Coulmas, F.** (1983). Writing and literacy in China. In F. Coulmas and K. Ehlich (Eds.), *Writing in focus* (pp. 239–253). Mouton.
- DeFrancis, J.** (1984). *The Chinese language*. University of Hawaii Press.
- Erdener, V. D., & Burnham, D. K.** (2005). The role of audiovisual speech and orthographic information in nonnative speech production. *Language Learning*, *55*(2), 191–228.
- Escudero, P.** (2015). Orthography plays a limited role when learning the phonological forms of new words: The case of Spanish and English learners of novel Dutch words. *Applied Psycholinguistics*, *36*(1), 7–22.
- Escudero, P., Hayes-Harb, R., & Mitterer, H.** (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics*, *36*(2), 345–360.
- Gottfried, T. L., & Suiter, T. L.** (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, *25*(2), 207–231.
- Hao, Y. C.** (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, *40*(2), 269–279.
- Hao, Y.-C.** (2018). Second language perception of Mandarin vowels and tones. *Language and Speech*, *61*(1), 135–152.
- Hayes-Harb, R., & Cheng, H. W.** (2016). The influence of the Pinyin and Zhuyin writing systems on the acquisition of Mandarin word forms by native English speakers. *Frontiers in Psychology*, *7*, 785. doi: [10.3389/fpsyg.2016.00785](https://doi.org/10.3389/fpsyg.2016.00785).
- Hayes-Harb, R., & Hacking, J.** (2015). The influence of written stress marks on native English speakers' acquisition of Russian lexical stress contrasts. *The Slavic and East European Journal*, *59*(1), 91–109.
- Hayes-Harb, R., Nicol, J., & Barker, J.** (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech*, *53*(3), 367–381.
- Haynes, M., & Carr, T. H.** (1990). Writing system background and second language reading: a component skills analysis of English reading by native speaker-readers of Chinese. In T. H. Carr (Ed.), *Reading and its development: Component skills approaches* (pp. 375–421). Academic Press.
- Kirilloff, C.** (1969). On the auditory perception of tones in Mandarin. *Phonetica*, *20*(2–4), 63–67.
- Koda, K.** (1990). The use of L1 reading strategies in L2 reading: Effects of L1 orthographic structures on L2 phonological recoding strategies. *Studies in Second Language Acquisition*, *12*(4), 393–410.
- Leck, K. J., Weekes, B. S., & Chen, M. J.** (1995). Visual and phonological pathways to the lexicon: Evidence from Chinese readers. *Memory & Cognition*, *23*(4), 468–476.
- Liu, N.** (2013). Implicit priming effects in Chinese word recall: The role of orthography and tones in the mental lexicon. *Computational Linguistics & Chinese Language Processing*, *18*(3), 1–20.
- Liu, Y., Yao, T.-c., Bi, N.-P., Ge, L., & Shi, Y.** (2009). *Integrated Chinese*. Cheng & Tsui Company.
- MacMillan, N. A., & Creelman, C. D.** (2004). *Detection theory: A user's guide*. Lawrence Erlbaum Associates.
- Mathieu, L.** (2016). The influence of foreign scripts on the acquisition of a second language phonological contrast. *Second Language Research*, *32*(2), 145–170.
- Mori, Y.** (1998). Effects of first language and phonological accessibility on kanji recognition. *The Modern Language Journal*, *82*(1), 69–82.
- Perfetti, C. A., & Liu, Y.** (2005). Orthography to phonology and meaning: Comparisons across and within writing systems. *Reading and Writing*, *18*(3), 193–210.

- Perfetti, C. A., Zhang, S., & Berent, I.** (1992). Reading in English and Chinese: Evidence for a “universal” phonological principle. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology and meaning* (pp. 227–248). North-Holland.
- Pytlyk, C.** (2011). Shared orthography: Do shared written symbols influence the perception of L2 sounds? *The Modern Language Journal*, *95*(4), 541–557.
- Shen, X. S.** (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, *24*(3), 27–47.
- Showalter, C. E.** (2018). Impact of Cyrillic on native English speakers’ phono-lexical acquisition of Russian. *Language and Speech*, *61*(4), 565–576.
- Showalter, C. E.** (2020). Russian phono-lexical acquisition and orthographic input: Naïve learners, experienced learners, and interventions. *Studies in Second Language Acquisition*, *42*(2), 255–277.
- Showalter, C. E., & Hayes-Harb, R.** (2013). Unfamiliar orthographic information and second language word learning: A novel lexicon study. *Second Language Research*, *29*(2), 185–200.
- Showalter, C. E., & Hayes-Harb, R.** (2015). Native English speakers learning Arabic: The influence of novel orthographic information on second language phonological acquisition. *Applied Psycholinguistics*, *36*(1), 23–42.
- Simon, E., Chambless, D., & Alves, U.K.** (2010). Understanding the role of orthography in the acquisition of a non-native vowel contrast. *Language Sciences*, *32*(3), 380–394.
- So, C. K., & Best, C. T.** (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273–293.
- So, C. K., & Best, C. T.** (2014). Phonetic influences on English and French listeners’ assimilation of mandarin tones to native prosodic categories. *Studies in Second Language Acquisition*, *36*(2), 195–221.
- Turnage, T. W., & McGinnies, E.** (1973). A cross-cultural comparison of the effects of presentation mode and meaningfulness on short-term recall. *The American Journal of Psychology*, *86*(2), 369–381.
- Van den Bosch, A., Content, A., Daelemans, W., & De Gelder, B.** (1994). Measuring the complexity of writing systems. *Journal of Quantitative Linguistics*, *1*(3), 178–188.
- Wang, M., & Koda, K.** (2005). Commonalities and differences in word identification skills among learners of English as a second language. *Language Learning*, *55*(1), 71–98.
- Wang, M., Koda, K., & Perfetti, C. A.** (2003). Alphabetic and nonalphabetic L1 effects in English word identification: A comparison of Korean and Chinese English L2 learners. *Cognition*, *87*(2), 129–149.
- Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A.** (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America*, *106*(6), 3649–3658.
- Young-Scholten, M.** (2002). Orthographic input in L2 phonological development. In P. Burmeister, T. Piske, & A. Rohde (Eds.), *An integrated view of language development: Papers in honour of Henning Wode* (pp. 263–279). Wissenschaftlicher Verlag Trier.
- Young-Scholten, M., & Langer, M.** (2015). The role of orthographic input in second language German: Evidence from naturalistic adult learners’ production. *Applied Psycholinguistics*, *36*(1), 93–114.

Appendix The 16 target words used in the word-learning experiment and their segmental and tonal mismatches.

Character	Pinyin	English meaning	IPA	Segmental mismatch	Tonal mismatch
知觉	zhījué	consciousness	tʃi1 tɕyɛ2	tei1 tɕyɛ2	tʃi4 tɕyɛ2
生理	shēngli	physiology	ʃəŋ1 li3	səŋ1 li3	ʃəŋ4 li3
人工	réngōng	man-made	ɹən2 koŋ1	lən2 koŋ1	ɹən3 koŋ1
打听	dǎtīng	inquire about	tɑ3 tʰiŋ1	tɑ3 tiŋ1	tɑ2 tʰiŋ1
不时	bùshí	frequently	pu4 ʃi2	pu4 ei2	pu4 ʃi3
对比	duìbǐ	contrast	tue4 pi3	tu4 pi3	tue4 pi2
人道	réndào	humanitarianism	ɹən2 tao4	lən2 tao4	ɹən2 tao1
本事	běnsì	capability	pən3 ʃi4	pən3 si4	pən3 ʃi1
分红	fēnhóng	bonus	fən1 xoŋ2	fan1 xoŋ2	fən1 xoŋ3
空想	kōngxiǎng	day dream	kʰoŋ1 ciəŋ3	kʰoŋ1 ʃəŋ3	kʰoŋ1 ciəŋ2
学说	xuéshuō	theory	ɕyɛ2 ʃuo1	ɕyɛ2 ʃuo1	ɕyɛ2 ʃuo4
手工	shǒugōng	hand-made	ʃou3 koŋ1	ʃuo3 koŋ1	ʃou3 koŋ4
告别	gàobié	say goodbye to	kaʊ4 piɛ2	kaʊ4 pei2	kaʊ1 piɛ2
要点	yàodiǎn	main point	jaʊ4 tian3	jaʊ4 tʰian3	jaʊ1 tian3
学问	xuéwèn	knowledge	ɕyɛ2 wən4	ʃue2 wən4	ɕyɛ3 wən4
请教	qǐngjiào	seek advice	teiŋ3 tejaʊ4	teiŋ3 jaʊ4	teiŋ2 teiaʊ4

Cite this article: Hao, Y.-C. and Yang, C.-L.M. (2021). The effect of second-language orthographic input on the phonological encoding of Mandarin words. *Applied Psycholinguistics* 42, 887–906. <https://doi.org/10.1017/S0142716421000114>