


STUDY

Embodied music training can help improve speech imitation and pronunciation skills

Yuan Zhang¹ , Florence Baills² and Pilar Prieto³

¹Zhejiang University, Hangzhou, China, ²University of Lleida, Lleida, Catalonia and ³Universitat Pompeu Fabra, Barcelona, Catalonia

Corresponding author: Yuan Zhang; Email: yuan.zhang@zju.edu.cn

(Received 29 November 2022; revised 9 September 2024; accepted 14 September 2024)

Abstract

Recent research has shown that music interventions involving body movements are beneficial for reinforcing music learning. Given the reported positive transfer between music training and phonological skills, we investigated, for the first time, the value of an embodied music training program for improving pronunciation skills. In a classroom experiment, 48 Chinese adolescent learners of English participated in three 40-minute sessions of either Dalcroze-inspired embodied music training or treatment-as-usual non-embodied music training. Participants in the embodied music group were involved in a series of activities designed to develop their rhythmic and melodic skills through bodily experiences. Participants in the non-embodied music group followed music lessons designed by their music teacher, appreciating music pieces and receiving music knowledge. Before and after training, participants were tested with a language imitation task using six unfamiliar languages and an oral reading task in second language English. Results show that the embodied music group significantly outperformed the non-embodied music group in both tasks. Overall, our findings suggest that an embodied music training program has beneficial effects on pronunciation skills, supporting an embodied-based language teaching approach.

1. Introduction

In second language (L2) classrooms, teachers tend to view the use of music as a teaching instrument favorably (e.g., Ajibade & Ndububa, 2008; Engh, 2013). In recent decades, scholars have suggested that using music and songs in the language classroom can trigger an increase in both students' language learning and motivation (e.g., Lo & Li, 1998). Particularly in the case of teaching L2 pronunciation, L2 teachers frequently use embodied rhythmic activities such as clapping and tapping to teach syllabification, word stress, and sentence rhythm (Smotrova, 2017). There is evidence that embodied engagement triggers beneficial effects on natural cognitive development (e.g., Wellsby & Pexman, 2014) and is especially relevant in education (e.g., Ionescu & Vasc, 2014; Macedonia, 2019; Shapiro & Stolz, 2019). Notably, embodied learning has been proven useful for enhancing different linguistic abilities (e.g., Kiefer & Trumpp, 2012; Madan & Singhal, 2012).

The positive experience of the use of embodied musical activities in the L2 classroom contrasts with the paucity of experimental evidence regarding the effectiveness of these activities in improving L2 pronunciation. Moreover, despite the positive relationship between music and phonological language skills (see Chobert & Besson, 2013 for a review), there is still very little empirical evidence regarding the role of musical interventions involving listening to and moving with instrumental music without any speech input in the target language on phonological learning (but see François et al., 2013). To our knowledge, no previous study has assessed the specific contribution of a non-linguistic embodied

© The Author(s), 2024. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.



music training program (i.e., involving only instrumental music and unrelated to any language task) to the phonological productive abilities of non-native language learners. The present study focuses on addressing this gap by assessing whether a Dalcroze-inspired embodied music training program, involving playful exercises that couple body movements with the perception of music rhythm and melody, can boost speech imitation skills and L2 English oral reading skills. Following the hypothesis that perceptive and productive phonological skills can be transferred between music and language (e.g., Besson & Schön, 2001; Besson et al., 2015; Christiner & Reiterer, 2013, 2015; Milovanov et al., 2008), and recent claims surrounding the benefits of the embodied music cognition paradigm on music learning, which states that music cognition is strongly determined by corporeally mediated interactions (Leman et al., 2018), we hypothesized that an embodied music intervention involving motor-based rhythmic and melodic activities might trigger beneficial effects on speech production skills. Importantly, we assessed whether a transfer effect or a causal effect on speech production skills may take place when training musical rhythmic and melodic skills.

1.1 Similarities between music and language

While on the surface, music and language may seem to be two distinct human abilities, research has highlighted the strong similarities between the two domains, and at different levels of analysis (e.g., McMullen & Saffran, 2004; Patel, 2003, 2010; Patel et al., 2006). Acoustically, both music and language are organized temporally, with sounds unfolding in time along melodic and rhythmic patterns (e.g., Patel, 2003; Patel et al., 2006). In a review article, Schön and Morillon (2019) emphasized that both speech and music can be characterized as multi-timescale structured signals, involving a hierarchical organization. That is, while during speech, segmental and syllabic units are organized into larger units that typically involve prosodic components (Giraud & Poeppel, 2012); in the case of music, musical notes are organized into a rhythmic and melodic structure that incorporates meter, scale, and harmony (Vuust & Witek, 2014). From a developmental perspective, several studies have demonstrated that infants under one year of age can identify both linguistic and musical patterns by detecting sound properties, such as a pitch decline or the lengthening of a final note, which are key in marking structural boundaries (Patel, 2010; Saffran et al., 2006).

Research in the field of neuroscience has also found similarities between music and language (e.g., Besson & Schön, 2001; Peretz et al., 2015). Brain imaging studies have shown that for both adults and children (Perani et al., 2010; Schön et al., 2010), some specific regions of the brain (e.g., primary auditory regions, and secondary auditory regions) are activated simultaneously upon hearing and understanding words or listening to musical scales, as well as by auditory imagery processes and access to melodic representations (see Besson & Schön, 2001 for a review). Ogg and Slevc (2019) point out that there is strong evidence for neural overlap between music and language, and cross-domain plasticity in the brainstem and the auditory cortex, but less apparent evidence at a higher level of processing, which seems to be specialized and influenced by other domains and cognitive functions.

One of the theories that supports the idea that non-linguistic music training can drive cross-domain plasticity in speech processing networks is the OPERA hypothesis (see Patel, 2011, 2012, 2014). According to Patel, the OPERA hypothesis is designed to explain the cross-domain plasticity that exists between music training and speech processing. The hypothesis predicts that the benefits of music training on speech processing can happen when the following five conditions are met: overlap, precision, emotion, repetition, and attention. Under these conditions, a transfer of skills between music and language will take place.

1.2 Music and phonological skills

Research in recent decades has shown that musical expertise, musical aptitude, and musical training have a positive effect on phonological receptive skills (see Besson et al., 2015; Chobert & Besson, 2013 for reviews) and speech imitation skills (Christiner & Reiterer, 2013, 2015).

When investigating musical expertise, Besson and colleagues found that instrumentalist musicians have higher speech perception capacities than non-musicians, particularly in relation to the perception of rhythm and pitch and the identification of lexical tones (Besson et al., 2007; Magne et al., 2006; Marie et al., 2011). For instance, Magne et al. (2006) showed that musician children with more than three years of training playing an instrument outperformed non-musician children in detecting pitch violations in both the final notes of musical phrases and the final words in sentences. Regarding phonological production, Christner and Reiterer (2015) showed that the musical expertise of both instrumentalists and singers played a facilitative role in sentence imitation tasks of an unknown language and an L2. Interestingly, this study showed that within a group of professional musicians, singers obtained even higher scores than instrument players, most likely due to their extensive vocal practice.

In the case of musical aptitude, it has been shown that higher musical perceptive abilities in both young and adult populations are related to better phonological and reading skills, such as detecting prosodic violations, more accurately perceiving and producing speech in a foreign language, and improving spoken or written language deficits (see Milovanov et al., 2008 for a review). In relation to reading abilities, Strait et al. (2011) tested the first language (L1) reading skills of children and found that, regardless of music experience, children who obtained higher scores in a musical aptitude test performed better in silent reading, auditory memory, and attention tasks. Slevc and Miyake (2006) found that the music perception abilities of Japanese speakers predicted perceptive and productive phonological abilities in their L2 (English) when factors such as language exposure and working memory capacities were controlled for. A recent study showed that the music perception abilities of native Catalan speakers can also predict their speech imitation skills in a variety of unknown languages (Li et al., 2022).

Regardless of musicianship or musical aptitude, music training has been shown to enhance perceptive phonological abilities (e.g., François et al., 2013; Linnavalli et al., 2018). A number of neurolinguistic studies have tested the OPERA hypothesis and given more support to this argument. Superior scores in skills such as speech-in-noise discrimination (Hennessy et al., 2021), auditory novelty detection (Chobert et al., 2014), and pitch discrimination (Moreno et al., 2009; Tervaniemi et al., 2005) have been positively correlated with the participants being musicians or having undertaken intense musical practice. Compared with studies on musical aptitude that are mainly based on correlational analyses, studies on musical training typically involve short- or long-term training programs. Interestingly, the results of a two-year training study with eight-year-old children showed enhanced segmentation skills when discriminating four trisyllabic pseudo-words in children who followed musical training, compared with those who undertook painting training (François et al., 2013). A long-term educational intervention in a kindergarten added evidence in favor of music education benefitting the development of linguistic skills; Linnavalli et al. (2018) found that a two-year playschool program involving weekly music activities (such as singing, playing simple instruments, body percussions, along with listening and moving to music) contributed to children's phoneme processing and vocabulary learning, compared with those children who did not attend such lessons. Importantly, previous studies on music training have focused on its positive effects on perceptual phonological abilities, although, to our knowledge, no previous studies have investigated the effects of transfer between music training and phonological skills from a production perspective (see Schön & Morillon, 2019 for a review).

1.3 Embodied music training and musical skills: The embodied music cognition paradigm

When listening to music, both adults and children tend to naturally engage with it physically and perform body movements (such as head bobbing, foot-tapping, swaying, or more elaborate movements) that typically follow the rhythmic and/or the melodic properties of the musical piece (Janata et al., 2012; Zelechowska et al., 2020). Music can even influence natural body movements such as walking (Styns et al., 2007). Some motion capture studies (e.g., Gonzalez-Sanchez et al., 2018) have demonstrated that even when participants are forced to stand still, music provokes spontaneous body

movements, such as swaying to the beat of music, which, for example, often lead to significantly more head movements (e.g., head nodding).

In the last two decades, embodied music cognition has become an influential paradigm within music research (see Leman, 2007 for a review). An earlier praxial approach to music education proposed by Elliott (1995, 2005) also emphasized the use of bodily engagement and active practice. According to Leman et al. (2018), the cognitive processing of music is based on corporally mediated interactions. Several studies have shown how the engagement of body movements with music can enhance rhythm perception in music (e.g., Manning & Schutz, 2013; Su & Pöppel, 2012). We adopt an embodied music training approach that emphasizes the positive role of coupling bodily movements with music to enhance musical skills, by integrating sensory, motor, and cognitive processes.

Previous studies have also shown the benefits of involving body movements (both part- or whole-body movements) in music education, using the Dalcroze approach (see Juntunen, 2020 for a review and for a set of practical ways to enhance embodied learning in music). Émile Jaques-Dalcroze was a Swiss composer, musician, and music educator who developed the Dalcroze Eurhythmics approach: a multisensory music training approach involving aural, visual, tactile, and muscular senses aimed at developing and improving target skills that are vital to a competent musician, such as a sense of rhythm or finesse of hearing (Juntunen, 2016).¹ Typically, the exercises used in a Dalcroze teaching sequence include functional (e.g., showing the height of pitch with the hand), rhythmic, creative, dramatic, and dance exercises. The repertoire of musical body movements ranges from simple movements, such as walking, turning, and running to starting-and stopping games, expressing music with fingers, hands, or arms, moving to the beat, and includes activities with tools, such as hoops, balls, or scarves.

In the field of music education, studies have provided empirical evidence for the supporting role of Dalcroze-inspired interventions involving body movements in the boosting of music skills (e.g., Leman, 2007; Leman et al., 2018). Reflecting on her own experience, Larsen (2016) showed that the Dalcroze approach was effective for strengthening her understanding of rhythm and harmony, and for improving her performance as a professional pianist. The implementation of the Dalcroze methodology may help music students improve their musical abilities, improvisation, expressive performance, as well as increase enjoyment, creativity, self-confidence, and risk-taking (e.g., Daley, 2013; Juntunen & Hyvönen, 2004; see also Sedar, 1997 for a review).

In addition, a set of studies reported in doctoral dissertations have assessed the positive effects of Dalcroze exercises on rhythmic abilities in kindergarten and first- and second-grade children. For example, O'Dell (2007) showed that first-graders who undertook a set of Dalcroze-based lessons improved their ability to perform a steady beat significantly more than the control group, who received the same lessons without the movement activities. Crumpler (1982) found a significant improvement between pre- and posttest on the pitch discrimination abilities of first-graders who participated in an eight-week Dalcroze music training program, whereas a control group of first-graders who for an equivalent amount of time took music lessons involving music concepts and sang singing songs without movement did not show any improvement. Interestingly, Holme (2009) emphasized that although focused on music, Jaques-Dalcroze saw his educational ideas as unrestricted to music. Holme (2009) also suggested that internalizing music with body movements may stimulate thoughts regarding the rhythmic nature of languages, and especially in the case of learners whose L1 is tonal (e.g., Mandarin), who might face difficulties in learning stress-timed languages such as English. Despite these observations, to our knowledge no empirical evidence has been put forth to assess whether Dalcroze's music training methods can be applied successfully when improving phonological productive skills.

1.4. Music and embodied music activities for L2 phonological learning

It is a shared view of practitioners in the L2 community that music is a helpful teaching tool for both young and adult learners alike, as it helps not only to foster enjoyable environments but also to

improve learners' speaking, writing, and listening comprehension skills (e.g., Pérez Niño, 2010). In L2 classrooms, there is a growing tendency to incorporate musical activities involving songs or other melodic and rhythmic activities so as to promote more active and enjoyable classes (Selinker & Gass, 2008). Indeed, some classroom studies have suggested that musical activities boost motivation, aid vocabulary memorization, increase learners' attention, and lead to enhanced learning (e.g., Fonseca-Mora & Gant, 2016; Kao & Oxford, 2014). Some studies have also empirically tested the role of singing and listening to songs in developing L2 pronunciation, with somewhat mixed results (e.g., for positive results, see Baills et al., 2021; Good et al., 2015; Ludke et al., 2014; Zhang et al., 2023; for mixed results, see Nakata & Shockey, 2011; Nemoto et al., 2016).

Recent research has found that hand-clapping, a rhythmic activity that is closely related to music, helps improve pronunciation in a foreign language for both children and adults (e.g., Baills & Prieto, 2021; Lee et al., 2020; Zhang et al., 2020). Hand-clapping involves the simultaneous activation of seeing, hearing, touching, and motor experience executed by the arms and the hands, which have to be synchronized with musical beats and rhythm (Baker, 2014; Chan, 2018). Synchronizing these arm and hand movements with speech rhythm requires a rigorous interaction of the vocal and motor systems (Sulkin & Brodsky, 2007) and, therefore, training this synchronization ability may, in turn, have a positive effect on speech.

Crucially, the above-mentioned experimental and classroom training studies have long used music (e.g., songs) or rhythmic activities (e.g., hand-clapping) in combination with the use of the target L2 languages.

1.5 Goals of the present study

The goal of the present study is to assess whether an embodied music training may have a positive effect on non-native phonological productive skills. Considering the well-known benefits of Dalcroze-based embodied music training on music-related abilities, it is interesting to assess the potential effects of a Dalcroze-inspired music embodied training program on abilities related to non-native speech production. This is a relevant question that has potential implications for L1 and L2 phonological learning. Given the strong connections between language and music, we expect to uncover a new transfer effect between music and speech by showing that training music skills through an embodied approach can trigger an improvement in phonological language skills.

With this goal in mind, we designed three music training sessions with embodied activities inspired by the Dalcroze approach and compared their effects with those of three treatment-as-usual music sessions of the same duration. All activities in the embodied music training sessions were inspired by those proposed by the Dalcroze approach to music teaching, incorporating exercises that couple music with body movement. The treatment-as-usual music sessions were designed by the music teacher according to the Chinese teaching curriculum, which follows an aesthetic approach to music education (Reimer, 1989). The activities consisted of listening to pieces of traditional Chinese songs (instrumental or orchestral versions) and learning about each piece of music (historical context, composer, relevant musical instruments, etc.).

Adolescent participants were recruited for the present study. While previous studies have primarily focused on children or adults, there is a lack of studies focusing on adolescents. Adolescence is a crucial developmental stage marking a period of transition from both a cognitive and a physical perspective (Engelhard, 2014).

All in all, the research question of this study states as follows:

Does an embodied music program have a positive effect on adolescents' L2 phonological productive skills:

- (a) in a speech imitation task with unfamiliar foreign languages, and
- (b) in an oral reading task in L2 English?

First, the capacity to imitate and repeat strings of a certain length in foreign speech has been positively related to L2 pronunciation learning (e.g., Hao & de Jong, 2016; Jia et al., 2006). The participants' productions in the imitation task were perceptually evaluated for accentedness, which is generally defined as the perceived distance between a foreign speaker's speech and the target native speech (Trofimovich & Isaacs, 2012). Accentedness is one of the most frequently used perceptive measures for the assessment of L2 pronunciation and has been related to pronunciation accuracy measures in terms of segmental and suprasegmental errors (Saito et al., 2016). As for the evaluation of the English oral-reading task, we followed the claim by Saito and Plonsky (2019) that a truly comprehensive assessment of L2 pronunciation should take into account measurements at both a holistic and specific level. Besides accentedness, measures of comprehensibility (defined as the ease with which one understands the meaning of what is uttered, see Derwing & Munro, 2009) and fluency (which refers to native speakers' perceptual impression as to whether the speech planning and production of the language is functioning smoothly and efficiently, and is primarily associated with speed of delivery and pausing behavior (e.g., De Jong et al., 2013; Segalowitz, 2010), were also included. Moreover, the accuracy of suprasegmental and segmental features was also evaluated perceptively (e.g., Saito et al., 2016; Suzukida & Saito, 2021; Trofimovich & Isaacs, 2012). In total, five dimensions were selected to assess participants' pronunciation in the L2 reading task: accentedness, comprehensibility, fluency, segmental accuracy, and suprasegmental accuracy.

In addition, previous research has illustrated (a) that working memory plays an important role in L2 vocabulary learning and phonological processing (e.g., Darcy et al., 2015), and (b) that musical aptitude may influence L2 perceptive and productive phonological abilities (e.g., Milovanov et al., 2008, 2010; Slevc & Miyake, 2006; Strait et al., 2011). For this reason, we controlled for potential differences between the groups regarding these individual differences and checked if they could have an effect on the pronunciation measures.

2. Material and methods

The present study is a between-subject classroom intervention with a pre- and posttest design, in which a group of Chinese adolescent learners of English received training over three music sessions in one of the following two conditions: (1) embodied music condition, which involved the group performance of embodied musical activities (henceforth, EMG for Embodied Musical Group), and (2) non-embodied music condition, which involved treatment-as-usual music lessons (henceforth, NEMG for Non-Embodied Musical Group). Before and after training, participants were asked to perform two non-native speech production tasks: (1) an imitation task involving two sentences in each of the six unfamiliar languages, and (2) an oral reading task in English involving three phrases.

2.1 Participants

A total of 50 hearing Chinese adolescents aged between 13 and 15 years ($M = 13.7$, 25 females), all belonging to the same eighth-grade group class in a secondary school in Shandong Province, China, were initially recruited for this study.

All the participants took part in the experiment on a voluntary basis and digitally signed a consent form. In coordination with the school administrators, the training sessions for both groups were scheduled during school hours as an extra activity to the regular music classes. The regular music classes took place once a week with a duration of 45 minutes, as per the teaching curriculum at the school. During the week of training, neither group attended their regular music class.

Due to the absence of two students in one of the three training sessions, the data from only 48 of the participants could be included in the final analysis (EMG: $n = 25$, 15 females, $M = 13.8$ years, $SD = 0.577$; NEMG: $n = 23$, 10 females, $M = 13.7$ years, $SD = 0.559$). An a priori power analysis was conducted using G*power3.1 to test the interaction between groups and tests (ANOVA: repeated

measures, within-between interaction; medium target effect size $\eta^2 = 0.04$, $\alpha = 0.05$). Results showed that the sample of 48 participants was expected to achieve a power of 0.72.

Information about participants' language and musical background was self-reported by means of a questionnaire (an English translation is provided in Supplementary Appendix S1). All participating students were Mandarin speakers learning English on a weekly basis at their school and reported learning English (attending English classes, self-studying, and homework) at an average of five hours per week. Based on a recent study by Peng et al. (2021), the mean English proficiency of eighth-graders in China corresponds to the A1 to B1 levels of the Common European Framework of Reference for Languages (CEFR), that is, low-intermediate proficiency. None of them reported speaking a third language nor being musicians or amateur musicians.

2.2 Materials and measures

This section describes the materials and measures used in the three training sessions, the pre- and posttest tasks, and the control tasks (see Figure 1). The training sessions took place on three consecutive days. Pretests took place on the same day as the first training session, and posttests occurred on the same day of the final training session. All sessions were carried out outside of the school curriculum. The planned duration of both conditions was 45 minutes for the first session (5 minutes introducing the whole program + 40 minutes training) and 40 minutes for the following two sessions. As the embodied music training program required more time for preparation and arrangement (i.e., transition between each activity, arranging participants' positioning for each activity, distributing accessories, etc.), each session in EMG included approximately five additional minutes. Overall, the teachers accurately followed the instructions and the mean duration was 47.93 minutes for EMG sessions and 41.53 minutes for NEMG sessions (see Supplementary Appendix S5 for detailed lesson design).

2.2.1 Control measures

To ensure that the two groups were homogeneous, individual control measures of English vocabulary size, music perception skills, and short-term memory span were assessed by performing several control tasks. All the control tests were conducted individually in quiet computer rooms three days before the pretest and the whole procedure took approximately 50 minutes to complete. The computer rooms were equipped with multiple computers and earphone headsets, enabling participants to take tests individually and simultaneously without any disturbance.

2.2.1.1 Basic English vocabulary test. To assess the participants' command of basic English vocabulary, a set of 75 English words was chosen from three seventh- and eighth-grade English textbooks that are frequently used in Chinese high schools. Each correct answer was given the score "1", and the final score of each participant was the sum of the correct answers. The test was carried out with an online survey platform *Wenjuanxing* (see Supplementary Appendix S2).

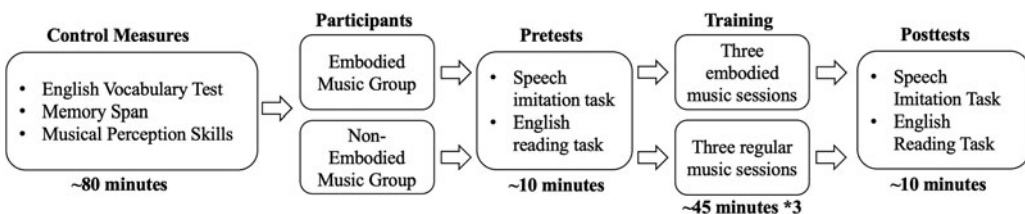


Figure 1. Overall experimental procedure.

2.2.1.2 Working memory. To measure participants' working memory, an adaptation of the forward digit span task by Woods et al. (2011) was used. The test was realized by a software developed in PsychoPy3. The individual scores were automatically generated, calculated through the recalling of an increasing (ranging from 3 to 16) sequence of digits in a forward order. The task lasted approximately 5 minutes.

2.2.1.3 Music perception skills. Musical aptitude was assessed using four subtests of the free-access Profile of Music Perception Skills test (PROMs) by Law and Zentner (2012). The MELODY, PITCH, ACCENT, and RHYTHM perception subtests were chosen. The scores were obtained automatically by the program. The test took around 20 minutes.

2.2.1.4 Satisfaction survey. After each session, a brief satisfaction survey was given to each student. We asked the participants to respond to the following three questions using a nine-point Likert scale ("1" means the lowest level and "9" means the highest level): (1) Preference: "How much did you like this session?"; (2) Difficulty: "How difficult do you think this session was?"; (3) Effectiveness: "Do you think that this session improved your musical abilities?"

2.2.2 Pre- and posttest tasks: Materials, procedure, and assessment

To assess the effect of embodied music training on non-native speech production skills, two tasks were administered before and after intervention: (1) a speech imitation task involving the oral imitation of sentences in six unfamiliar languages, and (2) an oral reading task in English. Participants performed the tasks individually using a computer in a silent room. They started the tasks after the audio-recording was activated by the assistant teacher. The whole pre- and posttest procedure lasted approximately 10 minutes.

2.2.2.1 Speech imitation task. To test each participant's ability to imitate non-native speech, they undertook a modified version of the speech imitation task in Zhang et al. (2020) both at pre- and posttest. The test involved listening to 12 short sentences (from five to 12 syllables) in six unfamiliar foreign languages (see Supplementary Appendix S3: two sentences per language; the languages involved were Catalan, Hebrew, Japanese, Russian, Turkish, and Vietnamese) and imitating them as closely as possible. The test took approximately 5 minutes and all the participants' oral productions were audio-recorded.

Three native speakers of each language were recruited to evaluate participants' production in their respective native languages (a total of 1,152 oral productions). Prior to the evaluation phase, the raters participated in a one-hour training session per language to acquaint themselves with the rating procedure and practice the rating with sample items. The raters evaluated the recordings through the online survey platform *Alchemer*. They were asked to evaluate the participants' accentedness on a scale from 1 to 9, where "1" corresponded to "extremely accented" and "9" indicated "not accented at all". For each item, the raters first listened to the sentence pronounced by a native speaker, and then to two oral productions, which corresponded to the pretest and posttest renditions of the sentence produced by the same participant at pre- and posttest, but in random order. Each rater spent around one hour on the rating task. The final speech imitation score per participant was calculated by computing the mean score for each sentence across the three raters. Inter-rater reliability among the three raters for each language was assessed with Intraclass Correlation Coefficient (ICC). The scores were 0.789 (Hebrew), 0.757 (Catalan), 0.907 (Japanese), 0.919 (Russian), 0.897 (Turkish), and 0.831 (Vietnamese), which are interpreted as good or excellent reliability (Koo & Li, 2016).

2.2.2.2 Oral reading task in English. Before and after the intervention, participants were asked to read three English phrases (10–17 syllables). These phrases contained target words and structures that were adapted to their English proficiency. In order to choose three adequate sentences, a total of 30 sentences were pre-selected and sent to five eighth-grade students who did not take part in the

experiment. They were asked to audio-record themselves reading the sentences and their oral production. First, the fluency and comprehensibility of each sentence were assessed by the three authors using a nine-point Likert scale to ensure that the participants would be able to pronounce the sentences. Second, it was ensured that the vocabulary appearing in the target sentences was part of the English textbook used by the learners. The final three selected sentences met the following two conditions: (1) they included no more than two words that did not appear in their target vocabulary, and (2) their comprehensibility and fluency were above five, respectively (see Supplementary Appendix S4).

Three native English speakers rated the participants' pronunciation (1,440 audio recordings) for each sentence in terms of accentedness, comprehensibility, fluency, the accuracy of segmental features (consonants and vowels), and the accuracy of suprasegmental factors (rhythm, accent, and intonation) based on a Likert-scale from 1–9, where “1” corresponded to the least accurate performance and “9” corresponded to the most accurate performance. Similar to the previous task, the raters participated in an online training session in which the raters received the detailed rating procedure, listened, and practiced with sample audio files that corresponded to different levels of proficiency. After their assessment, inter-rater reliability was calculated and the ICC score (for all five measures together) was 0.847, indicating good reliability.

2.2.3 Training sessions

A total of three sessions were prepared for both groups. While the lesson plans for the EMG were designed by the current team of authors, the lesson plans for the NEMG were designed by the music teacher from the participating school to fit the eighth-grade music curriculum. The music pieces selected for training were different across groups. While the music activities were designed by the authors for the EMG group, the music teacher designed the activities for the NEMG group. The detailed lesson plans and the audiovisual materials can be downloaded from OSF at: <https://osf.io/bm9c8/>.

2.2.3.1 Embodied Music Group. The three sessions for the EMG consisted of a set of rhythmic and melodic music activities that were inspired by the Dalcroze music-teaching approach (see Figure 2). For each session, the first author created a detailed presentation of the activities for the music teacher in charge of the EMG.



Figure 2. Screenshots of activities in the EMG.

Note: Ordered from top left to bottom right: Move to rhythmic patterns (S1), Rhythmic body percussion Part 2 (S2), Move to the changes of the music scale (S2), Balloon games: tapp to the beat (S3)

Each session comprised three or four embodied musical activities for training rhythm and melodic skills. The full training program included the practice of four music concepts related to rhythm and melody: (1) beat, that is, the basic unit of music, or ‘pulse’ that occurs at periodic intervals, which can be defined as a series of approximate periodic time points recognizable to rhythm listeners; (2) tempo, defined as the speed or pace of a given piece, and measured in beats per minute; (3) rhythm, which is the serial pattern of durations marked by musical notes and silences, and a grouping of accented and unaccented beats; and (4) melody, defined as a linear succession of individual musical tones that form a single entity (see McAuley, 2010).

Before the intervention, the music teacher of the EMG was trained by the first author over two 90-minute meetings, and performed one 45-minute pilot session with a different group of 25 eighth-graders. The procedure for each session was as follows: at the beginning of each session, the music teacher introduced the target music concepts to the participants and, if deemed necessary, showed them example videos of the activity. The first session also started with a brief introduction to the Dalcroze methodology. The second and the third sessions started with short warm-up activities that briefly repeated an activity from the previous session. At the end of each session, the teacher provided the participants with a brief summary of what had been learned.

All the sessions were video recorded with four cameras (AVA AE-A6 Recording and Playing System) to ensure training fidelity and the involvement of the students. The first author inspected the videos with the three training sessions from both the EMG and the NEMG and was able to confirm that the two music teachers had followed the lesson plans properly and that all participants engaged in the proposed activities, performing the appropriate body movements and paying attention to the teacher’s instructions.

2.2.3.2 Non-embodied music group. The lesson plan involved music appreciation and an introduction to musical instruments, composers, and music pieces. Each session revolved around one main theme, with one piece of traditional Chinese music. The three pieces were selected from the participants’ eighth-grade music textbook. The main piece of music for the first session was *Liangzhu: The Butterfly Lovers*, which is one of the most famous pieces for violin orchestras in China, and which describes the legend of the love story between Liang Shanbo and Zhu Yingtai. The music selected for the second session was *Colorful Clouds Chasing the Moon*. Finally, the music selected for the third session was *Blossoms on a Moonlit River in Spring*, which is an orchestral piece with traditional Chinese instruments.

2.2.4 Statistical analyses

All the statistical analyses were conducted in IBM SPSS Statistics, *Version 26.0*. To check for any statistical differences between the EMG and NEMG in terms of English vocabulary, memory span, and music perception skills, three *t*-tests were applied.

To assess the effect of training on the speech imitation task, a Generalized Linear Mixed Model (henceforth GLMM) was run with accentedness as the dependent variable (e.g., the mean rating scores for the six languages). The fixed factors were group (two levels: EMG and NEMG), test (two levels: pretest and posttest), and their interaction, Group \times Test. One random effects block was specified, with participant and item intercepts.

To assess the potential gains in the oral reading task in English, five GLMMs were run with the mean rating scores obtained for the three phrases of the following five dependent variables: accentedness, comprehensibility, fluency, segmental accuracy, and suprasegmental accuracy. The fixed factors were group (two levels: EMG and NEMG), test (two levels: pretest and posttest), and their interactions: Group \times Test. One random effects block was specified, with participant and item intercepts. Importantly, a collinearity analysis (Dormann et al., 2013) found that no collinearity issues were detected among the variables.

Finally, to assess the impact of the different types of training on the participants’ three self-reported satisfaction scores (i.e., preference, difficulty, and effectiveness), three GLMMs were run with

preference, difficulty, and effectiveness as dependent variables and group (two levels: EMG and NEMG), session (three levels: S1, S2, S3), and their interactions (Group \times Session) as the fixed factors. Participant, age, and gender were included as the random effects.

3. Results

The scores from three independent *t*-tests confirmed that there was no significant difference between the two groups in: (1) vocabulary size: $t(46) = -0.713, p = 0.479$; (2) memory span: $t(46) = -0.987, p = 0.329$; and (3) music perception skills: $t(46) = 1.241, p = 0.221$. In addition, we ran a GLMM with random effects that confirmed no significance in terms of memory span ($p = 0.146$) or music perception skills ($p = 0.382$).

3.1 Speech imitation task

Table 1 shows the descriptive statistics of the accentedness rating scores for the speech imitation task. The result of the GLMM with accentedness as the dependent variable shows a significant effect of test ($p < 0.001$), a significant interaction of Group \times Test ($p < 0.001$) and no significant effect of group (see Table 2).

Post-hoc analyses indicate that both the EMG and NEMG improved significantly from pretest to posttest (EMG: *contrast estimate* = 0.756, $p < 0.001$; NEMG: *contrast estimate* = 0.230, $p = 0.025$). A significant difference between the groups was found at posttest (*contrast estimate* = 0.425, $p = 0.041$), but not at pretest (*contrast estimate* = -0.101, $p = 0.626$).

3.2 Oral reading task in English

Table 3 shows the descriptive statistics of the mean rating scores for the English reading task. Table 4 summarizes the results of the five GLMMs for the mean rating scores across all five measures (see also Figure 3). We found a main effect of test ($p < 0.001$) and significant interactions Group \times Test for all five measures, and no significant main effect of group.

Post-hoc analyses showed significant improvements in the EMG group between pre- and posttest in terms of accentedness (*contrast estimate* = 0.423, $p = 0.001$), comprehensibility (*contrast estimate* = 0.717, $p < 0.001$), fluency (*contrast estimate* = 0.733, $p < 0.001$), segmental accuracy (*contrast estimate* = 0.515, $p < 0.001$), and suprasegmental accuracy (*contrast estimate* = 0.533, $p < 0.001$). By contrast, the NEMG showed non-significant changes in accentedness (*contrast estimate* = 0.005, $p = 0.971$), comprehensibility (*contrast estimate* = 0.135, $p = 0.384$), fluency (*contrast estimate* = 0.087, $p = 0.537$), segmental accuracy (*contrast estimate* = 0.050, $p = 0.695$), and suprasegmental accuracy (*contrast estimate* = 0.008, $p = 0.952$).

Post-hoc analyses further revealed that there were no significant differences between the two groups at both pretest and posttest in terms of accentedness (*contrast estimate* = 0.051, $p = 0.862$; *contrast estimate* = 0.478, $p = 0.106$), comprehensibility (*contrast estimate* = -0.067, $p = 0.822$; *contrast estimate* = 0.515, $p = 0.085$), fluency (*contrast estimate* = -0.242, $p = 0.412$, *contrast estimate* = 0.404, $p = 0.171$), and suprasegmental accuracy (*contrast estimate* = -0.121, $p = 0.687$; *contrast estimate* = 0.420, $p = 0.164$). Regarding segmental accuracy, there was a significant difference between the two groups at posttest (*contrast estimate* = 0.619, $p = 0.023$), but not at pretest (*contrast estimate* = 0.053, $p = 0.854$).

3.3 Satisfaction with training

Table 5 shows the descriptive statistics of the scores for the satisfaction questionnaire. Table 6 summarizes the results of the GLMMs for the three satisfaction variables. A main effect of group was found for preference ($F(1, 138) = 22.938, p < 0.001$), difficulty ($F(1, 138) = 33.580, p < 0.001$), and effectiveness ($F(1, 138) = 25.390, p < 0.001$). Post-hoc analyses revealed that the EMG yielded

Table 1. Mean, std. deviation, std. error, and 95% confidence interval for the rating scores of accentedness at pretest and posttest across groups

Variable	Test	EMG				NEMG			
		<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95%CI</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95%CI</i>
Accentedness	Pre	2.57	1.36	0.08	[2.41, 2.72]	2.71	1.24	0.07	[2.57, 2.83]
	Post	3.31	1.57	0.09	[3.13, 3.49]	2.94	1.24	0.07	[2.80, 3.08]

Table 2. GLMM: Fixed effects of the mean accentedness rating scores in the speech imitation task

Fixed factors	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Group	0.687	1	1121	0.407
Test	49.371	1	1121	<0.001
Group × Test	14.057	1	1121	<0.001

significantly higher scores than the NEMG in terms of preference (*contrast estimate* = 1.616, $p < 0.001$), difficulty (*contrast estimate* = 2.107, $p < 0.001$), and effectiveness (*contrast estimate* = 2.045, $p < 0.001$).

4. Discussion

The present study aimed to explore a transfer effect between embodied music training and non-native speech production skills. With this objective in mind, we assessed the potential gains of a three-session Dalcroze-inspired embodied music intervention involving no foreign language input on non-native productive speech abilities. A total of 48 Chinese teenagers participated, either in three treatment-as-usual music lessons, in which participants remained seated while learning about and listening to different pieces of music, or in three embodied music sessions, in which participants experienced musical rhythm and melody through body movements. Participants' speech production skills were assessed with a speech imitation task in six unfamiliar languages and with an English oral reading task. The results of the speech imitation task indicated that the EMG participants improved their speech imitation skills in terms of accentedness significantly more than the participants in the NEMG. Regarding the oral reading task in L2 English, only the EMG participants significantly improved their pronunciation at posttest in terms of accentedness, comprehensibility, fluency, segmental accuracy, and suprasegmental accuracy. The incorporation of body movements in the music training may have led to a stronger engagement and motivation by the participants, which might have resulted in higher preference ratings in the EMG. Our satisfaction survey (rated on a scale of 1 to 9) indicated that participants in the EMG showed significantly higher scores than the NEMG in terms of enjoyment ($M = 8.25$ vs. $M = 6.64$), and effectiveness ($M = 7.81$ vs. $M = 5.77$), despite finding the sessions more challenging ($M = 4.77$ vs. $M = 1.67$). Since positive emotions have been found to correlate with L2 motivation (MacIntyre & Vincze, 2017), this might have been a modulating factor in our results. Further research will be needed to further assess how embodied training can influence learners' emotional behavior and its concomitant relationship with L2 learning.

The results of the present study reveal an important transfer effect between the improvements produced by a Dalcroze-inspired embodied intervention and direct improvements in non-native language production. Our results add new evidence supporting the embodied music cognition paradigm by showing the value of an embodied music intervention (with no foreign language input) as a means of enhancing non-native speech production skills. To date, research on embodied music training has primarily focused on its direct effect on the development of musical abilities in the field of music education (e.g., Crumpler, 1982; Larsen, 2016; O'Dell, 2007) and to our knowledge, the application of this paradigm for the improvement of non-native speech production skills had not been tested before. Moreover, while previous research has shown the role of musical expertise, aptitude, and training on phonological language perception (e.g., Hennessy et al., 2021; Magne et al., 2006; Slevc & Miyake, 2006), the present study adds further evidence on the effects of music training on productive language skills.

In our view, the transfer effects also highlight the strong resemblance between music and language in terms of temporal and structural organization, as well as physiological and neural processing (e.g., Besson & Schön, 2001; Giraud & Poeppel, 2012; McMullen & Saffran 2004; Ogg & Slevc 2019; Patel, 2010). Our view is that the transfer obtained between embodied music training and non-native speech

Table 3. Mean, std. deviation, std. error, and a 95% confidence interval for the rating scores at pretest and posttest across groups

Variable	Test	EMG				NEMG			
		<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95%CI</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95%CI</i>
Accentedness	Pre	4.64	0.96	0.19	[4.25, 5.04]	4.59	1.17	0.24	[4.09, 5.10]
	Post	5.07	0.89	0.18	[4.70, 5.43]	4.56	1.11	0.23	[4.08, 5.04]
Comprehensibility	Pre	5.29	1.00	0.20	[4.88, 5.71]	5.37	1.26	0.26	[4.82, 5.91]
	Post	6.01	0.78	0.16	[5.69, 6.33]	5.47	1.10	0.23	[4.99, 5.94]
Fluency	Pre	4.80	0.96	0.19	[4.40, 5.20]	5.04	1.21	0.25	[4.52, 5.57]
	Post	5.54	0.86	0.17	[5.18, 5.89]	5.11	1.10	0.23	[4.63, 5.58]
Segmental accuracy	Pre	4.74	0.90	0.18	[4.37, 5.11]	4.68	1.16	0.24	[4.19, 5.19]
	Post	5.25	0.81	0.16	[4.92, 5.59]	4.59	1.08	0.23	[4.12, 5.06]
Suprasegmental accuracy	Pre	4.81	0.95	0.19	[4.42, 5.20]	4.93	1.23	0.26	[4.40, 5.47]
	Post	5.34	0.85	0.17	[4.99, 5.70]	4.89	1.16	0.24	[4.39, 5.39]

Table 4. Results of the five GLMMs for the mean rating scores

Dependent variables	Fixed factors	<i>F</i>	<i>df</i> ₁	<i>df</i> ₂	<i>p</i>
Accentedness	Group	0.898	1	94	0.346
	Test	5.547	1	94	<0.05
	Group × Test	5.793	1	94	<0.05
Comprehensibility	Group	0.663	1	94	0.417
	Test	15.488	1	94	<0.001
	Group × Test	7.211	1	94	<0.05
Fluency	Group	0.087	1	94	0.769
	Test	17.218	1	94	<0.001
	Group × Test	10.658	1	94	<0.05
Segmental accuracy	Group	1.582	1	94	0.212
	Test	6.696	1	94	<0.05
	Group × Test	9.916	1	94	<0.05
Suprasegmental accuracy	Group	0.276	1	94	0.601
	Test	8.317	1	94	<0.05
	Group × Test	8.827	1	94	<0.05

production supports the OPERA hypothesis (see Patel, 2010, 2011, 2014). Our embodied music program fitted the required conditions (overlap, precision, positive emotion, high engagement with repetitions, and focused attention), which may explain why it could lead to an enhanced adaptive plasticity in speech processing networks. Our study offers new insights into the effects of music-to-language transfer, building upon previous research demonstrating the facilitative role of long-term musical experience in speech perception (Patel, 2011, 2014). Our study showed that music training with body movements may trigger positive effects on L2 pronunciation skills. As the training did not involve any foreign language input, only musical melodic and rhythmic activities, and the EMG participants performed significantly better after training in terms of imitation skills and English pronunciation, our results suggest a positive music-to-language transfer effect. While previous studies found evidence of transfer effects after long-term training, the present study shows that a short training program can also trigger transfer effects after embodied rhythmic and melodic exposure for L2 pronunciation learning. The cross-domain effects found on pronunciation skills might stem from the common acoustic cues between music and language, in particular pitch and rhythm, and shared neural mechanisms (Patel, 2003; Steinbeis & Koelsch, 2008). These mechanisms do not only concern auditory processing areas but also several other brain regions such as the supramarginal gyrus, which is involved in both phonological processing and the processing of melody and pitch (e.g., Deschamps et al., 2014; Schaal et al., 2015), and the cerebellum, which is traditionally associated with motor control and contributes to timing and coordination aspects of both language and music (Krause et al., 2010; Martin et al., 2006). Transfer effects have been supported by neuroscientific evidence, as demonstrated by Dittinger et al. (2021), who noted that musicians could outperform non-musicians in long-term novel word memorization, possibly attributable to superior auditory processing. Interestingly, as Langus et al. (2023) suggest, the perception of speech prosody might be facilitated by cross-domain effects of musical experience. In their paper, the authors found that German-learning monolingual infants at six months with high exposure to music and infant-directed speech at home were more likely to exhibit an enhanced categorical perception of prominence in speech, whereas those with low exposure did not display a similar trend.

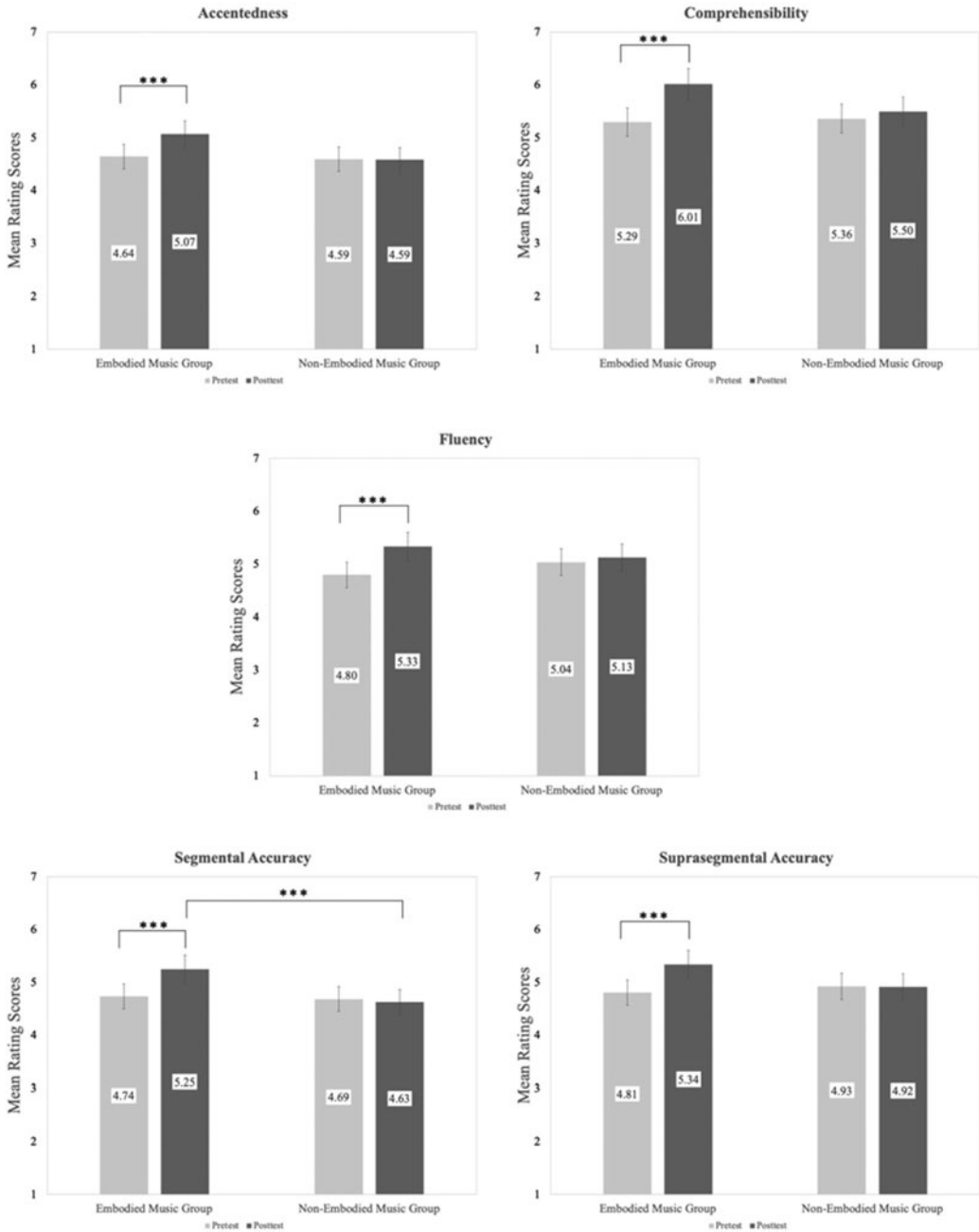


Figure 3. Estimated means across groups at pre- and posttest in terms of accentedness, comprehensibility, fluency, segmental accuracy, and suprasegmental accuracy.

On a practical level, our results have pedagogical implications for L2 pronunciation instruction. Previous studies have underlined the positive role of musical and embodied musical activities in the foreign language classroom, such as listening to songs and singing and hand-clapping (e.g., Baills et al., 2021; Good et al., 2015; Lee et al., 2020; Ludke, 2018; Zhang et al., 2020). In this context, our results confirm for the first time the importance of training musical rhythm and melody through

Table 5. Mean, std. deviation, std. error, and a 95% confidence interval for the satisfaction questionnaire across groups

Variables	Session	EMG				NEMG			
		<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95%CI</i>	<i>M</i>	<i>SD</i>	<i>SE</i>	<i>95%CI</i>
Preference	S1	8.32	1.11	0.22	[7.86, 8.78]	6.78	1.56	0.33	[6.11, 7.46]
	S2	8.28	1.34	0.27	[7.73, 8.83]	6.39	1.88	0.39	[5.58, 7.20]
	S3	8.16	1.28	0.26	[7.63, 8.69]	6.74	1.81	0.38	[5.95, 7.52]
Difficulty	S1	4.72	1.67	0.33	[4.03, 5.41]	2.30	1.26	0.26	[1.76, 2.85]
	S2	4.64	1.89	0.38	[3.86, 5.42]	2.70	1.64	0.34	[1.99, 3.40]
	S3	4.96	2.01	0.41	[4.12, 5.80]	3.00	1.76	0.37	[2.24, 3.76]
Effectiveness	S1	7.84	1.40	0.28	[7.26, 8.42]	6.00	1.62	0.34	[5.30, 6.70]
	S2	7.72	1.17	0.24	[7.24, 8.20]	5.83	2.29	0.48	[4.84, 6.82]
	S3	7.88	1.27	0.25	[7.36, 8.40]	5.48	1.88	0.39	[4.67, 6.29]

Table 6. Summary of the three GLMMs: Fixed effects of the scores for the satisfaction questionnaire

Dependent variables	Fixed factors	<i>F</i>	<i>df1</i>	<i>df2</i>	<i>p</i>
Preference	Group	22.938	1	138	<0.001
	Session	0.402	2	138	0.670
	Group × Session	0.513	2	138	0.600
Difficulty	Group	33.580	1	138	<0.001
	Session	1.283	2	138	0.280
	Group × Session	0.405	2	138	0.668
Effectiveness	Group	25.390	1	138	<0.001
	Session	0.665	2	138	0.516
	Group × Session	1.803	2	138	0.341

body movement to improve L2 pronunciation. Crucially, our findings show that even without any linguistic input, embodied music training has the capability to improve non-native language production skills. On the whole, our results also highlight the relevance of implementing embodied music training (which would typically belong to the area of music education) to benefit not only music learning but also language learning. In light of this, future educational curricula would need to further integrate the language and music training programs.

Importantly, our results also expand on previous insights showing that the integration of music (and embodied musical activities) into language classrooms tends to increase positive emotions and willingness to participate (e.g., Juntunen, 2020). In this regard, the results from our satisfaction survey revealed that the participants of the EMG self-reported that they liked the session more than the NEMG did, despite having evaluated the sessions as being more difficult. These positive side-effects of music may have triggered the students' involvement in the embodied music training sessions, which, in turn, might have helped by encouraging more positive outcomes. Further research would be needed to assess the effects of the positive emotional component induced by embodied music training, and specifically whether this component might affect the training outcomes.

The present study has the following limitations. First, to strengthen the findings of the present study, which only consists of three sessions with a limited number of participants, further research involving more participants and a longer embodied music intervention would be needed. Notably, we could not find a difference between the two groups at posttest, which may show that an improved intervention design is needed to find more clear-cut results. Second, it is important to mention that our experimental paradigm did not disentangle the effects of training musical melody and rhythm with or without embodied activities. In order to do this, a future study should include a control condition in which participants should be trained on the same musical melodic and rhythm skills involving non-embodied techniques, or at least exposed to the same musical input. As one of the reviewers suggested, adding another control group involving sports without music could provide information on the value of non-musical physical activity. In general, further research would be needed to assess the benefits of a non-embodied music training program that is more directly comparable to the one implemented in our embodied training group. Third, a more fine-grained acoustic analysis of participants' oral productions at pre- and posttest could be performed in order to help determine whether the improvements in speech imitation and L2 pronunciation are mainly due to a specific improvement in the production of the target prosodic features or whether they come from a combination of segmental and suprasegmental features, as the perceptual assessment seems to suggest. Fourth, future research could explore the mediating role of individual differences such as musical aptitude and musical expertise during embodied and non-embodied music training (e.g., Christiner & Reiterer, 2015; Christiner et al., 2022).

5. Conclusion

The present study suggests a transfer effect between embodied music training (i.e., without any foreign language input) on non-native productive phonological abilities. These results not only highlight the strong relationship between music and language but also the value of (embodied) music instruction paradigms for improving both music skills and language skills, thus supporting a more holistic educational approach that integrates music and language education more fully.

Supplementary material. The supplementary material for this article can be found at: <https://doi.org/10.1017/S0261444824000363>.

Note

¹ As a multidimensional approach, the Dalcroze method also incorporates solfège (ear training) and improvisation (Juntunen, 2016).

References

- Ajibade, Y., & Ndububa, K. (2008). Effects of word games, culturally relevant songs, and stories on students' motivation in a Nigerian English language class. *TESL Canada Journal*, 25(2), 27–48. <https://eric.ed.gov/?id=EJ806790>
- Baills, F., & Prieto, P. (2021). Embodying rhythmic properties of a foreign language through hand-clapping helps children to better pronounce words. *Language Teaching Research*, 27(6), 1576–1606. doi:10.1177/13621688209867
- Baills, F., Zhang, Y., Cheng, Y., Bu, Y., & Prieto, P. (2021). Listening to songs and singing benefitted initial stages of second language pronunciation but not recall of word meaning. *Language Learning*, 71(2), 369–413. doi:10.1111/lang.12442
- Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *Tesol Quarterly*, 48(1), 136–163. doi:10.1002/tesq.99
- Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Sciences*, 930(1), 232–258. doi:10.1111/j.1749-6632.2001.tb05736.x
- Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, 25(3-4), 399–410. <https://content.iospress.com/articles/restorative-neurology-and-neuroscience/rnn253423>
- Besson, M., Chobert, J., François, C., Astésano, C., & Marie, C. (2015). Influence of musical expertise on the perception of pitch, duration and intensity variations in speech and harmonic sounds. In C. Astésano & M. Jucla (Eds.), *Neuropsycholinguistic perspectives on language cognition: Essays in honour of Jean-Luc Nespoulous* (pp. 88–102). Psychology Press, Taylor & Francis Group.
- Chan, M. J. (2018). Embodied pronunciation learning: Research and practice. *CATESOL Journal*, 30(1), 47–68. <https://eric.ed.gov/?id=EJ1174234>
- Chobert, J., & Besson, M. (2013). Musical expertise and second language learning. *Brain Sciences*, 3(2), 923–940. doi:10.3390/brainsci3020923
- Chobert, J., François, C., Velay, J. L., & Besson, M. (2014). Twelve months of active musical training in 8- to 10-year-old children enhances the preattentive processing of syllabic duration and voice onset time. *Cerebral Cortex*, 24(4), 956–967. doi:10.1093/cercor/bhs377
- Christiner, M., & Reiterer, S. M. (2013). Song and speech: Examining the link between singing talent and speech imitation ability. *Frontiers in Psychology*, 4, 874. doi:10.3389/fpsyg.2013.00874
- Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in Human Neuroscience*, 9, 482. doi:10.3389/fnhum.2015.00482
- Christiner, M., Serrallach, B. L., Benner, J., Bernhofs, V., Schneider, P., Renner, J., Sommer-Lolei, S., & Groß, C. (2022). Examining individual differences in singing, musical and tone language ability in adolescents and young adults with dyslexia. *Brain Sciences*, 12(6), 744. doi:10.3390/brainsci12060744
- Crumpler, S. E. (1982). *The effect of Dalcroze Eurhythmics on the melodic musical growth of first grade students* [Doctoral dissertation, Louisiana State University and Agricultural & Mechanical College]. LSU Scholarly Repository.
- Daley, C. (2013). *Moved to learn: Dalcroze applications to choral pedagogy and practice* [Doctoral dissertation, University of Toronto]. TSpace Repository.
- Darcy, I., Park, H., & Yang, C. L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63–72. doi:10.1016/j.lindif.2015.04.005
- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34(5), 893–916. doi:10.1017/S0142716412000069
- Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490. doi:10.1017/S026144480800551X

- Deschamps, I., Baum, S. R., & Gracco, V. L. (2014). On the role of the supramarginal gyrus in phonological processing and verbal working memory: Evidence from rTMS studies. *Neuropsychologia*, 53, 39–46. doi:10.1016/j.neuropsychologia.2013.10.015
- Dittinger, E., Korke, B., & Besson, M. (2021). Evidence for enhanced long-term memory in professional musicians and its contribution to novel word learning. *Journal of Cognitive Neuroscience*, 33(4), 662–682. doi:10.1162/jocn_a_01670
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J. R. G., Gruber, B., Lafourcade, B., Leitão, P. J., Münckmüller, T., McClean, C., Osborne, P. E., Reineking, B., Schröder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, 36(1), 27–46. doi:10.1111/j.1600-0587.2012.07348.x
- Elliott, D. J. (1995). *Music matters: A new philosophy of music education*. Oxford University Press.
- Elliott, D. J. (2005). Musical understanding, musical works, and emotional expression: Implications for education. *Educational Philosophy and Theory*, 37(1), 93–103. doi:10.1111/j.1469-5812.2005.00100.x
- Engelhard, E. S. (2014). Dance/movement therapy during adolescence – Learning about adolescence through the experiential movement of dance/movement therapy students. *The Arts in Psychotherapy*, 41(5), 498–503. doi:10.1016/j.aip.2014.08.005
- Engh, D. (2013). Why use music in English language learning? A survey of the literature. *English Language Teaching*, 6(2), 113–127. <https://eric.ed.gov/?id=EJ1076582>
- Fonseca-Mora, M. C., & Gant, M. (Eds.). (2016). *Melodies, rhythm and cognition in foreign language learning*. Cambridge Scholars Publishing.
- François, C., Chobert, J., Besson, M., & Schön, D. (2013). Music training for the development of speech segmentation. *Cerebral Cortex*, 23(9), 2038–2043. doi:10.1093/cercor/bhs180
- Giraud, A. L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, 15(4), 511–517. doi:10.1038/nn.3063
- Gonzalez-Sanchez, V. E., Zelechowska, A., & Jensenius, A. R. (2018). Correspondences between music and involuntary human micromotion during standstill. *Frontiers in Psychology*, 9, 1382. doi:10.3389/fpsyg.2018.01382
- Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music*, 43(5), 627–640. doi:10.1177/0305735614528833
- Hao, Y. C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics*, 54, 151–168. doi:10.1016/j.wocn.2015.10.003
- Hennessy, S., Wood, A., Wilcox, R., & Habibi, A. (2021). Neurophysiological improvements in speech-in-noise task after short-term choir training in older adults. *Aging*, 13(7), 9468–9495. doi:10.18632/aging.202931
- Holme, R. (2009). *Cognitive linguistics and language teaching*. Palgrave Macmillan.
- Ionescu, T., & Vasc, D. (2014). Embodied cognition: Challenges for psychology and education. *Procedia-Social and Behavioral Sciences*, 128, 275–280. doi:10.1016/j.sbspro.2014.03.156
- Janata, P., Tomic, S. T., & Haberman, J. M. (2012). Sensorimotor coupling in music and the psychology of the groove. *Journal of Experimental Psychology: General*, 141(1), 54–75. doi:10.1037/a0024208
- Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *The Journal of the Acoustical Society of America*, 119(2), 1118–1130. doi:10.1121/1.2151806
- Juntunen, M.-L. (2016). The Dalcroze approach. In C. R. Abril & B. M. Gault (Eds.), *Teaching general music: Approaches, issues, and viewpoints* (pp. 141–168). Oxford University Press.
- Juntunen, M.-L. (2020). Ways to enhance embodied learning in Dalcroze-inspired music education. *International Journal of Music in Early Childhood*, 15, 39–59. doi:10.1386/ijmec_00011_1
- Juntunen, M.-L., & Hyvönen, L. (2004). Embodiment in musical knowing: How body movement facilitates learning within Dalcroze Eurhythmics. *British Journal of Music Education*, 21(2), 199–214. doi:10.1017/S0265051704005686
- Kao, T. A., & Oxford, R. L. (2014). Learning language through music: A strategy for building inspiration and motivation. *System*, 43, 114–120. doi:10.1016/j.system.2014.01.003
- Kiefer, M., & Trumpp, N. M. (2012). Embodiment theory and education: The foundations of cognition in perception and action. *Trends in Neuroscience and Education*, 1(1), 15–20. doi:10.1016/j.tine.2012.07.002
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163. doi:10.1016/j.jcm.2016.02.012
- Krause, V., Schnitzler, A., & Pollok, B. (2010). Functional network interactions during sensorimotor synchronization in musicians and non-musicians. *Neuroimage*, 52(1), 245–251. doi:10.1016/j.neuroimage.2010.03.081
- Langus, A., Boll-Avetisyan, N., van Ommen, S., & Nazzi, T. (2023). Music and language in the crib: Early cross-domain effects of experience on categorical perception of prominence in spoken language. *Developmental Science*, 26(5), e13383. doi:10.1111/desc.13383
- Larsen, M. (2016). *The pianist's embodiment of rhythm and harmony: Incorporating body movement into learning Manuel de Falla's Fantasia Bética* [Doctoral dissertation, New York University]. ProQuest Dissertation and Theses.
- Law, L. N., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS One*, 7(12), e52508. doi:10.1371/journal.pone.0052508

- Lee, B., Plonsky, L., & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, 88, 102185. doi:10.1016/j.system.2019.102185
- Leman, M. (2007). *Embodied music cognition and mediation technology*. MIT Press.
- Leman, M., Maes, P. J., Nijs, L., & Van Dyck, E. (2018). What is embodied music cognition? In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 747–760). Springer.
- Li, P., Zhang, Y., Fu, X., Baills, F., & Prieto, P. (2022). Melodic perception skills predict Catalan speakers' speech imitation abilities of unfamiliar languages. Proceedings of the 11th Speech Prosody Conference. University of Lisbon, Portugal.
- Linnavalli, T., Putkinen, V., Lipsanen, J., Huotilainen, M., & Tervaniemi, M. (2018). Music playschool enhances children's linguistic skills. *Scientific Reports*, 8(1), 1–10. doi:10.1038/s41598-018-27126-5
- Lo, R. S. M., & Li, H. C. F. (1998). Songs enhance learner involvement: Materials development. *English Teaching Forum*, 36(3), 8–11. <https://eric.ed.gov/?id=EJ597338>
- Ludke, K. M. (2018). Singing and arts activities in support of foreign language learning: An exploratory study. *Innovation in Language Learning and Teaching*, 12(4), 371–386. doi:10.1080/17501229.2016.1253700
- Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition*, 42(1), 41–52. doi:10.3758/s13421-013-0342-5
- Macedonia, M. (2019). Embodied learning: Why at school the mind needs the body. *Frontiers in Psychology*, 10, 2098. doi:10.3389/fpsyg.2019.02098
- MacIntyre, P. D., & Vincze, L. (2017). Positive and negative emotions underlie motivation for L2 learning. *Studies in Second Language Learning and Teaching*, 7(1), 61–88. doi:10.14746/ssllt.2017.7.1.4
- Madan, C. R., & Singhal, A. (2012). Using actions to enhance memory: Effects of enactment, gestures, and exercise on human memory. *Frontiers in Psychology*, 3, 507. doi:10.3389/fpsyg.2012.00507
- Magne, C., Schön, D., & Besson, M. (2006). Musician children detect pitch violations in both music and language better than nonmusician children: Behavioral and electrophysiological approaches. *Journal of Cognitive Neuroscience*, 18(2), 199–211. doi:10.1162/jocn.2006.18.2.199
- Manning, F., & Schutz, M. (2013). “Moving to the beat” improves timing perception. *Psychonomic Bulletin & Review*, 20(6), 1133–1139. doi:10.3758/s13423-013-0439-7
- Marie, C., Delogu, F., Lampis, G., Belardinelli, M. O., & Besson, M. (2011). Influence of musical expertise on segmental and tonal processing in Mandarin Chinese. *Journal of Cognitive Neuroscience*, 23(10), 2701–2715. doi:10.1162/jocn.2010.21585
- Martin, T., Houck, J. M., Bish, J. P., Kičić, D., Woodruff, C. C., Moses, S. N., Lee, D. C., & Tesche, C. D. (2006). MEG reveals different contributions of somatomotor cortex and cerebellum to simple reaction time after temporally structured cues. *Human Brain Mapping*, 27(7), 552–561. doi:10.1002/hbm.20200
- McAuley, J. D. (2010). Tempo and rhythm. In M. R. Jones, A. N. Popper, & R. R. Fay (Eds.), *Music perception* (pp. 165–199). Springer.
- McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, 21(3), 289–311. doi:10.1525/mp.2004.21.3.289
- Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P. A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research*, 1194, 81–89. doi:10.1016/j.brainres.2007.11.042
- Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P. A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, 20(1), 56–60. doi:10.1016/j.lindif.2009.11.003
- Moreno, S., Marques, C., Santos, A., Santos, M., Castro, S. L., & Besson, M. (2009). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex*, 19(3), 712–723. doi:10.1093/cercor/bhn120
- Nakata, H., & Shockey, L. (2011). The effect of singing on improving syllabic pronunciation-vowel epenthesis in Japanese. Proceedings of the International Conference of Phonetic Sciences. Hong Kong, China.
- Nemoto, S., Wilson, I., & Perkins, J. (2016). Analysis of the effects on pronunciation of training by using song or native speech. *The Journal of the Acoustical Society of America*, 140(4), 3343. https://u-aizu.ac.jp/~jperkins/Saori_hawaii_8.pdf
- O'Dell, K. L. (2007). *The effect of movement-based instruction on the steady beat performance of first-grade children* [Doctoral dissertation, University of Kansas]. ProQuest Dissertations & Theses.
- Ogg, M., & Slevc, L. R. (2019). Neural mechanisms of music and language. In G. Zubicaray & N. Schiller (Eds.), *Oxford handbook of neurolinguistics* (pp. 907–952). Oxford University Press.
- Patel, A. D. (2003). Rhythm in language and music: Parallels and differences. *Annals of the New York Academy of Sciences*, 999(1), 140–143. doi:10.1196/annals.1284.015
- Patel, A. D. (2010). *Music, language, and the brain*. Oxford University Press.
- Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? the OPERA hypothesis. *Frontiers in Psychology*, 2, 142. doi:10.3389/fpsyg.2011.00142
- Patel, A. D. (2012). The OPERA hypothesis: Assumptions and clarifications. *Annals of the New York Academy of Sciences*, 1252(1), 124–128. doi:10.1111/j.1749-6632.2011.06426.x

- Patel, A. D. (2014). Can nonlinguistic musical training change the way the brain processes speech? The expanded OPERA hypothesis. *Hearing Research*, 308, 98–108. doi:10.1016/j.heares.2013.08.011
- Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, 119(5), 3034–3047. doi:10.1121/1.2179657
- Peng, C., Liu, J., & Cai, H. (2021). Aligning China's standards of English language ability with the common European framework of reference for languages. *The Asia-Pacific Education Researcher*, 31, 667–677. doi:10.1007/s40299-021-00617-2
- Perani, D., Saccuman, M. C., Scifo, P., Spada, D., Andreolli, G., Rovelli, R., ... Koelsch, S. (2010). Functional specializations for music processing in the human newborn brain. *Proceedings of the National Academy of Sciences*, 107(10), 4758–4763. doi:10.1073/pnas.0909074107
- Peretz, I., Vuvan, D., Lagrois, M.É., & Armony, J. L. (2015). Neural overlap in processing music and speech. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370, 1664. doi:10.1098/rstb.2014.0090
- Pérez Niño, D. F. (2010). The role of music in young learners' oral production in English. *Profile Issues in Teachers Professional Development*, 12(1), 141–157. <https://revistas.unal.edu.co/index.php/profile/article/view/13994>
- Reimer, B. (1989). Music education in China: An overview and some issues. *Journal of Aesthetic Education*, 23(1), 65–83. doi:10.2307/3332889
- Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In D. Kuhn, R. S. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology: Cognition, perception, and language* (pp. 58–108). John Wiley & Sons, Inc.
- Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69(3), 652–708. doi:10.1111/lang.12345
- Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2), 217–240. doi:10.1017/S0142716414000502
- Schaal, N. K., Williamson, V. J., Kelly, M., Muggleton, N. G., Pollok, B., Krause, V., & Banissy, M. J. (2015). A causal involvement of the left supramarginal gyrus during the retention of musical pitches. *Cortex*, 64, 310–317. doi:10.1016/j.cortex.2014.11.011
- Schön, D., & Morillon, B. (2019). Music and language. In M. H. Thaut & D. A. Hodges (Eds.), *The Oxford handbook of music and the brain* (pp. 391–416). Oxford University Press.
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J. L., & Besson, M. (2010). Similar cerebral networks in language, music and song perception. *Neuroimage*, 51(1), 450–461. doi:10.1016/j.neuroimage.2010.02.023
- Sedar, S. (1997). *The use of music and creative movement as a tool in language development in elementary children: A review of related literature* (ED422230). ERIC. <https://files.eric.ed.gov/fulltext/ED422230.pdf>
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. Routledge.
- Selinker, L., & Gass, S. M. (2008). *Second language acquisition. An introductory course* (3rd ed.). Routledge.
- Shapiro, L., & Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory and Research in Education*, 17(1), 19–39. doi:10.1177/1477878518822149
- Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, 17(8), 675–681. doi:10.1111/j.1467-9280.2006.01765.x
- Smotrova, T. (2017). Making pronunciation visible: Gesture in teaching pronunciation. *Tesol Quarterly*, 51(1), 59–89. doi:10.1002/tesq.276
- Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18(5), 1169–1178. doi:10.1093/cercor/bhm149
- Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions*, 7, 44. doi:10.1186/1744-9081-7-44
- Styns, F., van Noorden, L., Moelants, D., & Leman, M. (2007). Walking on music. *Human Movement Science*, 26(5), 769–785. doi:10.1016/j.humov.2007.07.007
- Su, Y. H., & Pöppel, E. (2012). Body movement enhances the extraction of temporal structures in auditory sequences. *Psychological Research*, 76(3), 373–382. doi:10.1007/s00426-011-0346-3
- Sulkin, I., & Brodsky, W. (2007). The effects of hand-clapping songs training on temporal-motor skills among elementary school children. In K. Overy (Ed.), *Proceedings of the summer workshop on music, language, and movement*. Edinburgh: Institute for Music in Human and Social Development, University of Edinburgh.
- Suzukida, Y., & Saito, K. (2021). Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research*, 25(3), 431–450. doi:10.1177/1362168819858246
- Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, 161(1), 1–10. doi:10.1007/s00221-004-2044-5
- Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, 15(4), 905–916. doi:10.1017/S1366728912000168

- Vuust, P., & Witek, M. A. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, 5, 1111. doi:10.3389/fpsyg.2014.01111
- Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, 5, 506. doi:10.3389/fpsyg.2014.00506
- Woods, D. L., Kishiyama, M. M., Yund, E. W., Herron, T. J., Edwards, B., Poliva, O., Hink, R. F., & Reed, B. (2011). Improving digit span assessment of short-term verbal memory. *Journal of Clinical and Experimental Neuropsychology*, 33(1), 101–111. doi:10.1080/13803395.2010.493149
- Zelechowska, A., Gonzalez Sanchez, V. E., Laeng, B., Vuoskoski, J. K., & Jensenius, A. R. (2020). Who moves to music? Empathic concern predicts spontaneous movement responses to rhythm and music. *Music & Science*, 3, 1–16. doi:10.1177/2059204320974216
- Zhang, Y., Baills, F., & Prieto, P. (2020). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*, 24(5), 666–689. doi:10.1177/1362168818806531
- Zhang, Y., Baills, F., & Prieto, P. (2023). Singing songs facilitates L2 pronunciation and vocabulary learning: A study with Chinese adolescent ESL learners. *Languages*, 8(3), 219. doi:10.3390/languages8030219

Yuan Zhang is a New Hundred Talents Program researcher at Zhejiang University, China. Her research interests include second language acquisition and the interaction between music and language, especially the role of prosody and body movement on pronunciation skills. She is currently investigating the role of AI-based language learning tools on pronunciation skills.

Florence Baills is a lecturer/assistant professor at the University of Lleida in Catalonia, where she teaches French and translation. Her research interests lie in the interdisciplinary study of prosody, gesture/body movements, and second language acquisition, particularly oral pragmatic skills and pronunciation. Her primary contribution has been to extend the embodied cognition paradigm to phonological processing and demonstrate the positive effect of embodied prosodic techniques that incorporate hand gestures and kinaesthetic movements on L2 pronunciation. Currently, she is exploring the pragmatic role of head gestures and how these non-manual gestures are paired with prosody to encode prominence and make meaning.

Pilar Prieto is an ICREA Research Professor at the Department of Translation and Language Sciences at Universitat Pompeu Fabra, Barcelona, Catalonia. Her research focuses on the communicative role of prosody and gesture in language, as well as their significance in language development and learning. She currently serves as associate editor of the journals *Language and Speech* and *Frontiers in Communication*.