Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words

Yuan Zhang and Florence Baills
Universitat Pompeu Fabra, Catalonia, Spain

Pilar Prieto
Institució Catalana de Recerca i Estudis Avançats (ICREA), Catalonia, Spain

Abstract
Though research has shown that rhythmic training is beneficial for phonological speech processing, little empirical work has been carried out to assess whether rhythmic training in the classroom can help to improve pronunciation in a second language. This study tests the potential benefits of hand-clapping to the rhythm of newly learned French words for the acquisition of pronunciation patterns by Chinese adolescents. In a between-subjects training experiment with a pretest/posttest design, 50 Chinese adolescents either repeated new French words while clapping out their rhythmic structure (clapping condition) or only repeated the words (non-clapping condition). Participants’ oral production before and after training was (1) perceptually rated for accentedness by two French native speakers and (2) acoustically analysed for final syllable duration. While the results showed an only near-significant improvement from pretest to posttest in accentedness ratings for the clapping group, a significant improvement was obtained for acoustic durational measures. Individual musical abilities did not interact significantly in either of the two analyses, and working memory interacted significantly only with accentedness. These results show that a short training session in which clapping is used to highlight the prosodic structure of words can help improve pronunciation in a foreign language.

Corresponding author:
Yuan Zhang, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Ramon Trias Fargas, 25–27, Catalonia, 08002, Spain.
Email: yuan.zhang@upf.edu
Keywords
Hand-clapping, prosodic structure, rhythm, rhythmic training, second language phonology, second language pronunciation

I Introduction

Rhythm is an important shared feature of music and language. It represents the organization of acoustic events in a time scale with regular succession of strong and weak elements (beat, meter and tempo). Even though the metrical structure of speech is not as regular as music (Patel, 2010), the salient and non-salient syllables of speech create the metrical organization of utterances that allow finding a specific degree of rhythmic regularity and predictability. And, in both domains, rhythm allows a listener to predict what will happen next.

There is growing evidence that rhythmic priming is beneficial for different aspects of language related to phonological awareness and phonological language processing. Various priming studies by Cason, Schön and collaborators have shown that the phonological processing of speech is enhanced by the temporal expectations generated by a rhythmic prime that matches the rhythmic structure of the language. Cason and Schön (2012) tested 20 French participants with trials in which they heard a rhythmic (percussion) prime followed by a pseudo-word whose prosody either matched or mismatched the metrical structure of the prime in terms of both number of segments and stress placement. Then, in a phonemic detection task, the participants were asked to say whether a target phoneme which appeared on the screen had been present or not in the pseudo-word. Both reaction time and electrophysiological (ERP) results revealed a significant difference between the rhythmic/metrical matching and mismatching conditions, thus suggesting that using a music-like rhythmic prime matched to the prosody of the pseudo-word enhanced phonological processing. More recently, Cason, Astésano and Schön (2015) conducted an experiment to assess whether musical rhythmic priming would facilitate the phonological perception of real speech. Seventeen adult French-speaking participants listened to rhythmic primes that matched or mismatched the syllabic structure of sentences and then were asked to detect whether a target vowel had been present in the final syllable of the sentence. Participants were divided into two groups: those in an ‘audio-only’ group listened to the stimuli whereas those in an ‘audio-motor’ group heard the stimuli but additionally underwent at several points during the experiment a short training session where they were asked to mimic the rhythmic prime vocally. When reaction times for the two groups were compared, the results showed that if a rhythmic prime matched the syllable structure of a target, reaction time was significantly shorter, and this effect was even more pronounced by audio-motor training. Regarding production skills, Cason et al. (2015) tested the effect of vocal rhythmic priming (as in Cason et al., 2015) with 14 hearing-impaired children with cochlear implants. First, the children listened to and repeated the target sentences, then they were asked to listen to and vocally mimic the rhythmic prime, and finally they repeated the target sentences once more. As in the previous experiments, the primes matched or mismatched the metrical structure of the target sentences. The results showed significantly improved phonological accuracy...
in terms of both vowel and consonant production in the matching condition relative to baseline, suggesting that the rhythmic priming enhanced phonological production.

More extensive rhythmic intervention studies have also shown that rhythmic training can produce beneficial effects in adults’ and children’s phonological abilities and reading skills. Bhide, Power and Goswami (2013) carried out a two-month program of phonological and rhythmic training for 19 children who were considered poor readers. Nine of the children were trained using GraphoGame Rime, a child-friendly computerized reading program in which they heard sounds, rhymes and words spoken by the computer and had to match the spelling or rhyme of the words. The other ten children followed a musical training program which involved numerous activities such as mimicking short rhythms, clapping and marching to the beat of a song, or learning to chant and play hand-clapping games. A subsequent comparison of the two groups’ phonological abilities and literacy skills (non-word reading, word reading, spelling and phoneme deletion) showed that the musical training program was as effective as the GraphoGame Rime reader training program. Habib et al. (2016) investigated the effects of the Cognitive-Musical Training (CMT) method, a series of musical training exercises that focus on pitch, duration, tempo, pulsation and rhythm intended to enhance both perception and production. In their study, an 18-hour training session was given to ten-year-old children with dyslexia either intensively in three days or spread over six weeks. Their results showed that CMT training yielded higher rates of improvement in these children in terms of categorical perception relative to normal children. Moreover, the intensive training mode produced greater improvement in a syllabic lengthening task than the six-week mode.

The theory of embodied learning also backs up our hypothesis that hand-clapping will favor pronunciation learning. According to this theory, perceiving through multiple sensory channels, here visual and auditory, and producing the reinforcing element, will boost learning (Besson, Chobert, & Marie, 2011; Bidelman, Gandour, & Krishnan, 2011; Wilson & Foglia, 2017). When we feel the beat or when we hear the melody in music, people move their bodies to the rhythms of music. Phillips-Silver and Trainor (2005) provided evidence that the experience of body movements played an important role in musical rhythm perception in 7-month-olds. Feeling the rhythm of music and observing the rhythm of body movements, activate the auditory system and, hence, allow a learner to perceive accurate input of music or language (e.g., Baker, 2014; Smotrova, 2017).

However, little is known about whether rhythmic training activities can enhance phonological awareness or pronunciation in a second language. Most classroom pronunciation trainings tend to center around segmental instruction (that is, it focuses solely on specific speech sounds) and second language prosody is often overlooked. However, recent work has pointed to the need for L2 prosodic instruction, given that having incorrect prosody in the L2 may result in higher ratings of accentedness comprehensibility and intelligibility issues (for a review, see Anderson-Hsieh, Johnson & Koehler, 1992; Derwing & Munro, 2009). Several studies have highlighted the importance of suprasegmental instruction for improving learners’ overall fluency and comprehensibility and reducing their foreign accent (see, for example, Behrman, 2014; Derwing & Rossiter, 2003; Derwing, Munro & Wiebe, 1998; Gordon, Darcy & Ewert, 2013). Yet while suprasegmental training has proven to be successful in improving second language
learners’ overall fluency and comprehensibility, almost no work has tested the efficacy of suprasegmental training paradigms on specific pronunciation issues.

To our knowledge, only three studies from very different domains of research have been conducted on the potential benefits of rhythmic training on L2 pronunciation. In a qualitative study involving six advanced learners of English from various language backgrounds, Fischler (2009) explored the effects of a four-week training program on the learning of English sentence and word stress through activities related to rhythm and rap music. Both before and after the training program, the number of errors that participants made in stress placement in reading and narrative-picture tasks was counted. Additionally, their oral productions were rated by three English native speakers for intelligibility. Although not statistically backed up, an improvement in both intelligibility and stress placement was reported for the reading task only. In a second study, Wang, Mok and Meng (2016) tested a computer application that automatically generated a percussive rhythm for any given text in English with a within-subjects design. Twenty Chinese learners of English were asked to first pronounce 15 English sentences naturally, then introduce the same sentences in the interface, listen to the rhythm that was automatically generated and finally say the sentences out loud immediately afterwards. Sentences pronounced after a rhythmic cue obtained better ratings in terms of native-likeness, but only for speakers that performed worse in the initial unprimed condition. Finally, Gluhareva and Prieto (2017) found positive effects of a short training session that involved observing rhythmic beat gestures – simple up-and-down or back-and-forth hand movements naturally coordinated with the prominent parts of speech – on L2 English pronunciation by Catalan learners. At posttest, participants that had observed the beat gestures significantly improved their accentedness ratings on a set of difficult items related to the trained content.

The three above mentioned studies employed different types of procedures (rap music, a percussive rhythm generator, beat gestures) that are not necessarily easily applicable to teaching suprasegmental aspects of pronunciation in the language classroom. Perhaps surprisingly, the possible facilitating effect of hand-clapping, an activity that can be easily carried out in language learning contexts, has not yet been investigated. Clapping, like tapping one’s foot or dancing to musical rhythms, is a common way for people to use body movements to mimic the temporal structure of music and speech. A variety of hand-clapping games exist across cultures in which hand-clapping is performed to keep the rhythm with the words of a song (Cameron & Grahn, 2014). School teachers use hand-clapping activities with children to help them be more aware of syllabic and metrical structure of words (see Romero Naranjo & Romero Naranjo, 2013).

However, little is known about whether using clapping activities would have beneficial effects for learning L2 pronunciation. Our hypothesis is that clapping, an activity that is strongly linked to rhythm, can facilitate the auditory perception of new words, increase phonological awareness and ultimately lead to better pronunciation. Just as beat gestures that accompany L2 speech help speakers externalize the prosodic features of a foreign language (McCafferty, 2006), we hypothesize that hand-clapping will make more salient the metrical structure of words by acoustically highlighting it with different intensity and duration patterns. In the study by Gluhareva and Prieto (2017), beat gestures were used to mark the temporal structure of speech at a phrasal level (i.e., prominences marked by beat gestures corresponded to word or phrasal stress). By contrast,
hand-clapping highlights the prosodic properties of every syllable in a systematic way. In this study, our hypothesis is that promoting an understanding of linguistic rhythm through hand-clapping can lead to improved pronunciation of a second language. This would be consistent with recent work by Roncaglia-Denissen, Roor, Chen, and Sadakata (2016) in which an enhanced ability to perceive musical rhythm was seen to correlate with more rapid mastery of an L2 whose rhythmic properties differed from those of the L1, suggesting a cognitive transfer between the rhythmic knowledge of language and musical domains. Chen, Fan and Lin (1996) suggested that when teachers tapped, clapped or played rhythm instruments to mimic strong or weak prosodic beats, this helped Chinese students detect word and sentence stress in English.

The focus of our research will be the acquisition of pronunciation by Chinese speakers of words in French, a linguistically distant language which contrasts sharply in terms of the realization of prosody. It is well known that, as speakers of a tonal language, Chinese learners of Germanic and Romance languages frequently experience difficulties in the placement and strengthening of word and sentence stress (e.g., Bian, 2013; Chen, Fan & Lin, 1996; Li, 2012). For instance, Bian (2013) analysed the difficulties of Chinese learners of English as a foreign language (EFL) and reported frequent inappropriate placement of stress in their English words. In French, stress placement tends to be placed on the last syllable of the prosodic group, which can also cause problems for Chinese-speaking learners (see Astésano, 2001). The typical final stress placement found in French is absent from the Chinese phonological system (which links duration to different tone types) and represents a potential problem for Chinese speakers. By way of illustration, the two graphs in Figure 1 show the waveform, spectrogram and F0 contour of the French word confiture [kɔ̃.fi.tyʁ] ‘jam’ as spoken by a native speaker of French (left panel) and as rendered by a Chinese adolescent with no French background (right panel). It will be seen that one of the main acoustic differences between the two lies in the durational characteristics of the final target syllable. While the final syllable as spoken by the French speaker lasts much longer than the preceding syllables, syllable durations are roughly equal in the Chinese speaker’s pronunciation of this word.

Figure 1. Waveform, spectrogram and F0 contour of the French word confiture [kɔ̃.fi.tyʁ] ‘jam’ as spoken by a French native speaker (left panel) and as imitated by a Chinese adolescent with no French background (right panel).
The aim of the current study is to test the effect of clapping the rhythm of novel words in French on the learning of L2 pronunciation by Chinese adolescents. In a between-subjects training experiment, Chinese adolescents were randomly assigned to either a clapping or a non-clapping group. They were asked to repeat a set of 20 new French words under one of two conditions: they either orally repeated the words while clapping out their rhythmic structure (clapping condition) or only repeated the words (non-clapping condition). We hypothesized that observing and performing hand-clapping will lead to a) a general improvement in participants’ pronunciation of the French words and b) an appropriately lengthened production by participants of the words’ final syllables. In order to control for individual differences between participants, a set of individual measures related to working memory, speech imitation skills and musical abilities were also obtained.

II Methods

The experiment consisted of a between-subjects training study with two conditions (clapping vs. no-clapping) and a pretest/posttest design.

1 Participants

A total of fifty 13- to 15-year-old Chinese adolescents (mean age = 13.6; SD = 0.535, 16 females/34 males), all students at the Zhangqiu Experimental Middle School in Shandong Province, China, took part in the experiment on a voluntary basis and submitted written consent prior to the experiment.

All of the students were monolingual Mandarin speakers. The participants took English classes every week in their school and reported using English on average five hours per week, their level of English was proved to be at an initial level by their English teacher. None of them spoke a third language. Data about their language and musical background were self-reported by means of a questionnaire (an English translation of this questionnaire is provided in Appendix 1. Musical expertise was coded on the basis of their responses to the questionnaire, with participants labeled ‘1’ if they reported having had at least two years of training in singing or music and ‘0’ if they had had less than that amount.

2 Materials

a Audiovisual materials for the training session. Materials for the training session consisted of 20 videos prepared at the professional broadcasting studio of the Universitat Pompeu Fabra in Barcelona. Each video was designed to teach one French word. A total of 20 French words (see Table 1) were video-recorded in the two conditions (clapping and non-clapping) by two native French speakers (see Figure 2). The target words included a variety of consonant and vowel sounds in the target language, as well as different prosodic structures (bisyllabic and trisyllabic words, as well as longer words).

The video-recordings were carried out in a soundproof room and then edited with Adobe Premiere Pro CC 2015 and Audacity 2.0.5. Figure 2 shows still images from two sample
Table 1. Target French words selected for the training session.

<table>
<thead>
<tr>
<th>Disyllabic</th>
<th>Trisyllabic</th>
<th>More than 3 syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>balcon ‘balcony’</td>
<td>ambulance ‘ambulance’</td>
<td>television ‘television’</td>
</tr>
<tr>
<td>tambour ‘drum’</td>
<td>crocodile ‘crocodile’</td>
<td>ordinateur ‘computer’</td>
</tr>
<tr>
<td>musique ‘music’</td>
<td>biberon ‘baby bottle’</td>
<td>hélicoptère ‘helicopter’</td>
</tr>
<tr>
<td>purée ‘puree’</td>
<td>céréales cereals</td>
<td>aspirateur ‘vacuum cleaner’</td>
</tr>
<tr>
<td>oreille ‘ear’</td>
<td>confiture ‘jam’</td>
<td></td>
</tr>
<tr>
<td>pizza ‘pizza’</td>
<td>spaguettis ‘spaghetti’</td>
<td></td>
</tr>
<tr>
<td>avion ‘airplane’</td>
<td>éléphant ‘elephant’</td>
<td></td>
</tr>
<tr>
<td>famille ‘family’</td>
<td>mandarine ‘tangerine’</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Still images from video-recordings of the two native French-speakers.
Notes. Upper panels show the non-clapping condition and lower panels show the clapping condition.

videos of the two native French speakers in the non-clapping condition (upper panels) and clapping condition (lower panels). In the clapping condition, both instructors were asked to produce the target French words and simultaneously try to very carefully replicate by means of hand-clapping the intensity and the duration of the syllables of each word.
Each of the live recordings was then embedded in a longer video clip involving a similar set of sequences. First, a black and white line drawing illustrating the French word appeared on the screen. Three seconds later, a video clip showed one of the live recordings, that is, one of the two speakers saying the word while either clapping her hands (clapping condition), or remaining completely still (non-clapping condition). This was followed by a black screen, which lasted for five seconds (see Figure 3). The black screen was intended to provide time for the viewer of the video (the participant) to either (1) repeat the word and clap their hands, or (2) just repeat the word, depending on the experimental group to which they had been assigned (for the procedure, see Figure 4).

For both conditions, the training videos were organized as follows: first, a native speaker of Mandarin Chinese introduced the participants to the task and gave instructions on what they had to do. Then, the twenty sequences containing the training items were displayed in two consecutive blocks. Items were presented in a balanced way so that each item was said by a different instructor between the two blocks. In addition, one of six videos with different orders of items’ presentation was assigned randomly to the participants in order to avoid any primacy or recency effect.

b Audio materials for the pretest and posttest. For both pretest and posttest, participants performed an imitation task (Christiner & Reiterer, 2015) in which they needed to repeat after audio files recorded by a native French speaker. L2 acquisition research has shown the imitation task to be a reliable measure of L2 proficiency, reflecting the ease with which spoken language is comprehended and produced. In this case, materials for the pretest and posttest consisted of 14 French words of different syllabic lengths, ten of which had appeared during the training session and four of which had not (see Table 2). The 14 target words were audio-recorded by a female native French speaker, who was also one of the two speakers featured in the training stimuli, with a high quality HyperX Cloud II headset with microphone. The
14 resulting files were set up in an online survey platform, where their playback order would be automatically randomized.

3 Control measurements

In order to take into account individual differences, a series of control tests were created to measure participants’ cognitive, rhythmic and language imitation abilities.

a Working memory test. According to Henry, Messer, Luger-Klein, & Craine (2012), memory span – or ‘working memory’ – can be measured in terms of the maximum number of words (sequence of numbers, letters or words) from a list that one can recall. To measure participants’ working memory, a word memory span test was used which consisted of a set of 36 short lists of basic Mandarin words of the sort used in frequently
daily life (see Appendix 2). The first four lists containing three words were followed by four containing four words, and the number of per words per list increased in this fashion until the last four lists, which each contained nine words. The length of each list at each level contained from one to three syllables. Short video-recordings were made at the Universitat Pompeu Fabra of the author saying these word lists in Chinese, and the recordings were then embedded in a PowerPoint presentation for playback online in an application that would record the test-takers’ output and save the file for subsequent analysis. The score obtained by a participant for working memory corresponded to the maximum number of words s/he could recall after repeating four different lists with that number of words.

b **Speech imitation test.** As suggested by Nardo and Reiterer (2009), a talent for speech imitation and pronunciation skills in a foreign language may be highly interdependent. To test the participant’s ability to spontaneously imitate non-native speech, we created an imitation task with target short sentences in six languages, all completely unknown to the participants, namely German, Hebrew, Tagalog, Russian, Turkish and Greek. A total of 12 sentences of different lengths (between 6 and 12 syllables; 2 in each language) were recorded in a soundproof room by native speakers of those languages. The recordings were converted to .mp3 files and uploaded in an online survey platform, which automatically played them back in random order. Participants were asked to listen to each sentence twice and repeat it immediately while the program recorded their performance. The first author of this paper rated participants’ oral productions by comparing them with the native pronunciation of the target word on a scale from 1 (very different) to 7 (no difference at all).1

c **Musical ability: Discrimination tests for rhythm and accent.** Following Law and Zentner (2012), musical abilities were assessed using the online Rhythm and Rhythm Accent perceptive subtests from the PROMS (Profile of Music Perception Skills) test. In these two subtests, participants were asked to listen twice to a short audio file featuring a particular sequence of music. They then listened to another audio clip twice and were asked to decide if the second sequence exhibited the same rhythm as the first one or not. The standard Rhythm subtest consists of simple musical patterns with constant intensities over two bars. The comparison stimuli have one or more notes added or subtracted. The Rhythm Accent subtest assesses skills in discerning the relative emphasis given to certain notes in a rhythmic pattern. Whereas the rhythms of the two sequences remain identical, the pattern of accents may be varied in the comparison stimulus by reducing the intensity of particular notes. The program itself generated a combined score, with a maximum of 18 indicating a strong ability to discriminate between rhythms and accents in music.

4 **Procedure**

Participants were first randomly assigned to one of the two conditions, namely the clapping group ($n = 25$, 8 females) and the non-clapping group ($n = 25$, 8 females). They were then tested individually using a laptop computer in a quiet room at Zhangqiu
Experimental Middle School under the supervision of a research assistant, a female teacher at the school who was fully informed about the experimental procedure to be followed. As noted above, before starting the experiment participants were asked to sign an informed consent form and fill out a language and musical background questionnaire.

The same experimental procedure was followed for each participant, with each individual session lasting approximately 45 minutes. The sequence of tasks is illustrated in Figure 4 below. First, working memory and speech imitation skills were measured by means of the instruments described above. Second, the pretest was followed by the training session and then the posttest. Finally, the two PROMS musical subtests were administered.

a **Control measures: Working memory and speech imitation.** Participants were asked to follow the link to the online survey platform. Once there, they initiated the working memory task, which involved listening to the recordings of word lists as described above and then attempting to replicate each list. The research assistant was responsible for assigning them a score as explained in section 2.3.1. The test lasted around 5 minutes.

As soon as they had finished the memory span test, participants moved on to the speech imitation test by means of a new link. As described above, this involved listening to short sentences in foreign languages and attempting to replicate them all the while being recorded. This data was scored later as described above by the first author. This test also lasted around 5 minutes.

b **Pretest.** As noted above, the procedure followed for the pretest and posttest was identical and consisted of an online imitation test (see section 2.2.2) which participants accessed by opening a link on the survey online platform. Once they had accessed the site, their audio output was recorded. Having first silently read an explanation of the task in Chinese, they clicked on each audio file twice and repeated the words. For both pretest and posttest, the order of the items was randomized. A total of 700 experimental responses were obtained (50 participants × 14 items). The pretest lasted about 7 minutes.

c **Training session.** Depending on which experimental group the participant had been assigned to, the research assistant then launched either the clapping or the non-clapping training video. Participants first watched an introduction to the task recorded by a native Mandarin speaker (the first author). Then they were given two trials to familiarize them with the task (these involved the words *pizza* and *éléphante*). After watching and listening to each item, participants either repeated the words while clapping their hands (clapping condition) or only repeated the words (non-clapping condition). Figure 5 shows still images of participants recorded during the training session. Regardless of the group, each participant saw a total of 40 videos, always presented randomly. The training session lasted around 10 minutes.

d **Posttest.** This was identical to the pretest. Here again, a total of 700 experimental responses were obtained (50 participants × 14 items), and this part of the procedure likewise lasted about 7 minutes.
e **Musical ability tests.** By means of yet one more link, participants accessed the rhythm and accent sub-tests (for details see 2.3.3 above). After they completed the test, the results were uploaded to the system automatically and exported. The part of the procedure lasted around 10 minutes.

5 **Pronunciation ratings**

As noted, a total of 1,400 recordings were obtained (700 from the pretest and 700 from the posttest). The quality of participants’ pronunciation of French in these recordings was rated by two female native French speakers (mean age = 34, SD 4.2), both PhD students from the department of Translation and Language Sciences at the Universitat Pompeu Fabra. One rater was one of the two instructors and the other was not.

Each rater played back and evaluated all 1,400 recordings by means of an online survey platform. The raters were asked to evaluate the accentedness of the words on a scale.
from 1 to 7, where 1 corresponded to ‘not accented’ and 7 indicated ‘extremely accented’, as illustrated in Figure 6. For each item, the raters first listened to the word pronounced by a native speaker, and then to two oral productions, which corresponded to the pretest and posttest renditions of the word as produced by a single participant, though pretest and posttest versions were ordered randomly. The application allowed them to play any audio file as many times as they wished. The raters reported that the entire rating process took them about 4.5 hours to complete.

Inter-rater reliability was assessed with Cohen’s Kappa for each pre- and posttest item. This procedure was realized with the program SPSS version 23. The score of Kappa was .791, clustering the rating scores by each one rating point difference of both the rater (as the measurement of the rating was with a large scale). The result is interpreted as substantial agreement (McHugh, 2012).

6 Acoustic analysis

In order to detect any improvement in the acoustic realization of rhythm and specifically the expected lengthening of the final syllable in French, the first author of the study labeled all the sound files by segmenting the initial and final boundaries of each syllable using Praat (Boersma & Weenink, 2013). As Figure 7 shows, the following three tiers were created: (1) an orthographic tier, (2) the starting and end points of each syllable,
including the final syllable (labeled $fs$ in Figure 7, see below), and (3) the starting and end points of the rhyme of the final syllable (the vowel followed by the final consonant; labeled $vc$ in Figure 7).

To extract the duration of the audio files, we used a modified version of the script by Dan McCloy (original version by Mietta Lennes) which extracts duration, f0, f1 and f2 data of different tiers. Three durational measures were collected, namely word duration (in seconds), duration of the final syllable (in seconds) and duration of the final rhyme (in seconds). For purposes of analysis, they were then transformed into two relative measures, namely (1) relative duration of the word-final syllable (i.e., duration of the final syllable divided by total word duration); and (2) relative duration of the word-final vowel-consonant rhyme (i.e., duration of the rhyme divided by total word duration).

7 Statistical analyses

Two sets of General Linear Mixed Model analyses (henceforth GLMM) were run with the data using IBM SPSS Statistics 23. The first GLMM was applied to assess the effect of the type of training (clapping vs. non-clapping) on participants’ accentedness. The average of the two raters’ scores for each item was computed and used as an index of the variable Accentedness for subsequent statistical analysis. Accentedness was set as the dependent variable. Group (two levels: non-clapping vs. clapping), Session (two levels: pretest and post-test), Group*Session, Relatedness (two levels: trained word vs. new word) and Group*Session*Relatedness were set as fixed factors. The control measures, WorkingMemory (scaled from 3 to 6), SpeechImitation (scaled from 1.25 to 4.5) and MusicalAbility (scaled from 0 to 18) were set as fixed factors, as were their interactions. One random effects block
was specified, in which we controlled for Participant and Item intercepts. Effect sizes were calculated from the means and SDs following Cohen’s (1988) model.

The second set of GLMMs was applied to assess the effect of the type of training (clapping vs. non-clapping) on participants’ acoustic measurements. In total, two GLMMs were run separately with the following dependent variables: \( fsd \) (final syllable duration proportion) and \( vcd \) (vowel consonant duration proportion). Fixed factors and random factor were the same as in the GLMM for accentedness ratings.

III Results

First, a set of independent samples t-tests were run to check that the two between-subjects (non-clapping/clapping) groups were not statistically different. The characteristics compared and the t-test results were as follows: (1) Age in days: \( t(42) = –.381, p = .705 \); (2) SentenceImitation: \( t(48) = .194, p = .847 \); (3) WorkingMemory: \( t(48) = .518, p = .607 \); (4) MusicalAbility: \( t(48) = –.256, p = .799 \); and (5) MusicExperience: \( t(48) = .760, p = .451 \). These results confirm that the two groups were similar in terms of these characteristics.

1 Accentedness ratings

Figure 8 shows the mean accentedness ratings across the two conditions, Group (non-clapping and clapping) and Session (pretest and posttest). Regarding descriptive measures, accentedness ratings improved more in the clapping condition (mean improvement .311) than in the non-clapping condition (mean improvement .168). Results of the GLMM analysis run with the accentedness ratings showed a significant main effect of Session (\( F(1, 2484) = 24.920, p < .001 \)) and WorkingMemory (\( F(3, 2484) = 2.976, p < 0.05 \)) as well as a marginal significant effect of Group*Session (\( F(1, 2484) = 2.879, p = .090 \)). Effect sizes for the clapping group were higher (\( d = 0.362 \)) than for the non-clapping group (\( d = 0.178 \)).

2 Duration analyses

In this section, we report on the results of two GLMM models that were applied to the following measures: (1) relative duration of the word-final syllable (i.e., duration of the final syllable divided by total word duration); and (2) relative duration of the word-final vowel-consonant rhyme (i.e., duration of the rhyme divided by total word duration).

a Relative duration of the word-final syllable. Figure 9 shows the mean relative duration of the final syllable across the two conditions, Group (non-clapping and clapping) and Session (pretest and posttest). Regarding descriptive measures, relative duration measures were higher in the clapping condition (mean .04) than in the non-clapping condition (mean .015). Results of the GLMM model with relative duration of the word-final syllable as a dependent variable revealed a significant main effect of Session (\( F(1, 2375) = 36.813, p < .001 \)) and two two-way interactions, between Group* Session (\( F(1, 2375) = 26.306, p < .001 \)) and Group*MusicalAbility (\( F(5, 2375) = 4.806, p < .001 \)). Post-hoc analyses of the interaction Group*Session can be assessed in two ways. First, a significant difference was found between the two groups (clapping vs. non-clapping) in the posttest (\( F(1, 2375) = 5.953, \))
p < .05), while no significant difference was found in the pretest ($F(1, 2375) = 0.603, p = .438$). On the other hand, a significant difference was found in the clapping group between pretest and posttest sessions ($F(1, 2375) = 58.609, p < .001$), but no difference was found in the non-clapping group ($F(1, 2375) = 0.473, p = .492$).

b **Relative duration of the word-final rhyme.** Results of the GLMM model with relative duration of the word-final rhyme as a dependent variable revealed a significant main effect of Session ($F(1, 2376) = 64.584, p < .001$) and a two-way interaction of Group*Session ($F(1, 2376) = 14.684, p < .001$). However, the post-hoc analyses revealed no differences between the groups in terms of their pretest and posttest scores for this parameter. See Table 3.

**IV Discussion and conclusions**

In this study, we aimed to explore whether hand-clapping to the rhythm of newly learned words in a foreign language (French) may benefit adolescent Chinese students in learning the pronunciation of those words. The results showed that the Chinese adolescents who participated in the hand-clapping training group obtained greater benefit for their pronunciation of newly learned French words than the students who participated in the non-clapping
training group. Regarding accentedness ratings, although the interaction Group*Session was only near-significant, an analysis of effect sizes revealed that the improvement in accentedness scores was higher in the clapping group than in the non-clapping group. Importantly, a detailed acoustic analysis of the data showed that the Chinese adolescents in the clapping condition improved significantly more than the participants in the non-clapping condition with respect to the durational realization of the last stressed syllable of the word (i.e., by lengthening significantly more the duration of the syllable).

While the impact on pronunciation improvement of this 10-minute training activity with hand-clapping was strong from an acoustic (durational) point of view, it was less

**Figure 9.** Mean relative duration of the final syllable across the two conditions, Group (non-clapping and clapping) and Session (pretest and posttest). Note. Error bars indicate standard error.

**Table 3.** Proportion of the duration of the final syllable (fs) and the final rhyme (vc) in relation to total word duration.

<table>
<thead>
<tr>
<th>Session</th>
<th>Percentage-fs (%)</th>
<th>Percentage-vc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clapping group:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>54.485</td>
<td>73.236</td>
</tr>
<tr>
<td>Posttest</td>
<td>57.267</td>
<td>75.886</td>
</tr>
<tr>
<td><strong>Non-clapping group:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretest</td>
<td>53.295</td>
<td>71.020</td>
</tr>
<tr>
<td>Posttest</td>
<td>53.529</td>
<td>72.563</td>
</tr>
</tbody>
</table>
strong in terms of native-speaker ratings of participant accentedness. A potential explanation for the relatively low effect on pronunciation ratings of pronunciation may have to do with the fact that participants were required to cope simultaneously with learning the meaning of new French words as well as with their phonetic realization. Their attention during training may thus have been directed more toward learning the meaning of the new vocabulary items rather than learning how to pronounce them. We hypothesize that the cognitive demands placed on semantic processing added a supplementary difficulty to the task, potentially diminishing the impact of training. Thus, more work is needed to assess the effects of rhythmic tasks with materials that are more semantically transparent for L2 learners, or with more experienced learners.

In this study, we extend previous findings on the benefits of rhythmic training sessions, which have been proven useful to enhance speech processing (see, amongst others, Cason & Schön, 2012; Cason et al., 2015) to the domain of second language phonological processing. Our results support the findings by Fischler (2009), Wang et al. (2016), and Gluhareva and Prieto (2017) on the beneficial effects of rhythmic training in the context of second language pronunciation learning by extending them to the use of hand-clapping. Our study shows that, apart from boosting students’ phonological awareness and reading skills in their native language (Bhide et al., 2013), short rhythmic hand-clapping activities can be especially beneficial in the context of second language pronunciation instruction. In this context, the findings of this study provide additional support for the importance of an explicit, global, suprasegmental approach to L2 pronunciation instruction (Behrman, 2014; Derwing & Rossiter, 2003; Derwing et al., 1998; Gordon et al., 2013).

From an educational point of view, we suggest that the use of hand-clapping activities can be not only a useful tool but also a fun classroom technique for teachers of a second language (see also Romero Naranjo & Romero Naranjo, 2013). The clapping technique could be implemented in combination with hand beat gestures to emphasize higher-level prominence patterns (Gluhareva & Prieto, 2017). All in all, hand-clapping to the rhythm of new words in a second language seems to be an effective tool to improve second language pronunciation, at least at an early stage of learning. Besides addressing the need to more carefully disentangle semantic processing from pronunciation noted above, future research in this line should explore the effect of clapping-based training on more advanced learners.

Acknowledgements

We thank the students at Zhang Qiu Experimental Middle School who participated in the study. We are grateful to J. Li and J. Borràs-Comes for their assistance and to C. Monnier for her help recording the items. Also, I would like to express our gratitude to J.C. Mora, L. Viladot, J. Carrera and F. Clement for all of their helpful comments which improved the quality of the research.

Funding

This research was supported by funding from the Spanish Ministry of Economy and Competitiveness (FFI2015-66533-P) and the Generalitat de Catalunya projects 2014 SGR-925 and 2017 SGR _ 971).

Note

1. Since the purpose of the task is not to assess pronunciation in second language but rather assess speech imitation abilities, we believe that the fact that the rater is not a native speaker
of these languages may be considered as an advantage. Indeed, by comparing the model sentence to the participant’s production, the rater is able to concentrate strictly on the phonological differences between the two, without having a personal model on how the speech string sounds like. Moreover, the rater is a linguist specialized in phonetics and phonology who can be considered as highly qualified to carry out such task.

References


**Appendix 1.**

*Language and music background questionnaire (English translation).*

**Language background**

1. What language do you speak at home?
2. Do you speak a second language in your daily life? Yes or No.
3. Have you ever studied French?
4. How long have you been studying English?
   a. Less than 1 year
   b. 2 to 3 years
   c. 3 to 5 years
   d. More than 5 years
   e. I have never studied English
5. If you have studied English, how much do you speak it every day?
6. Have you ever studied any other foreign languages? If so, please indicate the language and your mastery level.

Music background

1. Have you had any training in music outside obligatory music lessons at school?
2. Do you have any training in singing outside obligatory music lessons at school?
   If you answered ‘yes’ to 2, answer questions 3 and 4.
3. How long have you had training in singing?
   a. less than half a year
   b. half a year to one year
   c. one to two years
   d. three to five years
   e. above five years
4. How many hours do you practice singing every day?
   a. 1 to 3 hours
   b. 3 to 6 hours
   c. 6 to 9 hours
   d. more than 9 hours
   e. I don’t practice.
5. Do you play a musical instrument?
   If your answer is ‘yes’, answer questions 6–9.
6. What instrument do you play?
7. How old were you when you started playing the instrument?
8. How long have you played the instrument?
   a. less than half a year
   b. half a year to one year
   c. one to two years
   d. three to five years
   e. above five years
9. How many hours do you practice the instrument every day?
   f. 1 to 3 hours
   g. 3 to 6 hours
   h. 6 to 9 hours
   i. more than 9 hours
   j. I don’t practice.
### Appendix 2. Working memory test word lists.

<table>
<thead>
<tr>
<th>Number of words in the list</th>
<th>Words in Chinese</th>
<th>Total number of syllables in the list</th>
<th>Translation into English</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>胡子,果汁,表弟</td>
<td>6</td>
<td>mustache, juice, cousin</td>
</tr>
<tr>
<td></td>
<td>圆珠笔,香蕉,门</td>
<td>6</td>
<td>ball-pen, banana, door</td>
</tr>
<tr>
<td></td>
<td>椅子,脖子,纸张</td>
<td>6</td>
<td>chair, neck, paper</td>
</tr>
<tr>
<td></td>
<td>未婚夫,蛋糕,眼镜</td>
<td>7</td>
<td>fiancé, cake, glasses</td>
</tr>
<tr>
<td>4</td>
<td>房子,语言,电话,苹果</td>
<td>8</td>
<td>house, language, telephone, apple</td>
</tr>
<tr>
<td></td>
<td>鞋子,吉他,太阳,厨房</td>
<td>8</td>
<td>shoes, guitar, sun, kitchen</td>
</tr>
<tr>
<td></td>
<td>葡萄,铅笔,帽子,草莓</td>
<td>8</td>
<td>grape, pencil, hat, strawberry</td>
</tr>
<tr>
<td></td>
<td>蜜蜂,脚,小丑,书本</td>
<td>7</td>
<td>bee, foot, clown, book</td>
</tr>
<tr>
<td>5</td>
<td>国王,单词,工作,电,女孩</td>
<td>9</td>
<td>king, word, job, electricity, girl</td>
</tr>
<tr>
<td></td>
<td>文化,武器,例子,水,时间</td>
<td>9</td>
<td>culture, weapon, example, water, time</td>
</tr>
<tr>
<td></td>
<td>原因,道路,水平,命令,公里</td>
<td>10</td>
<td>reason, road, level, order, kilometer</td>
</tr>
<tr>
<td></td>
<td>心脏,社会,现实,帮忙,人类</td>
<td>10</td>
<td>heart, society, reality, help, human</td>
</tr>
<tr>
<td>6</td>
<td>周期,期限,来源,情况,世纪,点</td>
<td>11</td>
<td>period, limit, origin, situation, century, point</td>
</tr>
<tr>
<td></td>
<td>国王,嘴,概念,颜色,血,仪式</td>
<td>10</td>
<td>king, mouth, concept, color, blood</td>
</tr>
<tr>
<td></td>
<td>知识,科技,地方,大海,戏剧,游戏</td>
<td>12</td>
<td>knowledge, technology, location, sea, drama</td>
</tr>
<tr>
<td></td>
<td>意愿,位置,健康,注意力,关系,性格</td>
<td>13</td>
<td>volunteer, situation, health, concentration, relationship, character</td>
</tr>
<tr>
<td>7</td>
<td>身体,数量,地址,国家,世纪,行动,丈夫</td>
<td>14</td>
<td>body, amount, address, country, century, action, husband</td>
</tr>
<tr>
<td></td>
<td>照相机,单元,战争,意识,位置,小时,点</td>
<td>14</td>
<td>camera, unit, war, conscious, location, hour, point</td>
</tr>
<tr>
<td></td>
<td>协议,重要性,活动,阴影,年龄,图像,街道</td>
<td>15</td>
<td>agreement, importance, activity, shadow, age, image, street</td>
</tr>
<tr>
<td></td>
<td>脚,钱,问题,功能,瞬间,木材,危险</td>
<td>12</td>
<td>foot, money, question, function, moment, wood, danger</td>
</tr>
<tr>
<td>Number of words in the list</td>
<td>Words in Chinese</td>
<td>Total number of syllables in the list</td>
<td>Translation into English</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------------</td>
<td>--------------------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>8</td>
<td>山，关系，教堂，火，喜欢，存在，空间，纸</td>
<td>13</td>
<td>mountain, relation, church, fire, like, existence, space, paper</td>
</tr>
<tr>
<td></td>
<td>作者，系统，花，问题，思想，语言，时间，情况</td>
<td>15</td>
<td>author, system, flower, question, mind, language, time, situation</td>
</tr>
<tr>
<td></td>
<td>表达，单词，时期，水，法律，石头，影响，城市</td>
<td>15</td>
<td>expression, word, period, water, law, stone, influence, city</td>
</tr>
<tr>
<td></td>
<td>爱，瞬间，原则，方面，案件，真实，善良，葡萄酒，生产</td>
<td>16</td>
<td>love, moment, principle, aspect, case, reality, modality, wine, produce</td>
</tr>
<tr>
<td>9</td>
<td>元素，然而，人民，月亮，空气，脖子，过程，政府，移动</td>
<td>18</td>
<td>element, however, population, moon, air, neck, process, government, move</td>
</tr>
<tr>
<td></td>
<td>小组，精神，节日，历史，城镇，安静，技巧，大部分，床</td>
<td>18</td>
<td>group, spirit, festival, history, city, silence, skill, majority, bed</td>
</tr>
<tr>
<td></td>
<td>家庭，灵魂，因此，人民，语言，经验，手，皮肤，场景</td>
<td>17</td>
<td>family, soul, thus, people, language, experience, hand, skin, scene</td>
</tr>
<tr>
<td></td>
<td>服务，树叶，夜晚，学习，脚，点子，本性，阶级，次数</td>
<td>17</td>
<td>service, leaf, night, study, foot, idea, nature, class, times</td>
</tr>
</tbody>
</table>
Speech imitation task sentences

1. Greek
   a. To proi pino himo portokali.
      English: In the morning, I drink orange juice.
   b. Hriazome diakopes.
      English: I need holidays.

2. Russian
   a. Mi rabotaem v ofise.
      English: We are working in the office.
   b. Eta gazeta lezit na stole.
      English: This newspaper is on the table.

3. German
   a. Spinat hat mir noch nie geschmeckt.
      English: Spinach has never tasted well to me.
   b. Ich möchte eine Schachtel Schokolade.
      English: I want a box of chocolates.

4. Tagalog
   a. Ano ang pangalan mo?
      English: What is your name?
   b. Pupunta ako sa tindahan.
      English: I will go to the store.

5. Hebrew
   a. šalom. šmi alon v'ani talmid.
      English: Hello. My name is Alon and I am a student.
   b. Hašom łom lefe, všahjemež zošvχat.
      English: today is a beautiful day, and the sun is shining.

6. Turkish
   a. Özge ona çarpılmıştı.
      English: Özge had been lovestruck by him.
   b. Ali hayır dedi.
      English: Ali said no.