



# Phonological Training and Word Learning in a Novel Language

Yixun Li<sup>1</sup>, Min Wang<sup>2\*</sup>, Chuchu Li<sup>3</sup> and Man Li<sup>4</sup>

<sup>1</sup>Department of Early Childhood Education, The Education University of Hong Kong, Hong Kong SAR, China, <sup>2</sup>Department of Human Development and Quantitative Methodology, University of Maryland, College Park, MD, United States, <sup>3</sup>Department of Psychiatry, University of California, San Diego, CA, United States, <sup>4</sup>Second Language Acquisition Program, University of Maryland, College Park, MD, United States

In reading Chinese words, learners may process segment and tone either separately or as an integral unit, as evidenced in previous research. The present study compared two ways—Segmental versus Whole-Syllable-Based Training—for improving learners' phonological and word learning in Chinese as a novel language, while controlling for learners' musical ability, an important factor that may contribute to phonological learning. Forty-two American college students learned Chinese words represented by Pinyin, a Romanized script which denotes the pronunciation of Chinese characters. Before the training, all participants were introduced to the phonology and Pinyin system. Then, they were trained on the pronunciation and meaning of the Pinyin words with or without an emphasis on separating the tonal from segmental information. All participants' musical ability was assessed using a musical ability test. Learning outcomes were measured through tasks of same-different phonological judgment, tone identification, and word comprehension. Results showed the equal success of the two training methods, probably due to the consistent involvement of Pinyin and learner's reliance on segment and tone as an integral unit rather than separate cues in phonological and word learning. Furthermore, musical ability seems to play a role in phonological and word learning among novel learners of Chinese.

**Keywords:** musical ability, segmental versus whole-syllable-based training, word learning, phonological training, Chinese L2 learners

## OPEN ACCESS

### Edited by:

David Saldaña,  
Sevilla University, Spain

### Reviewed by:

Kelvin Fai Hong Lui,  
Lingnan University, China  
Heather Winskel,  
Southern Cross University, Australia

### \*Correspondence:

Min Wang  
minwang@umd.edu

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 14 November 2020

**Accepted:** 11 January 2022

**Published:** 17 February 2022

### Citation:

Li Y, Wang M, Li C and Li M (2022)  
Phonological Training and Word  
Learning in a Novel Language.  
Front. Commun. 7:629349.  
doi: 10.3389/fcomm.2022.629349

## INTRODUCTION

A great deal of research has provided overwhelming evidence that phonological information is activated and involved in reading (e.g., Tan and Perfetti, 1997; Luo et al., 1998; Xu et al., 1999; Spinks et al., 2000; Braun et al., 2009; Li et al., 2013) and phonological awareness—the ability to reflect on and manipulate sounds in speech—is an important predictor of learning to read in Chinese (e.g., Ho and Bryant, 1997; Shu et al., 2008; Ruan et al., 2018; for reviews see Ziegler and Goswami, 2005). Experience with Pinyin, a transparent alphabetic transcription of the pronunciations of written characters, can improve native Chinese-speaking children's phonological awareness and character reading (Shu et al., 2008; Lin et al., 2010; Li et al., 2020). Pinyin knowledge also supports phonological processing among skilled native readers (e.g., Read et al., 1986; Leong et al., 2005; Zhu et al., 2009). Furthermore, research also suggests that Pinyin can facilitate phonological learning among learners of Chinese as a second language (L2) (e.g., Liu et al., 2011; Showalter and Hayes-Harb, 2013) and their subsequent word reading (e.g., Lü, 2017). However, previous research has paid little attention to the specific means by

which the use of Pinyin can better support non-native learners' phonological and word learning of Chinese.

The current study used Pinyin to train Chinese L2 learners on phonological perception and processing and spoken word recognition. Also, learners' musical ability was taken into consideration, since it is an important factor that may contribute to L2 learners' phonological learning based on previous research (e.g., Milovanov et al., 2008; Milovanov et al., 2010; Marie et al., 2011). The current study aims to build upon previous work by examining 1) the direct effects of two specific training methods not yet investigated in the literature; 2) the potential role of musical ability in L2 phonological and word learning.

Chinese phonology includes both segmental and tonal information. Traditionally, segments refer to discrete units that can be identified in the stream of speech (e.g., consonants and vowels) (Crystal, 2008). However, segments are often not completely discrete due to the overlap in cues for adjacent segments. Tonal information refers to suprasegments, which are identified as the vocal effect that extends over individual phonetic segments to syllables and larger units of speech. Stress, intonation and rhythm are examples of suprasegments (Gross, 2017). A lexical tone must be attached to a segmental syllable and cannot exist independently. In spoken Mandarin Chinese, a syllable consists of an onset (consonant) or without an onset, in addition to a rime (vowel + consonant or vowel only). In Mainland China, Pinyin has been used to aid children's learning of the pronunciations of the written characters. Pinyin has a transparent orthography that explicitly marks the segmental information and the tone over the vowel. For example, the Pinyin representation of the word 妈, *mother* is *mā*, where the pitch contour of high-level tone is clearly marked above the vowel (i.e., the tone marker).

Given the four different tones in Mandarin Chinese, the same segmental syllable can represent four different meanings, depending on the tone it carries (Li and Thompson, 1989). For example, for the same segment *ma*, *mā* with the high-level tone means "mother," *má* with the high-rising tone means "hemp," *mǎ* with the low-dipping tone means "horse," and *mà* with the high-falling tone means "scold". Therefore, in learning Chinese, it is important to associate the segmental syllable with the correct lexical tone in order to have access to the correct lexical item. The use of Pinyin symbols could help with this association between segmental syllable and lexical tone.

Although both segmental and tonal information play an important role in lexical access, readers can process and activate these two components separately (Spinks et al., 2000). Tone seems to be represented more poorly, and plays a secondary role in lexical processing as compared to segmental information, as demonstrated in previous research on visual Chinese character recognition among skilled native adult readers, using both online and off-line experiments (e.g., Taft and Chen, 1992; Chen et al., 2002; Malins and Joanisse, 2010; Li et al., 2013; Sereno and Lee, 2015; Wang et al., 2015). For example, Sereno and Lee (2015) used an online auditory lexical decision task among native Mandarin-speaking adults and their results supported for the separate roles of tone and segmental information in Chinese word

processing. Significant priming effects were shown when both tonal and segmental information or only segmental information overlapped, but no priming effects when only tones matched. Consistently, Chen et al. (2002) found priming effects for syllable-alone prime (i.e., tone differed) but no priming effects for tone-alone prime (i.e., syllables differed) using an implicit priming paradigm among native adult Mandarin speakers. However, there is some evidence that segmental and tonal information may also be processed as an integral unit among native speakers in a task such as primed word naming (Wang et al., 2015) or auditory word eye-tracking (Malins and Joanisse, 2010).

Compared to native readers, tone is even more poorly represented than segmental syllable among Chinese L2 learners. In spoken language, it is more difficult for non-native Chinese speakers who have an atonal native language to process tonal information in auditory Chinese word recognition than native Chinese speakers (Gottfried and Suiter, 1997). A similar L2 disadvantage has been demonstrated in visual Chinese word recognition (Li et al., 2018). In the work by Li et al. (2018), native Chinese readers and Chinese L2 learners were instructed to complete a homophone judgment task. Both groups performed worse when two visual Chinese characters shared the same segmental syllable but different tonal information (i.e., S + T-) than other three conditions (e.g., homophones S + T+, different segments with the same tone S-T+, and different segments with the different tone, S-T-), as indicated in both analyses of accuracy rates and reaction time. There was clearly a disadvantage in tonal processing; however, this disadvantage was much greater among L2 learners compared to the native peers. Note that this disadvantage could be largely reduced by presenting one of the two characters in Pinyin for Chinese L2 learners.

Previous research suggests that 2 weeks of tone training helped Chinese L2 learners better perceive tonal information as indicated by an average 21% improvement in identification accuracy, and the training effect lasted for at least 6 months (Wang et al., 1999). The effect of training on tone perception has also been supported at the neural level: it may lead to a cortical modification in the area of the left-hemisphere among Chinese L2 learners (Van Lancker and Fromkin, 1973; Wang et al., 2001; Wang Y. et al., 2003). Note that the aforementioned studies trained participants auditorily using real Chinese spoken words with different tones, and did not use Pinyin in their tonal training. Later training studies have shown that the presence of Pinyin facilitates Chinese L2 learners' tone perception and processing, especially when the explicit tone markers are provided (e.g., Liu et al., 2011; Showalter and Hayes-Harb, 2013). These studies also compared Chinese L2 learners' tone learning outcomes under the training conditions with and without visually presenting tone markers, and found the facilitation of visually presenting tone markers on tone learning. For instance, Showalter and Hayes-Harb (2013) trained their participants to map eight Chinese nonwords onto Pinyin either with or without tone marks and pictures illustrating meanings; they found that the presence of tone marks helped participants associate tone with the new words ( $\eta_p^2 = 0.18$ ). Thus, the effectiveness of providing tone markers together with Pinyin segment symbols (i.e., Whole-Syllable-Based Training) has some support from prior research. Building on this previous

work, Pinyin symbols were presented with tone markers consistently across trials in our study to support participants' learning.

Furthermore, prior work suggests that directing L2 learners' attention to tone markers support their improvement on tone perception (e.g., Liu et al., 2011; Godfroid et al., 2017). In addition to the previous findings that skilled native readers activate and process the two elements separately, we hypothesize that L2 learners may benefit from the separation of these two elements across a staggered training period. In the present study, we extended previous work to explicitly emphasize the separation between these two components—segments and tones—in the Segmental Training method. We then compared the two different training methods on phonological and word learning. Note that we also extended previous research beyond phonological perception to word learning and acquisition, and we were interested in both accuracy and processing speed. It was anticipated that the Segmental Training method would yield better outcomes, given that it was designed to scaffold learners' segment and tone engagement.

In addition to the role of Pinyin in Chinese L2 learning, previous research has shown that high musical ability has a positive effect on L2 learning in general. There is a close relationship between musical ability and L2 perception and production skills (e.g., Milovanov et al., 2008; Milovanov et al., 2010; Marie et al., 2011). For example, French adult musicians detected segmental variations in unknown Mandarin words more accurately than non-musicians (Marie et al., 2011). Similarly, Finnish native speakers with higher musical ability, including both children and college students, read aloud English words containing difficult phonemes more accurately compared to those with low musical ability (Milovanov et al., 2008; Milovanov et al., 2010). Musicians, therefore, appear to be more sensitive to variations in phonetic attributions in the process of building precise phonological representations associated with a novel language (Dittinger et al., 2016).

Research has also shown that high musical ability is associated with good perception and production of L2 lexical tones (Wong and Perrachione, 2007; Cooper and Wang, 2012; Li and DeKeyser, 2017). For instance, among adults who speak an atonal language, the successful learning of pitch patterns to identify artificial tone words (English pseudowords with pitch patterns resembling Mandarin tones) was linked to the learners' ability to discriminate musical pitch in a non-linguistic context (Wong and Perrachione, 2007). Li and DeKeyser (2017) showed that native English speakers with better musical ability achieved higher accuracy in tone word perception and production in Chinese. In the present study, we took musical ability into consideration when testing the effectiveness of the two different language-training methods and included participants' musical abilities as a covariate in the final analyses. Our goal was to control for participants' musical abilities to enable us to examine the unique contribution of the Segmental Training method, in the meantime, we can address the effect of musical ability on phonological and word learning.

In our experiment, both Segmental Training and Whole-Syllable-Based Training groups were introduced to the Chinese

phonological and Pinyin systems prior to the first training session. During the training session, participants learned through 16 Pinyin words with four tones. Immediately after the training sessions, both groups had exercise sessions to help strengthen what they had learned during the training sessions. The learning outcome was measured via a phonological judgment task, a tone identification task, and a word comprehension task.

It was hypothesized that the Segmental Training Group might yield a better learning outcome compared to the Whole-Syllable-Based Training Group, specifically in their use of segment and tone separately as cues for word pronunciation (e.g., Spinks et al., 2000). Consequently, learners' perception and processing of phonological information, especially at the tonal level, would be enhanced. Alternatively, it is possible that the two groups' segmental and tonal processing would be improved equally after the training using Pinyin, regardless of whether they utilized a segmental or whole-syllable-based approach. Learners may be able to rely on segment and tone as an integrated unit (e.g., Wang et al., 2015) and thus no further benefit from the explicit emphasis on the separation. Moreover, the exercises immediately after the training session were also thought to help enhance the mental representation of phonological units in a similar fashion. Specifically, as for the phonological judgment task, since previous work suggests that L2 learners of Chinese have particular difficulties in processing Chinese characters that share the same segmental syllable but different tonal information (S + T-) (Li et al., 2018) than the other three conditions (i.e., S + T+, S-T+, S-T-), we expected that our participants in both groups who were novel to Chinese might perform relatively poorly on S + T-condition even after the training. Based on evidence from previous research on the role of musical ability in L2 learning, it was hypothesized that musical ability would be a significant covariate, and there should be a general facilitative effect of musical ability in participants' phonological training and word learning.

## METHODS

### Participants

Prior to the study, ethical approval was granted by the Institutional Review Board at the University of Maryland, College Park. Participant consent was obtained from each participant before the experiment. Forty-two native English speakers without speech and hearing impairments were recruited, and all were paid for their participation. Participants had to meet each of the following requirements: 1) never been exposed to the Chinese language; 2) not bilingual; 3) no prior knowledge of a tonal language; 4) no more than 2 years of individual instruction of musical instruments and 5) be between 18 and 45 years old.

A background questionnaire was developed to collect information about participants' language and music experiences. The 42 participants (31 females) were students at a mid-Atlantic University (41 undergraduates; 1 graduate), aged 18–40 ( $M = 20.70$ ,  $SD = 3.49$ ). None spoke another language other than English: 21 participants reported no exposure to a second

language, and 21 reported some experience of learning a foreign language (e.g., Spanish, French) in a classroom setting, but none reported fluency in it. None had any prior knowledge of Mandarin or any other tonal languages (e.g., Cantonese, Thai, or Vietnamese, etc.). With regard to the musical experience, 31 participants had not had any formal musical training in any instrument, and 11 reported having limited musical training experiences, having taken private lessons in an instrument for less than 2 years. All participants thus could be considered as non-musicians, following Wong and Perrachione (2007) definition.

The participants were randomly assigned to either the Segmental Training or Whole-Syllable-Based Training group. The two groups were comparable in terms of age (Whole-Syllable-Based Training Group,  $M = 21.62$ ,  $SD = 4.39$ ; Segmental Training Group,  $M = 19.81$ ,  $SD = 1.99$ ,  $t(40) = 1.72$ ,  $p > 0.05$ ) and gender (18 females in the Whole-Syllable-based Training Group; 13 females in the Segmental Training Group). As for previous experience in studying foreign languages, nine in the Whole-Syllable-Based Training Group and 12 in the Segmental Training Group reported no exposure to a second language. Among those who had experiences learning a foreign language in a classroom setting (12 in the Whole-Syllable-Based Training Group and nine in the Segmental Training Group), the average age of starting to learn a foreign language was comparable between the two groups (11.75 years in the Whole-Syllable-Based Training Group and 10.78 years in the Segmental Training Group).

## Design

All participants were introduced to the standard Mandarin phonology and Pinyin system first, and they learned 16 vocabulary items used for training sessions via Pinyin syllable and picture association on Day 1. Both groups went through a training and then an exercise session on Day 2, Day 3, and Day 4. In the training and exercise sessions, participants learned to map the learned vocabulary to the correct pronunciation. For the Segmental Training Group, the critical manipulation was that they were trained to separate the segmental and tonal information in the vocabulary word via visual cues, in order to enhance their fine-grained metalinguistic awareness of phonological units. On the contrary, the Whole-Syllable-Based Training Group received training and exercise on a whole-word based phonological learning. Musical abilities were assessed for all participants.

## Training Materials and Measures

### Target Words

The target words were 16 monosyllabic Chinese words, which are comprised of four monosyllables (ma, tu, bao, and han) with four Mandarin tones (high-level, high-rising, low-dipping, and high-falling). These segmental syllables are representative of the existing Mandarin syllable structures (CV, CVV, and CVC), forming real Mandarin words with all four tones. In addition, the 16 words are easy to illustrate pictorially. See **Supplementary Appendix C** for a full list of the syllable words with the written Pinyin spelling, English translation, and a picture for each word.

## Pre-Training Test

We designed a Same-Different Phonological Judgment task to measure participants' ability to make a phonological judgment in Mandarin before training. In this task, participants were asked to judge whether the pair of spoken stimuli they heard via headphones sounded the same as accurately and fast as possible. There were four conditions with 15 pairs of items in each. The pairs either shared: 1) the same segmental syllable and tone (S + T+, e.g., *hàn* and *hàn*), 2) the same segment but different tone (S + T-, e.g., *hān* and *hán*), 3) the same tone but different segments (S-T+, e.g., *mā* and *tū*), and 4) different segments and tone (S-T-, e.g., *báo* and *hǎn*). In addition, 30 filler items sharing the same segment and tone were added to balance the "same" (S + T+) and "different" (all other conditions) answers. Those fillers included five monosyllables (ba, hu, mao, tan, and hao) with four tones, and the following ten additional syllables: *tā*, *tǎ*, *tà*, *bān*, *bàn*, *mú*, *mǔ*, *mù*, and *mǎn*. Note that different tokens were used in each condition. All pairs of items were randomly presented to the participants without blocking in conditions. The participants were asked to press the right SHIFT key for the "same" response, and left SHIFT key for the "different" response. Two practice trials were given before the formal testing.

## Explicit Phonology and Pinyin Instruction

An explicit instruction sheet was developed to provide participants with an introduction to standard Mandarin phonology and Pinyin system. The participants were introduced to the Mandarin syllable structure, including the segment (sequence of consonants and vowels) and tone. The participants were instructed that the tones differentiate word meaning in Chinese. Alphabetic Pinyin symbols were also introduced to the participants. All the four tone patterns were presented visually to introduce the pitch contour (see **Supplementary Appendix D**), and four concrete examples were provided to demonstrate how changing the tone would influence the meaning (e.g., *mì* can have four different meanings depending on the tone, *mǐ* (nap), *mí* (lost), *mǐ* (rice), *mì* (secret)). In addition, the instruction explained that the consonants and vowels in the syllable word are referred to as segmental syllable. Segmental syllable and tone together form a complete pronunciation of a Chinese syllable. The instruction was printed on paper and also audio-recorded by a female native Mandarin speaker.

## Musical Ability Test

We used the Pitch Change Test, a subtest from the Wing Measures of Musical Talents (WMMT) (Wing, 1968), to estimate participants' musical ability broadly. This task measures the individual differences in pitch perception in a nonlinguistic context and to examine its potential relationship with the ability to learn Chinese words. For each item in the test, participants heard two musical chords and were asked to judge whether they sounded the same; if they sounded differently, participants were asked to judge whether the altered note moved up or down. There were 30 items in total. Three practice trials were provided before formal testing. Cronbach's

**TABLE 1** | The training schedule for both groups.

	Segmental training group	Whole-syllable-based training group
Day 1		Informed consent (3 min) Background questionnaire (3 min) Pre-training Same-Different Phonological Judgment Task (6–8 min) Explicit Pinyin instruction (5 min) Vocabulary Learning (10 min) Vocabulary Testing (5–35 min)
Day 2	Vocabulary Review Test (10–20 min) Training 1: Separating tonal from segmental information Exercises 1: Segment and tone judgment task (30 min)	Training 1: Whole pronunciation Exercises 1: Whole pronunciation judgment task (30 min)
Day 3	Vocabulary Review Test (10–20 min) Training 2: Separating tonal from segmental information Exercises 2: Segment and tone judgment task (25 min) Musical ability tests (8 min)	Training 2: whole pronunciation Exercises 2: Whole pronunciation judgment task (25 min)
Day 4	Training 3: Separating tonal from segmental information Exercise 3: Segment and tone judgment task (20 min)	Training 3: Whole pronunciation Exercise 3: Whole pronunciation judgment task (20 min)
		Post-Tests (12 + 3+3–20 min) Post-training Same-Different Phonological Judgment Task Tone Identification task Word Comprehension task

alpha was calculated for the internal consistency of this test, and it reached 0.81.

### Post-Training Tests

A set of three tests was administered to all participants immediately after the last training session to assess the learning outcomes.

#### *Same-Different Phonological Judgment Task*

This task was the same as the pre-training same-different phonological judgment task except that 48 new items (using segmental syllables de, di, fang, kai, ke, ni with four tones) were added. The new items were used to test the generalization of the training effects. For those newly added items, there were eight items (pairs) in each of the four experimental conditions and 16 fillers to balance yes and no answers.

#### *Tone Identification Task*

This task was designed to test the participants' ability to identify tones. In this test, participants heard a word for each item and were asked to identify its tone by selecting the Pinyin with the correct tone mark from a choice of four options that constituted the minimal tonal quadruplet. In addition to the 16 learned words, 16 new Chinese words were tested, including four new syllables (wei, pi, hong, fu), each with four tones. Four highlighted buttons on the keyboard (F, J, C, and M) served as the response buttons and corresponded to the positions of the Pinyin spellings on the screen. The four Pinyin spellings only differed in tone. The participants were instructed to press the button that indicates the position of the Pinyin of their choice as quickly and accurately as they can, although they were given 30 s to respond for each item. The learned and new words were presented in two separate blocks with a break given in between. To familiarize them with the task format,

participants were given eight practice items using English words at the beginning of the test.

#### *Word Comprehension Task*

This post-test was designed to test participants' ability to comprehend the meaning of the 16 words they learned in the present study. In this test, participants heard one word at a time and were shown four pictures for words that constitute the minimal tonal quadruplet. They were asked to identify the meaning of the word they heard by choosing the correct picture via pressing the button that indicates the position of the picture of their choice. Though given 30 s to respond for each item, they were asked to respond as quickly and accurately as possible. Eight practice items using English examples were given before formal testing.

### Procedure

Each subject participated in four sessions—each about an hour in length—on four separate, consecutive days within 1 week. **Table 1** presents the training schedule.

#### Day 1

All participants were first informed about the study and asked to fill out the participant background questionnaire (e.g., demographic information, language, and music experiences). Immediately after, the pre-training Same-Different Phonological Judgment Task was conducted. Participants then started the explicit phonology and the Pinyin instruction phase. The participants listened to the audio instruction on a computer, using headphones, while reading the printed instruction sheet.

After the explicit instruction, participants moved on to the Vocabulary Learning phase, in which they learned the 16 target words via PowerPoint slides on the computer screen. Before they started, each participant was told that their task was to learn the

meaning of 16 Chinese words and focus on remembering the mapping between the Pinyin symbols and the pictures. They were also informed that they would be tested on the mapping from Pinyin symbols to pictures and vice versa. Participants were then presented with the target word in Pinyin (e.g., *mā*) along with a picture to illustrate its meaning and its English translation, one at a time, on the slide. Each word remained on the screen for 5 seconds. The computer screen then moved on to the next word on the same slide automatically. The four vocabulary words occupied the four corners of the screen. The previous words stayed on the screen when presenting the next word on the same slide. This was to emphasize the similarities and differences between the four target words at the same time and to draw participants' attention to one syllable word at a time. Each slide showed the four words with the same segments, but different tones in the order of high-level, high-rising, low-dipping, and high-falling tones. Each word was repeated six times, which resulted in six rounds of these four slides in total. The first three rounds were repetitions of the four slides (participants thus saw the same stimuli three times). In the fourth round, each slide presented the four words with the same segmental syllables but in randomized tone order. The final two rounds involved randomized words across segments and tones (all 16 words were thus randomized, regardless of their segments or tones).

Immediately after the vocabulary study phase, participants entered the Vocabulary Testing phase, during which they were tested in both directions from Pinyin to pictures and from pictures to Pinyin until they reached the criterion performance (i.e., less than two errors for the 16 words in the last round of repetition) in both directions. More specifically, they were tested first from Pinyin to pictures. After they reached the criterion or up to five rounds in this direction, they were tested in the other direction from pictures to Pinyin, until they reached criterion or up to five rounds. When testing from Pinyin to pictures, the experimenter presented each word in Pinyin on the PowerPoint slide, one at a time, along with all 16 pictures, and the participants were asked to identify its meaning by selecting the correct picture for the word. When testing from pictures to Pinyin, the experimenter presented one picture at a time on the PowerPoint slide, along with all 16 words in Pinyin, and the participants were asked to identify the correct Pinyin for the picture. In both tasks, the experimenter provided feedback each time after the participant made a choice by indicating the correct answer with a red circle.

## Day 2

All participants first took a Vocabulary Review Test of the 16 words they learned on Day 1. They were tested until they made less than two errors or up to five rounds in both directions (picture to Pinyin; Pinyin to picture). They were then randomly assigned to either the Segmental or Whole-Syllable-Based training group. In both conditions, each training session consisted of two phases, a training phase containing five rounds and then an exercise phase with four rounds.

### *Segmental Training Group*

The participants were trained to strengthen their awareness of tonal and segmental information. In the Training phase, for each

presentation of a word, they heard the pronunciations of the word together with the visual presentation of its picture and Pinyin on the screen and were asked to learn to associate the pronunciations with the Pinyin and the picture. The 16 words were presented in five rounds. The first two rounds involved the four words with the same segments but different tones in the order of high-level, high-rising, low-dipping, and high-falling tones on each of the four slides. The third and fourth rounds involved the four words with the same tone (in the order of high-level, high-rising, low-dipping, and high-falling tones) but different segments on each slide. The fifth round involved syllables with different segments and tones on each slide. The segments of Pinyin syllables were presented in blue, and the tones in red.

In addition, an animation technique in PowerPoint was used: when each word was presented with its pronunciation, the tone mark independently moved apart from the segments. These experimental treatments were designed to highlight the difference between segments and tones visually, to encourage learners to treat them as different parts of the syllable pronunciation in order to enhance their metalinguistic awareness of the different phonological units. Furthermore, participants were asked to pay close attention to the tone of the words (shown in red) in the first two rounds, and then to the segmental syllable of the words (shown in blue) in the third and fourth rounds. In the fifth round, participants were asked to pay attention to the pronunciation of the whole word.

Immediately after the training phase that aimed to strengthen their segmental and tonal awareness, the participants moved on to the Exercise phase, in which they were required to complete a segment and tone judgment task. This exercise phase was designed to provide the learners with an opportunity to practice what they learned in the training phase—namely, to tease apart the segmental and tonal information in a spoken word. There were four rounds; each round consisted of 16 items, 64 items in total. Specifically, for each item, participants first heard a target word twice and saw the corresponding picture illustrating the meaning of the word on the screen. They then heard each of three choice words once along with their written Pinyin syllables. The participants were asked to select the item from three options that shared the same segment (in the first two rounds) or the same tone (in the third and fourth rounds) with the target word. Feedback was given each time after the participants made a choice, by indicating the correct answer with a red circle on the screen.

### *Whole-Syllable-Based Training Group*

The participants were trained to associate pronunciations with Pinyin symbols and pictures as a whole without color-coding or animation of the segmental and tonal components of the Pinyin for each word. For the Training phase, the same five rounds of 16 words were used as in the Segmental Training Group, with the same length of training time. However, participants were asked to pay attention to the pronunciation as a whole across all five rounds rather than highlighting the separation of the segment and tone mark in Pinyin (i.e., there was no animation available, and segments and tones were presented in the same black color).

Following the training phase, the participants moved on to the Exercise phase, in which they were tested in a similar manner to

the judgment task used in the Segmental Training Group. The only difference was that the participants were asked to select the item from the three options that sounded most similar to the word heard instead of explicitly judging whether they shared the same segment or tone. For example, for the item *hàn*, the three options were *bào*, *hàn*, and *tù*. In this case, *hàn* was the most similar choice since it shared the same segment of *hàn*, whereas the other two options shared neither the segment nor the tone. For another item *tǔ*, among the three alternatives—*bào*, *mǎ*, and *hàn*—*mǎ* shared the same tone of *tǔ*, and thus it was the most similar choice. In each of the four rounds with 16 items each round, the correct answer of half of the items shared the same segments, and that of the remaining half shared the same tone. Feedback was given in the same manner as in the Segmental Training Group.

### Day 3

All participants first took a Vocabulary Review Test of the 16 target words in both directions, just as they did on Day 2. The training session (including both the training and exercise phases) was repeated, both in the Segmental and in the Whole-Syllable-Based conditions. After the training session, all participants took the formal musical ability test.

### Day 4

All participants repeated the training session and then completed the set of three post-training tests. The post-training tests were computer-administered using DMDX software (Forster and Forster, 2003). Both accuracy and response time were recorded. The post-training Same-Different Phonological Judgment task was administered first, which was then followed by the Tone Identification task and the Word Comprehension task. The order of the Tone Identification and Word Comprehension tasks was counterbalanced across participants.

## RESULTS

### Musical Ability

The percentage accuracy rate was calculated for each group (Whole-Syllable-Based:  $M = 0.67$ ,  $SD = 0.19$ ; Segmental:  $M = 0.62$ ,  $SD = 0.15$ ). No significant difference was shown between the two groups ( $t(40) = 1.00$ ,  $p = 0.32$ ). The Whole-Syllable-Based Training Group had a slightly higher musical ability than the Segmental Training Group. As four participants did not complete the final outcome test of phonological judgment<sup>1</sup>, after excluding those four, the Whole-Syllable-Based Training Group again had slightly higher accuracy than the Segmental Training Group; however, the difference was still not statistically significant,  $t(36) = 1.11$ ,  $p = 0.28$ .

<sup>1</sup>Four participants (one from the Whole-Syllable-Based group) did not complete the posttest of phonological judgment task, and three of them (one from the Whole-Syllable-Based group) did not complete the pretest of phonological judgment task either. All participants completed the other two outcome measures: Tone Identification and Word Comprehension

### Phonological Judgment

All analyses were carried out in R, an open-source programming environment for statistical computing (R Core Team, 2013), with the lme4 package (Bates et al., 2015) for linear mixed-effects modeling (i.e., Response Time (RT) analyses) and logistic regression modeling (i.e., error rates analyses), both with the “bobyqa” optimizer. RTs of incorrect responses were excluded. In each condition (i.e., S + T+, S + T-, S-T+, and S-T-), we compared participants’ performance in the pre-training vs. post-training on original items, as well as that in post-training on original items and new items separately. For the linear mixed-effects models for both comparisons, contrast-coded fixed effects included group (Whole-Syllable-Based vs. Segmental were coded as -0.5 vs. 0.5), centered musical ability (participants’ own musical ability score minus the mean), training effect (pre-training vs. post-training as -0.5 vs. 0.5), and all the two- and three-way interactions among these factors. Participants and items were entered as two random intercepts with related random slopes (i.e., item type for the subject; group, musical ability, and their two-way interaction as the slopes for the item). The same fixed effects and random intercepts were included in the logistic regression for error rate analyses, in which correct and incorrect trials were coded as 0 vs. 1, respectively. The significance of each fixed effect was assessed via likelihood ratio tests (Barr et al., 2013).

### Comparison Between Pre- and Post-training on the Original Test Items

For the original items, out of all the four conditions (see Tables 2, 3 for basic statistics on all the four conditions, and Supplementary Appendix E,F for figure illustrations), we focused on the S + T-<sup>2</sup>. Previous research has shown that this condition poses a particular challenge to Chinese as L2 learners (Wang M. et al., 2003; Li et al., 2018). For error rates in the S + T-condition, participants overall made significantly fewer errors in the post-training than in the pre-training test ( $M = 2.6$  vs. 14.0%;  $\beta = -1.24$ ,  $SE \beta = 0.61$ ,  $\chi^2 = 4.42$ ,  $p = 0.04$ ). No significant effects were shown for any other main effects or interactions ( $ps > 0.10$ ). For RTs, participants were overall faster in the post-training than in the pre-training test ( $M = 1,040$  ms vs. 1,386 ms;  $\beta = -352.45$ ,  $SE \beta = 49.68$ ,  $\chi^2 = 34.64$ ,  $p < 0.001$ ). In addition, participants with higher musical ability, in general, responded faster than those with the lower musical ability ( $\beta = -561.79$ ,  $SE \beta = 279.50$ ,  $\chi^2 = 4.23$ ,  $p = 0.04$ ). All other fixed effects (i.e., group and all the two-way and three-way interactions) failed to show significant results ( $ps > 0.10$ ).

For all the other three conditions, the analyses of error rates were not no further model analysis was conducted for error rates due to the overall low rate of errors, thus floor effects occurred (ranged from 0.00–0.03; see Table 2 for details). Consistent with the results on the S + T-items, the analyses of RTs of all the three conditions consistently showed a significant main effect of musical ability (higher musical ability led to faster RTs) and

<sup>2</sup>We did not identify any evidence for an accuracy-RT tradeoff for the original items in the S + T-condition in the pre-test, as the correlation between accuracy and RT was not significant,  $r(37) = -0.01$ ,  $p = 0.950$

**TABLE 2 |** Error Rates (SDs in parentheses) on All Conditions and Item Types in the Same-Different Phonological Judgment Task in Both Segmental and Whole-Syllable-Based Training Groups.

Item_type	Condition	Segmental training	Whole-syllable-based training
Pre_Test	S + T+	0.01 (0.12)	0.02 (0.13)
	S + T-	0.13 (0.34)	0.15 (0.36)
	S-T+	0.03 (0.18)	0.02 (0.13)
	S-T-	0.02 (0.15)	0.03 (0.16)
Post_Test Original Items	S + T+	0.02 (0.15)	0.01 (0.10)
	S + T-	0.02 (0.14)	0.03 (0.18)
	S-T+	0 (0)	0 (0.06)
	S-T-	0.01 (0.09)	0 (0.06)
Post_Test New Items	S + T+	0.01 (0.12)	0.01 (0.08)
	S + T-	0.17 (0.38)	0.14 (0.35)
	S-T+	0 (0)	0 (0)
	S-T-	0 (0)	0 (0)

The training effect refers to the difference between the post-test vs. pre-test (i.e., reduced error rates on the original items or on the new items).

**TABLE 3 |** RTs (SDs in parentheses) on All Conditions and Item Types in the Same-Different Phonological Judgment Task in Both Segmental and Whole-Syllable-Based Training Groups.

Item_type	Condition	Segmental training	Whole-syllable-based training	JZS bayes factor (BF 10 <sup>a</sup> ) of group difference on training effect
Pre_Test	S + T+	1,229.2 (460.92)	1,141.13 (419.39)	N/A
	S + T-	1,402.3 (516.1)	1,371.4 (546.64)	N/A
	S-T+	1,240.24 (434.27)	1,114.23 (435.84)	N/A
	S-T-	1,273.53 (518.14)	1,143.14 (414.22)	N/A
Post_Test Original Items	S + T+	985.84 (457.83)	913.74 (319.48)	0.39
	S + T-	1,068.49 (357.49)	1,015.57 (335.86)	0.36
	S-T+	956.55 (438.28)	859.63 (314.27)	0.38
	S-T-	937.93 (349.8)	864.74 (299.84)	0.33
Post_Test New Items	S + T+	996.5 (332.42)	953.86 (284.76)	0.36
	S + T-	1,016.13 (314.02)	960.94 (429.79)	0.38
	S-T+	918.59 (315.94)	874.49 (260.22)	0.33
	S-T-	895.11 (317.23)	853.53 (256.08)	0.32

The training effect refers to the difference between the post-test vs. pre-test (i.e., reduced error rates on the original items or on the new items).

<sup>a</sup>BF, 10 values equal or smaller than .33 (1/3) favor null hypotheses.

training effect (faster RTs in the post-training test) ( $ps < 0.03$ ), while all other effects (i.e., group and all the two-way and three-way interactions) were not significant ( $ps > 0.10$ ).

### Comparison Between Original and New Test Items on Post-test

Comparing the original items and new items in post-test only, for S + T-condition<sup>3</sup>, participants overall made significantly fewer errors with the original than new items ( $M = 2.6$  vs.  $15.5\%$ ;  $\beta = -2.60$ ,  $SE \beta = 2.12$ ,  $\chi^2 = 4.27$ ,  $p = 0.039$ ). No significant effects were shown for any other main effects or interactions ( $ps > 0.10$ ). For RTs, the only significant effect was that: participants with higher musical ability in general responded faster than those with lower musical ability ( $M =$

$2.6$  vs.  $15.5\%$ ;  $\beta = -560.31$ ,  $SE \beta = 260.67$ ,  $\chi^2 = 5.97$ ,  $p = 0.01$ ). All other fixed effects (i.e., group, item type, and all the two-way and three-way interactions) failed to show significant results ( $ps > 0.10$ ).

For all the three other conditions, again, the analyses of error rates were not further conducted due to the floor effects (ranged from  $0.00$ – $0.03$ ; see **Table 2** for details). Similar to the results on S + T-items, for RTs, participants with higher musical abilities showed significantly faster RTs than those with lower abilities when responding to S+T+ items ( $ps < 0.05$ ). All other fixed effects (i.e., group, item type, and all the two-way and three-way interactions) failed to show significant results ( $ps > 0.10$ ).

### Vocabulary Learning During Testing and Review

**Table 4** shows the mean accuracy rate in the Vocabulary Testing (Day 1) and Review (Days 2 and 3) phase, including the overall accuracy for all rounds on each day and the accuracy rate for the

<sup>3</sup>We did not identify any evidence for an accuracy-RT tradeoff for the new items in the S + T-condition in the post-test, as the correlation between accuracy and RT was not significant,  $r(36) = -0.23$ ,  $p = 0.161$

**TABLE 4** | The accuracy rate of vocabulary testing and review in both segmental and whole-syllable-based training groups.

—	Pinyin-picture		Picture-pinyin	
	Segmental training	Whole-syllable-based training	Segmental training	Whole-syllable-based training
D1-overall	0.73 (0.24)	0.79 (0.20)	0.82 (0.18)	0.89 (0.10)
D1-last	0.83 (0.23)	0.84 (0.23)	0.88 (0.12)	0.93 (0.06)
D2-overall	0.85 (0.16)	0.91 (0.10)	0.88 (0.15)	0.95 (0.06)
D2-last	0.87 (0.15)	0.93 (0.07)	0.92 (0.13)	0.96 (0.05)
D3-overall	0.92 (0.12)	0.98 (0.03)	0.96 (0.08)	0.99 (0.03)
D3-last	0.95 (0.06)	0.98 (0.03)	0.96 (0.05)	0.99 (0.03)

Note. D1-overall: Overall accuracy rate for all rounds in the vocabulary testing on Day 1. D1-last round: The accuracy rate for the last round in the vocabulary testing on Day 1.

last round on each day. A simple linear regression with Musical Ability, Group (Segmental vs. Whole-Syllable-Based) and Day (1, 2 vs. 3) being factors was conducted in R with the nlme package (Pinheiro et al., 2017) for both the overall accuracy and the accuracy in the last round of Pinyin-Picture and Picture-Pinyin identification tasks.

### Overall Accuracy

For the Pinyin-Picture task, the results showed a significant main effect of Day,  $F(2, 74) = 30.32, p < 0.001$ . Participants' performance increased significantly from Day 1 to Day 2, Day 1 to Day 3, Day 2 to Day 3,  $ps < 0.05$ . However, the main effect of Musical Ability and Group were both non-significant,  $F(1, 40) = 2.47, p = 0.12$ , and  $F(1, 74) = 0.98, p = 0.33$ . None of the two- or three-way interactions was significant, all  $ps > 0.10$ . Likewise, for the Picture-Pinyin task, there was a significant main effect of Day,  $F(2, 74) = 21.01, p < 0.001$ , and the participant's performance increased significantly from Day 1 to Day 3,  $p < 0.05$ . However, the main effects of Musical Ability and Group were non-significant, both  $ps > 0.09$ . The two- or three-way interactions were all non-significant, all  $ps > 0.40$ .

### Accuracy in the Last Round

For the Pinyin-Picture task, results showed a significant main effect of Day,  $F(2, 74) = 11.11, p < 0.001$ , and the participants' performance improved significantly from Day 1 to Day 2, Day 1 to Day 3, Day 2 to Day 3,  $ps < 0.05$ . Neither the main effect of Musical Ability nor Group was significant,  $F(1, 40) = 3.09, p = 0.09$ , and  $F(1, 74) = 0.70, p = 0.40$ . The Group  $\times$  Day, Group  $\times$  Musical Ability, or Group  $\times$  Day  $\times$  Musical Ability interactions were not significant, all  $ps > 0.6$ . There was a significant interaction between Day and Musical Ability,  $F(2, 74) = 3.29, p = 0.04$ . Participants with higher musical ability showed a trend of higher accuracy than those with lower musical ability in Day 1,  $p = 0.06$ , but not in Days 2 or 3,  $ps > 0.10$ .

For the Picture-Pinyin task, there was a significant main effect of Day,  $F(2, 74) = 13.51, p < 0.001$ , and the participant's performance increased significantly from Day 1 to Day 3,  $p < 0.05$ . No significant effect was shown for Group,  $F(1, 74) = 2.60, p = 0.11$ , nor Musical Ability,  $F(1, 40) = 3.71, p = 0.06$ . The two- and three-way interactions were all non-significant, all  $ps > 0.30$ .

We also listed in **Supplementary Appendix G** the number of rounds it took the participants to reach the criterion and the number of participants who did not reach criterion (0.88

accuracy) by the end of the last round in vocabulary testing in Days 1, 2, and 3.

## Exercises Across Training Sessions

### Segmental Training Group

**Table 5** presents the mean accuracy rates in the exercises across training sessions (Session 1 on Day 2, Session 2 on Day 3, Session 3 on Day 4), including the accuracy for each of the four rounds in each session. A simple linear regression with Musical Ability, Session (1, 2 vs. 3), and Judgment Type (1 and 2: segment judgment vs. 3 and 4: tone judgment) as independent variables was conducted. The main effect of Session was significant,  $F(2, 93) = 5.86, p < 0.01$ , participants' performance improved significantly from Session 1 to Session 3,  $p < 0.01$ . There was a marginally significant improvement from Session 2 to Session 3,  $p = 0.09$ , but not from Session 1 to Session 2,  $p = 0.18$ . No significant main effect was found for Judgment Type or Musical Ability, both  $ps > 0.20$ . None of the two- or three-way interactions was significant ( $ps > 0.10$ ).

### Whole-Syllable-Based Training Group

A simple linear regression with Musical Ability, Session (1, 2 vs. 3) and Round (1, 2, 3 vs. 4) as three independent variables revealed that there was no significant main effect for Session and Round, ( $ps > 0.60$ ). The effect of Musical Ability was significant,  $F(1, 19) = 9.39, p = .01$ . The interaction was significant between Session and Round,  $F(6, 209) = 2.94, p = 0.01$ . Only in first session participants showed difference across the Rounds,  $F(3, 80) = 3.32, p < 0.05$ , but not in the second or third sessions ( $ps > 0.49$ ), respectively. Specifically, in Session 1, participants' performance on Rounds 3 and 4 were both significantly better than Round 1,  $p = 0.03, p = 0.02$ , respectively, but no significant difference between Rounds 1 and 2, 2 and 3, 2 and 4 or 3 and 4, all  $ps > 0.10$ . The three-way interaction was significant,  $F(6, 209) = 2.79, p = 0.01$ . However, the interactions between Session and Musical Ability and between Judgment Type and Musical Ability were not significant, both  $ps > 0.20$ .

## Tone Identification and Word Comprehension

We compared the participants' performance in the Segmental vs. Whole-Syllable-Based Training Groups for both error rates (logistic regression models) and RTs (linear mixed-effects models) on the two post-training tests: Tone Identification and

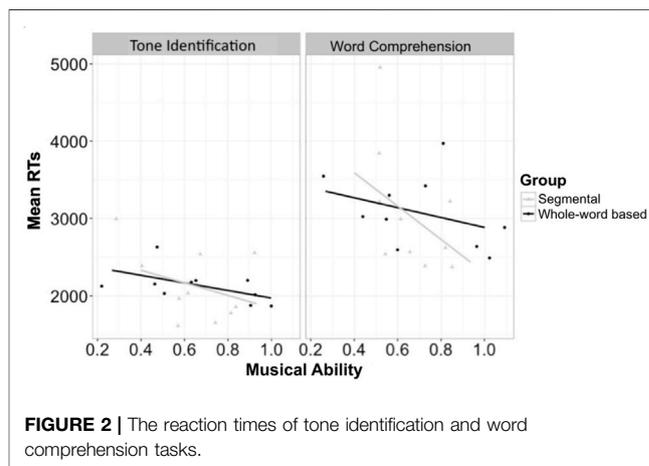
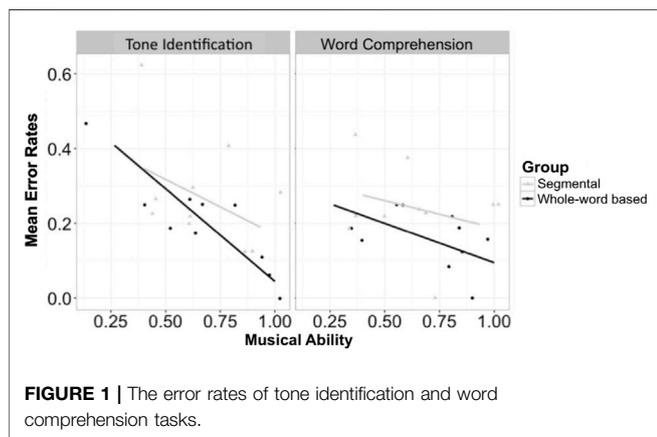
**TABLE 5 |** The accuracy rates in the exercises across training sessions in both segmental and whole-syllable-based training groups.

—	Segmental training				Whole-syllable-based training			
	Round 1	Round 2	Round 3	Round 4	Round 1	Round 2	Round 3	Round 4
Session 1	0.90 (0.14)	1.00 (0.01)	0.97 (0.05)	0.98 (0.04)	0.92 (0.12)	0.96 (0.10)	0.98 (0.10)	0.98 (0.04)
Session 2	0.96 (0.06)	1.00 (0.01)	0.96 (0.18)	0.99 (0.03)	0.96 (0.14)	0.97 (0.14)	0.96 (0.14)	0.96 (0.14)
Session 3	0.99 (0.14)	1.00 (0.14)	0.99 (0.14)	1.00 (0.14)	0.97 (0.14)	0.98 (0.10)	0.95 (0.19)	0.96 (0.20)

**TABLE 6 |** The Mean Accuracy Rates and RTs (SDs in parentheses) in the Tone Identification and Word Comprehension Tasks in Both Segmental and Whole-Syllable-Based Training Groups.

—		Segmental training	Whole-syllable-based training	JZS bayes factor (BF 10 <sup>a</sup> )
Tone Identification	Accuracy Rate	0.74 (0.44)	0.80 (0.40)	0.32
	RTs	2,195.29 (1,289.83)	2,136.43 (1,061.73)	0.37
Word Comprehension	Accuracy Rate	0.76 (0.43)	0.82 (0.38)	0.32
	RTs	3,369.43 (2,104.35)	3,064.34 (1,513.81)	0.32

<sup>a</sup>BF, 10 values equal or smaller than 0.33 (1/3) favor null hypotheses.



Word Comprehension (See **Table 6** and **Figures 1, 2** for the results). For both analyses, contrast-coded fixed effects included Group (Whole-Syllable-Based vs. Segmental), Musical Ability, and the interactions of these factors. Participants and items were entered as two random intercepts with related random slopes (i.e., Musical Ability, Group, and their two-way interaction as the slopes for the item).

For the Tone Identification task, there was a significant main effect of Musical Ability in error rates (higher musical ability led to lower error rates;  $\beta = -2.51$ ,  $SE \beta = 0.95$ ,  $\chi^2 = 6.70$ ,  $p = 0.01$ ), and a marginally significant main effect of Musical Ability in RTs (higher musical ability tended to yield lower RTs;  $\beta = -1,039.64$ ,  $SE \beta = 611.39$ ,  $\chi^2 = 3.05$ ,  $p = 0.08$ ). None of the other fixed effects showed significant results ( $ps > 0.24$ ). We also compared the original and new items in this task. There was no significant difference, and the interactions between the item type and other factors were not significant ( $ps > 0.20$ ).

For the Word Comprehension task, the only significant result was the main effect of Musical Ability in RTs (higher musical ability led to lower RTs;  $\beta = -2095.80$ ,  $SE \beta = 1,084.32$ ,  $\chi^2 = 3.90$ ,

$p < 0.05$ ). For both error rates and RTs, none of the other fixed effects showed significant results ( $ps > 0.20$ ).

Finally, since no significant group effect was observed on phonological judgment, tone identification, and word comprehension tasks, the Bayes Factor was calculated to address whether an effect is substantial or not. The JZS Bayes Factor (BF10) is the ratio of the Bayesian probability of the alternative hypothesis and the null hypothesis occurring, and the odds equal or smaller than one third can be considered substantial evidence in favor of the null hypothesis (Rouder et al., 2009). In the three tasks, all the BF 10 values were below or close to 1/3, providing support for the null group effects (details see **Tables 2, 3, and 6**).

## DISCUSSION

In the present study, the ways in which we can improve phonological and word learning in a novel language were investigated. A training

method that focused on enhancing learners' metalinguistic awareness of segmental and tonal information was compared to an integrated syllable approach. The role of musical ability in learning a new language was also taken into consideration, as a covariate in the final analyses, based on previous research having established its role in language performance. The present findings replicated those of previous studies that have established that short-term training can indeed facilitate L2 learners' acquisition of Chinese phonology and extended this finding to word learning. In our study, with as few as 2 hours of training across 3 days, both groups showed significant improvement in phonological judgment as well as spoken word recognition (error rates <26% in both spoken tone identification and word comprehension tasks for both groups in the post-test). Moreover, participants also improved their scores in the vocabulary review and test performance across the training days.

As expected, the processing of tonal information poses a particular challenge for Chinese L2 learners. In the same-different phonological judgment task, a significant improvement in both accuracy and speed was seen on the original items comparing pre- and post-test results. However, the accuracy rate of new items was significantly lower than that of original items in the post-tests, suggesting low generalizability of tone skill. These results suggest that the training effects limited to the original items. As a matter of fact, the accuracy rate of new items was similar to that of the original items in pre-tests (see **Supplementary Appendix F**), when participants had not received any training, despite the fact that the participants achieved a faster response time for these new items in the post-test compared to the original items in the pre-test probably due to their becoming familiar with the segmental and tonal information in the novel language.

The two groups did not show a significant difference in any of the post-training tests (i.e., phonological judgment, tone identification, and word comprehension), suggesting that the two training methods are equally effective for learning Chinese phonology. One exception is that the learners in the Segmental Training Group showed significant improvement across the exercise sessions during the training days, whereas the Whole-Syllable-Based Training Group did not show such improvement. Lastly, learners with higher musical ability, in general, had better performance in all three post-training tests, supporting the role of musical ability for phonological and word learning, replicating previous work, which has provided some insight into the role of musical ability in language learning.

## The Benefit of Pinyin in Phonological Training

The two groups of participants' segmental and tonal processing were improved equally after both training types in the present study. Being trained to process segmental and tonal information separately did not yield more benefits, as was originally hypothesized. It is unlikely a floor or ceiling effect, as both groups did show significant improvement in tone perception in the homophone judgment task (i.e., lower error rates and faster reaction times in the different-tone homophone condition, at least for original items). However, there is clearly still room for improvement (i.e., they still had the lowest accuracy rate and slowest reaction times in the different-tone

homophone condition than in the other conditions). We speculate that the explicit phonology and Pinyin instruction at the beginning of the training program might have been effective. The learners might have successfully acquired concrete knowledge about the four tones and their associated visual pitch contours early in training. Later, when the participants learned vocabulary words, they might be able to rely on segment and tone as an integrated unit (e.g., Wang et al., 2015) instead of using them as separate cues (e.g., Spinks et al., 2000). Therefore, no further benefit from the explicit emphasis on the separation. Furthermore, the exercise sessions during the training days might have contributed to the enhanced phonological representation across the two groups in a similar way. As a result, the presence of Pinyin throughout the entire program might have been sufficient for assisting the participants in processing segmental and tonal information for all participants.

Even with the support from Pinyin, Chinese L2 learners did not show significant improvement in accurately judging different-tone homophone spoken word pairs for the new items in the post-test of the phonological judgment task. Their accuracy rates were significantly lower than that of the original items which they had heard in the pre-test. Nevertheless, the participants did demonstrate improvement in their reaction times on the new items compared to the original ones included in the pre-test. In addition, the learners performed on the new items as accurately and rapidly as the original ones in the post-test of the tone identification task. It is likely that the learners might have established a deep understanding of the tonal variation in Chinese after the training. Therefore, they were able to generalize it to the new items when matching the auditory information to the written Pinyin. Taken together, these findings seem to suggest that although it is difficult for Chinese L2 learners to fully generalize the tone perception to new materials (as shown in the difference between original vs. new items in the homophone judgment task), with Pinyin symbols, they were able to identify the tonal information in new spoken words as accurately and as rapidly as those trained items.

## Musical Ability Contributes to Phonological and Word Learning

Participants' musical ability, more specifically, pitch perception ability, was included in the current study as a covariate. As hypothesized, a clear effect of musical ability was shown in phonological and word learning of a novel language. The effect of musical ability was significant in the reaction time in the different-tone homophone condition in the post-test of the phonological judgment task for both the original and the new items. The effect of musical ability was also shown in the accuracy rates for the post-test of the tone identification task and in the reaction time for the word comprehension task. These results provide converging evidence that musical ability contributes to learning a novel language. Li and DeKeyser (2017) showed that musical ability was highly correlated to the accuracy not only in tone word perception but also in production. It is possible that musical ability is linked to language learning in general, but maybe particularly related to learning a tonal language. In the present study, our learners have clearly benefitted from their musical ability in perceiving and processing the pitch variations

and contours embedded in the Chinese tonal system. Finally, since musical ability is a significant covariate and the two training methods are equally effective in the current study, it is not surprising that musical ability contributes to the learning in a similar fashion across the two groups.

## Limitations and Future Directions

The current findings need to be processed with some caution. One limitation of the present study is the limited number of items included in the training modules, given that one initial aim was to maintain a relatively brief and simple training schedule. We only trained the novel language learners on four basic segmental syllables combined with four tones, thus introducing 16 different syllables in total. Therefore, it is likely that learners may be able to achieve better generalization of tonal knowledge after a lengthier training with more syllables. A longer training time will also allow for collection of more data from the participants to show a longer-term training effect. Future research can test this hypothesis with the inclusion of more training items.

In addition, all of our training items were monosyllabic. Training on di-syllabic or multi-syllabic words can improve the diversity of the materials that participants learn and thus enhance the learner's generalizability of tonal knowledge. Moreover, the di-syllabic or multi-syllabic items present a greater challenge to learners, and so the training strategies used in the Segmental Training Group to enhance the learners' awareness of tonal and segmental information may begin to show some advantages. Future research could aim to develop a more extensive training program that involves monosyllabic, di-syllabic, and multi-syllabic words to address how best to improve learners' metalinguistic awareness of the segmental and tonal information. Relatedly, future research can consider using high-variability training instead of the same word repetition and using a more active paradigm that engages participants in production beyond the more implicit training heavily focused on phonological perception. Also, our participants made overall very few errors in the phonological judgment task. In future research, to further increase the difficulty of this task to get more errors, researchers could consider using different speakers for the same token stimuli when repeated or degrading the stimuli (e.g., put them in noise).

Moreover, our participants were all native English speakers, thus the generalizability of the training effect could not be addressed. Future work can recruit participants with different backgrounds of native languages to address the generalizability issue and potential learning differences. Lastly, note that all of our participants had never learned a tonal language and thus Pinyin and Chinese were novel to them prior to our study. We acknowledge that if we have included a comparable non-Pinyin training group, our experiment would have been even more rigorous, which is a limitation that could be addressed in future work.

## CONCLUSION

In the present study, two training methods aiming to help native English-speakers learn sixteen Chinese words with assistance

from Pinyin symbols were designed, and their relative effectiveness was compared. One method emphasized the separation between segmental and tonal information of the Pinyin symbols throughout the training, whereas the other used Pinyin symbols as an integral unit. Results demonstrated that the two different training methods were equally effective in supporting Chinese L2 learners' phonological learning, tone identification, and spoken word comprehension. The equal success via the two training methods may be due to the consistent involvement of Pinyin throughout the training and learner's reliance on segment and tone as an integral unit rather than separate cues in learning Chinese words.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board at the University of Maryland, College Park. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

MW and CL designed the study. YL, MW and CL reviewed articles and wrote the paper. YL, CL and ML collected and analyzed the data. All authors discussed the results and commented on the manuscript.

## FUNDING

This study was supported by a Professional Development fund from the University of Maryland to MW. The writing of this article was also supported by the Spencer Foundation awarded to MW under Grant number 201900053.

## ACKNOWLEDGMENTS

We thank Michelle Yang, an undergraduate RA, for assisting with the data collection. We thank Hailey Gibbs at the University of Maryland, College Park, for constructive feedback and proofreading.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2022.629349/full#supplementary-material>

## REFERENCES

- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random Effects Structure for Confirmatory Hypothesis Testing: Keep it Maximal. *J. Mem. Lang.* 68 (3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2015). lme4: Linear Mixed-Effects Models Using Eigen and S4. R package version 1, 1–7.
- Braun, M., Hutzler, F., Ziegler, J. C., Dambacher, M., and Jacobs, A. M. (2009). Pseudohomophone Effects Provide Evidence of Early Lexico-Phonological Processing in Visual Word Recognition. *Hum. Brain Mapp.* 30 (7), 1977–1989. doi:10.1002/hbm.20643
- Chen, J.-Y., Chen, T.-M., and Dell, G. S. (2002). Word-Form Encoding in Mandarin Chinese as Assessed by the Implicit Priming Task. *J. Mem. Lang.* 46 (4), 751–781. doi:10.1006/jmla.2001.2825
- Cooper, A., and Wang, Y. (2012). The Influence of Linguistic and Musical Experience on Cantonese Word Learning. *The J. Acoust. Soc. America* 131 (6), 4756–4769. doi:10.1121/1.4714355
- Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. 6th ed. Malden, MA: Wiley-Blackwell.
- Dittinger, E., Barbaroux, M., D’Imperio, M., Jäncke, L., Elmer, S., and Besson, M. (2016). Professional Music Training and Novel Word Learning: From Faster Semantic Encoding to Longer-Lasting Word Representations. *J. Cogn. Neurosci.* 28 (10), 1584–1602. doi:10.1162/jocn\_a\_00997
- Forster, K. I., and Forster, J. C. (2003). DMDX: A Windows Display Program with Millisecond Accuracy. *Behav. Res. Methods Instr. Comput.* 35 (1), 116–124. doi:10.3758/bf03195503
- Gottfried, T. L., and Suiter, T. L. (1997). Effect of Linguistic Experience on the Identification of Mandarin Chinese Vowels and Tones. *J. Phonetics* 25 (2), 207–231. doi:10.1006/jpho.1997.0042
- Godfroid, A., Lin, C. H., and Ryu, C. (2005). Hearing and seeing tone through color: An efficacy study of web-based, multimodal Chinese tone perception training. *Lang. Learn.* 67 (4), 819–857.
- Gross, H. S. (2017). Prosody. *Encyclopedia Britannica*. Available at: <https://www.britannica.com/art/prosody> (Accessed August 10, 2021).
- Ho, C. S.-H., and Bryant, P. (1997). Phonological Skills Are Important in Learning to Read Chinese. *Develop. Psychol.* 33 (6), 946–951. doi:10.1037/0012-1649.33.6.946
- Leong, C. K., Cheng, P. W., and Tan, L. H. (2005). The Role of Sensitivity to Rhymes, Phonemes and Tones in reading English and Chinese Pseudowords. *Read. Writ* 18 (1), 1–26. doi:10.1007/s11145-004-3357-2
- Li, M., and DeKeyser, R. (2017). Perception Practice, Production Practice, and Musical Ability in L2 Mandarin Tone-Word Learning. *Stud. Second Lang. Acquis* 39 (4), 593–620. doi:10.1017/s0272263116000358
- Li, C. N., and Thompson, S. A. (1989). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley, CA: University of California Press.
- Li, C., Lin, C. Y., Wang, M., and Jiang, N. (2013). The Activation of Segmental and Tonal Information in Visual Word Recognition. *Psychon. Bull. Rev.* 20 (4), 773–779. doi:10.3758/s13423-013-0395-2
- Li, C., Wang, M., Davis, J. A., and Guan, C. Q. (2018). The Role of Segmental and Tonal Information in Visual Word Recognition with Learners of Chinese. *J. Res. Reading* 42 (2), 213–238. doi:10.1111/1467-9817.12137
- Li, Y., Chen, X., Li, H., Sheng, X., Chen, L., Richardson, U., et al. (2020). A Computer-Based Pinyin Intervention for Disadvantaged Children in China: Effects on Pinyin Skills, Phonological Awareness, and Character Reading. *Dyslexia* 26 (4), 377–393. doi:10.1002/dys.1654
- Lin, D., McBride-Chang, C., Shu, H., Zhang, Y., Li, H., Zhang, J., et al. (2010). Small Wins Big. *Psychol. Sci.* 21 (8), 1117–1122. doi:10.1177/0956797610375447
- Liu, Y., Wang, M., Perfetti, C. A., Brubaker, B., Wu, S., and MacWhinney, B. (2011). Learning a Tonal Language by Attending to the Tone: An *In Vivo* experiment. *Lang. Learn.* 61 (4), 1119–1141. doi:10.1111/j.1467-9922.2011.00673.x
- Lü, C. (2017). The Roles of Pinyin Skill in English Chinese Bilingual Learning: Evidence from Chinese Immersion Learners. *Foreign Lang. Ann.* 50 (2), 306–322. doi:10.1111/flan.12269
- Luo, C. R., Johnson, R. A., and Gallo, D. A. (1998). Automatic Activation of Phonological Information in reading: Evidence from the Semantic Relatedness Decision Task. *Mem. Cogn.* 26 (4), 833–843. doi:10.3758/bf03211402
- Malins, J. G., and Joannis, M. F. (2010). The Roles of Tonal and Segmental Information in Mandarin Spoken Word Recognition: An Eyetracking Study. *J. Mem. Lang.* 62 (4), 407–420. doi:10.1016/j.jml.2010.02.004
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., and Besson, M. (2011). Influence of Musical Expertise on Segmental and Tonal Processing in Mandarin Chinese. *J. Cogn. Neurosci.* 23 (10), 2701–2715. doi:10.1162/jocn.2010.21585
- Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A. A., and Tervaniemi, M. (2008). Musical Aptitude and Second Language Pronunciation Skills in School-Aged Children: Neural and Behavioral Evidence. *Brain Res.* 1194, 81–89. doi:10.1016/j.brainres.2007.11.042
- Milovanov, R., Pietilä, P., Tervaniemi, M., and Esquef, P. A. A. (2010). Foreign Language Pronunciation Skills and Musical Aptitude: A Study of Finnish Adults with Higher Education. *Learn. Individual Differences* 20 (1), 56–60. doi:10.1016/j.lindif.2009.11.003
- Pinheiro, J., Bates, D., DebRoy, S., and Sarkar, D.; R Core Team (2017). nlme: Linear and Nonlinear Mixed Effects Models. R Package version 31-131.
- R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.
- Read, C., Zhang, Y. F., Nie, H. Y., and Ding, B. Q. (1986). The Ability to Manipulate Speech Sounds Depends on Knowing Alphabetic Writing. *Cognition* 24 (1-2), 31–44. doi:10.1016/0010-0277(86)90003-x
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., and Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bull. Rev.* 16 (2), 225–237. doi:10.3758/s13423-013-0395-2
- Ruan, Y., Georgiou, G. K., Song, S., Li, Y., and Shu, H. (2018). Does Writing System Influence the Associations between Phonological Awareness, Morphological Awareness, and reading? A Meta-Analysis. *J. Educ. Psychol.* 110 (2), 180–202. doi:10.1037/edu0000216
- Sereno, J. A., and Lee, H. (2015). The Contribution of Segmental and Tonal Information in Mandarin Spoken Word Processing. *Lang. Speech* 58 (2), 131–151. doi:10.1177/0023830914522956
- Showalter, C. E., and Hayes-Harb, R. (2013). Unfamiliar Orthographic Information and Second Language Word Learning: A Novel Lexicon Study. *Second Lang. Res.* 29 (2), 185–200. doi:10.1177/0267658313480154
- Shu, H., Peng, H., and McBride-Chang, C. (2008). Phonological Awareness in Young Chinese Children. *Develop. Sci.* 11 (1), 171–181. doi:10.1111/j.1467-7687.2007.00654.x
- Spinks, J., Liu, Y., Perfetti, C. A., and Tan, L. H. (2000). Reading Chinese Characters for Meaning: The Role of Phonological Information. *Cognition* 76 (1), B1–B11. doi:10.1016/s0010-0277(00)00072-x
- Taft, M., and Chen, H.-C. (1992). “Judging Homophony in Chinese: The Influence of Tones,” in *Language Processing in Chinese*. Editors H.-C. Chen and O. J. L. Tzeng (Amsterdam, Netherlands: Elsevier), 151–172. doi:10.1016/s0166-4115(08)61891-9
- Tan, L. H., and Perfetti, C. A. (1997). Visual Chinese Character Recognition: Does Phonological Information Mediate Access to Meaning? *J. Mem. Lang.* 37 (1), 41–57. doi:10.1006/jmla.1997.2508
- Van Lancker, D., and Fromkin, V. A. (1973). Hemispheric Specialization for Pitch and “Tone”: Evidence from Thai. *J. Phonetics* 1 (2), 101–109. doi:10.1016/s0095-4470(19)31414-7
- Wang, Y., Spence, M. M., Jongman, A., and Sereno, J. A. (1999). Training American Listeners to Perceive Mandarin Tones. *J. Acoust. Soc. America* 106 (6), 3649–3658. doi:10.1121/1.428217
- Wang, Y., Jongman, A., and Sereno, J. A. (2001). Dichotic Perception of Mandarin Tones by Chinese and American Listeners. *Brain Lang.* 78 (3), 332–348. doi:10.1006/brln.2001.2474
- Wang, Y., Sereno, J. A., Jongman, A., and Hirsch, J. (2003a). fMRI Evidence for Cortical Modification during Learning of Mandarin Lexical Tone. *J. Cogn. Neurosci.* 15 (7), 1019–1027. doi:10.1162/089892903770007407
- Wang, M., Perfetti, C. A., and Liu, Y. (2003b). Alphabetic Readers Quickly Acquire Orthographic Structure in Learning to Read Chinese. *Scientific Stud. Reading* 7 (2), 183–208. doi:10.1207/s1532799xsr0702\_4
- Wang, M., Li, C., and Lin, C. Y. (2015). The Contributions of Segmental and Suprasegmental Information in reading Chinese Characters Aloud. *PloS one* 10 (11), e0142060. doi:10.1371/journal.pone.0142060
- Wing, H. D. (1968). *Tests of musical ability and appreciation: An investigation into the measurement, distribution, and development of musical capacity* 2nd ed. London: Cambridge University Press.
- Wong, P. C. M., and Perrachione, T. K. (2007). Learning Pitch Patterns in Lexical Identification by Native English-Speaking Adults. *Appl. Psycholinguistics* 28 (4), 565–585. doi:10.1017/s0142716407070312

- Xu, Y., Pollatsek, A., and Potter, M. C. (1999). The Activation of Phonology during Silent Chinese Word reading. *J. Exp. Psychol. Learn. Mem. Cogn.* 25 (4), 838–857. doi:10.1037/0278-7393.25.4.838
- Zhu, Z. X., Liu, L., Ding, G. S., and Peng, D. L. (2009). The Influence of Pinyin Typewriting Experience on Orthographic and Phonological Processing of Chinese Characters. *Acta Psychologica Sinica* 41 (09), 785–792. doi:10.3724/sp.j.1041.2009.00785
- Ziegler, J. C., and Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled reading across Languages: A Psycholinguistic Grain Size Theory. *Psychol. Bull.* 131 (1), 3–29. doi:10.1037/0033-2909.131.1.3

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Li, Wang, Li and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*