THE BENEFITS OF MUSIC-BASED ACTIVITIES ON LEARNING PRONUNCIATION OF A SECOND LANGUAGE

YUAN ZHANG

Supervisor: Dr. PILAR PRIETO VIVES
ABSTRACT

Previous studies have reported that music may enhance language learning (e.g., Kao & Oxford, 2014; Besson, Chobert & Marie, 2011; De Groot, 2006). Researchers investigated how musical expertise, musical aptitude, and musical training may benefit language acquisition (e.g., Magne, Schön, & Besson, 2006; Pastuszek-Lipińska, 2007; Parbery-Clark, Tierney, Strait, & Kraus, 2012 for musical expertise; Milovanov, Pietilä, Tervaniemi, & Esquef, 2010 for musical aptitude; Tierney & Kraus, 2013; François, Chobert, Besson, & Schön, 2013 for musical training). However, little is known about how music may enhance second language pronunciation. Even though there is growing evidence that rhythmic and melodic training can be beneficial for different aspects of language development (e.g., speech segmentation, phonological skills, memorization, reading skills, etc.) in the native language (Schön, Magne, & Besson, 2004; see Besson, Chobert, Astésano, & Marie, 2015 for a review), less is known about their role in second language development. This PhD project will explore the potential beneficial effects of musical activities on learning second language pronunciation through three experimental studies involving Chinese adolescents (13 to 15 years old).

Based on the evidence that rhythm is one of the shared features of music and language, the first study consists of a short between-subject training with a pretest and posttest design. Following up on work that shows that rhythmic training can be beneficial for phonological processing and also for second language speech processing (Fischler, 2009; Wang, Mok, & Meng, 2016; Gluhareva & Prieto, 2017), the first study tests whether hand clapping to the rhythm of newly learned French words benefit the acquisition of pronunciation for Chinese adolescents. Fifty adolescent participants from China taught twenty target French words, where participants were divided randomly into two conditions: with or without hand clapping instructions. The improvement of pronunciation from pretest and posttest was assessed and compared across groups. Acoustic analyses showed that the clapping condition group got an overall greater accuracy in the production of French syllable-final duration. This finding suggests that the training made the prosodic structure of French syllables more salient for learners.
The second study will examine the effects of training with a musical instrument which highlights rhythm on pronunciation at the discourse level in a reading task. No previous studies have explored empirically how an instrument-based short rhythmic training may influence second language pronunciation. Observing and producing co-speech beat gestures has been proven to enhance speech fluency, reduce accent and aide in learning vocabulary (Gluhareva & Prieto, 2017; Kushch, 2018). However, no previous studies have investigated how a rhythmic training with instruments in conjunction with speech affect pronunciation acquisition. The study is a short between-subject training with a pre- and posttest design. Fifty adolescents will be assigned to two groups randomly, either playing an easy to handle instrument (in this study, a Chinese drum) or listening to instructions about English prosody in their native language during the training. The aim is to test for potential gains from the short instrument training on second language pronunciation.

The third study will be a classroom training study that will assess the effects of singing songs on pronunciation learning in a second language. Given the contradictory findings of previous studies regarding the use of songs on second language learning (e.g., for positive evidence see Barreiro, Estebas-Vilaplana, & Soto, 2005; Fischler, 2009; Nakata & Shockey, 2011; Toscano-Fuentes & Fonseca-Mora, 2012; Ludke, Ferreira & Overy, 2014; Tizian, 2016; Kilgour, Jakobson & Cuddy, 2000; Rukholm, 2011; Ludke, Ferreira & Overy, 2014; Yousefi, Yekta & Farahmandian, 2014; for negative evidence see Nemoto, Wilson, & Perkins, 2016; Racette & Peretz, 2007), this classroom training study will test whether listening to songs and singing may enhance English pronunciation by lower-intermediate Chinese students. The training involves an 8-session training course over the period of one month with a pretest and posttest between-subject design. One hundred teenagers will be randomly assigned to two training groups which will use the same texts, with one group listening to and singing songs (singing group) and the other group carrying out speech-only activities. Their vocabulary and pronunciation skills will be assessed before and after training.

With the goal of assessing whether the L2 speech and cognitive skills of the participants correlated with musical skills (musical aptitude, musical expertise, or experience), a set of control measures will be assessed for each participant
individually and for each study. We will control for imitation and working memory capacities, rhythm and melody abilities, and music and language background.
ABSTRACT ................................................................................................................................. i

1. INTRODUCTION .................................................................................................................. 1

1.1 OBJECT OF INVESTIGATION .............................................................................................. 1

1.2 STATE OF THE ART ................................................................................................................. 2

1.2.1 Music and language in the brain .................................................................................. 2

1.2.2 The effects of musical expertise on language learning ............................................. 3

1.2.3 The effects of musical aptitude on language learning .............................................. 4

1.2.4 The effects of musical training on language learning .............................................. 5

1.3 MUSIC-BASED ACTIVITIES AND THE TEACHING OF L2 PRONUNCIATION ................. 5

1.3.1 Rhythmic training and language learning ................................................................. 6

1.3.1.1 Rhythmic activities and second language pronunciation learning ...................... 8

1.3.2 Songs used in L2 learning and teaching ................................................................. 11

1.4 OBJECTIVES OF THE PHD THESIS ............................................................................... 14

1.5 THEORETICAL FRAMEWORK(S) ...................................................................................... 16

1.5.1 Music and Language .................................................................................................... 17

1.5.2 Embodied cognition ..................................................................................................... 17

1.5.3 Cross-domain auditory plasticity ............................................................................... 18

1.5.3.1 Music rhythm and speech rhythm ......................................................................... 19

1.5.3.2 Songs and cross-domain plasticity ......................................................................... 19

1.6 HYPOTHESES .................................................................................................................... 20

2. RESEARCH METHODOLOGY ............................................................................................. 22

2.1.1 Working memory ........................................................................................................... 22

2.1.2 Speech imitation ............................................................................................................ 22

2.1.3 Musical ability ................................................................................................................. 23

2.1.3.1 Discrimination tests for rhythm and accent ........................................................ 23

2.1.3.2 Discrimination tests for melody and rhythm-to-melody ....................................... 23

2.2 STUDY 1 .............................................................................................................................. 24

2.2.1 Aim of the study ............................................................................................................ 24

2.2.2 Methodology ................................................................................................................ 25

2.2.2.1 Participants ............................................................................................................. 25

2.2.2.2 Materials ................................................................................................................ 26

2.2.2.2.1 Audiovisual materials for the training session ................................................. 26
APPENDIX A. .............................................................................................................. 64
APPENDIX B.................................................................................................................. 65
APPENDIX C. .................................................................................................................. 67
1. **INTRODUCTION**

In this section, I will first explain the object of study for this PhD project. Second, I present an overview of previous studies in the fields of music and first and second language learning. In addition, several concepts that are related to this field of investigation will be introduced, such as musical expertise, musical aptitude, and musical training, as well as their interaction with language, first language acquisition and second language acquisition.

1.1 **Object of investigation**

Recent studies have shown a growing interest in investigating the relationship between music and language (e.g., Patel, 2010). Music and language share several common prosodic features like rhythm and melody. While previous studies have shown that long-term musical training is beneficial for language learning, less is known about its effects on second language learning. In the foreign language classroom, there is a growing tendency to incorporate musical activities in order to make the lessons more active and enjoyable (Gass & Selinker, 2001). This strategy has promoted a belief that music may have positive effects on learning a language. Language teachers tend to use music-based activities in their teaching plans, such as listening to songs in the target language, learning to sing the songs with lyrics and so on (e.g. Richard-Amato, 1988). Teachers suggest that these activities promote motivation and help vocabulary memorization, increasing learners’ attention and learning effectiveness (e.g., Kao & Oxford, 2014; Kang & Williamson, 2014; Besson, Chobert & Marie, 2011; De Groot, 2006; Fonseca Mora & Gant, 2016).

However, only a few classroom studies have empirically investigated the potential effects of using music-based activities in the second language classroom and thus, more empirical evidence is needed. In this PhD project, a set of three studies will explore the effects of different kinds of rhythmic and melodic training on language learning and specifically on pronunciation learning. These studies will also control for the potential effects of individual factors related to musical expertise, musical aptitude, language imitation abilities, and working memory.
1.2 State of the Art

1.2.1 Music and language in the brain

Music and language have been shown to be tightly linked in the brain. Neuroimaging studies suggest that music shares common neural networks with language (e.g., Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Koelsch, Gunter, Friederici, & Schröger, 2000; Koelsch et al., 2002; Besson & Schön, 2001). For instance, Besson and Schön (2001) showed in an ERP experiment that when people listen to music and speech, their brain reacts with similar expectations to the correctly produced prosody or lyrics and language sentences. By contrast, when the stimuli were sentences with incorrect syntax, or pieces of music with the wrong melody, the brain responses would also have a similar pattern. This suggests that music and language may interact during integration. According to Zatorre (2005), some functions of music and speech processing such as syntax may require common neural resources, as shown by neuroimaging studies, indicating that music performance, listening, and composition are embedded with cognitive abilities, and thus enhance cerebral plasticity and language abilities.

In cognitive science, the links between music and language learning have been studied from different perspectives. Some studies have revealed that better music skills might lead to better language abilities. Musicians have been shown to possess a greater capacity for acquiring new language (see sections 1.2.2 and 1.2.3 below). For instance, in speech perception, the experiment by Parbery-Clark et al. (2012) revealed that musicians performed better than non-musicians in a speech discrimination task. On the other hand, the role of rhythmic training has been shown to have potential benefits for people with dyslexia (e.g., Huss, Verney, Fosker, Mead, & Goswami, 2011; Corriveau & Goswami, 2009). Furthermore, music itself has been shown to be effective in reducing anxiety for people (e.g., White, 1992; Doğan, & Şenturan, 2012), which in turn may help second language perceiving and producing.

The tight relationship between language and music has been investigated in different areas. In the following sections, we will discuss how musical expertise (generally understood as the expertise in regularly playing an instrument or singing for more than two years with or without instruction; see Chobert & Besson, 2013; Sloboda, 2005),
**musical aptitude** (which refers to innate personal musical talent and is generally assessed by a set of discrimination tests of rhythm, pitch, melody, etc.; see Gordon, 1965), and **musical training** (referring to relatively short training of up to two years; see Chobert et al., 2014) influence language abilities.

### 1.2.2 The effects of musical expertise on language learning

Evidence has shown that first language (L1) learning may benefit from musical expertise. The study of Parbery-Clark et al. (2012) has summarized the behavioral and neurological effects of musical expertise on native language proficiency.

Previous studies have investigated the positive effects of musical expertise on phonological abilities such as speech perception, both in the first and second language. Musical expertise may facilitate lexical tone processing and segmental processing (Marie et al., 2011). A series of experimental studies carried out by Besson and her colleagues have shown that musicians are more sensitive to speech perception (Besson, Schön, Moreno, Santos, & Magne, 2007; Magne et al., 2006; Marques, Moreno, Castro, & Besson, 2007; Moreno & Besson, 2006; Schön et al., 2004). Schön, Magne, and Besson (2004) tested 18 French adult musicians and non-musicians on their ability to detect pitch violations in their native language. They found that musicians outperformed non-musicians in pitch violation detection. Moreover, musicians were found to be more sensitive at detecting weak pitch violation than non-musicians. Similar results were found by Magne et al. (2006) using the same experimental design in children.

Regarding pronunciation abilities in a second language, Pastuszek-Lipińska (2007) found evidence that musicians perform better than non-musicians in speech imitation. One hundred and six Polish native speakers participated in the experiment and were divided into two groups, musicians and non-musicians, as determined by a questionnaire about their music background. Their imitation ability for several foreign languages including Dutch, French, Italian, Spanish, and Japanese was tested while examining different components of phonology, syntax and intonation. Results showed that musicians globally performed better experiment and were able to correctly repeat more sentences in the target language and make fewer errors. Gottfried, Staby, and Ziemer (2004) tested 43 American students without any knowledge of a tonal
language, from which 25 were enrolled at the conservatory of music. They were asked to perform two tasks, a discrimination task (e.g., tell whether the target pair of Mandarin tones were the same or different) and an imitation task (e.g., imitate Mandarin l-vowel syllable, as well as target syllables in four tones). Results showed that the conservatory students were able to not only discriminate but also imitate the four Mandarin tones more accurately than the non-conservatory students.

1.2.3 The effects of musical aptitude on language learning

Previous studies indicate that musical abilities facilitate language skills by increasing attention, memory, creativity, and vocabulary (e.g., Hill-Clarke & Robinson, 2003; Silva, 2006; Paquette & Rieg, 2008). A study by Strait, Hornickel, and Kraus (2011) tested the L1 (English) reading skills of 42 children (mean age, 10.4 years) and found that regardless of music experience, children with higher musical aptitude performed better in silent reading, auditory memory and attention.

Similarly, musical aptitude allows for learning a second language more efficiently. Evidence shows that musical aptitude correlates with L2 phonological production and perception capacities, both for adult and children population (e.g., Slevc and Miyake, 2006; Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008). Milovanov et al. (2010) examined the pronunciation of English as a second language by 3 groups of 46 adults, namely non-musicians, choir members, and English philology students. While all the participants performed equally in a phonemic discrimination task, those with a higher musical aptitude had better English pronunciation.

Interestingly, a recent study by Christiner and Reiterer (2013) showed that good singing performance predicts good speech imitation. Forty-one adult German native speakers and singers with different levels of singing proficiency (from beginners to advanced singers) participated in their study. Their singing and musical ability, speech imitation and working memory were assessed. Results indicated that singing performance was the best indicator of the ability to imitate speech. Regarding perception, research by Yuan, González-Fuente, Baills and Prieto (2018) investigating the effects of pitch gestures on learning Spanish intonation by Mandarin speakers found that participants with higher musical abilities performed better in learning to discriminate the main intonation features of Spanish.
1.2.4 The effects of musical training on language learning

Previous research has reported that musical training leads to modification in brain plasticity and enhances speech abilities in a first language (see Besson et al., 2015, for a review) as well as phonological awareness and reading abilities (Tierney & Kraus, 2013). As for phonological abilities, musical training may enhance speech segmentation. François et al. (2013) carried out a 2-year training study with 24 French non-musician eight-year-old children. The children were assigned to two different groups, the music-training group and painting-training group. Researchers tested the children three times (before training, after the first year and after the second year) with a speech segmentation task which consisted of choosing the item that more closely resembled the pseudo-word. Both behavioral and electrophysiological results showed that only the music-training group improved their speech segmentation skills while the painting group performed equally in these three tests.

For learning a new language, long-term musical training has proved to be beneficial for verbal memory (e.g., Chan, Ho, & Cheung, 1998). The researchers tested sixty Chinese college students, of whom 30 had previous long-term musical training and the other 30 had received no previous musical training. They assessed the participants’ verbal memory via a recall task involving a 16-word list presented orally, and they found significantly higher scores in musical-training group.

Sorouri and Sarsarabi (2016) investigated the relationship between perception and production abilities in a second language and musical abilities. Sixty young Polish students were asked to repeat short English sentences in a speech production task. With a within subject and pretest/posttest design, the students took part in 15-minute training sessions containing aspects of both language and music. Training included tasks such as listening to music, fast reading, among others. In the posttest, they were asked to discriminate different sounds (e.g. feet, fit, feet). Results of this study revealed that students with higher musical intelligence performed better in the discrimination task at posttest.

1.3 Music-based activities and the teaching of L2 pronunciation
Many second language teachers incorporate music in the academic curriculum and believe in the potential benefits that music may bring to the students. A study by Seifried (2006) explored how attending guitar lessons behaviorally affected a group of high school students in the US. The students reported experiencing feelings of fun and freedom. Despite the commonly held beliefs regarding the negative impacts of extra hours devoted to “entertainment” on academic life, students’ grades went up, perhaps due to an increase in focus and a decrease in stress brought on after the guitar classes. Kang and Williamson (2014) tested the effects of background music on language learning. They took 32 UK participants learning two new languages, Mandarin and Arabic, and divided them into two groups according to the L2 (with 16 persons per group). Participants of each group were further divided into two learning conditions: using language-learning materials with or without background music. Results showed that students in the background music group on average showed better performance in recall, translation and pronunciation posttests. Finally, music has been proven to benefit young people’s development in regards to their personality, intelligence and creativity (Hallam, 2010). Consequently, researchers place great importance on musical education as well (see Aguilera, Vallverdú, & Pérez-Moreno, 2017 for a review)

1.3.1 Rhythmic training and language learning

Rhythm is a shared feature of music and speech. Rhythm is a very important aspect of spoken language (Patel, 2003) that is used to organize the sound stream by grouping acoustic events in both language and music. Rhythm is especially important in the initial stages of language learning (e.g., Roncaglia-Denissen, Roor, Chen, & Sadakata, 2016). Newborns are particularly sensitive to language’s rhythmic components and can distinguish their native speech from a non-native one (Nazzi, Bertoncini, & Mehler, 1998; Ramus & Steger, 2000). At the age of four months, they show preference to languages that have the same rhythmic class as their native language. 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, 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 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 a short training session at several points during the experiment 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 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 also showed 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 out over six weeks. The results showed that CMT training yielded higher rates of improvement in dyslexic children relative to normal children in terms of categorical perception. Moreover, the intensive training mode produced greater improvement in a syllabic lengthening task than the six-week mode, where children were asked to tell whether tri-syllabic words were pronounced with correct syllable length (e.g., *canapé*) or with an incongruous lengthening (e.g., *caaaapé*).

1.3.1.1 Rhythmic activities and second language pronunciation learning

Little is known about whether rhythmic training activities can enhance phonological awareness or pronunciation in a second language. Most classroom pronunciation training has tended 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 (see Anderson-Hsieh, Johnson & Koehler, 1992; Derwing & Munro, 2009, for a review). Several studies have highlighted the importance of suprasegmental instruction for improving learners’ overall fluency and comprehensibility and reducing their foreign accent (see for example Derwing, Munro & Wiebe, 1998; Derwing & Rossiter, 2003; Gordon, Darcy & Ewert, 2013; Behrman,
Yet while suprasegmental training has proven to be successful in improving second language learners’ overall fluency and comprehensibility, few studies have tested the efficacy of suprasegmental training paradigms on specific pronunciation issues.

Another often overlooked aspect of language instruction is the role of body movement. Indeed, teachers tend to use gestures or body movements for suprasegmental training (e.g., Smotrova, 2017 and Baker, 2017); however, little is known about the potential effects of using gestures in the second language classroom. Smotrova (2017) pointed out that teachers applied body movements to teach the students L2 pronunciation and that such body gestures serve to enhance suprasegmental features, as word stress and speech rhythm, ultimately facilitating pronunciation. According to Baker (2014), teachers tend to apply rhythmic strategies such as clapping, tapping the desk, or foot movements for L2 pronunciation instructions.

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 et al. (2016) tested a computer application that automatically generated a percussive rhythm for any given text in English. In a within-subjects design, 20 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 the 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) reported 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.
posttest, participants that had observed the beat gestures significantly improved their accentedness ratings on the set of difficult items.

The three above-mentioned studies employed different types of procedures (a percussive rhythm generator, beat gestures) that are not necessarily easily applicable in the language classroom. By contrast, 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 and educational systems (Cameron & Grahn, 2014; see also Romero Naranjo & Romero Naranjo, 2013). However, little is known about whether using clapping activities would have beneficial effects for learning L2 pronunciation.

Our hypothesis in Study 1 of this thesis will be that clapping, an activity linked strongly 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 the metrical structure of words more salient by acoustically highlighting it with different intensity and durational 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.

Study 2 of this thesis is also related to testing more lengthy rhythmic training in the classroom. Our objective will be to use percussion musical instruments in the classroom to work on general rhythmic knowledge before practicing the rhythm of speech. A designed exercise phase will be incorporated into the training session, training the participants with percussion musical instrument. If our hypothesis is confirmed, this would be consistent with recent work by Roncaglia-Denissen et al.
(2016) in which an enhanced ability to practice 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 rhythmic knowledge in the language and musical domains. Chen, Fan and Lin (1996) suggested that when teachers tapped, clapped or played rhythmic instruments to mimic strong or weak prosodic beats, these actions helped Chinese students detect word and sentence stress in English.

1.3.2 Songs used in L2 learning and teaching

One of the easiest ways of using music in the L2 classroom is using songs. Songs are easy to be included into the lessons as compared with instruments, which are heavy, costly and take more time to manage. The potential benefits of using songs have been pointed out by language professionals. Most students report that listening to songs and singing songs is an enjoyable experience, their confidence rises, and they feel more relaxed, ultimately allowing them to pay more attention to the second language (see Medina, 2002). In the third study of this thesis, our underlying hypothesis is that using songs in the classroom may make language learning easier.

Singing can be a valuable tool for learning a foreign language. Fonseca-Mora has argued in her (2000) article that using music or singing could be an effective way to access a second language, as using music encourages the students generally with more memorable vocabulary or improvement of pronunciation skills. Abbott (2002) suggested that using songs, such as pop or rock music, could be an effective tool for teaching a second language (particularly English). Researchers and teachers suggest that songs help learning language via rhythm, melody and lyrics. Particularly, music can increase a learner’s motivation and memorization for foreign language learning. For instance, according to Patel (2010), children prefer singing to speech in infant-directed style speaking. Spicher and Sweeny (2007) provided a series of strategies to applying folk songs in second language teaching under the consideration that using music (especially folk music) may help the students to develop the oral-aural abilities.

Previous studies have suggested the positive effects of songs on L2 pronunciation acquisition (e.g., Barreiro et al., 2005; Fischler, 2009; Nakata & Shockey, 2011; Toscano-Fuentes & Fonseca-Mora, 2012; Ludke, Ferreira & Overy, 2014; Tizian,
2016) and word learning (e.g., Kilgour et al.; Rukholm, 2011; Ludke et al., 2014; Yousefi, Yekta & Farahmandian, 2014). However, negative results been found as well for L2 pronunciation (e.g., Nemoto, Wilson, & Perkins, 2016) and word learning (e.g., Racette & Peretz, 2007).

Among the set of classroom studies assessing the use of songs in the L2 classroom, Fischler (2009) provided evidence that using rhythmic activities and rap songs may enhance stress placement of English. Tizian (2016) explored the effects of songs on L2 pronunciation with 10 Italian university students. In this study, she observed a higher pronunciation improvement of the non-linguistic background participants in the experimental group, which was trained with listening to English songs. However, the control group didn’t receive original English input. Moreover, the sample size was too small to draw reliable conclusions. Toscano-Fuentes and Fonseca (2012) tested 49 Spanish sixth-graders in a one-year English learning program where the students were trained listening to and singing songs. They claimed that the students benefited from songs and improved their L2 abilities as in aspects of pronunciation, communication and comprehension. However, it is hard to assess if the improvement is benefited through songs themselves or due to the language learning that occurred over the academic year, as there was no control group.

The study by Nakata and Shockey (2011) tested how a short singing training session affected English pronunciation by Japanese speakers. As Japanese learners of English tend to insert vowels between consonant clusters in English words, the aim was to find out whether the training might reduce this problem of pronunciation. After a three month training period, the experimental group of 16 participants was found to significantly reduce the rate of inserting vowels into consonant clusters compared to a control group of 11 Japanese participants. However, it is difficult to compare the two experimental groups, as participants from the control group hadn’t received any pronunciation-related activities whatsoever. Ludke et al. (2014) showed that singing can be an effective tool to learn a foreign language. They tested 60 participants without any background knowledge of Hungarian and divided them into three groups: speaking, rhythmic speaking, and singing conditions. Participants were asked to do two tasks. The first was a production task where the participants need to reproduce the 20 phrases in Hungarian that prepared by the researchers and presented by
condition (only spoken, spoken in the rhythm of the written melody or sung in time with the metronome). The second was a delayed-recall conversation task. Only the singing group performed significantly better compared to speaking or rhythmic group. In another study, Yoshida, Nose and Ito (2014) investigated 11 Japanese speakers’ English pronunciation by creating a corpus of recordings of singing two English popular songs. They evaluated participants’ pronunciation while singing and speaking and found that more singing experience (not only singing in English) may lead to a better English pronunciation. However, they found that when singing, speakers tended to make more pronunciation mistakes, this research didn’t include a scientific statistical analysis and the sample was too small. This suggests the importance of careful consideration when designing the experiment.

In contrast, a study by Nemoto et al. (2016) found negative results in their song training group. They carried out a listening to melody and speech training with 30 Japanese adult participants divided into two groups of 15 people in each condition, singing and speech only. The material was a 14-word sentence of a selected English song Alive in a 10 minute training session. The pronunciation of four phrases of this song was evaluated which was considered generally hard to pronounce for Japanese people. They used an online service where 108 native or near-native speakers evaluated the experimental recordings and they found that the singing group performed generally worse than speech training group. In their study, the native raters were not instructed on how to do the rating, nor were they selected by the researchers. Thus, the raters were not very reliable. In addition, though there were only 14 words in the experimental singing sentence, not all of the phrases or words were evaluated for pronunciation.

Similarly, while listening to songs has been shown to be an effective tool for lexical learning (Salcedo, 2010; Rukholm, 2011; Yousefi et al., 2014), this has not been found in all studies (Racette & Peretz, 2007). For instance, Rukholm (2011) recruited sixty-six beginning Italian learners of various native languages and tested their lexical acquisition improvements. The experimental group was trained by listening to either a song or a poem, while a control group only did the through pretest and posttest. The results showed that participants exposed to songs outperformed those who exposed to lyrics only. The findings suggest that listening to songs is helpful for lexical learning,
which further supports the idea of using songs for second language acquisition. For L2 vocabulary learning effectiveness, Yousefi et al. (2014) also showed positive results. Sixty junior high school female English learners from Iran were randomly divided into two groups (music group and no-music group). The target vocabulary was one hundred unfamiliar English words included in songs or speech only. The results revealed that music group participants outperformed the control group on their vocabulary test after the training. Thus, music proved to be helpful for second language lexical learning. Salcedo (2010) compared the experimental results of two groups of US students, where one group listened to a Spanish song and the other listened to the same lyrics in a speech audio reading by a native speaker. The participants were asked to fill forms to recall the text of the lyrics. Statistical analyses showed a significant increase in the music group for the recall task. Heidari and Araghi (2015) tested the effectiveness of songs and pictures as instructional tools for Iranian children EFL learners (all males). However, participants from the picture condition outperformed in the posttest, where gender was considering as one of the consequences. In contrast, Racette and Peretz (2007) did not find a positive effect of singing songs for vocabulary learning. They tested 18 non-musicians and 18 professional musicians who were French native speakers on text and melody recall in songs with or without lyrics. They did not find that learning a song’s lyrics while listening to the melody was more effective compared to only reading the lyrics without the melodies in both non-musician and musician group.

To sum up, previous studies have shown both positive and negative effects of listening or singing songs on either L2 pronunciation or vocabulary learning. However, little has been experimentally investigated in the classroom with well-planned singing sessions that can be strictly compared with the speech conditions. In addition, few studies have taken into consideration how musical aptitude and expertise may affect their experimental results.

1.4 Objectives of the PhD Thesis

In China, English serves as a very important language. English is one of the main subjects according to the nine-year compulsory education system of China, thus most Chinese students attach great importance to learning English. Chinese students are
always facing the problem of learning English pronunciation. The pronunciation of an English word is related to its orthographic structure while Chinese is characterized by the loose orthographic-phonemic correspondence (Schmitt, Pan, & Tavassoli, 1994). This means there are few cues of pronunciation when observing the word structure. Chinese learners of English usually suffer a lot when they begin to learn English pronunciation. According to the Test Taker Performance of 2016 of IELTS (The International English Language Testing System)¹, China performed in the lower level among the 40 top places by calculating the mean score of English skills. What’s more, Chinese’s average speaking and writing scores were worse than the listening and reading scores. One may blame the modern English education in China, which puts emphasis on listening and reading skills more than speaking and writing skills.

Given that musical experience can boost L2 phonological abilities, and particularly rhythmic abilities, our underlying question is whether training activities which highlight rhythmic and melodic structures of a second language can help boost L2 production abilities.

Typically, Chinese learners of Germanic and Romance languages experience common difficulties in the placement of primary stress and rhythm (e.g., Bian, 2013; Chen et al., 1996; Li, 2012). For instance, Bian (2013) analyzed the difficulties of Chinese EFL learners and reported frequent inappropriate placement of stress in their English words. Chen et al. (1996) provided some applicable tools for English teachers to use rhythm to help Chinese students on English pronunciation. They compared the different stress patterns of English and Chinese syllables and sentences while highlighting the contrast between the English stress-timed rhythm and the Chinese syllable-timed-rhythm. For example, one of the difficulties that Chinese speakers have is to strengthen the target stress of words like "e-co-NO-mics". The author suggests in this article that the teachers may tap, clap or play rhythm instruments to produce strong or weak beats thus to help students to find out the stress of words in sentences.

Rubin (1975) pointed out that motivation is essential for good language learning. However, Chinese students usually feel stressed by being pushed to learn English, as

---

it is one of the main obligatory subjects in the Chinese nine-year compulsory education system and for the Entrance Examination of College. Moreover, Chinese students tend to be shy and quiet, which works against them in regards to speaking. Further, the fear of making mistakes, failure or ridicule regarding their English speech may prevent them from advancing on their pronunciation. Our hypothesis is that music could work out as an effective tool for learning English. Music may improve motivation and make language learning more joyful (e.g., Patel, 2010; Kao & Oxford, 2014). Paquette and Rieg (2008) describe the beneficial aspects of incorporating music into daily class instructions while providing practical activities for improving children’s English learning, such as using songs for presentation, practicing phases, teaching vocabulary, pronunciation, rhythm and so on. Nevertheless, more experimental evidence is needed to give us more clear clues.

In sum, more empirical evidence is needed on the potential benefits of rhythmic and melodic trainings in the second language classroom. Given this, the main objective of this thesis is to investigate how training with music-based activities affects L2 pronunciation learning. Our underlying hypothesis is that rhythmic and melodic trainings can be beneficial for L2 pronunciation learning.

Summarizing, the main objectives of the three studies in this thesis are the following:

(a) whether hand-clapping to the rhythm of newly learned words help raise the awareness of word structure then improve pronunciation of foreign words for beginners of a foreign language (Study 1);

(b) whether a training session with playing percussion instruments to the speech rhythm can arouse the attention and perception of rhythm for intermediate students of a foreign language (Study 2);

(c) whether an 8-week training session with listening to and singing songs facilitate memorizing and producing foreign sentences by activating vocal-motor abilities, leading to better L2 pronunciation for intermediate students of a foreign language (Study 3).

1.5 Theoretical Framework(s)
1.5.1 Music and Language

Music and language share common features and differ in various perspectives. Their functions in human society are fundamental as well, especially during the first years of life (Trehub, 2003). Neuroscientists claim that music and language use a common network in the brain. For instance, with evidence from an ERP experiment, Besson and Schön (2001) have shown that when the brain reacts to violations of syntax or music melody, similar components are produced. Neuroimaging studies demonstrate that brain regions for language processing are active during music processing (e.g., Levitin, & Menon, 2003; Abrams et al., 2010). In addition, Zatorre (2005) revealed with neuroimaging evidence that some functions of music and speech require the same neural resources.

Due to the common properties shared by music and language, and the fundamental functions that they play for human society, there is a rising tendency for the research on their interaction. Central elements of in both music and language are rhythm/stress and timing, and melody/intonation. Moreover, they both follow determined rules of harmony/syntax. Recently, researchers in this field have successfully shown the benefits of musical aptitude, music expertise, and musical training on language abilities (see section 1.2 above).

1.5.2 Embodied cognition

The theory of embodied cognition is that many features of cognition are shaped by aspects of entire body organisms (see Wilson and Floglia 2017 for a review). According to the theory of embodied cognition, many features of cognition are embodied in that they are deeply dependent upon characteristics of the physical body of an agent, such that the agent's beyond-the-brain body plays a significant causal role, or a physically constitutive role, in that agent's cognitive processing, and moreover, suggest that sensorimotor processes contribute to language processing. For language acquisition, a set of factors may influence the proficiency of language learning, such as encouragement, environmental exposure, working memory and general cognition development. In first language acquisition, children learn vocabulary faster when directed joint focus of attention.
For early age language acquisition, gesture plays an important role for infants to acquire the lexicon (see Wilson & Foglia, 2017, for a review). Researchers strongly suggest that parents who use gestures frequently, broad embodiment of interactions, and also encourage the use of gestures can help infants in language development. Recently, Yuan et al. (2018) investigated how pitch gestures can help Mandarin Chinese speakers to learn Spanish intonation. A total of sixty-four participants were assigned to two training groups: a non-gesture group, which received intonation training without gestures, and a gesture group, which received the same training with pitch gestures representing nuclear intonation contours. Results showed that observing pitch gestures during the learning phase improved pronunciation skills significantly better than a training without gestures.

1.5.3 Cross-domain auditory plasticity

Cross-domain auditory plasticity refers to the idea that experience or training could modify aspects of brain function and structure (Herholz & Zatorre, 2012), in this case music and language. Bidelman, Gandour, & Krishnan (2011) provided evidence that Chinese (tonal language) speakers and English musicians outperformed English non-musicians in pitch-tracking accuracy, in both the music and language domains. Besson et al. (2011) have shown that with the involvement of music, musicians showed an enhanced auditory attention and working memory. The auditory system is crucial for musical training and, hence, one of the systems most altered by it. Musical experience influences functional and structural changes of auditory pathways of the brain (see Herholz & Zatorre, 2012 for a review).

Experimental and longitudinal research that examines the impact of musical training on neural processing of speech demonstrate the cross-domain neural plasticity from music to speech. For instance, Chobert, François, Velay and Besson (2012) have explored the functional consequences of neural change in speech processing through a pre- and posttest between-subject design. They randomly assigned participants to two longitudinal (two year) training conditions: musical training or painting training. In order to distinguish the changes in speech processing due to the musical training, a more active control group is crucial. As opposed to some of the training studies in the
previous review which use a passive control group with no training, in this present thesis, passive control group will be avoided.

1.5.3.1 Music rhythm and speech rhythm

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 or weak elements (beat, meter and tempo). For musical experience, feeling the beat is fundamental and tempo represents the speed of the beat (London, 2012). Beat refers to the spontaneous ability to perceive periodicities, for instance, periodic head nodding or foot tapping (Nozarandan, Peretz, & Mouraux, 2012). 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., Smotrova, 2017; Baker, 2014). Bhide et al. (2013) investigated the relationships between rhythmic training and speech abilities of 6-7 year old children and showed the connections between processing of rhythm and speech.

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.

1.5.3.2 Songs and cross-domain plasticity

Previous studies have shown that using songs may facilitate speech segmentation and word learning (e.g., De Groot, 2006; Heidari, & Araghi, 2015). Songs combine elements from speech (phonemes, syllables), and music (e.g., melodies built from musical scales and beat-based rhythms), thus the influence of singing-based training on speech abilities could be due to within-domain plasticity partly or wholly (Patel, 2014).
There are a growing number of longitudinal studies regarding the impact of music training (using songs) on speech processing. For instance, Chobert et al. (2012) studied 8-10 year old children randomly assigned to long-term training conditions, either musical training or painting training. They found that musical training group outperformed the painting training group in pre-attentive processing of syllable onset time. Thompson, Schellenberg, & Husain (2004) studied the impact of piano training on children's ability to decode affective speech prosody, and found that such training enhanced sensitivity to emotion.

1.6 Hypotheses

The three training studies of this PhD thesis are based on the underlying assumption that music-based activities may produce positive effects on second language pronunciation learning. The assumptions behind the three studies of this thesis are based on (a) evidence showing the significant correlations and interactions between music and language, (b) evidence showing that language learning can be enhanced by musical abilities (musical aptitude, expertise and training), and (c) evidence showing that rhythmic and melodic training which highlights prosodic structure might increase attention and verbal memory.

The hypotheses of the three experimental studies proposed will be the following:

(a) hand-clapping to the rhythm of newly learned words may help increase the awareness of word structure which in turn will improve pronunciation of foreign words for beginners of a foreign language (Study 1);

(b) a training session with playing percussion instruments to the speech rhythm can boost participant attention and perception of rhythm for intermediate students of a foreign language (Study 2);

(c) an 8-week training session with listening to and singing songs may help facilitate the memorization and production of foreign sentences by activating vocal-motor abilities, leading to better L2 pronunciation for intermediate students of a foreign language (Study 3).

In China, even though some English teachers frequently include listening to or singing
songs in their pedagogical plans, many of them are not sure about whether the effects are positive or not. Cheng (1998) pointed out that using songs is one of the most meaningful and motivating ways to teach English pronunciation, and that this strategy has proved to be successful in their pronunciation classes.

For the first study, French has been chosen as a target language because of its prosodic characteristics. For Studies 2 and 3, the target language will be English, which is known by the popularity of Chinese adolescent participants, considering in testing their pronunciation improvement at both the vocabulary and phrasal level.
2. RESEARCH METHODOLOGY

The three experimental studies in this PhD thesis will consist of three between-subjects training studies with two conditions (clapping vs. non-clapping; instrument training vs. non-instrument training; singing vs. non-singing) in a pretest/posttest design. For each of the three studies, a set of cognitive and linguistic measures will be taken for each participant and are explained below.

2.1 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. In all the three studies, the following set of control measures will be applied for each individual participant.

2.1.1 Working memory

According to Henry, Messer, Luger-Klein and 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 in daily life (see Appendix B). 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.

2.1.2 Speech imitation
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, see Appendix C) 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 author 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).

2.1.3 Musical ability

2.1.3.1 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.

2.1.3.2 Discrimination tests for melody and rhythm-to-melody
As the two subtests of rhythm and accent, for testing participants’ musical abilities on perceiving melody, two PROMS subtests of melody and rhythm-to-melody will be assigned to each participant of Study 3.

2.2 Study 1

2.2.1 Aim of the study

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 repeated the words orally 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.

The focus of Study 1 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 face difficulties in the placement and strengthening of word and sentence stress (e.g., Bian, 2013; Chen et al., 1996; Li, 2012). For instance, Bian (2013) analyzed the difficulties of Chinese EFL learners and reported frequent inappropriate placement of stress in their English words. In French, stress placement tends to be consistently placed on the last syllable of the prosodic group, which can also cause problems for Chinese-speaking learners (see Astésano, 2001). By way of illustration, the two graphs in Figure 1 show the waveform, spectrogram and F0 contour of the French word *confiture* [kɔ̃fi.tyʁ] ‘jam’ as spoken by a native speaker of French (left panel) and as rendered by a Chinese adolescent without any knowledge of French (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](image)

**Figure 1.** Waveform, spectrogram and F0 contour of the French word *confiture* [kɔfi.tys] ‘jam’ as spoken by a French native speaker (left panel) and as imitated by a Chinese adolescent with no French background (right panel).

### 2.2.2 Methodology

#### 2.2.2.1 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. 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 A.

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.
2.2.2.2 Materials

2.2.2.2.1 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 below). 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).

<table>
<thead>
<tr>
<th>Related (10 words)</th>
<th>disyllabic</th>
<th>trisyllabic</th>
<th>more than 3 syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>tambour</td>
<td>crocodile</td>
<td>aspirateur</td>
<td></td>
</tr>
<tr>
<td></td>
<td>mandarine</td>
<td>céréales</td>
<td></td>
</tr>
<tr>
<td></td>
<td>biberon</td>
<td>hélicoptère</td>
<td></td>
</tr>
<tr>
<td></td>
<td>confiture</td>
<td>television</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unrelated (4 words)</th>
<th>disyllabic</th>
<th>trisyllabic</th>
<th>more than 3 syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>peinture ‘painting’</td>
<td>sorcière ‘sorcerer’</td>
<td>calendrier ‘calendar’</td>
<td></td>
</tr>
<tr>
<td>parachute ‘parachute’</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Table 1.* Target French words selected for the training session.

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

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 (see Figure 4 for the Procedure).
Figure 3. Stills from a training video. A black and white illustration of an elephant is followed by a video of a speaker uttering the word in French (“éléphant”) in the clapping condition (upper panels) or in the non-clapping condition (lower panels).

For each condition, six different training videos were created to present the 20 items produced by the two speakers in different orders in order to avoid any primacy or recency effects.

2.2.2.2 Audio materials for the pretest and posttest

For both pretest and posttest, participants performed an adapted version of the elicited imitation task (EIT) (Van Moere, 2012). L2 acquisition research has shown the EIT 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 EIT 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.

<table>
<thead>
<tr>
<th>disyllabic</th>
<th>trisyllabic</th>
<th>more than 3 syllables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Related (10 words)</td>
<td>crocodile</td>
<td>aspirateur</td>
</tr>
<tr>
<td>tambour</td>
<td>mandarine</td>
<td>céréales</td>
</tr>
<tr>
<td></td>
<td>biberon</td>
<td>hélioptère</td>
</tr>
<tr>
<td></td>
<td>confiture</td>
<td>television</td>
</tr>
<tr>
<td>Unrelated (4 words)</td>
<td>peinture ‘painting’</td>
<td>sorcière ‘sorcerer’</td>
</tr>
<tr>
<td></td>
<td>parachute ‘parachute’</td>
<td>calendrier ‘calendar’</td>
</tr>
</tbody>
</table>

Table 2. French words selected for the pretest/posttest.
2.2.2.3 Procedure

Participants were first randomly assigned to one of the two conditions, namely the clapping group \((N = 25, 8 \text{ females})\) and the non-clapping group \((N = 25, 8 \text{ females})\). 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 in 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.

![Figure 4. Experimental procedure](image-url)
2.2.3.1 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 researcher. This test also lasted around 5 minutes.

2.2.3.2 Pretest

As noted above, the procedure followed for the pretest and posttest was identical and consisted of an online EIT (see section 2.2.2) which participants accessed by opening a link 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 items were presented randomly. A total of 700 experimental responses were obtained (50 participants × 14 items). The pretest lasted about 7 minutes.

2.2.3.3 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 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. (We may delete this figure) Regardless of the group, each participant saw a total of 40 videos, always presented randomly. The training session lasted around 10 minutes.
2.2.2.3.4 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.

2.2.2.3.5 Musical ability tests

By means of yet one more link, participants accessed the rhythm and accent sub-tests (for details see 2.2.3 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.

2.2.2.4 Pronunciation ratings

As noted, a total of 1400 EIT 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.
Each rater played back and evaluated all 1400 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.

![Figure 6. Sample page from the online rating survey](image)

### 2.2.2.5 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 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 6 shows, the following three tiers were created: (a) an orthographic tier, (b) the starting and end points of each syllable, including the final syllable (labeled $f_s$ in Figure 6, see below), and (c) the
starting and end points of the rhyme of the final syllable (the vowel followed by the final consonant; labeled vc in Figure 6).

Figure 7. Sample page of the Praat analysis of a participant producing the French word “calendrier” from a pretest or posttest recording, showing the three additional tiers: orthography, syllable boundaries (final syllable labeled fs), and final syllable rhyme (vc).

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).

2.2.2.6 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. Inter-rater
reliability was assessed using an intraclass correlation analysis for both pretest and posttest items. The Cronbach’s Alpha score was .777 which is over the accepted average measure of .7 (McDowell, 2006). ACCENTEDNESS was set as the dependent variable. GROUP (two levels: non-clapping vs. clapping), SESSION (two levels: pretest and posttest), 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.

2.2.3 RESULTS of Study 1

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.

2.2.3.1 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 near 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).

Figure 8. Mean accentedness ratings across the two conditions, GROUP (non-clapping and clapping) and SESSION (pretest and posttest).

2.2.3.2 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).

2.2.3.2.1 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 \( (F(1, 2375) = 0.473, p = .492) \).

![Figure 9](image.png)

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

2.2.3.2.2 *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.
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>Group</th>
<th>Session</th>
<th>Percentage-fs (%)</th>
<th>Percentage-vc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clapping</td>
<td>Pretest</td>
<td>54.485</td>
<td>73.236</td>
</tr>
<tr>
<td>group</td>
<td>Posttest</td>
<td>57.267</td>
<td>75.886</td>
</tr>
<tr>
<td>Non-clapping</td>
<td>Pretest</td>
<td>53.295</td>
<td>71.02</td>
</tr>
<tr>
<td>group</td>
<td>Posttest</td>
<td>53.529</td>
<td>72.563</td>
</tr>
</tbody>
</table>

2.2.4 DISCUSSION and CONCLUSION

In Study 1, 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 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 (Cason & Schön, 2012; Cason et al., 2015; and others) to the domain of second language phonological processing. Our results support the findings by Fischler (2009), Wang, Mok, and Meng (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, Power & Goswami, 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 (Derwing, Munro & Wiebe, 1998; Derwing & Rossiter, 2003; Gordon, Darcy & Ewert, 2013; Behrman, 2014).

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.

2.3 Study 2
2.3.1. Aim of the study

Previous findings have shown that musicians outperform non-musicians in phonological abilities (e.g., Pastuszek-Lipińska, 2007; Magne et al., 2006; Besson et al., 2007). Still, no previous studies have empirically tested the potential effects of a short rhythmic training session with instruments on second language pronunciation. Study 2 will investigate whether Chinese adolescent students improve English pronunciation after training with or without percussion instruments. The training will consist of two main parts: first, the instrument group will use percussion musical instruments in the classroom to work on general rhythmic knowledge before practicing the rhythm of speech; second, a designed exercise phase will be used in which rhythmic exercises will follow the structure of the target language. To design the first part of the training on general rhythmic knowledge, we will use the CMT methodology. According to Habib et al. (2015), the Cognitive-Musical Training (CMT) method was designed by speech therapists and is based on widely recognized principles of effective rhythmic intervention. Based on this method, we will design several exercises adapted from the CMT methodology that cover different dimensions of music such as tempo, duration, pulsation, and rhythm. They will be aimed at developing both speech perception and production (e.g., learn to play pieces of rhythm, tapping or clapping to speech rhythm in English, and also incorporating body movements, see Figure 10). The second part of the training session for the instrument group will be playing a percussion instrument (Chinese drum), following the speech rhythms found in a narrative recording of easy English fairy tales (see Figure 11).

In contrast, the non-instrument group will train on general pronunciation abilities followed by specific practices that will be comparable to the instrument group. We hypothesize that an enhanced ability to practice musical rhythm with instruments will correlate with more rapid mastery of an L2 whose rhythmic properties differed from those of the L1, suggesting a cognitive transfer between rhythmic knowledge in the language and musical domains.
2.3.2. Methodology

The experiment is a short training study with a pre- and posttest design. Participants will be tested individually.

2.3.2.1 Participants

A total of fifty Chinese adolescents (a group of first graders from Zhangqiu Experimental Middle School) will be recruited for this experiment. All of them will be low-intermediate learners of English and none of them will have previous background knowledge of a third language.

2.3.2.2 Materials
The pre- and posttest tasks will assess a variety of phonological abilities. The materials will consist of: (a) audio recordings of pairs of English sentences with minor errors in speech stress for the discrimination test, recorded by native English speaker, (b) audio files of intermediate level English sentences and phrases (the participants are at an initial level), for the English repetition task, and (c) printed materials of selected paragraphs from the students' textbook, for a general pronunciation test.

For the first session of the training phase, the materials will be designed and prepared following the CMT method. For the second session, following the procedure used for recording the audiovisual materials for Study 1, two native English instructors will be audiovisual-recorded reading phrases and sentences for speech repetition for the training session in two versions: with or without accompanying a percussion instrument (see Figure 1) with speech. Sequences of phrases or sentences with the same length will be created for both training groups and subsequently combined in several random orders and assigned randomly to each participant.

![Figure 11. A picture of the potential drum that will be used for Study 2](image)

The phrase and sentence repetition task will also include items unrelated to the training session, which will be at different difficulty levels of English in order to test for potential generalization effects on pronunciation learning.

### 2.3.2.3 Procedure

In the pretest phase, a set of English tests will be used, including a perceptive test (sentences stress discrimination task), a productive test (English phrases and
sentences elicited repetition task, with items that are both related and unrelated to the training) and a general pronunciation test (English text book reading).

During the training session, the instrumental/speech-only training will last around 30 minutes. In the first 15 minutes, participants will be trained following the CMT methodology. For example, students learn to play pieces of rhythm, tapping or clapping to different speech rhythms, or incorporating body movements with speech. In the last 15 minutes, the training group will be given an exercise that includes playing with a percussion instrument (drum), following pieces of rhythms that reflect the speech rhythm of phrases and sentences in English. The phrases and sentences will be selected from the textbooks that the second-grade students are using (unknown to first graders) and fairy tales for beginners such as Little Red Riding Hood (see Figure 12). The control group will perform a parallel task while repeating English phrases and sentences but without using the percussion instrument.

![Figure 12](image-url)  
*Figure 12. Sample image of the fairy tale, Little Red Riding Hood.*

For posttests, in addition to performing the same three pretest tasks, there will be a vocabulary recall task testing the new learned words in the training.

During the training, participants will also be video recorded to monitor their behavior. Participants’ pre- and posttest productions will be presented in pairs for English native speakers to rate for their accentedness, using the same rating procedure as in Study 1 (see section 2.2.2.4 for more details). Comparing the rates of improvement of the two groups will show whether rhythmic training with an instrument is more effective.
Moreover, the analysis of unrelated items will test whether the potential effects are generalized or not.

2.3.2.4 Accentedness ratings

As in the first study, 5 English native speakers will rate the productions collected from the pre- and posttests. A comparison of the accentedness ratings will allow us to determine whether singing could be more effective for pronunciation improvement.

2.4 Study 3

2.4.1. Aim of the study

This study aims to investigate whether an 8-session training period with singing songs in a second language benefits second language pronunciation learning. The classroom study by Toscano-Fuentes and Fonseca-Mora (2012) suggested that a sound-music program involving listening to music and singing songs improved six-graders’ listening comprehension, speech production and motivation. However, their results are not reliable as they did not have a control group, nor did they perform any quantitative analyses. The study by Fischler (2009) also investigated the use of music in second language pronunciation learning via rap music. She suggests that rap music helped students by enhancing sentence structure. Nemoto et al. (2016) has shown a negative influence on English pronunciation with listening to sung speech; however, the stimuli was only one 14-word sentence. Further, no original speech input was given to experimental group. In contrast, Nakata and Shockey (2011) have shown that Japanese learners of English who trained with songs showed a significant improvement in consonant pronunciation accuracy. The aim of this study will focus on the effects of a one-month singing training period (2 sessions per week) on English pronunciation. The hypothesis is that an 8-week training session with listening to and singing songs facilitate memorizing and producing foreign sentences by activating vocal-motor abilities, leading to better L2 pronunciation for low-intermediate students of a foreign language.

2.4.2 Methodology
2.4.2.1 Participants

A total of fifty Chinese adolescents (a group of first graders from Zhangqiu Experimental Middle School) will be recruited for this experiment. All of them will be low-intermediate learners of English and none of them will have previous background knowledge of a third language.

2.4.2.2 Materials

2.4.2.2.1 Materials for pre- and posttests

For pre- and posttests, the materials will be (a) the printed lyrics, to be read aloud in order to assess reading skills, (b) audio recordings of English native speakers reading the texts, for an imitation task, and (c) lists of words from lyrics and textbooks for new vocabulary learning.

2.4.2.2.2 Songs for training

The songs we will select will be simple and easy to learn for low-intermediate Chinese learners of English. The song selection procedure will be based on (a) the difficulty of the songs from a melodic and linguistic point of view - they shouldn’t be too difficult to follow, as the participants have a low-intermediate level of English, (b) the familiarity of the songs - the songs should be unfamiliar to the participants, as generally known songs like “Twinkle Twinkle Little Star” or “Happy Birthday” will affect testing improvement. For the moment, two sample songs have been initially selected after speaking with the English teachers of the participants, namely “Big Big World” by Emilia Rydberg, and “Imagine” by John Lennon (for a paragraph of the lyrics from “imagine”, see below). To create the final selection, we will create a survey and send it to at least 20 students from other groups of the school asking their preference of a set of songs, including the questions about the clarity of the lyrics and emotional feelings.

“You may say I’m a dreamer

But I’m not the only one

I hope someday you’ll join us

And the world will be as one
Imagine no possessions

I wonder if you can

No need for greed or hunger

A brotherhood of man

Imagine all the people

Sharing all the world"

The final set of songs will be either sung or read by native English speakers with good singing abilities. Several words and sentences will be selected and recorded separately for the familiarization phase. For pre- and posttest, both related and unrelated sentences/words will be recorded by the same instructors.

The experiment is a training study in the classroom with a pre- and posttest design. Participants will be tested in small groups (5 participants per group).

2.4.3.2 Procedure

In the pretests, participants will be tested the following tests individually, (a) reading of the lyrics of the song/text; (b) imitation of the lyrics of the song/texts; (c) and vocabulary learning of the target texts/lyrics.

The training will last one month and consist of eight separate sessions. Each session will be around 45 minutes, more or less the same length of a normal class session. Before the training starts, the English pronunciation of each participant will be recorded via a text reading task and a target sentence repeating task. Additionally, a vocabulary test will be run.

Participants will be divided into two training groups: a singing and a speech-only group. In the first twenty minutes, participants from both groups will have a vocabulary learning phase which will include the familiarization of the lyrics; in the remaining twenty-five minutes, the participants will watch audiovisual materials recorded by English native speaker either singing the song or reading the song at the same speed. During the training, participants will be video recorded to monitor their behavior. After
the training, a posttest session will be carried out with the same tasks as in the pretest, namely (a) reading of the lyrics of the song/text; (b) imitation of the lyrics of the song/texts; (c) and vocabulary learning of the target texts/lyrics. After the 8-week training period, a general pronunciation assessment will be carried out.

2.4.3.3 Accentedness ratings

As in the Study 1, five English native speakers will rate the productions collected from the pre- and posttests. A comparison of the accentedness ratings will allow us to determine whether a singing intervention in the second language classroom can be more effective for pronunciation improvement over unsung speech.
3. WORKING SCHEDULE

April – May 2018

- Submission of the article based on the Treball Final de Màster to the journal *Language Teaching Research* (currently under revision). Title: Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. Coauthors: Florence Baills, Pilar Prieto

- Writing of the PhD research plan.

June 2018

- Writing and submission of PhD research plan


August – October 2018

- Cooperation with English teachers in China. Experimental methods and design for Study 2 and Study 3.

- Preparation of audio and audiovisual materials for Study 2.

- Participant recruitment for Study 2.

November – December 2018

- Control measures and second training study.

- Data collection for Study 2

January – March 2019

- Data analyses for Study 2

- Writing of Study 2
- Methods design for Study 3

April – June 2019

- Material preparation for Study 3

July 2019

- Cooperation with school in China
- Testing the training materials for Study 3

August – December 2019

- Data collection for Study 3
- Data preparation and analysis for Study 3
- Writing of Study 3

January – June 2020

- Writing of the PHD thesis

July 2020

- PHD defense
4. SELECTED REFERENCES


Ludke, Ferreira and Overy (2014) showed that a short training session with singing can facilitate short-term paired-associate phrase learning of an unfamiliar language, Hungarian. Sixty participants (30 male, 30 female) were randomly assigned to one of three learning conditions: speaking, rhythmic speaking, and singing. The training stimuli, a list of 20 phrases, were audio-recorded by native speakers of English and Hungarian in three conditions. The control of duration and rate of presentation was taking into consideration in this study. After a 15 minute learning phase in the three conditions, participants were asked to complete five tests, namely a Hungarian production test, a multiple-choice vocabulary posttest, an English recall test, a delayed-recall Hungarian conversation, and a Hungarian recognition test. Results show that the singing group performed significantly better in Hungarian production and delayed-recall Hungarian conversation tests. Their findings suggest that singing could be an effective tool for learning a second language in both the aspect of production and vocabulary learning. This study is important for this thesis, as it is one of the few short musical training studies applied to a second language.


Nemoto, Wilson, and Perkins (2016) tested whether training either with spoken or sung sentences improved English pronunciation by Japanese speakers. Thirty participants were divided into conditions with 15 persons each group. The material was a 14-word sentence from a selected English song, *Alive*. The training lasted 10 minutes. The
pronunciation of four phrases of this song was evaluated which was considered generally hard to pronounce for Japanese people. Their results showed the pronunciation performance of singing group was significantly worse than spoken condition group.

Due to its negative results, this study contributes to the improvement of the experimental design of this thesis in the following aspects. First, the singing training input will be even for two groups. Second, the singing group will be exposed not only to the songs’ melodies but also to speech melody. Third, the materials shouldn’t be only one sentence, but cover a more extensive group of materials.


Toscano-Fuentes and Fonseca-Mora (2012) implemented a classroom training program with forty-nine Spanish sixth-graders. They conducted a nine-month sound-music program (one academic year) adding music-related activities using vocal songs or instrumental music in the ESL classroom. After this academic year, the authors claimed that students benefited from the program and generally improved their English, and particularly in reading and auditory comprehension. In this study, the pronunciation of the students was evaluated by teachers during regular speaking activities. However, the lack of a control group makes it hard to tell if the improvements are due to the high engagement of musical activities of nine-month or to the normal improvement of one academic year’s language learning.

This long-term study offered motivation and inspiration for Study 3, as it consists of an implementation of singing activities in second language classroom. In Study 3 of this thesis, a more strict experimental procedure will be applied with careful selections of materials and training sessions as well as the inclusion of a control group. We will focus on pronunciation and vocabulary learning effectiveness.

Cason, Astésano, and Schön (2015) is a follow-up study of Cason and Schön (2012), in which showed evidence that rhythmic priming enhanced the phonological processing of bi- and trisyllabic pseudo-words with a stressed final-syllable when the prime meter matched the metrical speech structure of the words. Cason et al. (2015) realized a follow-up study in order to investigate this effect could also found with real French sentences. Thirty-four participants heard a musical rhythmic prime followed by a sentence with either a matching or mismatching prosodic structure and immediately performed a phoneme detection task. To check for the possible effect of active production, half the participants had received audio–motor training with the musical rhythms presented in the experiment. However, the terms of this audio-motor training cannot be assessed as they are not described in the article. Analysis of reaction times showed that phoneme detection was faster with a matching metrical prime. The researchers also found that the listeners may benefit from shorter, more repetitive rhythmic structures.

This research is important for the thesis because it underlines the link between music rhythm and speech rhythm through the effects of rhythmic priming and provides motivation for Study 1 and Study 2.


Gluhareva and Prieto (2017) conducted an experimental training study for twenty Catalan-dominant university students testing their pronunciation before and after the training. The participants were split into two training conditions, with or without
observing beat gestures. They found that the training with observing an English instructor producing a set of target sentences in English with co-speech hand gestures may positively affect participants’ accentedness rating score, that is, reducing foreign-like accents. The tested items were split into two difficulty categories, easy and difficult, a larger significant improvement was found in difficult items for beat gesture training group.

This study provides baseline evidence for Studies 1 and 2, since it shows evidence that beat gestures, which are hand gestures that mark the rhythm of speech, are effective for L2 pronunciation. It provides evidence in favor of the main objective of conducting training activities which highlight speech rhythm using music-based rhythmic activities.
5. GENERAL REFERENCES


Tierney, A., & Kraus, N. (2013). The ability to move to a beat is linked to the consistency of neural responses to sound. Journal of Neuroscience, 33(38), 14981-14988.


APPENDIX
Appendix A.
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 B.

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,</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>苹果</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>鞋子, 吉他, 太阳, 厨房</td>
<td>8</td>
<td>shoes, guitar, sun, kitchen</td>
<td></td>
</tr>
<tr>
<td>萄, 铅笔, 帽子, 草莓</td>
<td>8</td>
<td>grape, pencil, hat, strawberry</td>
<td></td>
</tr>
<tr>
<td>蜜蜂, 脚, 小丑, 书本</td>
<td>7</td>
<td>bee, foot, clown, book</td>
<td></td>
</tr>
<tr>
<td>国王, 单词, 工作, 电, 女孩</td>
<td>9</td>
<td>king, word, job, electricity, girl</td>
<td></td>
</tr>
<tr>
<td>文化, 武器, 例子, 水, 时间</td>
<td>9</td>
<td>culture, weapon, example, water, time</td>
<td></td>
</tr>
<tr>
<td>原因, 道路, 水平, 命令, 公里</td>
<td>10</td>
<td>reason, road, level, order, kilometer</td>
<td></td>
</tr>
<tr>
<td>心脏, 社会, 现实, 帮忙, 类人</td>
<td>10</td>
<td>heart, society, reality, help, human</td>
<td></td>
</tr>
<tr>
<td>周期, 期限, 来源, 情况, 世纪, 点</td>
<td>11</td>
<td>period, limit, origin, situation, century, point</td>
<td></td>
</tr>
<tr>
<td>国王, 嘴, 概念, 颜色, 血, 仪式</td>
<td>10</td>
<td>king, mouth, concept, color, blood</td>
<td></td>
</tr>
<tr>
<td>知识, 科技, 地方, 大海, 戏剧, 游戏</td>
<td>12</td>
<td>knowledge, technology, location, sea, drama</td>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<td>身体, 数量, 地址, 国家, 世纪, 行动, 夫</td>
<td>14</td>
<td>body, amount, address, country, century, action, husband</td>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<td>脚, 钱, 问题, 功能, 瞬间, 木头, 危险</td>
<td>12</td>
<td>foot, money, question, function, moment, wood, danger</td>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
<tr>
<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>
<td></td>
</tr>
</tbody>
</table>

**Appendix C.**

**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. שלום. אניalon ו ואני תلمיש
      English: Hello. My name is Alon and I am a student.
   b. היום יום_arefe, wangşemaf Zɔøeʃet
      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.