Article

LANGUAGE TEACHING RESEARCH

Language Teaching Research
1–31

DOI: 10.1177/1362168820986716
journals.sagepub.com/home/ltr
SAGE

Embodying rhythmic properties of a foreign language through hand-clapping helps children to better pronounce words

Florence Baills
Universitat Pompeu Fabra, Spain

Pilar Prieto
Universitat Pompeu Fabra, Spain; Institució Catalana de Recerca i Estudis Avançats (ICREA), Spain

Abstract
This study tested the effects of hand-clapping to the rhythm of newly learned French words on the pronunciation of these words by 7- to 8-year-old Catalan children. In a short training experiment with a pre- and posttest design, 28 children either repeated cognate words in French (e.g. French *aspirateur*, Catalan *aspirador* 'vacuum cleaner') while clapping to the rhythmic structure of those words or only repeated the words. Participants' oral productions before and after training were rated for accentedness by three French native speakers. Results showed that in both groups, participants' pronunciation improved after training, and crucially, children in the clapping group improved significantly more than those in the non-clapping group. Additionally, an acoustic analysis of the duration of word-final vowels indicated that only children in the clapping group significantly lengthened the final vowel after training, thus producing more target-like durational patterns. Our results suggest that a brief embodied intervention based on highlighting the rhythmic structure of words through hand-clapping has the potential to enhance pronunciation in a foreign language. The implications for second language teaching of pronunciation are discussed.

Keywords
Catalan-French cognates, embodied learning, hand clapping, L2 pronunciation, rhythm, rhythmic training

Corresponding author:
Florence Baills, Universitat Pompeu Fabra, Roc Boronat 138, Barcelona, 08018, Spain
Email: florence.baills@upf.edu

I Introduction

Many studies have shown the importance of rhythm perception in language development (Johnson & Jusczyk, 2001; Morgan & Saffran, 1995; for a review, see Gordon et al., 2015a) and language processing (Magne et al., 2007; Pitt & Samuel, 1990; Roncaglia-Denissen, Schmidt-Kassow & Kotz, 2013). The importance of rhythmic abilities for language learning has been assessed regarding various competencies, especially in children. For example, rhythmic abilities have been shown to influence children's syntactic competency (Gordon et al., 2015b). The accurate perception of language rhythmic structure has also been claimed to be crucial for phonological development and the processing of word metric structure (Goswami et al., 2002), as well as for speech intelligibility (Zion Golumbic, Poeppel & Schroeder, 2012). Further evidence shows that reading struggles in children are related to underlying difficulty in neural rhythmic entrainment, which can be detected by impaired auditory rhythm perception (Corriveau & Goswami, 2009; Goswami, 2011) and impaired musical beat perception (Goswami et al., 2013).

Below, we review the literature showing how rhythmic priming and rhythmic training can facilitate speech processing and help children improve phonological awareness (e.g. Cason et al., 2015b), as well as overcome reading difficulties (e.g. Bhide, Power & Goswami, 2013; Nelson, 2016). The present article assesses the potentially beneficial effects of rhythmic training through hand-clapping on another area of language learning which has been less investigated, namely the learning of foreign language pronunciation by children.

1 Effects of rhythmic priming on speech processing

There is growing evidence that rhythmic priming is beneficial for different aspects of language processing in adults. Falk, Lanzilotti and Schön (2017a) presented participants with sentences in French which were preceded by matching or non-matching musical rhythmic priming and observed that phase coupling, as measured by EEG (electroencephalography), was enhanced by the rhythmic auditory input when the latter was coupled with accented syllables. Their findings support the hypothesis that rhythmic cues mapping onto speech metrical structure enhance temporal expectancy and facilitate the processing of upcoming events in speech at predicted times (Falk & Dalla-Bella, 2016; Falk, Volpi-Moncorger & Dalla Bella, 2017b; Kotz & Gunter, 2015).

Other priming studies by Cason and collaborators have shown that the phonological processing of speech by adult participants is enhanced by the temporal expectancy generated by a musical rhythmic prime (Cason & Schön, 2012; Cason et al., 2015a). First, Cason and Schön (2012) presented French participants with matching and mismatching percussive rhythmic primes followed by nonwords respecting French phonotactics, and asked them to state whether a target phoneme had been pronounced in the nonword. Behavioral measures in the form of reaction times (RTs) showed that target phonemes were detected faster when positions matched the prime beat. Additionally, when a beat expectancy violation occurred, ERP measurements (event-related potentials, also obtained by EEG) showed a larger-amplitude and longer latency response at P300. These findings were successfully reproduced in a follow-up study (Cason et al., 2015a) with spoken sentences in French preceded by a prime musical meter to induce metrical expectancy about both stress patterns and the number of syllables. Additionally, in this study, a group of participants underwent a short audio-motor training session several times during the experiment (just before and halfway through each block) which consisted of repeating vocally the prime rhythm using different sounds to distinguish between strong and weak musical beats. The results revealed that the priming effect was enhanced by the audio-motor training.

2 Benefits of rhythmic training

The benefits of rhythmic training on children's developing phonological and reading skills have been investigated thoroughly. For example, Bhide et al. (2013) compared the effect of a two-month rhythmic nonverbal training program to the effect of rhyme-based training software on the reading and phonological skills of 19 children aged 6 and 7 who were considered poor readers. The rhythmic training consisted of activities such as tapping in time to a metronome, differentiating between tempos and rhythm, mimicking a rhythmic sequence, clapping or marching to a song or playing hand-clap games. The results showed that after intervention, the reading and phonological skills of participants in both training conditions improved with comparable effect sizes. Additionally, the authors found a strong correlation between children's improvement in rhythmic entrainment as an effect of the intervention and improvement in the overall reading score between pre- and posttest. These results suggest that interventions using purely musical rhythms may have a positive impact on reading skills. Similarly, Nelson (2016) integrated rhythmic activities into an 8-week literacy intervention program and found better results in rhyme awareness for the preschoolers that followed that program compared to those that followed regular classroom activities.

3 Rhythmic training for L2 phonological development

Second language teachers regard pronunciation as an important aspect of language to be mastered by learners in order to achieve successful communication (e.g. Nagle et al., 2018). Numerous studies on pronunciation instruction show the positive effect of overtly teaching pronunciation to foreign and second language learners (for reviews, see Lee, Jang & Plonsky, 2015; Saito, 2012). 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 (for a review, see Gordon & Darcy, 2016; Thomson & Derwing, 2015). However, recent work has pointed to the need for L2 prosodic instruction, as having non-target prosody in the L2 affects negatively accentedness, comprehensibility and intelligibility (Anderson-Hsieh, Johnson & Koehler, 1992; Kang, Rubin & Pickering, 2010). In this context, 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 et al., 1998; Derwing & Rossiter, 2003; Gordon et al., 2013; Behrman, 2014).

Little is known about whether rhythmic training activities can enhance phonological awareness and pronunciation in a second language teaching context. Several complementary lines of evidence lead us to think that a short rhythmic intervention can enhance second language production patterns, including pronunciation. First, various studies have demonstrated that musical aptitude, more particularly rhythmic receptive and productive abilities, are correlated with phonological abilities and pronunciation in a foreign language (e.g. Arellano & Draper, 1972; Cohrdes, Grolig & Schroeder, 2016; Gilleece, 2006; Milovanov et al., 2008; Morgan, 2004; Nardo & Reiterer, 2009; Slevc & Miyake, 2006). Second, there is evidence that rhythmic priming has immediate positive effects on the phonological production skills of hearing-impaired children speaking their first language. For example, Cason et al. (2015b) looked at the effect of rhythmic priming on the oral accuracy of 14 hearing-impaired children with cochlear implants. In this study, children had to repeat the prime vocally and then immediately pronounce the sentence. As in the previous experiments, the primes either matched or mismatched the metrical structure of the target sentences. A comparison of the children's oral production before and after the priming session showed significantly improved pronunciation accuracy for both vowels and consonants as well as syllable and word accuracy in the matching condition only, suggesting that rhythmic priming enhances phonological production.

More research is needed on the potential positive effects of rhythmic training on L2 pronunciation. To our knowledge, only a few studies from different domains of research have been conducted, exploring the potential benefits of rap music (Fischler, 2009), a computer-based rhythm generator (Wang, Mok & Meng, 2016), rhythmic beat gestures (Gluhareva & Prieto, 2017; Kushch, 2018) and hand-clapping (Iizuka, Nakatsukasa & Braver, 2020; Zhang, Baills & Prieto, 2018) on L2 pronunciation, with mixed results. During a four-week intensive course, Fischler (2009) taught sentence and word stress in English to six advanced adolescent learners with different L1 backgrounds through activities related to rhythm and rap music. A qualitative analysis of the number of errors in stress placement and of intelligibility during reading and narrative-picture tasks before and after training showed a general improvement for the reading task only. However, in the absence of a control group, the author could not claim that the participants benefited specifically from the training method. Following a different approach, Wang et al. (2016) tested the effect of a computer application that automatically generated a percussive rhythm on the pronunciation of sentences by 20 Chinese learners of English. Participants were asked to pronounce 15 English sentences before and after the rhythmic cue. Only those who obtained the lowest scores in terms of native-likeness before the rhythmic priming significantly improved their pronunciation. Adopting another approach, Gluhareva and Prieto (2017) tested whether the observation of rhythmic beat gestures, simple up-and-down or back-and-forth hand movements naturally coordinated with the prominent parts of speech, was beneficial for the pronunciation of English sentences by Catalan intermediate learners during a short training session. The results pointed to a positive effect of rhythmic beat training on elicited semi-spontaneous speech in terms of accentedness reduction.

The facilitating effect on the pronunciation of words of marking syllables by hand-clapping, an activity that lends itself very easily to the classroom context, has been investigated only recently in two studies, with mixed results. First, a study by Zhang et al. (2018) in which, during a short audiovisual training session, two groups of 25 Chinese adolescents repeated unknown French words while either clapping out their rhythmic structure or not. Accentedness ratings of participants' oral production before and after training showed only a near-significant difference in improvement between the two groups. However, acoustic analysis of final rhyme duration indicated that participants in the clapping group lengthened the final syllable more appropriately than did participants who were not trained to clap, indicating that hand-clapping helped participants acquire the rhythmic structure of the words. However, in this study, participants had to learn the meaning and pronunciation of words at the same time, rendering it not possible to determine whether the effects of clapping on pronunciation might not have been negatively impacted by cognitive overloading. Second, Iizuka et al. (2020) assessed the effect of watching and performing hand-clapping of Japanese moras on the perception and pronunciation of long vowels, geminates and moraic nasals presented in loanwords by adult English native speakers and found a significant benefit of hand-clapping for the perception of these segmental features in a delayed posttest. However in this study, despite reducing the cognitive load of meaning retrieval by using loanwords, the results of the production task failed to show a superior effect of repeating words with hand-clapping compared to repeating speech only. Overall, given the mixed results obtained in the literature, further research is needed to empirically test the effects of hand-clapping on L2 pronunciation. In addition, to our knowledge no previous study has assessed the role of a short hand-clapping training session on L2 pronunciation patterns in children.

II The present study

The aim of the present study was to assess whether a short training session using hand-clapping to highlight the rhythmic structure of words can improve the pronunciation of newly learned cognate words in French. The 28 Catalan-speaking children who participated in the training session had no prior knowledge of French. Crucially, the 20 items chosen for the training session were French-Catalan 'cognates', that is, words with identical meanings and similar forms, like *avion* / *avió* 'airplane'. This was done deliberately on the grounds that it would facilitate word recall (de Groot & Keijzer, 2000) and allow participants to focus exclusively on pronunciation rather than word meaning, thus avoiding the potential cognitive overload present in the study by Zhang et al. (2018) noted above. Importantly, while the transparency of meaning offered by cognates can facilitate comprehension and memorization, the similarity in phonological forms may enhance phonological transfer from their L1, thus penalizing pronunciation (Flege, 1987).

Catalan is considered a stress-accented language in which lexically stressed syllables generally serve as the main landing site for phrasal pitch accents. Word stress is realized on one of the last three syllables of the morphological word, and at the phrasal level, the last content word in the intonational phrase receives the main phrasal stress (Prieto et al., 2015). Unlike Catalan, French has no lexical stress and is considered to be an edge-prominence language. Stress is assigned at the phrasal level, as follows: (1) an obligatory phrase-final primary stress is generally assigned to the last metrical syllable of a content word and has a demarcative function which marks the right edge of a prosodic phrase; and (2) an optional secondary stress can be assigned phrase-initially (see, among others, Di Cristo & Hirst, 1993; Jun & Fougeron, 1995, 2000; Delais-Roussarie et al., 2015).

In Catalan and in French, as in many other Romance Languages, the nuclear accent falls at the end of the sentence (Nuclear Stress Rule: Halle & Vergnaud, 1987; see Frota & Prieto, 2015). However, regarding the phonetic properties of stress realization, there seems to be a basic difference between the two languages which affects the duration of the stressed syllable. In comparison with Catalan, French stress is realized by a more extreme lengthening of the stressed phrase-final syllable and more particularly of the full vowel (Astésano, 2001; Delattre, 1966; Di Cristo & Hirst, 1993; Fletcher, 1991; Vaissière, 1991). To our knowledge, although thus far no study has systematically analysed cross-linguistic differences in final lengthening between Catalan and French, two types of acoustic evidence point to a difference in final lengthening patterns between these two languages. First, acoustic comparisons between French and Spanish (a language with a rhythmic structure and durational patterns that are similar to Catalan; see Prieto et al., 2012) tend to point to more exaggerated final lengthening patterns in French. While Rao (2010) found that final lengthening patterns before a pause in three different varieties of Spanish may reach an average of 30%, studies scrutinizing final lengthening in French (Bartkova et al., 2012; Zellner, 1996) have found as much as a 50% increase in syllable-final durations. However, in these studies, syllable structure was not controlled for, the speakers' samples for the analysis were small, and, importantly, final lengthening was calculated for paroxytone words only. In a recent study with a large dataset (15 hrs of speech), Gendrot, Adda-Decker & Santiago (2019) compared the duration of final vowels produced at the right edges of Intonation Phrases in both French and Spanish. The results showed that, in oxytonic positions, French vowels tend to be longer than Spanish vowels. Second, for the purpose of this study, an exploratory acoustic analysis was carried out which compared the duration of stressed and unstressed syllables in 20 pairs of cognate words in Catalan and French (i.e. *balcó – balcon* 'balcony'). The Catalan words (N = 20) were pronounced by two 8-year-old speakers of Catalan and the French words (N = 20) were pronounced by two 8-year-old speakers of French (see Table 1). The mean ratio between stressed and unstressed syllables was calculated for each language and for the three stress positions (e.g. oxytonic, paroxytonic and proparoxytonic positions) in Catalan. We found that in French, accented syllables were 1.91 times longer than the preceding unstressed syllable, whereas in Catalan, accented syllables were 1.75 times longer than the preceding unstressed syllable for paroxytones, 1.28 times longer for paroxytones and 1.65 times longer than the following syllable for proparoxytones (for the description of the procedure, see Figure 1; see also Appendix A in supplemental material).

For the purposes of the present study, the realization of the more extreme lengthening patterns in the final stressed syllable may be of crucial importance for the production and perception of French prosodic phrasing for L2 learners. Schwab (2012) analysed the production of adult intermediate Spanish learners of French and found some evidence that they transferred Spanish stress realization when speaking French. In a follow-up study, Schwab (2013) showed that adult intermediate Spanish speakers were able to produce the intended stressed syllable at the right edge of the accentual phrase, marking the

Table 1. Target French words for the training session and their Catalan cognates.

French	Catalan: oxytone	English gloss	French	Catalan: paroxytone + proparoxytone	English gloss
balcon [balˈkɔ̃]	*balcó* [bəɫˈko]	'balcony'	*oreille* [ɔˈʁɛj]	*orella* [uˈɾɛʎə]	'ear'
tambour [tɑ̃ˈbuʁ]	*tambor* [təmˈbo]	'drum'	*famille* [faˈmij]	*familia* [fəˈmiljə]	'family'
purée [pyˈʁe]	*puré* [puˈɾe]	'puree'	*musique* [myˈzik]	*música* [ˈmuzikə]	'music'
avion [aˈvjɔ̃]	*avió* [əˈβio]	'airplane'	*pizza* [piˈdza]	*pizza* [ˈpid͡zə]	'pizza'
crocodile [kʁɔkɔˈdil]	*cocodril* [kukuˈðɾil]	'crocodile'	*confiture* [kɔ̃fiˈtyʁ]	*confitura* [kuɱfiˈtuɾə]	'jam'
biberon [bibˈʁɔ̃]	*biberó* [biβəˈɾo]	'baby bottle'	*spaghettis* [spageˈti]	*espaguetis* [əspəˈɣetis]	'spaghetti'
céréales [seʁeˈal]	*cereals* [səɾəˈals]	'cereals'	*éléphant* [eleˈfɑ̃]	*elefant* [eləˈfan]	'elephant'
aspirateur [aspiʁaˈtœʁ]	*aspirador* [əspiɾəˈðo]	'vacuum cleaner'	*mandarine* [mɑ̃daˈʁin]	*mandarina* [məndəˈɾinə]	'tangerine'
télévision [televiˈzjɔ̃]	*televisió* [tələβiˈzjo]	'television'	*ambulance* [ɑ̃byˈlɑ̃s]	*ambulància* [əmbuˈɫansjə]	'ambulance'
ordinateur [ɔʁdinaˈtœʁ]	*ordinador* [uɾdinəˈðo]	'computer'	*hélicoptère* [elikɔpˈtɛʁ]	*helicòpter* [əliˈkɔptəɾ]	'helicopter'

stressed syllable by means of variations in duration and F0. These results were confirmed by Santiago and Mariano's (2019) study analysing a corpus of adult intermediate Spanish learners of French. However, additional empirical studies are needed to examine the exact phonetic realization of French final lengthening by L2 learners. Learners with less knowledge of the target language or exposed to less input may benefit from specific training to speed up the acquisition of a more extreme durational production of word-final syllables in French. In addition, a series of recent studies have shown that L2 learners' general pronunciation may be improved by training them in the production of rhythmic prosodic features (e.g. Gluhareva & Prieto, 2017; Li, Baills & Prieto, 2020; Yuan et al., 2019), corroborating the idea that suprasegmental features count as a major factor in measures of accentedness and perception of oral proficiency (see, for example, Kang et al., 2010).

The rhythmic training proposed in this study consisted of audiovisually highlighting the rhythmic structure of words through hand-clapping. We hypothesized that hand-clapping can serve to acoustically and visually highlight the prosodic patterns of speech. Acoustically, the clapping sound will auditorily highlight the syllabic structure of the target words. Visually, the fact that the hands stay longer together on the stressed syllable call attention to the longer duration of this syllable. Embodying these prosodic patterns might reinforce the phonological learning process by increasing phonological awareness, which can ultimately lead to better pronunciation as measured by accentedness ratings and acoustic analysis. We surmise that the effect of the training might be detectable through an

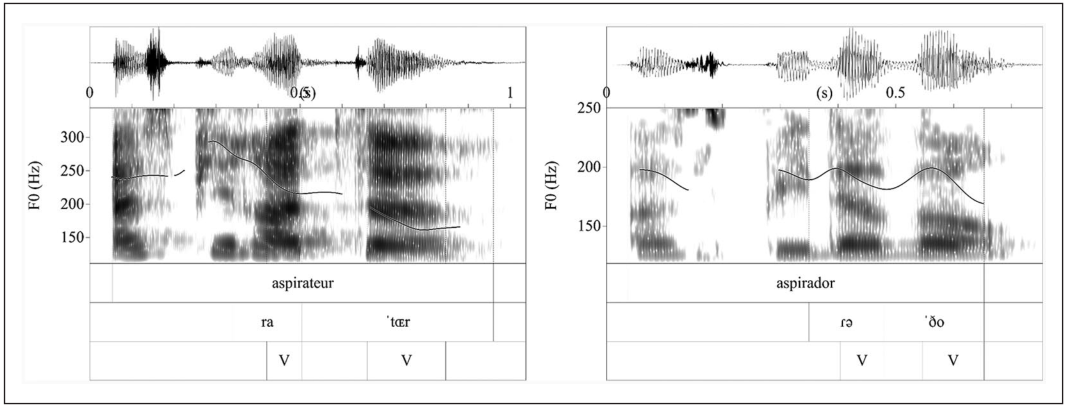

Figure 1. Acoustic representation of the French–Catalan cognate pair of words *aspirateur / aspirador* ‘vacuum cleaner’. The comparison of vowel duration measures show that while the French stressed syllable [tœʁ] in aspirateur is 2.22 times longer than the preceding syllable [ʁa] (left panel), the Catalan stressed syllable [ðo] is 1.4 times longer than the preceding syllable [ɾə] (right panel).

acoustic analysis that can assess a more target-like duration of the stressed syllable by L2 learners.

1 Hand-clapping as prosodic embodiment

Hand-clapping is intrinsically related to the concept of rhythm and, as such, falls simultaneously within the two domains of music and language. It is present throughout infancy in the form of hand-clapping games and songs, and can be observed in numerous cultures (Cameron & Grahn, 2014; see also Romero Naranjo, 2013). Clapping one's hands, like tapping one's foot or dancing to musical rhythms, is a natural way to express the temporal structure of music with body movements (Repp & Su, 2013). There are reasons to believe that the reinforcement by means of a motor action (e.g. clapping) can be helpful for L2 learners to better process and produce a prosodic feature of speech (in this case the rhythmic structure of words) that may be difficult for them to acquire.

According to the theory of grounded cognition, the activation of appropriate perceptual and motor interactions during learning should enhance the development of cognitive functions (Borghi & Caruana, 2015). Indeed, neuroscientific studies have shown that not only perception but also motor brain networks are activated when participants engage in different tasks involving abilities such as memory, knowledge, language and thought (for a review, see Barsalou, 2008). Embodied theories of language processing suggest that motor action and semantic processing are closely interrelated (see, e.g. Glenberg & Kaschak, 2003; Zwaan & Taylor, 2006) and that the execution of motor actions has a selective effect on the linguistic processing of words (Rueschemeyer et al., 2010). In the field of gesture studies, research has shown that learners achieve better results in different memory and cognitive tasks when producing hand gestures than when merely observing them (Goldin-Meadow, 2014; Goldin-Meadow, Cook & Mitchell, 2009; for a review of the effects of enactment and gestures on memory recall, see Madan & Singhal, 2012). There is also evidence from neurophysiological research that self-performing a gesture when learning verbal information favors the formation of sensorimotor networks that contribute to the representation and storage of words in a native language (Masumoto et al., 2006) as well as foreign language (Macedonia, Müller & Friederici, 2011).

The implications of the benefits of embodiment are crucial for education (for reviews, see Kiefer & Trumpp, 2012; Wellsby & Pexman, 2014). Embodied approaches to music pedagogy are a good illustration of how body movements facilitate the understanding and enhance the retention of complex musical concepts (Juntunen, 2016). In the field of the acquisition of L2 phonological patterns, a recent series of studies on pitch gestures also show the benefits for word recall and the perception of pitch information of watching and producing up-and-down hand movements that represent rising and falling pitch (Baills et al., 2019; Kelly, Bailey & Hirata, 2017; Morett & Chang, 2015). Such prosodic hand gestures also seem to facilitate the production of difficult pitch contours in a foreign language by tonal language speakers (Yuan et al., 2019).

2 Individual differences in pronunciation learning

Since the main goal of the present study was to assess the role of hand-clapping in second language pronunciation learning, three types of individual measures related to working memory, speech imitation skills and musical abilities were taken into account, as they have been shown to play an important role in L2 learners' pronunciation. First, phonological working memory has attracted attention as a contributing factor to pronunciation talent (Aliaga-Garcia, Mora & Cerviño-Povedano, 2011; Darcy, Park & Yang, 2015; Rota & Reiterer, 2009). Second, speech imitation talent and pronunciation skills in a foreign language have been shown to be highly interdependent in research by Nardo and Reiterer (2009). Reiterer et al. (2013) also found that speech-motor flexibility may be among the best predictors of speech imitation capacities, leading to better pronunciation in an unknown language. Speech imitation abilities would then be of major importance when assessing learners' pronunciation. Finally, some studies have shown that musical abilities are related not only to receptive but also to productive phonological learning skills in an L2 (Delogu & Zheng, 2020; Milovanov et al., 2008, 2010; Slevc & Miyake, 2006).

III Method

In a between-participants training study with a pre- and posttest design, a group of 28 Catalan children were asked to learn 20 new cognate French words under one of two audiovisual conditions: (1) training which involved observing and replicating the behavior of a native speaker simultaneously saying a word and clapping to highlight the prosodic structure of the word; or (2) training which involved observing and replicating a native speaker who merely spoke the word without clapping. We hypothesized that observing and subsequently performing hand-clapping would lead to a greater improvement in pronunciation of the French words both in terms of perceived accentedness ratings and in terms of acoustic patterns (e.g. a more native-like lengthened production of the words' final rhyme and final vowel).

1 Participants

Twenty-eight 7- to 8-year-old children from the city of Girona, Catalonia, took part in the experiment at their school premises after their parents signed a written consent form. They were all Catalan-Spanish bilinguals with Catalan as their dominant language (percentage of time Catalan used in their daily life: $M = 87\%$, as reported by participants' caregivers). None of them had any prior knowledge of French. They were informed that they would learn words in French and were randomly divided into two groups, namely the clapping group ($n = 14$, $M_{age} = 7.43$; $SD = 0.5$, 7 females) and the non-clapping group ($n = 14$, $M_{age} = 7.29$; $SD = 0.46$, range 7 females).

2 Materials

a Training session. Materials for the training session consisted of two 10-minute videos prepared at the professional broadcasting studio of the Universitat Pompeu Fabra in Barcelona. For both conditions, the videos were designed to teach the 20 target French words with two instructors (see Table 1).

The rationale for the selection of words was that (1) their meaning should be transparent to Catalan speakers (e.g. Catalan *avió* / French *avion* 'plane', Catalan *ordinador* /

French *ordinateur* 'computer'); and (2) they should name objects that would be easy to represent by means of a simple black and white line drawing in order to avoid any written input.[1] A variety of consonantal environments were proposed, mainly constrained by the obligation to work with cognates in the two languages and to maintain the number of syllables constant.

Crucially, though the target French words were all cognates, they included a variety of sounds in the target language that are not part of the Catalan sound inventory, such as the labiodental [v] and the uvular rhotic [ʁ] for consonant sounds, as well as the rounded front vowels [y] and [œ] and nasal vowels [ɑ̃] and [ɔ̃]. Importantly, apart from these segmental differences, a salient phonological feature of all the French words as compared to their corresponding Catalan cognates was the presence of a phonetically strong lengthened phrasal-final stress in French, which would compete with lexical stress in Catalan. For ten of the French items, stress was located in the same position as in their Catalan oxytone cognates (*balcon* – *balcó*). For the ten remaining items, the Catalan counterparts were nine paroxytones (*oreille* – *orella*) and one proparoxytone (*musique* – *música*). In these cases, the stress was either on the same syllable (6 items) or on a different syllable (4 items) (see Table 1). The words included two-syllable words (8 items, 4 oxytones, 3 paroxytones, 1 proparoxytone), three-syllable words (8 items, 3 oxytones, 5 paroxytones) and four-syllable words (3 oxytones, 1 paroxytone).

The two video stimuli (the non-clapping video and the clapping video) were prepared as follows. Two female native French speakers were video-recorded producing all 20 target words in Table 1 as if speaking to a class of learners in a very clear manner. A total of 80 videos were recorded in this fashion (20 words × 2 instructors × 2 conditions).

For the clapping stimuli, instructors first spoke one word without moving their hands, and immediately repeated the same word while simultaneously clapping once on each of the target syllables of the word and then returning their hands to their rest position. For each target word, each syllable was marked by a regular hand-clapping sound, highlighting its syllabic structure. Visually, the duration of the hand-claps highlighted the prosodic prominence patterns, with the syllables preceding the last stressed syllable not bearing any prosodic emphasis. A frame-by-frame analysis of the 20 clapping videos showed that the hands remained in contact longer in the last syllable before returning to the rest position (M = 0.499 sec, SD = 0.249, 95% CI [0.419, 0.578] for the final syllable, M = 0.069 sec, SD = 0.026, 95% CI [0.062, 0.074] for the other syllables), thus visually highlighting the longer duration of the final syllable. We asked the instructors to avoid using a higher clapping intensity (volume) on the stressed syllable for two reasons: first, because it might interfere with perception of the speech signal; and second, because stressed syllables in French are not characterized by higher intensity, and louder clapping might have cued participants to use intensity instead of duration to mark stress. Intensity was further measured at each hand-clap in the 40 target stimuli to ensure equivalent levels (M = 78.053 dB, SD = 3.225, 95% CI [76.802, 79.304] for the final syllables, M = 75.364 dB, SD = 6.521, 95% CI [73.470, 77.257] for the other syllables). We also checked that the sound produced by the clapping at no time masked the voice of the instructor.

For the non-clapping stimuli, the two instructors spoke the words twice without moving their hands. No pause was produced between syllables. Moreover, whether they were accompanying their speech with claps or not, the two instructors were asked to use natural head movements and facial expressions while they spoke but refrain from emphasizing stressed syllables by head, eyebrow or chin movements.

In order to check that speech rate and word duration did not differ between the spoken and hand-clapped words, the first author of this study extracted the soundtracks of the 80 items in the training videos produced by the two instructors for the two conditions and labeled those items in Praat (Boersma & Weenink, 2017). The values for speech rate and word duration were automatically extracted by using Praat scripts (de Jong & Wempe, 2009; Elvira-Garcia, 2014[2]). Potentially significant differences between the speech rate and syllable duration patterns across the two conditions (clapping vs. non-clapping) were tested by means of two independent sample *t*-tests. The speech rate of the instructors' production of the target items in the clapping condition (M = 1.96 sec, SD = 0.44) compared to the speech rate of the instructors for the non-clapping condition, M = 1.90 sec, SD = 0.42) did not differ significantly ($t(65) = -0.602$, $p = .549$, 95% CI [–0.271, –0.145]. Similarly, no significant differences were found for word duration, $t(65) = 0.888$, $p = .378$, 95% CI [–0.075, 0.197], between the clapping condition (M = 1.42, SD = 0.28, 95% CI [1.32,1.52]) and the non-clapping condition (M = 1.48 sec, SD = 0.27, 95% CI [1.38, 1.58]).

Each of these 80 video recordings was embedded between two still sequences. The first, which lasted for 3 seconds, showed a black and white drawing illustrating the target French word about to be spoken. The still sequence following the stimuli lasted for 5 seconds and showed a black screen with no image (see Figure 2). This 5-second blank screen was intended to give the viewer time to replicate what they had seen and heard, depending on the group to which they had been assigned. If they had been assigned to the clapping group, they would repeat the word and clap as they had seen it done (Figure 2, top panel). If they had been assigned to the non-clapping condition, they would merely repeat the word as they had heard it spoken (Figure 2, bottom panel).

In order to balance the presence of the two speakers, for each condition, two blocks were created with a total of 20 sequences/items with the two speakers appearing an equal number of times, making sure that each consecutive item was produced by a different instructor. All in all, the 20 target words were trained twice, each time with a different instructor. To ensure variability in order of presentation of the stimuli across participants, six videos with different orders of presentation of the target items were created for each condition.

b Pre- and posttest materials. To test participants' pronunciation and to avoid the influence of written input, a word repetition task was created. The materials required for the pre- and posttest word repetition task consisted of 15 French words which the participants would hear and then repeat. Ten of these words were words related to the training session (*balcon, tambour, musique, purée, ambulance, crocodile, biberon, mandarine, confiture, aspirateur*; see Table 1) and merely consisted of the audio tracks from the video recordings by the two instructors. The other five words were completely new (*calendrier* 'calendar', *garderie* 'kindergarten', *sportive* 'sporty', *imprimante* 'printer', *cheveu* 'hair', *râteau* 'rake') and were not cognate words. In this case new audio recordings were made by one of the two native speakers that featured in the videos.

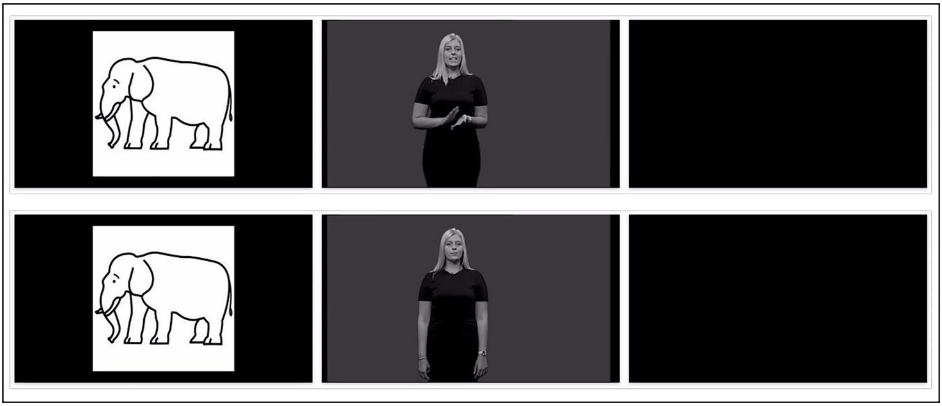

Figure 2. Stills from the training video for the word *éléphant* 'elephant' in the clapping condition (top panel) and non-clapping condition (bottom panel).

Sixteen different orders of presentation of the 15 words were created using a presentation software and participants were assigned to different combinations of these orders at pre- and posttest. The audio files were directly embedded in the presentation software.

c Individual measures. As noted above, in order to control for the effect of individual differences related to cognitive, linguistic and phonological abilities, the following five tasks were administered:

1. Short-term memory task: Short-term memory was assessed through a memory span task where participants had to repeat different lists of Catalan words (see Appendix B in supplemental material) ranging from three to six words (Bunting, Cowan & Saults, 2006).
2. Imitation talent task: Participants' imitation abilities were tested through a word repetition task involving 12 words in six different languages. The items were 2- or 3-syllable words containing segmental information that can be considered difficult for Catalan speakers and which are not part of the consonantal and vocalic inventory of the Catalan language (see Appendix C in supplemental material).
3. Phonological perceptual ability task: Participants undertook a standard phonological discrimination task for children whereby they had to listen to pairs of French nonwords and decide if they were the same or different (Macchi et al., 2013).
4. Rhythmic perceptual ability tasks: Participants undertook two standard discrimination tests for musical rhythm (8 items) and musical accent (10 items) extracted from a free musical perception test called PROMS that can be tailored for children in terms of the number and difficulty of items (Law & Zentner, 2012). The procedure followed for both tests was the same, namely the children listened to one sequence twice and then to a last sequence which they had to qualify as same, different, or unsure. While the rhythm subtest consists of discriminating among

simple patterns of quarter notes, eight notes, and sixteenth notes, the musical accent subtest assesses the ability to distinguish the relative emphasis given to certain notes in a rhythmic pattern. As such, it is related to the concepts of meter in music and stress in speech (for a detailed description of the subtests, see Law and Zentner, 2012).

5. Rhythmic production ability task: In this hand-clapping replication test, participants heard a rhythm sample and immediately had to replicate the rhythm by clapping (six samples, all 4/4 time, 2 measures).

3 Procedure

The experimental procedure, which consisted of pretest – training session – posttest, lasted 20 minutes (see Figure 3). The experiment was carried out with each child individually in a quiet room at their school. The child was seated in front of a tablet computer and wore a comfortable high quality headset equipped with a high quality microphone. The first author of the study remained in the room with the child to ensure that the training session was completed successfully. As noted above, prior to the training as such, each of the 28 participating children took a battery of five tests to measure their cognitive, linguistic and phonological abilities in a separate session that lasted around 40 minutes. A research assistant helped the first author to collect the individual measures.

The pretest consisted of a short word repetition task. Children were asked to touch a key to play audio recordings of 15 French words, one at a time, repeating each word before moving on to the next. They were thus able to set their own pace. The duration of the pretest was roughly 5 minutes. The participant's speech was recorded throughout the pretest.

After the pretest, participants watched one of the two training videos (clapping or non-clapping) depending on which group the child had been randomly assigned to. Children assigned to the clapping group were randomly exposed to one of the six variants of the clapping stimulus video. For each trained word in the clapping condition, children first saw the drawing depicting the word, then watched the video clip of the instructor producing the word while clapping her hands, and finally (as they viewed the empty black screen) repeated the word while also clapping their hands. By contrast, children in the non-clapping group saw the drawing, watched the instructor producing the word without clapping, and simply repeated the word. As the stimulus video consisted of two blocks, each child was exposed to each item twice.

The total duration of the training session was roughly 10 minutes. When the training session finished, children performed the posttest word repetition task, which was identical to the pretest task. Their verbal output was likewise recorded during the posttest. The full procedure, including the prior individual measures testing, pretest, training session and posttest, lasted a total of approximately 60 minutes.

4 Data coding

The collected data underwent two types of analyses, namely (1) perceived accentedness as judged by three native French speakers listening to the recordings, and (2) acoustic measures of the final rhyme duration and final vowel duration. The results of the five

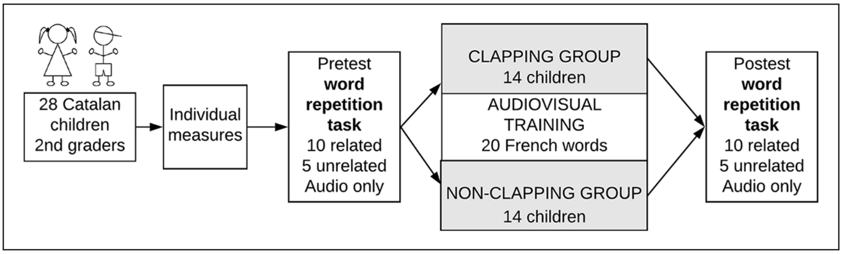

Figure 3. Experimental procedure.

tasks designed to collect individual measures (short-term memory, imitation ability, phonological perception, rhythmic perception and production) were coded and added to the database (see below).

a Perceived accentedness ratings. The 840 audio files of the children's oral productions during pre- and posttest (15 words × 2 tests × 28 children) were rated by three non-linguist French native speakers, who were unaware of the purpose of the experiment. To avoid fatigue effects, the audios were split into four different blocks and uploaded in four different rating surveys. The raters took between 45 and 60 minutes to complete each block. They were asked to complete the task within four consecutive days. For each word, the raters listened to the original audio prompt that featured in the test and then heard the children's productions in random pretest/posttest pairs. In other words, they did not know whether a particular item came from the pretest or the posttest. The raters were asked to compare the children's productions with the original auditory stimuli and evaluate the general accentedness of the target words on a scale from 1 'not accented' to 7 'extremely accented', that is, the degree to which their pronunciation approximated the native model (Munro, Derwing & Morton, 2006). We preferred to use an accentedness measure over a comprehensibility measure because it has been shown that accentedness scores by native listeners are more closely associated with target pronunciation features (e.g. vowels, consonants, stress errors) than comprehensibility scores, which have been found to be associated with non-phonological variables like lexis or grammar (for a review, see Saito, Trofimovich & Isaacs, 2017).

Inter-rater reliability was assessed using IBM SPSS Statistics 23 by calculating the Intraclass Correlation Coefficient based on the scores given by the three raters for each item. The results pointed to a high degree of reliability ($ICC = .97$, $F(839, 1678) = 99.87$, $p < .001$, 95% CI [0.95, 0.97]).

b Acoustic analysis. Since training only had an effect on the pronunciation of trained items, as measured by accentedness ratings (see Section IV), the acoustic analysis was carried out with the trained items only, for a total of 560 audio files (10 words × 2 tests × 28 children). In order to analyse the duration patterns of the target words, word boundaries, word-final rhyme boundaries and word-final syllable boundaries of the children's oral productions at pre- and posttest were manually annotated in Praat by the first author, following Machač and Skarnitzl's guidelines (2009). Absolute duration measures were then extracted with an automatic script (Dan McCloy, original version by Mietta Lennes). This process yielded, in seconds (sec), word duration, word-final rhyme duration and word-final vowel duration. Using this data, the ratio obtained by dividing the word-final rhyme duration and word duration, and the ratio obtained by dividing the word-final vowel duration and the word duration (as a %) were calculated in order to control for speakers' speech rate differences and for differences related to word duration.

c Individual measures. The memory span score corresponded to the number of words participants remembered in at least three lists with the same number of words, regardless of the order in which the words were recalled.

- Imitation talent: The first author, a phonetician, rated participants' oral production by comparing them to the native pronunciation on a scale between 1 ('very close to target pronunciation') and 7 ('very different from target pronunciation'). For each item, the rating consisted of listening to the word pronounced by the native speaker first and then immediately to the same word pronounced by the participant. The rater compared how close to the target the sounds were produced.
- Phonological perceptual ability: The score for this task corresponded to the number of correct answers, with a maximum of 36 points.
- Rhythmic perceptual ability: The final score was automatically calculated from the online software for the two subtests.
- Rhythmic production ability: Each sequence that was accurately replicated in terms of number of beats and rhythmic pattern was coded by the first author as 1. Inaccurate replications scored 0.

5 Statistical analyses

All statistical analyses were run using IBM SPSS Statistics 23. For each model, the tests of significance were two-tailed with an alpha level of .05, and post hoc comparisons were adjusted with the Bonferroni correction. A participant-sorted database was created displaying individual measure scores per participant, their mean accentedness rating at pre- and posttest, and their mean duration ratios for final rhyme and final vowel at pre- and posttest. The individual measures for each participant were used to test for (1) potential differences between the between-participant groups, (2) potential effects of individual differences on perceived accentedness scores and (3) potential effects of individual differences on acoustic measures.

First, in order to test for homogeneity between the two groups, we ran an independent sample *t*-test with individual measures (age, short-term memory, imitation, phonological discrimination, rhythm perception, rhythm production) as tested variables and group (two levels: clapping vs. non-clapping) as the grouping factor.

Then, to explore the potential effects of individual differences on our results, three stepwise multiple regression analyses were run with mean accentedness score, rhyme duration ratio and vowel duration ratio as the dependent variables and the individual measures short-term memory, imitation, phonological discrimination, rhythm perception and rhythm production as fixed factors. Consequently, the individual differences that were found to have an effect on the accentedness results were added as covariates[3] to the models testing the effect of training on accentedness and final lengthening (see sections below).

6 Effects of training with hand-clapping and item familiarity on perceived accentedness and acoustic measures

An item-sorted database was created, displaying the three raters' accentedness scores, the duration ratio for the final rhyme and the duration ratio for the final vowel for each of the 15 items at pre- and posttest for each participant.[4] For each item it was also indicated if the word was one of the 10 words which appeared in the training sequence (which we labeled 'trained') or was one of the five that were not ('new'), and the number of syllables and the stress position were specified.

To analyse the effect of the type of training (clapping vs. non-clapping) on the accentedness ratings of participants' pronunciation, a generalized linear mixed model (GLMM) was run with accentedness as the dependent variable. Training group (two levels: clapping vs. non-clapping), session (two levels: pretest and posttest), Training group × Session, familiarity (two levels: trained vs. new), number of syllables (three levels: two, three, or four), stress position (three levels: oxytone, paroxytone or proparoxytone) and the interactions Training group × Familiarity, Session × Familiarity, Training group × Session × Familiarity were set as fixed factors. One random effects block was specified with the variables participant and item, and the covariates imitation and phonological discrimination (the ones found to have a significant effect) were added to the model.

To analyse the effect of the type of training (clapping vs. non-clapping) on the acoustic measures assessing final lengthening, two GLMMs were run with the dependent variables word-final rhyme duration ratio and word-final vowel duration ratio. Fixed and random factors were the same as for the GLMM for accentedness ratings. In addition, the interactions Training group × N of syllables, Session × N of syllables, Training group × Session × N of syllables were set as fixed factors. Rhythm perception and phonological discrimination were added as covariates.

IV Results

1 Differences between groups

An independent sample *t*-test indicated that there was no significant difference between groups for any of the five individual measures short-term memory, imitation, phonological discrimination, rhythm perception and rhythm production. These results confirmed that children in the two groups were equally distributed in terms of individual aptitudes (see Table 2).

2 Effects of individual differences on accentedness and acoustic measures

Results of the multiple regression analysis revealed that imitation and phonological discrimination abilities were significant predictors of participants' accentedness scores.

Table 2. Participants' scores on individual measures (means, SDs and confidence intervals per condition).

	Non-clapping group				Clapping group			
	M	*SD*	95% Confidence Interval		*M*	*SD*	95% Confidence Interval	
Age	7.43	0.50	7.23	7.62	7.29	0.46	7.11	7.46
Short-term memory	4.36	0.83	4.04	4.68	4.29	0.60	4.05	4.52
Imitation	4.36	1.42	3.81	4.91	4.36	1,37	3.83	4.89
Phonological discrimination	23.14	4.35	21.46	24.83	22.50	2.27	21.62	23.38
Rhythm perception	17.43	3.41	17.95	20.34	19.14	3.08	17.95	20.34
Rhythm production	4.46	2.21	3.61	5.32	4.43	1.41	3.88	4.98

Imitation and phonological discrimination scores explained 38.5% of the variance (R^2 = 0.38; F(2, 7,830) = 1.71, p = .002). Mean accentedness decreased .15 points for each point of improvement in the imitation test (β = −0.46, p = .007, 95% CI [–0.25, –0.04]) and fell 0.05 points for each additional point in the phonological discrimination test (β = −0.39, p = .019, 95% CI [–0.83, –0.01]). Consequently, as explained above, these two variables were added as covariates to the model testing the effect of training on perceived accentedness.

The results of the multiple regression analyses revealed that rhythm perception and phonological discrimination abilities were significant predictors of participants' lengthening measures. For the final-rhyme duration ratio, rhythm perception and phonological discrimination explained 1.4% of the variance (R^2 = 0.01; F(2,550) = 3.86, p = .022). The final-rhyme duration ratio increased 0.1 points for each point of improvement in the rhythm perception test (β = 0.26, p = .031, 95% CI [0.02, 0.50]) and increased 0.1 points for each additional point in the phonological discrimination test (β = 0.40, p = .02, 95% CI [0.06, 0.73]). For the word-final vowel duration ratio, rhythm perception explained 0.7% of the variance (R^2 = .007; F(1,551) = 4.92, p = .027). The final-vowel duration ratio increased 0.09 points for each point of improvement in the rhythm perception test (β = 0.20, p = .027, 95% CI [0.02, 0.38]). Consequently, as explained above, these variables were added as covariates to the models testing the effect of training on the two acoustic measures.

3 Effects of type of training and item familiarity on perceived accentedness

Table 3 and Figure 4 show the mean accentedness scores at pre- and posttest for the non-clapping and clapping groups and for trained and new items. Results of the GLMM showed significant main effects of session, F(1, 2,505) = 54.69, $p < .001$), and Training group × Session, F(1, 2,505) = 10.51, p = .001, on accentedness scores. Post-hoc analyses revealed a significant effect of session for both groups, meaning that there was a significant decrease in perceived accentedness between pretest and posttest scores in

Table 3. Mean accentedness ratings for the word imitation task across groups (clapping, non-clapping), sessions (pretest, posttest) and familiarity (trained, new).

Group	Session	Familiarity	*M*	*SD*	*SE*	95% Confidence interval	
						Inf.	Sup.
Non-clapping	Pretest	trained	4.28	1.38	0.15	3.98	4.58
		new	4.94	1.31	0.17	4.61	5.26
	Posttest	trained	3.99	1.40	0.15	3.68	4.29
		new	4.81	1.24	0.17	4.48	5.14
Clapping	Pretest	trained	4.95	1.39	0.15	4.66	5.25
		new	4.91	1.27	0.16	4.59	5.23
	Posttest	trained	3.92	1.42	0.15	3.63	4.22
		new	4.87	1.22	0.16	4.54	5.19

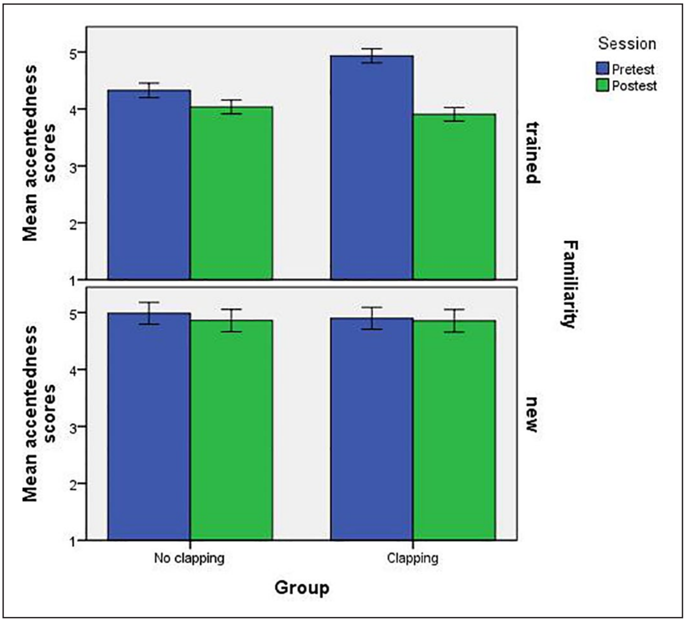

Figure 4. Mean accentedness ratings for the word imitation task across groups (clapping, non-clapping), sessions (pretest, posttest) and familiarity (trained, new).
Notes. Error Bars: +/− 2 SE.

both the non-clapping group, $F(1, 2{,}505) = 8.61, p = .003$) and the clapping group, $F(1, 2{,}505) = 56.64, p < 0.001$). The contrast estimates indicated a larger effect size in the clapping group ($\beta = 0.54, p < .001$) than in the non-clapping group ($\beta = 0.21, p = .003$). A significant difference between the clapping group and the non-clapping group

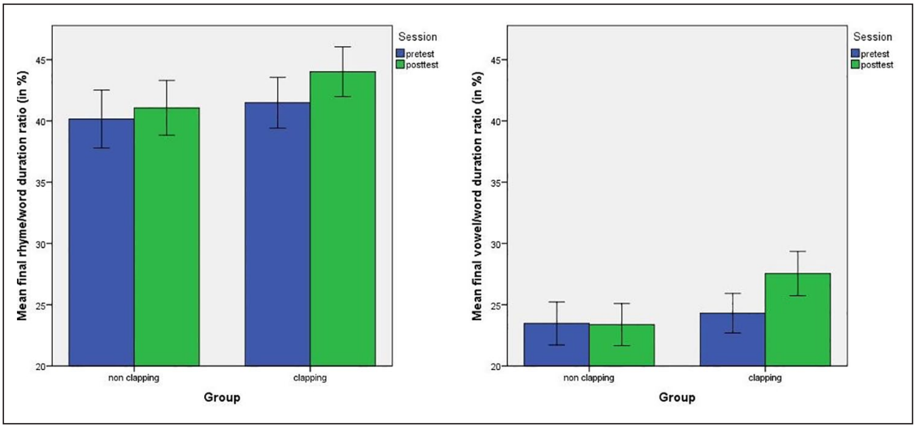

Figure 5. Mean ratio (in %) of (a) the relative duration of the final rhyme and (b) the relative duration of the final vowel, broken down by group and session.
Notes. Error Bars: +/− 2 SE.

was found at pretest, with significantly lower accentedness scores for the non-clapping group (F(1, 2,505) = 4.56, $p = .033$). However, no difference between groups was found at posttest. In other words, the non-clapping group were significantly better rated for accentedness at pretest than the clapping group; however, their improvement, although significant, did not reach the size of the improvement experienced in the clapping group.

Familiarity, $F(1, 2{,}505) = 8.57, p < .003$, Training group × Familiarity, $F(1, 2{,}505) = 7.91, p = .005$, Session × Familiarity, $F(1, 2{,}505) = 32.68$ $p < .001$, and Training group × Session × Familiarity, $F(1, 2{,}505) = 16.46, p < .001$, also had a significant effect on accentedness ratings. Post-hoc analyses revealed a significant improvement after training only for trained words in both the non-clapping group, $F(1, 2{,}505) = 12.11, p = .001$, and the clapping group, $F(1, 2{,}505) = 4.93, p < .001$, meaning that training had no effect in either group on participants' pronunciation of items they had not been trained for. Moreover, there was a significant difference between the clapping group and the non-clapping group at pretest for the trained items only ($F(1, 2{,}505) = 4.72, p = .03$).

No effect of number of syllables or stress position was found, revealing that differences between accentedness scores for the two groups were not related to word length or stress position.

4 Effects of type of training and word length on acoustic measures

The two graphs in Figure 5 show the mean final rhyme duration ratio (in %, left graph) and the mean final vowel duration ratio (in %, right graph) at pre- and posttest for the non-clapping and clapping groups (for descriptive results, see Table 4). Results of the GLMM with final rhyme duration as a dependent variable showed a significant effect of training group ($F(1, 539) = 7.62, p = .006$) and session ($F(1, 539) = 12.65, p < .001$).

Table 4. Mean duration ratios (in %) for the final vowel and final rhyme across groups (clapping, non-clapping) and sessions (pretest, posttest).

		M vowel duration ratio	*SD*	*SE*	95% CI		*M* rhyme duration ratio	*SD*	*SE*	95% CI	
					Inf.	Sup.				Inf.	Sup.
Non-clapping	Pretest	23.39	13.90	2.66	18.17	28.62	40.14	12.33	3.618	33.03	47.24
	Posttest	23.50	13.18	2.66	18.28	28.73	41.11	11.16	3.617	34.00	48.21
Clapping	Pretest	24.25	12.48	2.66	19.03	29.48	41.42	9.64	3.619	34.31	48.53
	Posttest	27.45	12.40	2.66	22.23	32.67	43.57	10.14	3.617	36.47	50.68

However, no significant Training group × Session interaction was found. As expected, a significant effect of number of syllables ($F(1, 539) = 11.88, p = .001$) revealed a significant difference in rhyme duration ratio depending on the length of the word. However, importantly, there was no significant interaction between Number of syllables × Training group or Number of syllables × Session.

By contrast, results of the GLMM with final vowel duration as a dependent variable revealed a significant effect of Training group ($F(1, 539) = 4.91, p = .027$), Session ($F(1, 539) = 7.69, p = .006$) and Training group × Session ($F(1, 539) = 4.62, p = .032$). Post-hoc analyses showed a significant difference between clapping and non-clapping at posttest only ($F(1,539) = 10.15, p =.002$) and a significant difference between pre- and posttest in the clapping group only ($F(1,539) = 19.19, p < .001$).

V Discussion

The results of the present investigation show that a short 20-minute training session involving hand-clapping during word learning in a second language helped children more than only audiovisual training in (1) reducing their accentedness as perceived by native speakers and (2) increasing the lengthening of the final vowel of the items, thus more closely approximating the way stress is phonetically realized in French. Two complementary sets of results back up our interpretation. First, regarding perceived accentedness, although both groups improved their pronunciation after training, our results show that the children in the clapping group reduced their accentedness scores significantly more than the children in the non-clapping group. Second, the results of the acoustic analyses show that the children belonging to the clapping group produced significantly longer (hence more target-like) final vowels[5] after training than the children in the non-clapping group. Since French phrasal stress/rhythm is essentially characterized by a significant lengthening of the final full vowel (Delais-Roussarie et al., 2015; Delattre, 1966; Di Cristo & Hirst, 1993; Fletcher, 1991; Vaissière, 1991) and it is considered a stable indicator of prominence in this language, an appropriate realization of final lengthening is indeed of crucial importance for the production of French rhythm.

All in all, the results of the study corroborate and expand previous results on the beneficial role of rhythmic training for phonological learning. From previous studies, we know that short training with rhythmic primes has positive effects on the phonological perception of speech in a first language (Cason & Schön, 2012; Cason et al., 2015). There is also extensive evidence that musical rhythmic activities increase children's phonological awareness and help develop their pre-reading skills (e.g. Herrera et al., 2011; Nelson, 2016; for a review, see Tierney & Kraus, 2013), and they can also be used as part of a method to help children with reading disorders (Flaugnacco et al., 2015; Habib et al., 2016; Overy, 2003). Rhythmic training has also been shown to have an immediate positive effect on the first-language phonological production of hearing-impaired children (Cason et al., 2015b).

In the context of the acquisition of second language pronunciation, our study backs up the general claim made by various researchers that suprasegmental (e.g. Darcy, Ewert & Lidster, 2012) or rhythmic training including a variety of activities can enhance the learning of prosody in a second language (e.g. on the use of rap music, see Fischler, 2009; on the use of beat gestures, see Gluhareva & Prieto, 2017; Kushch, 2018; on the use of hand-clapping, see Iizuka et al., 2020; Zhang et al., 2018). The present study thus extends the findings from the abovementioned studies by showing that a short audio-visual training session based on repeating words while hand-clapping their rhythmic structure can be of benefit in the second language classroom. First, while beat gestures may highlight higher levels of prosodic structure by marking nuclear pitch accents (Gluhareva & Prieto, 2017; Kushch, 2018), and beat gestures that accompany L2 speech help speakers externalize the prosodic features of a foreign language (McCafferty, 2006), hand-clapping lends itself to indicating durational structure at the syllabic level.

Importantly, our study complements previous mixed findings by Zhang et al. (2018) and Iizuka et al. (2020) on the beneficial effects of hand-clapping. First, in Zhang et al.'s study participants obtained a benefit from hand-clapping training, yet effects were smaller than those reported in the present study. While hand-clapping training was significantly helpful for the Catalan children who participated in the present study in terms of general pronunciation assessments, this was not the case with Chinese adolescents as reported by Zhang et al. (although a positive effect of acoustic lengthening was found). The reason for this difference might be that the participants in Zhang et al.'s study had to learn word pronunciation and meaning at the same time, while the present study avoided the potential cognitive overload that this more complex learning task may have entailed by using cognate words (French words that were similar in phonological form and meaning to their Catalan counterparts). The stronger effect found here might thus be due to the fact that while the Catalan children could fully direct their attention and efforts towards how to pronounce the words, Chinese adolescents had to also learn the meanings of the French words at the same time. That said, the two studies complement each other by both showing the beneficial effect of hand-clapping, regardless of the cognitive difficulty of the associated word learning task. Second, while participants in the hand-clapping group in Iizuka et al.'s (2020) study increased their perception abilities of segmental L2 features, they did not improve on their pronunciation of long vowels, geminates and moraic nasals. The fact that Iizuka et al.'s materials involved not only learning vowel duration patterns but also consonantal duration patterns may have made the task more difficult, and it may be that additional training was required to achieve appropriate pronunciation of these consonantal features. Moreover, since English loanwords in Japanese do not share the same phonological form as their English counterparts, a picture-naming task may have been too difficult to elicit pronunciation at an early stage of acquisition. Further research will be needed to tackle the role of hand-clapping on various aspects of L2 pronunciation at different stages of learning.

We believe the success of training with rhythmic hand-clapping may be explained by the properties of the hand-clapping sounds and movements. First, clapping on each syllable that makes up a word highlights the representation of the suprasegmental characteristics of the words through two perceptual channels, namely, the auditory and the visual channels. While both emphasize the temporal regularity of the syllables, the longer time spent with hands in contact emphasizes the stronger prominence present in the last syllable. Clapping to the rhythm of the words may thus have served to reinforce participants' ability to perceive the prosodic pattern of foreign speech and led to better overall pronunciation scores. In addition, although hand-clapping cannot visually represent duration in space the same way as pitch gestures do, training effects were also found on the realization of final lengthening. In addition, the subsequent embodied reproduction of hand-clapping by the participants themselves may have reinforced the overall perceptive effect even further. This would be consistent with the theory of embodied cognition (Barsalou, 2008), according to which the motor modality is closely linked to the perceptual modality (Borghi & Caruana, 2015). This suggests that appropriate sensory and motor interactions may make more efficient the development of human cognition, an idea that has crucial implications for learning and education (for reviews, see Kiefer & Trumpp, 2012; Wellsby & Pexman, 2014), including second language acquisition (Macedonia, 2019).

Regarding the effect of individual differences, an initial analysis revealed significant effects of language imitation aptitude and productive rhythmic skills, whereby better scores in these tasks predicted a reduction in the perceived accentedness of second language pronunciation. Good productive rhythmic skills, realized through hand-clapping, may have helped the children who ranked high in this aptitude to better follow the training session. By contrast, no significant effects were found for phonological and rhythmic perception scores, suggesting that rhythmic skills based on perception alone play a more limited role in production. Finally, perhaps surprisingly, we found that working memory did not play any role in our results. However, since the trained words were French/Catalan cognates, the pretest and posttest tasks did not place high demands on working memory, unlike in Zhang et al. (2018), where the two languages involved were entirely unrelated and the meanings of new words were thus totally opaque to the learners.

VI Limitations and future research

There are several limitations to this study. First, training with hand-clapping was found to improve the pronunciation of trained items only, while generalization to new, non-cognate items was not found. Perhaps a longer training period would be needed to detect generalization effects. Crucially, the untrained words were not cognate words; results might have been different if the new words had been cognates, as in the training session. In addition, results regarding the new words might also have been different if the training session had been carried out with non-cognate words. A follow-up study with non-cognate words could further assess the potential role of word familiarization when assessing the effectiveness of hand-clapping.

The small number of participants in this study is another limitation and suggests further testing with larger groups. Moreover, in the case of the analysis of accentedness, a significant difference between groups should have been observed at posttest only. With better scores in the non-clapping group at pretest, participants in this group may have had a narrower margin for improvement. We think that with a larger sample size, we might have been able to avoid the difference between groups at pretest. At this juncture, the acoustic analysis offered a more convincing type of evidence for the benefits of rhythmic training in this dataset. Importantly, further studies need to assess the effects of hand-clapping on other populations. For example, the effect of hand-clapping may show different outcomes depending on whether learners are still in a sensitive period for neural plasticity, as in the case of children, or not, as in the case of adults. Research has shown that adults tend to rely more on explicit knowledge to achieve learning than on a bottom-up process based only on direct experience (see, e.g. White et al., 2013). It would therefore be of interest to replicate this experiment using older second language learners (but see Iizuka et al., 2020; Zhang et al., 2018). Further, specific types of learners (e.g. kinaesthetic[6] learners) may benefit more from this type of training, and this aspect should be taken into account in future studies. Regarding language proficiency, future assessments might want to recruit participants who are actually in the process of learning the foreign language rather than merely exposed to a set of words for experimental purposes. Training could then be of a longer duration and findings extended to the learning of prosody across full sentences rather than single words.

In addition, the benefits of hand-clapping on L2 pronunciation should be further investigated with languages of different typologies. By highlighting the specific properties of a language in terms of syllabic structure, stress placement or prosodic patterns, hand-clapping could at the very least raise learner awareness of these elements and help improve pronunciation. Finally, in order to widen our understanding of the embodiment theory, more evidence is needed to assess the benefits of producing hand-clapping after watching the teacher as compared to merely watching the teacher and not replicating his/her behavior.

VII Conclusions

The results of the present study offer two complementary pieces of evidence – one perceptual, the other acoustic – of the usefulness of word-based rhythmic training with hand-clapping for the acquisition of L2 pronunciation. First, the study shows that rhythmic training with hand-clapping leads to greater improvement in accentedness scores. Second, acoustic analysis yields more specific evidence of the benefits of rhythmic training for the acquisition of final vowel lengthening. All in all, these results expand and complement the previous mixed evidence reported in Zhang et al. (2018) and Iizuka et al. (2020) on the role of rhythmic training through hand-clapping on pronunciation learning.

Learners' development of L2 pronunciation in an instructional context requires effective and practical tools. Given that mastering L2 suprasegmental features helps learners reduce their accentedness and improve their oral proficiency (e.g. Kang et al., 2010), training students in these prosodic features should be considered an important part of pronunciation instruction. In this context, from a practical perspective, hand-clapping is clearly a technique that would be easy to implement in the foreign language classroom as a tool to teach language rhythm patterns. Additionally, it would make the implementation of repetition-based drills or singing activities more engaging and pleasant and would potentially enhance motivation, especially with children. We surmise that the combination of acoustic and visual information channels, together with the motor experience involved in hand-clapping, can lead to a more optimized learning of the rhythmic structure of a novel language and consequently to better production of these rhythmic patterns.

Finally, from an educational point of view this approach would nicely mesh first with recent proposals advocating multisensory and embodied trainings in the classroom (e.g. Kiefer & Trumpp, 2012) and, second, a fuller integration of music and foreign language learning (for a review, see Viladot & Casals, 2018), which is consistent with an interdisciplinary approach to education (e.g. Jones, 2010). In our view, considerable work still remains to be done to develop and empirically test second language programs and educational tools based on rhythmic and melodic training which at the same time favor communicative situations and goal-based meaningful activities in the second language classroom.

Funding

The first author acknowledges a grand from the Department of Translation and Language Sciences at University Pompeu Fabra, Barcelona. This study is part of the project MULTIMODAL LANGUAGE LEARNING (MLL): Prosodic and Gestural Integration in Pragmatic and Phonological Development financed by the Spanish Ministry of Science, Innovation and Universities (ref.: PGC2018-097007-B-I00; 01/01/2019 - 31/12/2021).

ORCID iD

Florence Baills https://orcid.org/0000-0001-6743-9008

Supplemental material

Supplemental material for this article is available online.

Notes

1. The 20 target cognates were selected by having 35 adult Catalan native speakers listen to a longer preliminary list of 46 cognate French words and then rank each word on a five-point scale according to how close they thought it was to Catalan in terms of phonological transparency. The words used in this study were taken from the highest-scoring items, though we also took into account a set of phonological factors, like stress position within the words.
2. http://stel3.ub.edu/labfon/sites/default/files/prosodic_data-extraction-2.0.praat
3. In a GLMM, a variable that co-varies with the dependent variable can be controlled for in order to better define the effect of the fixed factors under scrutiny.
4. Because of the lack of an effect observed for the new items on accentedness, it was decided not to include the new items in the acoustic analysis.
5. Interestingly, in the present study, the acoustic analysis showed that the realization of final lengthening was realized more clearly at the vowel level than at the rhyme level, which may be explained by the fact that the pronunciation of neighboring consonants was at times far from the target and showed unnatural longer durations. Thus, vowel duration measures may be a more stable cue to investigate final lengthening patterns. Unsurprisingly, it has been frequently used as a phonetic measure for the assessment of final lengthening across studies (e.g. Gendrot et al., 2019; Santiago & Mairano, 2019; Schwab, 2012).
6. Kinaesthetic learners prefer to learn through active movements and experiences.

References

Aliaga-Garcia, C., Mora, J.C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, *47*, 1–14.

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, *42*, 529–555.

Arellano, S.I., & Draper, J.E. (1972). Relations between musical aptitudes and second-language learning. *Hispania*, *55*, 111–121.

Astésano, C. (2001). *Rythme et accentuation en français: Invariance et variabilité stylistique* [Rhythm and accentuation in French: Invariance and stylistic variability]. Paris: Editions L'Harmattan.

Baills, F., Suárez-González, N., González-Fuente, S., & Prieto, P. (2019). Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition*, *41*, 33–58.

Barsalou, L.W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Bartkova, K., Bonneau, A., Colotte, V., & Dargnat, M. (2012). Production of 'continuation contours' by French speakers in L1 (French) and L2 (English). In Ma, Q., Ding, H., & D. Hirst (Eds.), *Proceedings of the 6th speech prosody: Volume* (pp. 426–429). Shanghai: Tongji University Press.

Behrman, A. (2014). Segmental and prosodic approaches to accent management. *American Journal of Speech-Language Pathology*, *23*, 546–561.

Bhide, A., Power, A., & Goswami, U. (2013). A rhythmic musical intervention for poor readers: A comparison of efficacy with a letter-based intervention. *Mind, Brain, and Education*, *7*, 113–123.

Boersma, P., & Weenink, D. (2017). *Praat: Doing phonetics by computer: Version 6.0.33* [computer software]. Available at: http://www.praat.org (accessed January 2021).

Borghi, A.M., & Caruana, F. (2015). Embodiment theory. In Wright, J. (Ed.), *International encyclopedia of the social & behavioral sciences* (pp. 420–426). Amsterdam: Elsevier.

Bunting, M.F.M., Cowan, N., & Saults, J.S.S. (2006). How does running memory span work? *Quarterly Journal of Experimental Psychology*, *59*, 1691–1700.

Cameron, D.J., & Grahn, J.A. (2014). Enhanced timing abilities in percussionists generalize to rhythms without a musical beat. *Frontiers in Human Neuroscience*, *8*, 1003.

Cason, N., & Schön, D. (2012). Rhythmic priming enhances the phonological processing of speech. *Neuropsychologia*, *50*, 2652–2658.

Cason, N., Astésano, C., & Schön, D. (2015a). Bridging music and speech rhythm: Rhythmic priming and audio-motor training affect speech perception. *Acta Psychologica*, *155*, 43–50.

Cason, N., Hidalgo, C., Isoard, F., Roman, S., & Schön, D. (2015b). Rhythmic priming enhances speech production abilities: Evidence from prelingually deaf children. *Neuropsychology*, *29*, 102–107.

Cohrdes, C., Grolig, L., & Schroeder, S. (2016). Relating language and music skills in young children: A first approach to systemize and compare distinct competencies on different levels. *Frontiers in Psychology*, *7*, 1616.

Corriveau, K.H., & Goswami, U. (2009). Rhythmic motor entrainment in children with speech and language impairments: Tapping to the beat. *Cortex*, *45*, 119–130.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation 'toolbox'. In Levis, J., & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 93–108). Ames, IA: Iowa State University.

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.

de Groot, A.M.B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*, 1–56.

de Jong, N.H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*, 385–390.

Delais-Roussarie, E., Post, B., Avanzi, M., et al. (2015). Intonational phonology of French: Developing a ToBI system for French. In Frota, S., & P. Prieto (Eds): *Intonation in Romance* (pp. 63–100). Oxford: Oxford University Press.

Delattre, P. (1966). A comparison of syllable length conditioning among languages. *International Review of Applied Linguistics in Language Teaching*, *4*, 183–198.

Delogu, F., & Zheng, Y. (2020). Beneficial effects of musicality on the development of productive phonology skills in second language acquisition. *Frontiers in Neuroscience*, *14*, 618.

Derwing, T.M., & Rossiter, M.J. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, *13*, 1–17.

Derwing, T.M., Munro, M.J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393–410.

Di Cristo, A., & Hirst, D. (1993). Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français [Syllabic rhythm, melodic rhythm and hierarchical representation of the prosody of French]. *Travaux de l'Institut de Phonétique d'Aix*, *15*, 9–24.

Elvira-Garcia, W. (2014). Prosodic data extraction (v2.2) [Praat]. Available at: https://www.ub.edu/phoneticslaboratory/praat-scripts.html

Falk, S., & Dalla Bella, S. (2016). It is better when expected: aligning speech and motor rhythms enhances verbal processing. *Language, Cognition and Neuroscience*, *31*, 699–708.

Falk, S., Lanzilotti, C., & Schön, D. (2017a). Tuning neural phase entrainment to speech. *Journal of Cognitive Neuroscience*, *29*, 1378–1389.

Falk, S., Volpi-Moncorger, C., & Dalla Bella, S. (2017b). Auditory-motor rhythms and speech processing in French and German listeners. *Frontiers in Psychology*, *8*, 395.

Fischler, J. (2009). The rap on stress: Teaching stress patterns to English language learners through rap music. *MInneTESOL Journal*, *26*, 1–23.

Flaugnacco, E., Lopez, L., Terribili, C., et al. (2015). Music training increases phonological awareness and reading skills in developmental dyslexia: A randomized control trial. *PLoS ONE*, *10*, e0138715.

Flege, J.E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47–65.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics*, *19*, 193–212.

Frota, S., & Prieto, P. (2015). Intonation in Romance: systemic similarities and differences. In Frota, S., & P. Prieto (Eds.), *Intonation in Romance* (pp. 392–418). Oxford: Oxford University Press.

Gendrot, C., Adda-Decker, M., & Santiago, F. (2019). Acoustic realization of vowels as a function of syllable position: A cross-linguistic study with data from French and Spanish. In Gil, J., & M. Gibson (Eds.), *Romance phonetics and phonology* (pp. 77–88). Oxford: Oxford University Press.

Gilleece, L.F. (2006). An empirical investigation of the association between musical aptitude and foreign language aptitude. Doctoral dissertation, Trinity College Dublin, Dublin, Ireland. *Dublin: Trinity Access to Research Archive*. Available at: http://hdl.handle.net/2262/77993 (accessed January 2021).

Glenberg, A.M., & Kaschak, M.P. (2003). The body's contribution to language. In Ross, B. (Ed.), *The psychology of learning and motivation* (pp. 93–126). Cambridge, MA: Academic Press.

Gluhareva, D., & Prieto, P. (2017). Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations. *Language Teaching Research*, *21*, 609–631.

Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1651), 1–11.

Goldin-Meadow, S., Cook, S.W., & Mitchell, Z.A. (2009). Gesturing gives children new ideas about math. *Psychological Science*, *20*, 267–272.

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners. *Journal of Second Language Pronunciation*, *1*, 56–92.

Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation teaching and learning: Phonetic instruction in the L2 classroom. In Levis, J., & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 194–206). Ames, IA: Iowa State University.

Gordon, R.L., Jacobs, M.S., Schuele, C.M., & Mcauley, J.D. (2015a). Perspectives on the rhythm–grammar link and its implications for typical and atypical language development. *Annals of the New York Academy of Sciences*, *1337*, 16–25.

Gordon, R.L., Shivers, C.M., Wieland, E.A., et al. (2015b). Musical rhythm discrimination explains individual differences in grammar skills in children. *Developmental Psychology*, *18*, 635–644.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences*, *15*, 3–10.

Goswami, U., Huss, M., Mead, N., Fosker, T., & Verney, J.P. (2013). Perception of patterns of musical beat distribution in phonological developmental dyslexia: Significant longitudinal relations with word reading and reading comprehension. *Cortex*, *49*, 1363–1376.

Goswami, U., Thomson, J., Richardson, U., et al. (2002). Amplitude envelope onsets and developmental dyslexia: A new hypothesis. *Proceedings of National Academy of Sciences of the U.S.A.*, *99*, 10911–10916.

Habib, M., Lardy, C., Desiles, T., et al. (2016). Music and dyslexia: A new musical training method to improve reading and related disorders. *Frontiers in Psychology*, *7*, 26.

Halle, M., & Vergnaud, J.-R. (1987). An essay on stress. *Phonology*, *7*, 171–188.

Herrera, L., Lorenzo, O., Defior, S., Fernández-Smith, G., & Costa-Giomi, E. (2011). Effects of phonological and musical training on the reading readiness of native- and foreign-Spanish-speaking children. *Psychology of Music*, *39*, 68–81.

Iizuka, T., Nakatsukasa, K., & Braver, A. (2020). The efficacy of gesture on second language pronunciation: An exploratory study of handclapping as a classroom instructional tool. *Language Learning*. Epub ahead of print 3 June 2020. DOI: 10.1111/lang.12415.

Johnson, E.K., & Jusczyk, P.W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*, 548–567.

Jones, C. (2009). Interdisciplinary approach: Advantages, disadvantages, and the future benefits of interdisciplinary studies. *ESSAI*, *7*, 26.

Jun, S.-A., & Fougeron, C. (1995). The accentual phrase and the prosodic structure of French. In Elenius, K., & P. Branderud (Eds.), *Proceedings of 13th International Congress of Phonetic Sciences: Volume 2* (pp. 722–725). Stockholm: KTH & Stockholm University.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In Botinis, A. (Ed.), *Intonation: Analysis, modelling and technology* (pp. 209–242). Dordrecht: Kluwer Academic.

Juntunen, M.-L. (2016). The Dalcroze approach: Experiencing and knowing music through the embodied exploration. In Abril, C.R., & B. Gault (Eds.), *Approaches to teaching general music: Methods, issues, and viewpoints* (pp. 141–167). Oxford: Oxford University Press.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, *94*, 554–566.

Kelly, S., Bailey, A., & Hirata, Y. (2017). Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts. *Collabra: Psychology*, *3*, 7.

Kiefer, M., & Trumpp, N. M. (2012). Embodiment theory and education: The foundations of cognition in perception and action. *Trends in Neuroscience and Education*, *1*, 15–20.

Kotz, S.A., & Gunter, T.C. (2015). Can rhythmic auditory cuing remediate language-related deficits in Parkinson's disease? *Annals of the New York Academy of Sciences*, *1337*, 62–68.

Kushch, O. (2018). Beat gestures and prosodic prominence: Impact on learning. Doctoral dissertation, Universitat Pompeu Fabra, Barcelona, Spain. *Tesis Doctorals en Xarxa*. Available at: http://hdl.handle.net/10803/463004 (accessed January 2021).

Law, L.N.C., & Zentner, M. (2012). Assessing musical abilities objectively: Construction and validation of the profile of music perception skills. *PLoS ONE*, *7*, e52508.

Lee, J., Jang, J., & Plonsky, L. (2015). The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics*, *36*, 345–366.

Lennes, M. (n.d.). SpeCT - The speech corpus toolkit for Praat. Available at: https://lennes.github.io/spect/

Li, P., Baills, F., & Prieto, P. (2020). Observing and producing durational hand gestures facilitates the pronunciation of novel vowel length contrasts. *Studies in Second Language Acquisition*. Epub ahead of print 6 April 2020. DOI: 10.1017/S0272263120000054.

López-Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., & Martínez, M. (2005). *Inventario del desarrollo MacArthur: Versión española*. Madrid: TEA.

Macchi, L., Vansteene, C., Timmermans, N., & Boidien, F. (2013). Épreuve lilloise de discrimination phonologique: Présentation et illustration par deux études de cas cliniques [Lille phonological discrimination test: Presentation and illustration with two clinical case studies]. *Les Cahiers de l'ASELF*, *10*, 3–22.

Macedonia, M. (2019). Embodied learning: Why at School the mind needs the body. *Frontiers in Psychology*, *10*, 2098.

Macedonia, M., Müller, K., & Friederici, A.D. (2011). The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping*, *32*, 982–998.

Machač, P., & Skarnitzl, R. (2009). *Principles of phonetic segmentation*. Prague: Epocha.

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.

Magne, C., Astésano, C., Aramaki, M., et al. (2007). Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cerebral Cortex*, *17*, 2659–2668.

Masumoto, K., Yamaguchi, M., Sutani, K., et al. (2006). Reactivation of physical motor information in the memory of action events. *Brain Research*, *1101*, 102–109.

McCafferty, S.G. (2006). Gesture and the materialization of second language prosody. *International Review of Applied Linguistics in Language Teaching*, *44*, 197–209.

Milovanov, R., Pietilä, P., Tervaniemi, M., & Esquef, P.A.A. (2010). Foreign language pronunciation skills and musical aptitude: A study of Finnish adults with higher education. *Learning and Individual Differences*, *20*, 56–60.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P.A.A., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research*, *1194*, 81–89.

Morgan, C. (2004). Musical aptitude and second-language phonetics learning: Implications for teaching methodology. Doctoral dissertation, Simon Fraser University, Burnaby, BC, Canada. *Summit Institutional Repository*. Available at: http://summit.sfu.ca/item/6519 (accessed January 2021).

Morett, L.M., & Chang, L.-Y. (2015). Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience*, *30*, 347–353.

Morgan, J.L., & Saffran, J.R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development*, *66*, 911–936.

Munro, M., Derwing, T., & Morton, S. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, *28*, 111–131.

Nagle, C., Sachs, R., & Zárate–Sández, G. (2018). Exploring the intersection between teachers' beliefs and research findings in pronunciation instruction. *Modern Language Journal*, *102*, 512–532.

Nardo, D., & Reiterer, S.M. (2009). Musicality and phonetic language aptitude. In Dogil, G., & S. Reiterer (Eds.), *Trends in applied linguistics: Language talent and brain activity* (pp. 213–256). Berlin: De Gruyter Mouton.

Nelson, S.D. (2016). The effects of an integrated rhythmic and literacy intervention on the development of phonological awareness and rhythm skills of preschoolers. Unpublished doctoral dissertation, Iowa State University, Ames, IA, USA.

Overy, K. (2003). Dyslexia and music: From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, *999*, 497–505.

Pitt, M.A., & Samuel, A.G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 564–573.

Prieto, P., Vanrell, M.M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm: Evidence from Catalan, English, and Spanish. *Speech Communication*, *54*, 681–702.

Prieto, P., Borràs-Comes, J., Cabré, T., et al. (2015). Intonational phonology of Catalan and its dialectal varieties. In Frota, S., & P. Prieto (Eds.), *Intonation in Romance* (p. 9–62). Oxford: Oxford University Press.

Rao, R. (2010). Final lengthening and pause duration in three dialects of Spanish. In Ortega-Llebaria, M. (Ed.), *Selected proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology* (pp. 69–82). Somerville, MA: Cascadilla Proceedings Project. Available at: http://www.lingref.com/cpp/lasp/4/paper2368.pdf (accessed January 2021).

Reiterer, S.M., Hu, X., Sumathi, T.A, & Singh, N.C. (2013). Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology*, *4*, 782.

Repp, B.H., & Su, Y.H. (2013). Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin and Review*, *20*, 403–452.

Romero Naranjo, F.J.R. (2013). Science and art of body percussion: A review. *Journal of Human Sport and Exercise*, *8*, 442–457.

Roncaglia-Denissen, M.P., Schmidt-Kassow, M., & Kotz, S.A. (2013). Speech rhythm facilitates syntactic ambiguity resolution: ERP evidence. *PLoS ONE*, *8*, e56000.

Rota, G., & Reiterer, S. (2009). Cognitive aspects of pronunciation talent. *Language talent and brain activity* (pp. 67–96). Berlin: De Gruyter Mouton.

Rueschemeyer, S.A., Lindemann, O., Van Rooij, D., Van Dam, W., & Bekkering, H. (2010). Effects of intentional motor actions on embodied language processing. *Experimental Psychology*, *57*, 260–266.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, *46*, 842–854.

Saito, K., Trofimovich, P., & Isaacs, T. (2017). Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*, *38*, 439–462.

Santiago, F., & P. Mairano. (2019). Prosodic effects on L2 French vowels: a corpus-based investigation. In Calhoun, S., Escudero, P., Tabain, M., & P. Warren (Eds.), *Proceedings of the19th International Congress of Phonetic Sciences* (pp. 1084–1088). Melbourne: Australasian Speech Science and Technology Association.

Schwab, S. (2012). Do native Spanish speakers transfer accentual acoustic properties from Spanish to French L2? In Ma, Q., Ding, H., & Hirst, D. (Ed.). *Proceedings of the 6th International Conference on Speech Prosody* (pp. 330–333). Shanghai: Tongji University Press.

Schwab, S. (2013). Apprenants hispanophones de FLE et accentuation en Français [Spanish-speaking learners of French as a foreign language and emphasis in French]. *Travaux Neuchâtelois de Linguistique*, *59*, 71–86.

Slevc, L.R., & Miyake, A. (2006). Individual differences in second-language proficiency. *Psychological Science*, *17*, 675–681.

Thomson, R.I., & Derwing, T.M. (2015). The effectiveness of L2 pronunciation instruction: A Narrative Review. *Applied Linguistics*, *36*, 326–344.

Tierney, A., & Kraus, N. (2013). Music training for the development of reading skills. *Progress in Brain Research*, *207*, 209–241.

Vaissière, J. (1991). Rhythm, accentuation and final lengthening in French. In Sundberg, J., Nord, L., & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 108–121). Houndsmill: Macmillan.

Viladot, L., & Casals, A. (2018). Rhyming the rhythm and measuring the metre: Pooling music and language in the classroom. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education*, *1*, 37–43.

Wang, H., Mok, P., & Meng, H. (2016). Capitalizing on musical rhythm for prosodic training in computer-aided language learning. *Computer Speech and Language*, *37*, 67–81.

Wellsby, M., & Pexman, P.M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*, 506.

White, E.J., Hutka, S.A., Williams, L.J., & Moreno, S. (2013). Learning, neural plasticity and sensitive periods: Implications for language acquisition, music training and transfer across the lifespan. *Frontiers in Systems Neuroscience*, *7*, 90.

Yuan, C., González-Fuente, S., Baills, F., & Prieto, P. (2019). Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition*, *41*, 5–32.

Zellner, B. (1996). Structures temporelles et structures prosodiques en français lu [Temporal structures and prosodic structures in read French]. *Revue Française de Linguistique Appliquée*, *1*, 7–23.

Zhang, Y., Baills, F., & Prieto, P. (2018). Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words. *Language Teaching Research*, *24*, 666–689.

Zion Golumbic, E.M., Poeppel, D., & Schroeder, C.E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language*, *122*, 151–161.

Zwaan, R.A., & Taylor, L.J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology General*, *135*, 1–11.