Applied Linguistics 2022: 1–29
https://doi.org/10.1093/applin/amac010

Embodied Prosodic Training Helps Improve Accentedness and Suprasegmental Accuracy

[1,*]FLORENCE BAILLS, [2]CHARLOTTE ALAZARD-GUIU, and [1,3]PILAR PRIETO

[1]Department of Translation and Language Sciences, Universitat Pompeu Fabra, C/Roc Boronat 138, Barcelona 08018, Spain, [2]Département d'Etudes du Français Langue Etrangère, Université Toulouse II, 5 allée Antonio Machado, Toulouse 31058, France, and [3]Institució Catalana de Recerca i Estudis Avançats, Passeig de Lluís Companys 23, Barcelona 08010, Spain
*E-mail: florence.baills@upf.edu

With increasing evidence in favour of embodied learning techniques, more research is needed to explore eventual applications in the field of second language acquisition, for example, the effect of embodied training on phonological learning. This study investigated how pronunciation was affected by visuospatial hand gestures depicting speech rhythm and intonation during the oral repetition of logatomes (i.e. a series of identical nonsense CV syllables that maintain prosodic structure intact). Seventy-five Catalan learners of French participated in three training sessions with short dialogues, in one of three conditions: speech-only, non-embodied logatome, and embodied logatome. Before and after the training period, participants carried out an identical pre- and posttest which consisted of reading four dialogues aloud. Their oral output was evaluated in terms of fluency, comprehensibility, accentedness, and suprasegmental and segmental features by three native French speakers. While all three groups significantly improved in all measures after training, the embodied logatome group improved significantly more in terms of accentedness and suprasegmental features compared with the speech-only group, while the non-embodied logatome group did not.

INTRODUCTION

There is increasing evidence of the integration of the perceptual and motor systems in the cognitive system (e.g. Barsalou 2008; Wilson and Foglia 2017; Keily 2019) and of the benefits of embodied learning in education, notably through the use of hand gestures (e.g. Macedonia 2019; Shapiro and Stolz 2019). In the field of foreign language acquisition, however, since Atkinson's call for an embodied approach to SLA (2010), relatively little work has been carried out to put this claim into perspective. Regarding phonological learning, despite numerous studies confirming the positive role of pronunciation instruction—in particular, prosodic training (e.g. Gordon and Darcy 2016; Zhang and Yuan 2020)—and the important role of prosody in pronunciation evaluations (e.g. Trofimovich and Baker 2006; Kang 2010), there is a clear need for concrete, research-based, pronunciation teaching techniques that focus on highlighting L2 prosody. Based on previous evidence that prosodic features can be successfully depicted by hand movements (e.g. Connell *et al.* 2013; Dolscheid *et al.* 2014; Biau 2015), the present study explored the gains of training prosody using embodied techniques on L2 pronunciation.

LITERATURE REVIEW

Embodied cognition theory, embodied learning, language, and prosody

According to Embodied Cognition Theory, sensory-motor processes and the physical body are an integral part of human cognition and modulate cognitive processing (e.g. Barsalou 2008; Wilson and Foglia 2017; Keily 2019). This theory is based on the mutual effects of perception and actions on one another and their joint effect on mental representation and is claimed to have special relevance for education (e.g. Kiefer and Trumpp 2012; Ionescu and Vasc 2014; Macedonia 2019; Shapiro and Stolz 2019).

The discovery of mirror neurons, a group of motor neurons that are activated upon watching another person perform a behaviour, has led scholars to propose that these neurons may play a crucial role in understanding other peoples' actions and may be necessary for imitative learning (Rizzolatti and Craighero 2004). They stand as a potential explanation for the positive effects of active engagement and communicative gestures. Sullivan (2018) argued that instructors' movements and their use of representational gesture stimulate mental imitation by activating the mirror neurons, which may lead to an improvement in students' academic outcomes. Meanwhile, more empirical research about embodied cognition and learning has primarily focused on how increasing students' own motor involvement during instruction increases learning outcomes (e.g. Bahnmueller *et al.* 2014; Smith *et al.* 2014).

There is evidence that language, in particular, is embodied. Research has shown that the language areas in the brain activate during sensorimotor action (e.g. Desai *et al.* 2010) and conversely, motor areas activate during speech (e.g. Hauk *et al.* 2004), including when processing non-literal action language (e.g. Yang and Shu 2016). Gestures, which are closely tied to speech (e.g. McNeill 1992) and develop together in infancy (e.g. Iverson and Goldin-Meadow 2005), may stem from spatial representations and mental images and may arise from an embodied cognitive system, as proposed by Hostetter and Alibali's Gestures as Simulated Action framework (2008). Several studies lend evidence to the theory. For example, Rieser et al. (1994) found that linguistic tasks related to spatial orientation are facilitated by the mental representation of movement both in children and adults. Descriptions of spatial associations are comprehended faster than those of spatial dissociations (Glenberg *et al.* 1987) and words with high 'body–object interaction' ratings (Siakaluk *et al.*

2008) or related to manipulable objects (Rueschemeyer *et al.* 2010) are recognized faster, providing further evidence of the role of motor actions on lexical–semantic processing. Moreover, there is ample evidence of the effect of actions, gestures, and exercise on memory (see Madan and Singhal 2012 for a review). For example, lessons with gestures are shown to promote deeper reasoning, synthesis, and information retention than lessons that do not feature gestures (Goldin-Meadow and Alibali 2013). Interestingly, some work has also focused on the important role of prosody in syntax and syntax learning through the lens of embodied interaction (e.g. Bergmann *et al.* 2012; Kreiner and Eviatar 2014; Matsumoto and Dobs 2017).

Embodied learning in SLA

Research in the field of Conversation Analysis has documented how cognitive states are expressed during interaction, not only through speech but also via gaze, facial gesture, hand gesture, posture shift, and the manipulation of documents and objects and how these embodied cognitive states participate in the management of peer interaction (e.g. Goodwin and Goodwin 1986; Drew 2006; Mori and Hasegawa 2009; Eskildsen and Wagner 2013, 2015; Cekaite 2015; Kääntä 2015; Majlesi 2015; Jakonen 2020). To give a few examples, Mori and Hasegawa (2009) showed how two students organized themselves in a word search activity by simultaneously using different semiotic resources, such as language, body, and the structures of their textbooks and notebooks for language learning. Jakonen (2020) suggested that teachers use their body as a pedagogical device and analysed teachers' movement trajectories and body positioning in content and language-integrated learning (CLIL) classrooms. The analysis showed that walking through the classroom allowed the teacher to monitor student individual and group progress during a task, to display availability, and to invite students' interaction. Eskildsen and Wagner (2015) analysed how gesture–speech combinations are created by L2 learners to create a common understanding of new words and how they are reused on later occasions. Interestingly, Eskildsen and Wagner (2013) observed that the imitation of a speaker's gesture acts as a communicative resource for achieving and maintaining understanding in spontaneous conversations between pairs and with the teacher. Studies in the field of gesture are further exploring, describing, and classifying teachers' and learners' gestures as part of their linguistic conceptualization and expression (e.g. Gullberg and McCafferty 2008; Smotrova 2014; Wang and Loewen 2016). However, very few studies have been conducted to test empirically the effects of embodied learning strategies on second language acquisition. Rather, most of these studies have looked at the effect of spontaneous and nonspontaneous gestures on word recall [see Macedonia (2014) and Morett (2018) for reviews; for the effect of gestures on grammar learning, see Nakatsukasa (2016)].

Regarding phonological learning, strong evidence for a tight relationship between prosody and gesture (e.g. Loehr 2012; Biau 2015; Ferré 2018) suggests a positive role of embodied strategies on the learning of an L2 phonological system, especially on pronunciation. Chan (2018) advocates the integration of body movements and gestures to enhance the perception, pronunciation, and retention of L2 phonological features. There are several reasons to support her claim. First, research on embodied approaches to music education has shown that body movement can enhance the acquisition of musical rhythmic and melodic patterns (e.g. Juntunen 2016). In view of the close resemblance between musical and prosodic structure (e.g. Heffner and Slevc 2015), we surmise that, in a similar way, hand and arm movements may help the acquisition of speech rhythm and melody. In addition, from the field of sign language, there is evidence of the existence of a visuospatial 'phonological loop' in working memory, similar to the phonological loop for speech, which is structured uniquely by language (e.g. Wilson and Emmorey 1997). In that sense, the form of a gesture may be processed in a similar way to speech sounds and associated with the corresponding phonological feature. Finally, there is evidence that the mental representation of pitch is visuospatial in nature (e.g. Connell *et al.* 2013; Dolscheid *et al.* 2014), indicating that making pitch directions and movements visible to the learners may help them process foreign language prosody. In the following section, we review the literature on the benefits of prosodic pronunciation instruction, with a focus on embodied techniques.

Benefits of prosodic pronunciation instruction

Prosody plays an important role in pronunciation. There is evidence that transfer from a first to a second language takes place in the prosodic domain (e.g. Ueyama 2000; Trofimovich and Baker 2006; Lomotey 2013), as well as evidence that suprasegmental patterns play a crucial role in the perception of non-native pronunciation patterns [e.g. Kang *et al.* 2010; see Wang (2020) for a review] and seem to weigh more in the perception of foreign accentedness (e.g. Anderson-Hsieh *et al.* 1992; De Mareüil and Vieru-Dimulescu 2006; Trofimovich and Baker 2006).

A growing body of evidence has shown that pronunciation instruction focusing on speaking rate, intonation, rhythm, and word and sentence stress may improve overall measures of pronunciation more than segmental training or no training at all in sentence repetition, read-speech, and spontaneous speech tasks (e.g. Gordon *et al.* 2013; Saito and Saito 2017; Zhang and Yuan 2020). In a meta-analytic review, Thomson and Derwing (2015) found that 52 per cent of the studies on pronunciation instruction included in their analysis investigated segmental training, 18 per cent focused on suprasegmental training, and 30 per cent dealt with both, usually in combined lessons but occasionally as separate comparison groups. Unfortunately, these studies including long suprasegmental instruction paradigms used a varied set of techniques that ranged from explicit instruction involving theoretical presentation–practice–production sequences (e.g. Gordon *et al.* 2013) to more implicit techniques involving musical and rhythmic activities (e.g. Derwing *et al.* 1998), making a full synthesis of the results difficult. Moreover, as pointed out by the authors and also by Lee *et al.* (2015), most of the studies failed to provide a sufficiently thorough description of the training activities involved.

To our knowledge, only a small set of implicit prosodic training techniques involving music- and prosodic-based activities (some with visual feedback) have been empirically tested to assess their value for second language pronunciation improvement. Some studies have found that musical activities highlighting the rhythmic and melodic properties of language are helpful in improving L2 pronunciation. In a 12-week instruction study with a pre- and posttest design, Derwing *et al.* (1998) used rhythmic chants to train learners to count the number of syllables and stresses, tap out the beats, and use nonsense syllables to focus on rhythm. The authors found this type of training more beneficial than segmental training on the comprehensibility and fluency of spontaneous speech in a narrative task. More recently, Good *et al.* (2015) found a positive effect of teaching a short passage in a sung modality compared with spoken modality on the pronunciation of L2 vowels sounds. Ludke (2018) compared L2 instruction with singing and song listening activities to L2 instruction with visual arts (drawing and creating cartoons) and drama activities and found higher performance on intonation and flow of speech in the singing and song group. In a different approach, computer-assisted learning based on the development of speech analysis technology can also be used to teach L2 suprasegmental features by allowing learners to compare the visual representation of target pitch contours produced by native speakers to their own output and try to adjust it accordingly (e.g. de Bot 1983; Hardison 2004; Ramírez Verdugo 2006; Hincks and Edlund 2009; Tanner and Landon 2009; Liu and Tseng 2019) or by juxtaposing a computerized set of percussive sounds over target sentences to provide additional rhythmic cueing (Wang *et al.* 2016).

Embodied prosodic instruction in the classroom

In practice, it is not uncommon to see teachers spontaneously use co-speech gestures when explaining difficult pronunciation features. For example, based on the observation of audiovisual corpus, Tellier (2008) gave a description of the pedagogical gestures employed to teach pronunciation, in particular gestures that enable the students to visualize and feel the prosodic characteristics of speech. She mentioned that language teachers use flat, rising, and falling hand movements to imitate sentence intonation (see also Smotrova 2014). Hudson (2011) described horizontal movements of the hands or lateral movements of the body to represent vowel duration. Finally, beat gestures, tapping, or clapping rhythms function as a way to distinguish syllables or to indicate stress position (Hudson 2011; Baker 2014; Chan 2018).

The essential haptic-integrated English pronunciation (EHIEP) framework developed by Acton *et al.* (2013) proposes coupling speech with systematic hand movement, kinesthetic and tactile techniques to teach segmental (vowels and consonants) and suprasegmental (stress, rhythm, intonation) features. For example, tapping on one's own shoulder and arm with different intensities when uttering stressed and unstressed syllables (*butterfly technique*, Burri and Baker 2016) or performing boxing-like movements to physically experience rhythm and stress prominence (*rhythmic fight club technique*, Burri *et al.* 2016).[1]

Another approach to pronunciation teaching is known as the *verbotonal method* (henceforth, VT, e.g. Renard 2002; Guberina 2008), which is based on the notion that prosody acts as a frame for pronunciation development and should be taught from the first stages of language learning. This is achieved notably through the repetition of *logatomes* combined with visuospatial hand gestures that mimic the intonation and rhythm of the sentence (Billières 2002). A logatome is a series of same consonant–vowel nonsense sequences (e.g. /dadada/) that remove any target segmental information but keep the prosodic structure of the sentence intact. Repeating meaningless CV syllables in this fashion allows learners to focus on the suprasegmental features of target utterances while keeping the segmental content controlled [see Billières (2002) for a full explanation of the use of logatomes in the VT method]. In addition, the role of the body as a supporting tool is fundamental to this approach, as stated by Guberina (1965, p. 151):

> L'ensemble acoustique de toutes les langues contient certains facteurs structuraux qui sont immanents à notre être biologique. La tension, l'intensité, le rythme les tonalités sont des formes biologiques de l'homme. [The acoustic ensemble of all languages contains certain structural factors that stem from our biological nature. Tension, intensity, rhythm and tonality are all products of human biology]

Billières (2002) further describes the benefits of accompanying logatomes with hand gestures—what we will henceforth refer to as *embodied logatomes*—to mimic the intonation, rhythm, and stress patterns of the target sentence for learning purposes. The repetition of embodied logatomes is generally performed before the repetition of full target sentences, as the repetition of embodied logatomes is believed to have a priming effect and thereby augment the saliency of the target sentences' prosodic features.

Benefits of embodied prosodic training: empirical evidence

To our knowledge, only a few studies have empirically assessed the effects of using hand gestures on pronunciation, however not directly within the framework of embodied learning. For example, the perception and production of rhythmic movements such as simple up-and-down or back-and-forth motions of the hands—also called beat gestures—have been found to aid Catalan learners' accentedness and fluency in English (Gluhareva and Prieto 2017; Llanes-

Coromina *et al.* 2018). Recent studies have investigated the role of handclapping in second language pronunciation and found it beneficial for the perception of Japanese long vowels in English speakers (Iizuka *et al.* 2020) and the accentedness of young, Catalan, and Chinese naïve learners of French (Zhang *et al.* 2020; Baills and Prieto 2021). Hand gestures depicting specific suprasegmental properties such as vowel duration in Japanese (durational gesture, Li *et al.* 2020) and intonation contours in Spanish (pitch gesture, Yuan *et al.* 2019) have also shown positive effects on the pronunciation of these features.[2]

Regarding more classroom-based approaches, only a few empirical studies have assessed the potential beneficial effects of embodied prosodic training for L2 pronunciation through the EHIEP and VT techniques. Mister *et al.* (2021) tested the *rhythmic fight club* technique during a five-day workshop with 16 adult learners of English. They found that drawing attention to word stress patterns through kinesthetic/tactile training during both controlled and more spontaneous productive activities. Results indicate that learners increasingly improved the recall and the correct stress placement of the target words over the course of the intervention. However, this study did not include a control group and it remains unclear whether the technique employed in the training would outperform other kinds of non-embodied techniques. Alazard *et al.* (2010) found that eight weeks of global VT phonetic training sessions improved learners' fluency in L2 French more than training sessions based on reading aloud, text comprehension, and creative writing. However, it remains unclear whether the gains were due to the listen-and-repeat tasks, the use of the logatomes, or the use of hand gestures. Later, Alazard (2013) compared the VT method with the *articulatory method*, which involves the explicit teaching of segments' articulatory properties, and found that after four weeks, participants following the VT method showed significantly higher gains in their fluency, in particular when their French pronunciation was worse at the outset. However, this advantage disappeared after eight weeks of training. According to the author, the introduction of written activities during the second half of the course, and more specifically the intellectualization that goes with this type of activities, instead of improving reading fluency, may have led to a decline in pronunciation performance. These results indicate that it may also be necessary to practice oral-reading pronunciation.

All in all, while previous research has been mostly centred on testing specific prosodic aspects in laboratory settings, classroom-based studies remain scarce and reveal inconsistent results. Therefore, more empirical research is needed to assess the effects of embodied pronunciation teaching, in particular research with more classroom-based, learner-oriented experimental designs. Importantly, training pronunciation with visuospatial gestures mimicking the prosodic features of the target language does not only have pedagogical implications but also allows for the testing of the predictions of Embodied Cognition Theory for phonological learning.

The present study

The present study aimed to assess the efficacy of embodied pronunciation training on oral reading through visuospatial hand gesture movements mimicking the melodic and rhythmic patterns of target sentences. We hypothesized that embodied prosodic training involving the imitation of logatomes and gestures would yield greater improvements in oral-reading pronunciation than prosodic training that involved repeating logatomes, compared with a baseline condition where participants repeated speech only.

The participants for the present study were bilingual Catalan–Spanish speakers learning French as an additional language. Despite the close relationship between the Romance languages, Catalan learners face clear challenges in the acquisition of French prosody. Unlike French, Catalan does not have a phrasal-marked Accentual Phrase (AP) constituent (Prieto *et al.* 2015). In French, the AP may group together more than one lexical word plus the accompanying clitics, and it is characterized by the presence of an obligatory final pitch accent and an optional initial rise, which have a demarcative function (Delais-Roussarie *et al.* 2015). This means that while stress functions on a phrasal level in French, marking right—and optionally left—phrase boundaries (see, among others, Di Cristo and Hirst 1993; Jun and Fougeron 1995, 2000; Delais-Roussarie *et al.* 2015), in Catalan, as in Spanish, it works on a lexical level (MacPherson 1975; Mascaró 1976). As a consequence of the lack of lexical stress in French, there is a strong syncretism between accentuation, phrasing, and intonation, while, by contrast, Catalan and Spanish generally group two or three prosodic words, with no initial or final demarcative tonal features (Nibert 2000; Prieto *et al.* 2015). A second basic difference between the two languages lies in the phonetic properties of stress realization, which mainly affects the duration of the stressed syllable. French stress is realized by a more extreme lengthening of the stressed phrase-final syllable, and more particularly of the full vowel, than what is seen in Catalan (Fletcher 1991; Vaissière 1991; Di Cristo and Hirst 1993; Astésano 2001). This was demonstrated by Baills and Prieto (2021), who carried out an exploratory acoustic analysis comparing the duration of sentence-final stressed and unstressed syllables in 20 pairs of cognate words in Catalan and French (e.g. balcó—balcon 'balcony') and detected significantly longer phrase-final syllable durations in the French words.

Reading aloud is a common language classroom practice in any number of activities, despite the fact that it may hinder comprehension [e.g. Gabrielatos 2002; but see Gibson (2008) for exceptions]. For the purpose of teaching pronunciation, it may strengthen the grapho-phonemic correspondences of the L2 (e.g. Gibson 2008) and improve learners' fluency (e.g. Klomjit 2013). Riquelme Gil *et al.* (2017) tested the repeated reading method[3] by asking young Spanish learners of English to read short passages from a story book both silently and aloud over the course of six weeks. They found that, after the intervention, participants produced less pronunciation errors in three different tasks carried out at pre- and posttest: re-reading of the original text, reading of an unknown text, and spontaneous speech. In this study, we adopted the repeated reading method as a way to introduce our stimuli and training materials.

Finally, oral proficiency in a language is most often rated in terms of comprehensibility, fluency, and accentedness (e.g. Munro and Derwing 2015). However, more concrete features, from a wide range of suprasegmental features (e.g. Munro 1995; Trofimovich and Isaacs 2012; Saito *et al.* 2016) to segments with high functional load (e.g. Suzukida and Saito 2019), can also be rated to evaluate pronunciation accuracy. Thus, five dimensions were selected here to assess participants' pronunciation before and after training: comprehensibility, fluency, accentedness, segmental accuracy, and suprasegmental accuracy. This is in keeping with the view expressed by Saito and Plonsky (2019) that a truly comprehensive assessment of the effects of pronunciation instruction on L2 speech must take into account both holistic and specific levels of measurement.

METHODS

Participants

Seventy-five first- or second-year students doing undergraduate degrees in Translation and Interpreting or Applied Language at the Universitat Pompeu Fabra in Barcelona participated in this study. They were all enrolled in an intermediate-level French course, which consisted of 90 min of language theory and 60 min of language practice (including a variety of oral and written activities) per week over a four-month term. This pronunciation training study was incorporated into the French course and took place over five weeks. Participation was therefore mandatory for all students. The actual pronunciation training was carried out by the first author.

All of the students reported themselves to be Catalan–Spanish bilinguals. Results of a preliminary questionnaire showed that as a group they used Catalan 61% of the time on average in their daily lives (SD = 28.4). Participants self-reported their French proficiency to be between CEFR levels A2 and B1. They also reported that they had studied English as a foreign language to one extent or another. Prior to participation in this study, they all signed a form consenting to the use of audio recordings of their speech for the purposes of this research.

The 75 participants were randomly assigned to one of the three conditions such that the speech-only group contained 27 participants (M_{age} = 20.04, SD = 2.87 2 males), the non-embodied logatome group contained 22 (M_{age} = 19.79, SD = 1.32, 4 males), and the embodied logatome group contained 26 (M_{age} = 19.80, SD = 1.37 2 males).

An a priori power analysis was conducted using G*power3 to test the interaction between groups and tests (ANOVA: repeated measures, within-between factors; medium target effect size $\eta^2 = 0.04$, alpha = 0.05). Results showed that a total sample of 66 participants was required to achieve a power of 0.95.

Audiovisual stimuli for the pronunciation training sessions

All the materials used in this experiment can be seen in Online Appendix A and are openly available at https://osf.io/93pdw.

The materials used in the training sessions consisted of dialogues taken from a French-language textbook that focuses on teaching oral skills through meaningful, enjoyable texts (Martins and Mabilat 2003). Nine dialogues were used in the training sessions, with a different set of three employed in each of the three sessions. While the intention was to select target dialogues that did not include novel vocabulary, a short glossary in Catalan was provided adjacent to each text to be read clarifying any words that might be unfamiliar to lower-level participants. In addition, dialogues were chosen such that the oral performance of the dialogues would include a variety of intonation contours arising from different situational contexts.

A total of five sentences in each dialogue were selected (around 42% of the total number of sentences) to be target stimuli for repetition during the training sessions. Video recordings were then made of three instructors performing these five stimuli in the three experimental conditions. The instructors (two female and one male) were two specialists in the VT method and the first author of this study. Recording took place over 4 h at the second author's university broadcasting studio with professional equipment and help from a technical assistant (see Online Appendix B for a detailed description of the recording procedure).

In all recordings, the frame of the image was set to show the upper half of each instructor's body to allow a clear view of the face and all hand movements. For the speech condition, the instructors simply pronounced the target sentences clearly while standing still. For the non-embodied logatome condition, the logatome consisted of pronouncing the syllable 'da' instead of the phrase's syllables, but without changing the intonation of the phrase. As for the embodied logatome condition, as the logatome was uttered, the right hand, palm open facing downwards, made a sweeping left-to-right movement across the body at chest level that mimicked through upwards and downwards movements the rises and falls of the pitch contours of their oral utterance as they spoke. Importantly, these movements served to depict not only intonational pitch movements but also the rhythmic features of their speech by increased or decreased velocities and short pauses in the hand's movement. Figure 1 shows sequences of video stills from a sample stimulus trial in the non-embodied (top panel) and embodied logatome condition (middle panel), as well as the pitch contour and corresponding logatome 'da' syllables (bottom panel).

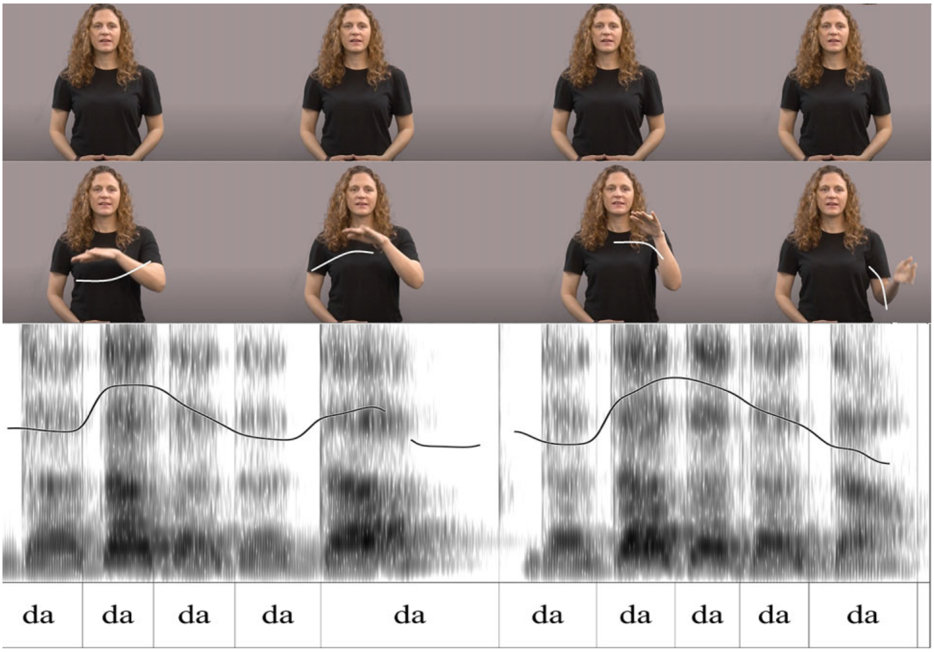

Figure 1: Stills from stimulus videos showing an instructor performing in the non-embodied (top panel) and embodied logatome conditions (middle panel). In this case, the target sentence is Je suis désolée, votre lettre n'est pas là 'I am sorry, your letter is not here'. Acoustic data and the intonation pattern of the logatome sequence are shown in the bottom panel.

The video clips were edited in Adobe Premiere Pro 13 to create three sets of stimulus materials corresponding to one of the experimental conditions (speech-only, non-embodied logatome plus speech, and embodied logatome plus speech). Figure 2 shows the training sequence for each sentence. The instructor pairs varied throughout the stimuli for the nine dialogues; however, the combination for each dialogue was consistent across the three conditions. Therefore, all participants were able to listen to the three instructors throughout the course of the training.

Finally, the nine dialogues were acted out by amateur actors in different appropriate locations, either in France or in Catalonia (but using French native speakers), and video-recorded. After each dialogue had been trained, the participants would be shown the corresponding enactment as a kind of wrap-up activity (these video files are available at https://osf.io/93pdw).

Thus, the final material for each session consisted of three training videos (five sentences each) in one of the experimental conditions, each one followed by the enactment of the full source dialogue as a wrap-up. This material was embedded in online presentation format using Alchemer[3] software,

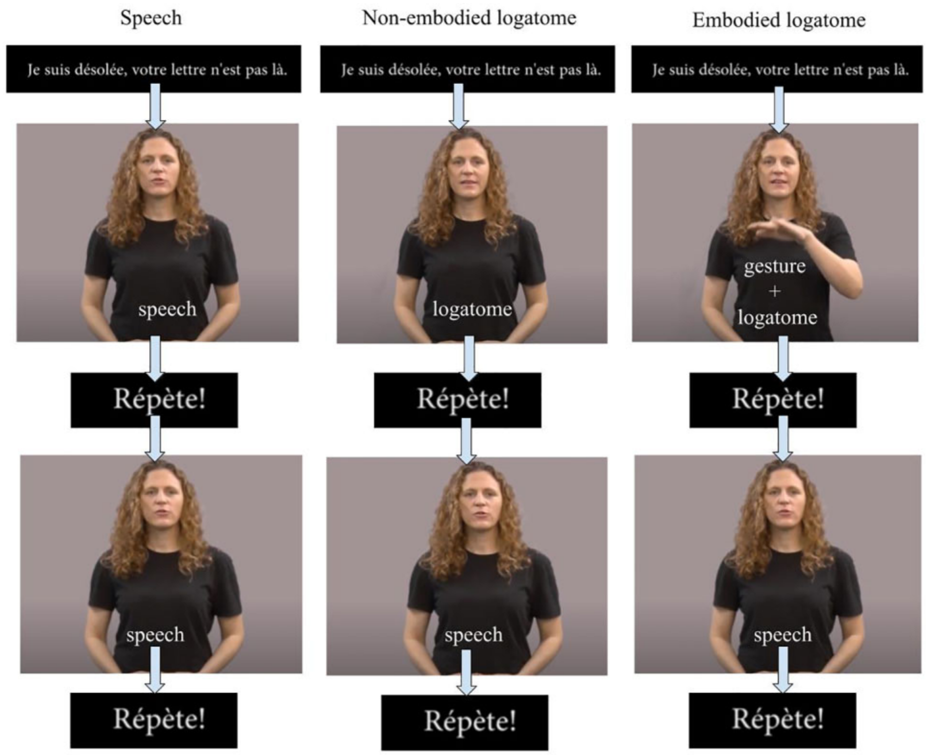

Figure 2: Audiovisual training sequence for each sentence.

accompanied by written instructions. Since training involved three separate sessions, three such presentations were prepared for each condition.

Pretest and posttest materials and control measures

Participants' pronunciation was tested before and after training by means of a dialogue-reading task. The pretest and posttest were identical and consisted of four dialogues to be read aloud, three of them also appearing in a training session (one dialogue from each set of three used in the training sessions) and the fourth being untrained. These materials as well as corresponding instructions were uploaded to the same online presentation platform as the training materials and could be accessed by a link provided by the teacher.

Two sets of data were gathered to control for potential differences between groups. The first set covered participants' self-reported proficiency in French and prior experience learning that language. This questionnaire yielded four separate scores per participant: the number of years spent learning French; the number of months spent learning French as an extracurricular activity (outside school/university); the number of months spent abroad in a

French-speaking country; and a nominal value from 1 to 6 indicating self-reported proficiency in French (A1 = 1, A2 = 2, B1 = 3, B2 = 4, B2 = 4, C1 = 5, and C2 = 6).

The second set of data, which was gathered at the end of the posttest, was a rating of the participants' satisfaction with regard to the pronunciation training they had received. This online questionnaire asked the participants to rate their satisfaction with the pronunciation training they had received by reacting to the following statements on a scale from 1 ('I strongly disagree') to 9 ('I strongly agree'): (i) I liked these pronunciation training sessions; (ii) I improved my pronunciation; and (iii) I would like to repeat this kind of activity with other texts.

Procedure

Figure 3 provides an overview of the experimental design of the three-session training programme with pre- and posttest.

A week prior to the first training session, participants received from their respective French language instructors a link to the website containing the materials for the pretest task, which consisted of video-recording themselves as they read aloud four dialogues. The full task took on average 10 min. They were required to complete the task and upload the resulting video files to a shared folder within three days of having received the link from their instructor. Participants were asked to carry out the pretest using their own computer and headset in a quiet environment. The purpose of video recording was to ensure that the tasks were performed properly and uploaded student recordings were regularly checked by the first author for this purpose. The audio tracks from the recordings were then extracted and saved for further analysis.

The experimental training took place in three separate sessions over three weeks during the regular class period of participants' French course. All sessions took place on the university premises in individual soundproofed booths equipped with computers and microphones. Before starting the first training session, the students answered the language questionnaire. The teacher then emailed a link to one of three separate sets of training materials, depending on the experimental group to which the participant had been randomly assigned previously. After reading some initial instructions, participants then completed the training procedure individually at their own pace, recording their speech output throughout using Audacity software. The training procedure consisted of completing a set of subtasks associated with three dialogues. Figure 4 shows the sequence of subtasks related to one dialogue. Participants moved from one step in the process to the next by clicking on their keyboard. The order of presentation of the dialogues was randomized automatically by the software. Each full training session lasted roughly 30 min, about 10 min per dialogue unit.

Once they had completed all three training tasks, participants stopped the recording process and uploaded the resulting audio file to a shared folder.

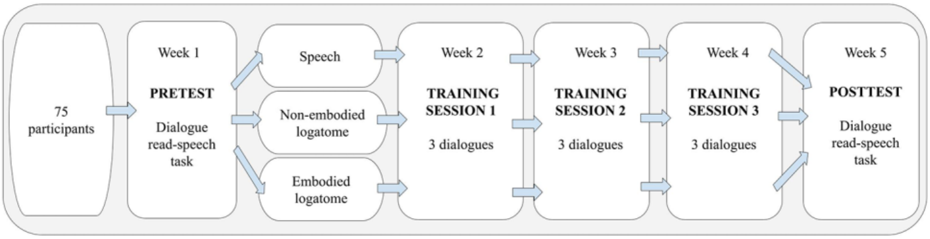

Figure 3: Diagram of the experimental design of the training programme.

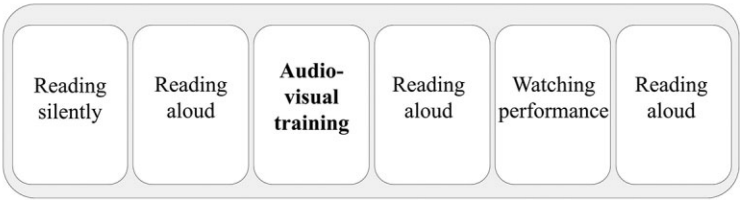

Figure 4: Procedure of a trial involving one dialogue. Each full training session consisted of three such sequences.

While the training session was in progress, the instructor monitored participant behaviour from outside the individual booths, particularly to ensure that participants in the embodied logatome training were duly performing the required hand movements. Because the class period was longer than the time required for the training session, once they had completed the training task, participants then proceeded to complete other language-learning activities assigned by their instructor.

One week after the third and last training session took place, participants took the posttest, which, like the pretest, consisted of recording themselves reading four dialogues aloud and then uploading the recordings to a shared folder.

Pronunciation assessment

The pronunciation assessment was carried out by three raters (two female and one male), all native speakers of French with extensive L2 teaching experience with Catalan learners. They took part in a 1-h training session to receive detailed explanations about the five dimensions they were to evaluate and instructions on how to apply the nine-point assessment scales (1 = worst score, 9 = best score) with which they would rate participant output on each of the five dimensions. They then individually practised applying these scales using five sample dialogues read by the participants and the first author provided feedback to ensure a clear understanding of the five dimensions.

Each rater evaluated the totality of speech samples taken from participant-recorded pretest and posttest audio files [(4 pretest dialogues + 4 posttest dialogues) × 75 participants = 600 audio files] for each of the five pronunciation dimensions, giving a total of 3,000 scores. The speech samples consisted of the full dialogues (durations in s: $M = 27.86$, SD = 6.07 for dialogue 1, $M = 41.28$, SD = 7.78 for dialogue 2, $M = 27.61$, SD = 4.83 for dialogue 3, and $M = 27.39$, SD = 4.84 for dialogue 4) and were randomized and grouped into 16 different batches using Alchemer online software. Each batch took about 1 h to rate. While we advised the raters to take only short breaks during each batch rating, so as not to lose the data, we recommended they rested as much as needed between each batch in order to avoid listener fatigue. The raters rated the dataset at home over the course of seven days by completing one to three batches per day. They received monetary compensation for their work.

The items' internal consistency was checked by means of Cronbach's alpha and satisfactory coefficients were obtained (0.93 for comprehensibility, 0.92 for fluency, 0.79 for accentedness, 0.92 for segmental accuracy, and 0.88 for suprasegmental accuracy). Interrater reliability was assessed by calculating the intraclass correlation coefficient (two-way random, absolute agreement, see Landers 2015), showing moderate to good agreement among the raters: ICC = 0.56, $F(599\ 1198) = 2.29$, $p < 0.001$, 95% CI [0.50, 0.62] for comprehensibility; ICC = 0.64, $F(599\ 1198) = 2.79$, $p < 0.001$, 95% CI [0.59, 0.69] for fluency; ICC = 0.73, $F(599\ 1198) = 3.76$, $p < 0.001$, 95% CI [0.69, 0.77] for accentedness; ICC = 0.61, $F(599\ 1198) = 2.59$, $p < 0.001$, 95% CI [0.56, 0.66] for segmental accuracy; and ICC = 0.72, $F(599\ 1198) = 3.64$, $p < 0.001$, 95% CI [0.68, 0.76] for suprasegmental accuracy.

Statistical analyses

Statistical analyses were carried out with IBM SPSS 23. Two databases were set up, one sorted by participants and the other sorted by items (i.e. stimulus sentence). In order to test for homogeneity across the three groups, the participant-sorted database was used to show individual scores for the self-reported French language proficiency measures and satisfaction with the training experience. As the measures showed a skewed distribution, differences between groups and mean satisfaction scores across groups were explored by means of a non-parametric Kruskal–Wallis *H* test.

The item-sorted database was used to analyse the effect of type of training (speech versus logatome versus embodied logatome) on participant pronunciation measures. Five general linear mixed models (GLMMs) were run, each with one of the following dependent variables: comprehensibility, fluency, accentedness, segmental accuracy, and suprasegmental accuracy. For all these variables, Shapiro–Wilk tests showed that the scores were positively skewed. Therefore, an inverse Gaussian distribution with a log function was specified in each model. Group (three levels: speech-only, logatome, and embodied logatome), session (two levels: pretest and posttest), and Group × Session and familiarity (two levels: trained and untrained items) were set as fixed factors; random intercepts were set for participants and for items. Sequential Bonferroni pairwise comparisons were used.

RESULTS

Homogeneity across groups

Results of the Kruskal–Wallis *H* test showed that there was no significant difference between the three groups in terms of age, $\chi^2(2) = 0.62$, $p = 0.73$; years of learning French, $\chi^2(2) = 0.36$, $p = 0.84$; months of extra-curricular French lessons, $\chi^2(2) = 0.80$, $p = 0.77$; months of stay abroad, $\chi^2(2) = 0.45$, $p = 0.80$; and self-assessed French proficiency, $\chi^2(2) = 3.02$, $p = 0.22$ (see Table 1).

Training effects

A general improvement between pre- and posttest was observed in all the measures. A general view of the results is presented in Figure 5. The descriptive results are gathered in Table 2. Below, we report only the significant results. All the inferential statistical results are available in Online Appendix C.

The result of the GLMM with comprehensibility as the dependent variable showed a significant effect of session, $F(1, 1793) = 18.63$, $p < 0.001$, $\eta^2 = 0.01$, 90% CI [0.004, 0.02]. No significant effect of group, Session × Group, or familiarity was found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 4.28$, $p = 0.04$, $\eta^2 = 0.002$, 90% CI [0.0076, 0.0079] for the speech-only group, $F(1, 1793) = 3.57$, $p = 0.059$, $\eta^2 = 0.002$, 90% CI [0, 0.007] for the non-embodied logatome group, and $F(1, 1793) = 12.56$, $p < 0.001$, $\eta^2 = 0.007$, 90% CI [0.002, 0.015] for the embodied logatome group.

The result of the GLMM with fluency as the dependent variable showed a significant effect of session, $F(1, 1793) = 96.50$, $p < 0.001$, $\eta^2 = 0.05$, 90% CI [0.973, 0.976]. No significant effect of group, Group × Session, or familiarity was found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 28.03$, $p < 0.001$, $\eta^2 = 0.01$, 90% CI [0.007, 0.026] for the speech-only group, $F(1, 1793) = 25.84$, $p < 0.001$, $\eta^2 = 0.01$, 90% CI [0.006, 0.025] for the non-embodied logatome group, and $F(1, 1793) = 44.57$, $p < 0.001$, $\eta^2 = 0.02$, 90% CI [0.01, 0.04] for the embodied logatome group.

The result of GLMM with accentedness as the dependent variable showed a significant effect of session, $F(1, 1793) = 68.14$, $p < 0.001$, $\eta^2 = 0.03$, 90% CI [0.02, 0.05], and Group × Session, $F(2, 1793) = 7.38$, $p = 0.001$, $\eta^2 = 0.008$, 90% CI [0.002, 0.016]. No significant effect of group or familiarity was found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 9.28$, $p = 0.002$, $\eta^2 = 0.005$, 90% CI [0.001, 0.012] for the speech-only group, $F(1, 1793) = 11.53$, $p = 0.001$, $\eta^2 = 0.006$, 90% CI [0.002,

Table 1: Descriptive statistics and rank mean values for age and French proficiency measures in each group

	Group	*M*	SD	SE	95% CI	Rank mean
Age	Speech only	20.04	2.91	0.55	[18.91 21.16]	35.64
	Non-embodied logatome	19.82	1.37	0.29	[19.21 20.42]	39.93
	Embodied logatome	19.76	1.30	0.26	[19.22 20.30]	38.94
Formal instruction in French	Speech only	4.25	2.58	0.49	[3.25, 5.25]	36.20
	Non-embodied logatome	4.41	2.01	0.43	[3.51, 5.30]	39.73
	Embodied logatome	5.08	3.84	0.76	[3.49, 6.66]	38.50
Extra-curricular instruction in French	Speech only	0.71	1.01	0.19	[0.32 1.11]	38.20
	Non-embodied logatome	0.81	1.11	0.24	[0.31 1.30]	39.95
	Embodied logatome	0.73	1.36	0.27	[0.17 1.29]	36.06
Stay abroad	Speech only	0.69	1.53	0.29	[0.09 1.28]	36.91
	Non-embodied logatome	0.84	1.72	0.36	[0.08 1.60]	37.14
	Embodied logatome	0.66	1.26	0.25	[0.14 1.18]	39.98
Self-reported proficiency	Speech only	3.71	0.71	0.13	[3.44, 3.99]	35.57
	Non-embodied logatome	3.73	0.93	0.20	[3.31, 4.14]	34.52
	Embodied logatome	4.04	0.89	0.18	[3.67, 4.41]	43.78

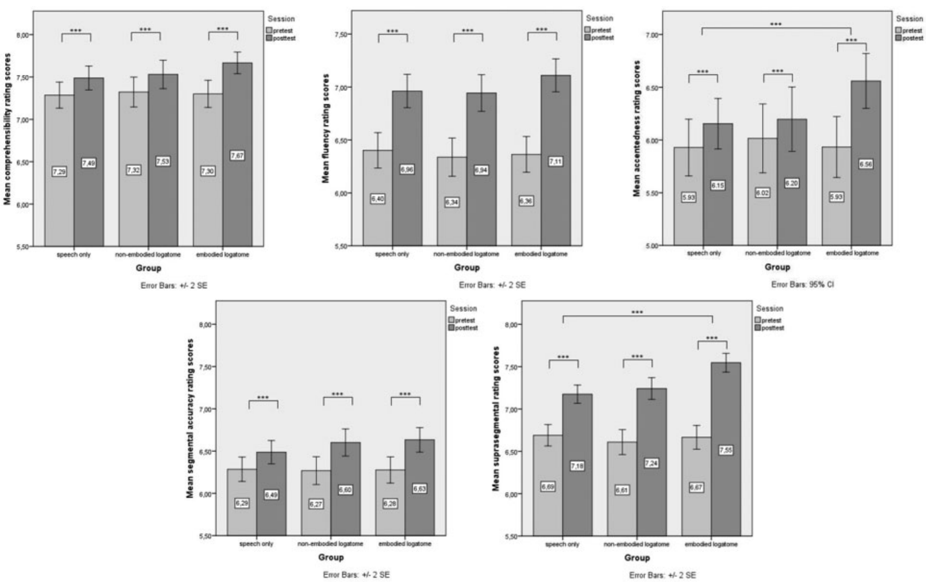

*Figure 5: Mean rating scores at pre- and posttest for the five pronunciation measures. Significant contrasts are labelled with asterisks (***p < 0.001).*

Table 2 Mean, standard deviation, standard error, and 95% confidence intervals at pre- and posttest for the speech only group, the non-embodied logatome group, and the embodied logatome group, and for trained and untrained items (Familiarity) in the five pronunciation assessment measures

		Comprehensibility				Fluency			
		Mean	SD	SE	95% CI	Mean	SD	SE	95% CI
Speech only	Pretest	7.29	1.41	0.08	[7.13, 7.44]	7.29	1.41	0.08	[7.13, 7.44]
	Posttest	7.49	1.30	0.07	[7.35, 7.63]	7.49	1.30	0.07	[7.35, 7.63]
Non-embodied logatome	Pretest	7.32	1.43	0.09	[7.15, 7.50]	7.32	1.43	0.09	[7.15, 7.50]
	Posttest	7.53	1.36	0.08	[7.36, 7.70]	7.53	1.36	0.08	[7.36, 7.70]
Embodied logatome	Pretest	7.30	1.39	0.08	[7.30, 7.46]	7.30	1.39	0.08	[7.30, 7.46]
	Posttest	7.67	1.10	0.06	[7.54, 7.79]	7.67	1.10	0.06	[7.54, 7.79]
Familiarity	Trained	7.44	1.35	0.04	[7.37, 7.51]	7.44	1.35	0.04	[7.37, 7.51]
	Untrained	7.40	1.33	0.06	[7.28, 7,52]	7.40	1.33	0.06	[7.28, 7,52]
		Accentedness				Segmental accuracy			
		Mean	SD	SE	95% CI	Mean	SD	SE	95% CI
Speech only	Pretest	5.93	1.27	0.07	[5.80, 6.07]	6.29	1.31	0.07	[6.14, 6.43]
	Posttest	6.18	1.23	0.07	[6.05, 6.31]	6.49	1.26	0.07	[6.35, 6.62]
Non-embodied logatome	Pretest	5.94	1.22	0.07	[5.79, 6.09]	6.27	1.34	0.08	[6.11, 6.43]
	Posttest	6.25	1.15	0.07	[6.11, 6.39]	6.60	1.30	0.08	[6.35, 6.62]
Embodied logatome	Pretest	5.92	1.16	0.07	[5.79, 6.06]	6.28	1.34	0.08	[6.12, 6.43]
	Posttest	6.61	1.05	0.06	[6.49, 6.73]	6.63	1.25	0.07	[6.49, 6.78]
Familiarity	Trained	6.14	1.20	0.03	[6.08, 6.20]	6.42	1.31	0.04	[6.35, 6,49]
	Untrained	6.13	1.23	0.06	[6.02, 6.24]	6.43	1.31	0.06	[6.31, 6.55]
		Suprasegmental accuracy							
		Mean	SD	SE	95% CI				
Speech only	Pretest	6.69	1.16	0.06	[6.57, 6.81]				
	Posttest	7.18	0.99	0.05	[7.07, 7.28]				
Non-embodied logatome	Pretest	6.61	1.20	0.07	[6.46, 6.76]				
	Posttest	7.24	1.04	0.06	[7.12, 7.37]				
Embodied logatome	Pretest	6.67	1.22	0.07	[6.53, 6.80]				
	Posttest	7.55	0.95	0.05	[7.44, 7.66]				
Familiarity	Trained	7.01	1.17	0.03	[6.95, 7.07]				
	Untrained	6.93	1.08	0.05	[6.83, 7.03]				

0.014] for the non-embodied logatome group, and $F(1, 1793) = 61.91$, $p < 0.001$, $\eta^2 = 0.03$, 90% CI [0.02, 0.05] for the embodied logatome group, as well as a significant difference between groups at posttest only, $F(2, 1793) = 3.50$, $p = 0.03$, $\eta^2 = .004$, 90% CI [0.0001, 0.009], with significantly higher improvement in the embodied logatome group than in the speech-only group, $p = 0.04$.

The result of the GLMM with segmental accuracy as the dependent variable showed a significant effect of session, $F(1, 1793) = 27.44$, $p < 0.001$, $\eta^2 = 0.01$, 90% CI [0.007, 0.02]. No significant effect of group, Session × Group, or familiarity was found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 4.79$, $p = 0.03$, $\eta^2 = 0.003$, 90% CI [0.0001, 0.008] for the speech-only group, $F(1, 1793) = 10.21$, $p = 0.001$, $\eta^2 = 0.006$, 90% CI [0.001, 0.01] for the non-embodied logatome group, and $F(1, 1793) = 12.28$, $p < 0.001$, $\eta^2 = 0.007$, 90% CI [0.002, 0.01] for the embodied logatome group.

The result of the GLMM with suprasegmental accuracy as the dependent variable showed a significant effect of session, $F(1, 1793) = 197.89$, $p < 0.001$, $\eta^2 = 0.10$, 90% CI [0.08, 0.12], and Session × Group, $F(2, 1793) = 6.25$, $p = 0.002$, $\eta^2 = .007$, 90% CI [0.002, 0.01]. No significant effect of group or familiarity was found. Post hoc analyses revealed a significant effect of session for the three groups, $F(1, 1793) = 39.59$, $p < 0.001$, $\eta^2 = 0.02$, 90% CI [0.01, 0.03] for the speech-only group, $F(1, 1793) = 52.89$, $p < 0.001$, $\eta^2 = 0.03$, 90% CI [0.02, 0.04] for the non-embodied logatome group, and $F(1, 1793) = 116.32$, $p < 0.001$, $\eta^2 = 0.06$, 90% CI [0.04, 0.08] for the embodied logatome group, as well as a significant difference between the embodied logatome group and the speech-only group at posttest, $F(2, 1793) = 3.40$, $p = 0.03$, $\eta^2 = 0.004$, 90% CI [0.0001, 0.009].

In sum, results showed a significant improvement in read–speech comprehensibility, fluency, accentedness, segmental accuracy, and suprasegmental accuracy after training in all the three groups, with higher effect sizes for the embodied logatome group in all the measures. In addition, the embodied logatome group improved significantly more than the speech-only group in terms of accentedness and suprasegmental accuracy, while the non-embodied logatome group did not (see Online Appendix C for fixed effects and contrast estimates). No significant differences were found between trained and untrained items, showing that participants improved equally in their pronunciation of French when reading a text aloud regardless of whether they had received prior training with that particular text or not.

Satisfaction with training

Results of the general satisfaction questionnaire showed high degrees of satisfaction for the three measures, as shown in Table 3. The Kruskal–Wallis *H* test showed that there was no statistically significant difference among the groups in terms of likeability of the activity [$\chi^2(2) = 4.18$, $p = 0.12$, with a mean rank score of 79.72 for the speech-only group, 64.31 for the non-embodied logatome group, and 68.41 for the embodied logatome group], self-perception of improvement [$\chi^2(2) = 3.58$, $p = 0.17$, with a mean rank score of 79.54 for the speech group, 68.17 for the non-embodied logatome group, and 65.11 for the

Table 3 Mean results for the satisfaction questionnaire across the three groups based on a 1–9 scale, from 1 ('I strongly disagree') to 9 ('I strongly agree')

	Group	Mean	SD	SE	95% CI for mead
I liked the pronunciation training sessions	Speech only	8.30	0.82	0.11	[8.07, 8.52]
	Non-embodied logatome	7.81	1.19	0.18	[7.44, 8.18]
	Embodied logatome	7.74	1.58	0.23	[7.27, 8.21]
I think I improved my pronunciation	Speech only	6.93	1.40	0.19	[6.54, 7.31]
	Non-embodied logatome	6.48	1.67	0.26	[5.96, 7.00]
	Embodied logatome	6.30	1.64	0.24	[5.82, 6.79]
I would like to repeat this kind of activity with other texts	Speech only	7.30	1.47	0.20	[6.89, 7.70]
	Non-embodied logatome	6.71	2.10	0.32	[6.06, 7.37]
	Embodied logatome	6.87	2.05	0.30	[6.26, 7.48]

embodied logatome group], and interest in repeating the activity [$\chi^2(2) = 76.06$, $p = 0.53$, with a mean rank score of 76.06 for the speech group, 66.83 for the non-embodied logatome group, and 70.41 for the embodied logatome group].

DISCUSSION

The present study explored the effects of embodied prosodic training via visuospatial gestures depicting rhythm and intonation on overall (comprehensibility, fluency, and accentedness) and specific (segmental and suprasegmental accuracy) measures of pronunciation with Catalan intermediate learners of French. This embodied training was embedded in repeated reading and oral imitation activities, while the effects on pronunciation were assessed through an oral-reading task. One week after the last session of our intervention, the speech-only group, the non-embodied prosodic group (logatome only), and the embodied logatome group (logatome and gesture) significantly improved in all the measures compared with pretest. Our results revealed that participants in the embodied logatome group obtained significantly higher gains compared with the speech-only group in terms of accentedness and suprasegmental accuracy, while the non-embodied logatome group did not. Nonetheless, the difference between the non-embodied and embodied logatome groups was not significant in any of the measures, despite systematic larger effect sizes in the improvement between pretest and posttest for the embodied logatome group. These results demonstrate that only when accompanied by a gesture did the logatome result in a superior effect on learning outcomes.

Our results provide evidence that the embodiment of a phonological feature in a foreign language helped learners process this specific feature more efficiently: embodying prosody directly improved the scores on suprasegmental accuracy. The motoric action provided by the perception and the production of the gesture may have reached the visuospatial phonological loop (Wilson and Emmorey 1997) and may have been associated to the adequate mental representation of rhythmic and melodic patterns, facilitating the processing and the acquisition of such features. In the absence of the visuospatial gesture, the ability of prosodic training with only logatomes to convey the saliency of suprasegmental features may not have been sufficient to make a difference. Our study thus supports the claims in favour of embodied techniques for teaching pronunciation (e.g. Billières 2002; Acton *et al.* 2013; Chan 2018) and sheds a new light on the mechanism behind the positive effects of rhythmic embodied training involving rhythmic gestures (Gluhareva and Prieto 2017; Llanes-Coromina *et al.* 2018) or hand clapping (Iizuka et al. 2020; Zhang *et al.* 2020; Baills and Prieto 2021), and embodied gesture training focusing on specific suprasegmental features (Yuan *et al.* 2019; Li *et al.* 2020).

At the practical level, our findings offer additional evidence of the efficacy of one of the main features of the VT method, namely the use of embodied logatomes. Though our findings confirm previous results that the use of embodied logatomes may not provide an advantage for fluency measures (Alazard 2013), they indicate that this method is able to boost pronunciation learning in terms of accentedness and suprasegmental accuracy. It is of interest to note in the present study the high level of participant satisfaction in all groups, indicating that they felt at ease with a repetition paradigm involving short dialogues, whether this included logatomes and embodiment or not. This suggests that the introduction of novel, maybe unusual methodologies, including using one's body was not a hindrance to learning—on the contrary. Hence, the use of embodied techniques may be of particular interest for language teachers who detect the need to improve their learners' pronunciation at any time during their class, without requiring any materials or heavy preparation.

The lack of any difference between the three training conditions for comprehensibility and fluency measures could be explained by the fact that suprasegmental features may weigh less in these measures than in the accentedness measure (e.g. Trofimovich and Isaacs 2012; Saito *et al.* 2016). However, the larger effect sizes obtained for embodied prosodic training in both measures may also point to a certain advantage for this type of training, which might be amplified if the duration of the training period was extended. As Alazard (2013) pointed out, a longer training period may be necessary to widen the differences among the groups with respect to fluency scores. In addition, regarding comprehensibility scores, it may be the case that these scores were already too high to be able to detect sufficiently large differences between groups and that effects may have been observed with learners of lower proficiency.

In line with previous research that demonstrated the value of pronunciation instruction, our results showed that 30 min of pronunciation training once a week over three weeks helped improve significantly comprehensibility, fluency, and accentedness in L2 read speech regardless of the training method, as our three experimental groups obtained significantly higher scores in those measures at posttest. Moreover, following the recommendation by Saito and Plonsky (2019), this study encompasses the three traditional overall measures of pronunciation, as well as specific segmental and suprasegmental measures, whereas previous literature tends to focus on only one of these aspects. Furthermore, this improvement in pronunciation was evident even in the one read dialogue for which they had not been trained, showing that participants may have been able to generalize what they learned during training to an untrained item and adding some evidence on the generalization of pronunciation gains after pronunciation training, an issue that is seldom raised (e.g. Levis and Pickering 2004).

There are several limitations to the present study. First, we only obtained moderate inter-rater reliability between the three raters. We think that perceptively evaluating long samples of read-speech (between 20 and 40 s) on a scale from one to nine might have allowed for more variability than evaluating single sounds or words. Second, our results are restricted to pronunciation in read speech. Although our findings on oral reading can be considered useful for improving learners' pronunciation—notably, because oral reading is a common task in the second language classroom—it is not clear whether the benefits of the embodied logatome technique would extend to spontaneous speech. As suggested by Saito and Plonsky (2019), more evidence is needed on the effect of perceptive and productive phonologic training on learners' pronunciation skills in spontaneous speech. In order to broaden the scope of the present findings, future studies should take into account spontaneous speech at both the training and testing stages through, for example, picture description tasks. Third, in the present case, the posttest took place one week after training. In light of research showing that gestures aid vocabulary and grammar retention (Macedonia and Klimesch 2014; Nakatsukasa 2016) and phonological learning over time (Li *et al.* 2021), it is likely that a delay longer than one week between training and posttest would provide important information about the durability of the benefits of embodied prosodic training on the development of learners' pronunciation.

Our study does not disentangle the respective benefits of producing and observing the gesture in the embodied logatome group, that is, the effects of training with gesture as opposed to just observing the models an equal number of times. Despite Eskildsen and Wagner's (2013) observation that imitating a speaker's gesture may induce and sustain understanding of the item being learned, the positive effects of gestures may stem from seeing the gesture performed by the instructor rather than from making the gesture. In that respect, there are few empirical studies directly comparing the effects of gesture perception and production. While Baills et al. (2019) did not find any difference between the perception and the production of pitch gestures for learning Mandarin tones and words, Li *et al.* (2021) showed that gesture production can be more beneficial only when the learner performs the gesture correctly.

Hence, further research should look at learners' gestural performance as an important factor when comparing gesture perception and production. For these reasons, it would have been interesting to add a gesture-observation group to the study and to control for individual differences in terms of gesture production accuracy.

Finally, the design of this study did not allow for any interaction with or feedback from the instructors, it was essential to strictly control potential differences between the groups. However, it is highly likely that individual feedback would have enhanced the pronunciation learning outcomes, as previous evidence suggests (Gordon *et al.* 2013; Lee *et al.* 2015). Most importantly, the role of gesture in corrective feedback may be highly relevant (Nakatsukasa 2016; Wang and Loewen 2016; Thompson and Renandya 2020).

CONCLUSION

The Embodied Cognition paradigm has already opened up many possibilities in the field of education, thanks in particular to the proven effects of embodiment on memory for language learning (Madan and Singhal 2012; Kiefer and Trumpp 2012; Macedonia 2019). In the field of second language learning, gestures and movements embodying actions or objects in a foreign language help learners retain new vocabulary (e.g. Quinn-Allen 1995; Tellier 2008; Macedonia and Klimesch 2014). All in all, the results of the present study confirm the predictions of the Embodied Cognition hypothesis for phonological learning and thus favour the embodying of phonological prosodic features in the teaching of pronunciation. In particular, we demonstrate the value of *embodied oral reading* in the development of L2 reading skills. Adding visuospatial gestures depicting prosody and probably other phonological features should be added to the toolkit of the second language teacher.

ACKNOWLEDGEMENTS

Many thanks to the members of the Prosodic Studies Group at UPF for their comments and advice during internal meetings. We would like to particularly thank Michel Billières for his participation in the creation of the stimuli and his expert eyes on the verbotonal method. A special thank goes to Bruno Bastard for the technical assistance during the videorecording of the stimuli at Toulouse II University. We are very grateful to the students of the FLE (French as a Foreign Language) Master programme at Toulouse II University for their interpretation of the dialogues. We are also indebted towards UPF French teachers Élisabeth Miche and Gilhem Naro for their kind cooperation during the experiment. Finally, we would like to thank all the participants in this study as well as our raters.

FUNDING

The first author acknowledges a grant from the Department of Translation and Language Sciences at the 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). This study was supported by the Pla de Recerca de Catalunya (ref: 2017 SGR_971).

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

Conflict of interest statement. None declared.

ENDNOTES

1 https://hipoeces.blogspot.com/, https://www.actonhaptic.com/basic
2 There is also ample evidence that pitch gestures enhance the perception of intonation and lexical tones by L2 learners (e.g. Morett and Chang 2015; Hannah *et al.* 2017; Kelly *et al.* 2017; Baills et al. 2019). https://www.alchemer.com/
3 Originally proposed by Samuels (1979, in (Taguchi et al., 2004)).

REFERENCES

Acton, W., **A. Ann Baker**, **M. Burri**, and **B. Teaman**. 2013. 'Preliminaries to haptic-integrated pronunciation instruction' in J. Levis and K. LeVelle (eds): *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference*. Iowa State University, pp. 234–44.

Alazard, C. 2013. 'Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Etrangère, doctoral dissertation, Université Toulouse le Mirail—Toulouse II, HAL theses.

Alazard, C., **C. Astésano**, and **M. Billières**. 2010. 'The implicit prosody hypothesis applied to foreign language learning: From oral abilities to reading skills' in ISCA Archive (ed.): *Proceedings of the 5th International Conference on Speech Prosody*, Chicago, 100648.

Anderson-Hsieh, J., **R. Johnson**, and **K. Koehler**. 1992. 'The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure,' *Language Learning* 42/4: 529–55.

Astésano, C. 2001. *Rythme et Accentuation en Français: Invariance et Variabilité Stylistique*. L'Harmattan.

Atkinson, D. 2010. 'Extended, embodied cognition and second language acquisition,' *Applied Linguistics* 31/5: 599–622.

Bahnmueller, J., **T. Dresler**, **A. C. Ehlis**, **U. Cress**, and **H. C. Nuerk**. 2014. 'NIRS in motion—unraveling the neurocognitive underpinnings of embodied numerical cognition,' *Frontiers in Psychology* 5: 743.

Baills, F. and **P. Prieto**. 2021. 'Embodying rhythmic properties of a foreign language through hand-clapping helps children to better pronounce words,' *Language Teaching Research* 136216882098671. https://doi.org/10.1177/1362168820986716

Baker, A. 2014. 'Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions,' *TESOL Quarterly* 48/1: 136–63.

Barsalou, L. W. 2008. 'Grounded cognition,' *Annual Review of Psychology* 59: 617–45.

Bergmann, P., J. Brenning, M. Pfeiffer, and E. Reber (eds). 2012. *Prosody and Embodiment in Interactional Grammar*. De Gruyter.

Biau, E. 2015. *Beat Gestures and Speech Processing: When Prosody Extends to the Speaker's Hands*, Doctoral dissertation, Universitat Pompeu Fabra, Tesis Doctorals en Xarxa.

Billières, M. 2002. 'Le corps en phonétique corrective' in R. Renard (ed.): *Apprentissage D'une Langue Étrangère Seconde 2. La Phonétique Verbo-Tonale*, De Boeck Université, pp. 35–70.

Burri, M. and **A. Baker**. 2016. 'Teaching rhythm and rhythm grouping: The butterfly technique,' *English Australia Journal: The Australian Journal of English Language Teaching* 31/2: 72–7.

Burri, M., **A. Baker**, and **W. Acton**. 2016. 'Anchoring academic vocabulary with a "hard-hitting" haptic pronunciation teaching technique' in T. Jones (ed.): *Pronunciation in the Classroom: The Overlooked Essential*. Teachers of English to Speakers of Other Languages, pp. 17–26.

Cekaite, A. 2015. 'The coordination of talk and touch in adults' directives to children: Touch and social control,' *Research on Language and Social Interaction* 48: 152–75.

Chan, M. J. 2018. 'Embodied pronunciation learning: Research and practice,' *The Catesol Journal* 30/1: 47–68.

Connell, L., **Z. G. Cai**, and **J. Holler**. 2013. 'Do you see what I'm singing? Visuospatial movement biases pitch perception,' *Brain and Cognition* 81/1: 124–30.

de Bot, K. 1983. 'Visual feedback of intonation: Effectiveness and induced practice behavior,' *Language and Speech* 4/4: 331–50.

Delais-Roussarie, E., **C. Buthke**, **B. Post**, and **A. F. di Cristo**. 2015. 'Intonational phonology of French: Developing a ToBI system for French' in S. Frota and P. Prieto (eds): *Intonation in Romance*. Oxford University Press, pp. 63–100.

De Mareüil, P. B. and **B. Vieru-Dimulescu**. 2006. 'The contribution of prosody to the perception of foreign accent,' *Phonetica* 63/4: 247–67.

Derwing, T. M., **M. J. Munro**, and **G. Wiebe**. 1998. 'Evidence in favor of a broad framework for pronunciation instruction,' *Language Learning* 48/3: 393–410.

Desai, R. H., **J. R. Binder**, **L. L. Conant**, and **M. S. Seidenberg**. 2010. 'Activation of sensory-motor areas in sentence comprehension,' *Cerebral Cortex* 20/2: 468–78.

Di Cristo, A. and **D. Hirst**. 1993. 'Rythme syllabique, rythme mélodique et représentation hiérarchique de la prosodie du français,' *Travaux de L'Institut de Phonétique D'Aix* 15: 9–24.

Dolscheid, S., **R. M. Willems**, **P. Hagoort**, and **D. Casasanto**. 2014. 'The relation of space and musical pitch in the brain,' *The 36th Annual Meeting of the Cognitive Science Society* 3: 421–6.

Drew, P. 2006. 'When documents "speak": Documents, language and interaction' in P. Drew, G. Raymond, and D. Weinberg (eds): *Talk in Interaction in Social Research Methods*. Sage Publications, pp. 63–80.

Eskildsen, S. W. and **J. Wagner**. 2013. 'Recurring and shared gestures in the L2 classroom: Resources for teaching and learning,' *European Journal of Applied Linguistics* 1: 139–61.

Eskildsen, S. W. and **J. Wagner**. 2015. 'Embodied L2 construction learning,' *Language Learning* 65/2: 268–97.

Ferré, G. 2018. 'Gesture/speech integration in the perception of prosodic emphasis' in Klessa, K. et al. (eds.): *Proceedings of the 9th International Conference on Speech Prosody*. Poznań, pp. 35–9.

Fletcher, J. 1991. 'Rhythm and final lengthening in French,' *Journal of Phonetics* 19/2: 193–212.

Gabrielatos, C. 2002. 'Reading loud and clear: Reading aloud in ELT.' ERIC Document: ED477572.

Gibson, S. 2008. 'Reading aloud: A useful learning tool?,' *ELT Journal* 62: 29–36.

Glenberg, A. M., **M. Meyer**, and **K. Lindem**. 1987. 'Mental models contribute to foregrounding during text comprehension,' *Journal of Memory and Language* 26/1: 69–83.

Gluhareva, D. and **P. Prieto**. 2017. 'Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations,' *Language Teaching Research* 21/5: 609–31.

Goldin-Meadow, S. and **M. W. Alibali**. 2013. 'Gesture's role in speaking, learning, and creating language,' *Annual Review of Psychology* 64: 257–83.

Good, A. J., **F. A. Russo**, and **J. Sullivan**. 2015. 'The efficacy of singing in foreign-language learning,' *Psychology of Music* 43/5: 627–40.

Goodwin, M. H. and **C. Goodwin**. 1986. 'Gesture and coparticipation in the activity of searching for a word,' *Semiotica* 62/1–2: 51–76.

Gordon, J. and **I. Darcy**. 2016. 'The development of comprehensible speech in L2 learners,' *Journal of Second Language Pronunciation* 2/1: 56–92.

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

Guberina, P. 1965. 'La méthode audio-visuelle structuro-globale,' *RPA* 1: 35–64.

Guberina, P. 2008. *Retrospección*. Éditions du CIPA—Asociación Española Verbotonal.

Gullberg, M. and **S. G. McCafferty**. 2008. 'Introduction to gesture and SLA: Toward an integrated approach,' *Studies in Second Language Acquisition* 30/2: 133–46.

Hannah, B., **Y. Wang**, **A. Jongman**, **J. A. Sereno**, **J. Cao**, and **Y. Nie**. 2017. 'Cross-modal association between auditory and visuospatial information in Mandarin tone perception in noise by native and non-native perceivers,' *Frontiers in Psychology* 8: 2051.

Hardison, D. M. 2004. 'Generalization of computer-assisted prosody training: Quantitative and qualitative findings,' *Language Learning & Technology* 8/1: 34–52.

Hauk, O., **I. Johnsrude**, and **F. Pulvermüller**. 2004. 'Somatotopic representation of action words in human motor and premotor cortex,' *Neuron* 41/2: 301–7.

Heffner, C. C. and **L. R. Slevc**. 2015. 'Prosodic structure as a parallel to musical structure,' *Frontiers in Psychology* 6: 1962.

Hincks, R. and **J. Edlund**. 2009. 'Promoting increased pitch variation in oral presentations with transient visual feedback,' *Language Learning and Technology* 13/3: 32–50.

Hudson, N. 2011. 'Teacher gesture in a post-secondary English as a second language classroom: A sociocultural approach', Doctoral dissertation, University of Nevada (UNLV theses, dissertations, professional papers, and capstones).

Iizuka, T., **K. Nakatsukasa**, and **A. Braver**. 2020. 'The efficacy of gesture on second language pronunciation: An exploratory study of handclapping as a classroom instructional tool,' *Language Learning* 70: 1054–90.

Ionescu, T. and **D. Vasc**. 2014. 'Embodied cognition: Challenges for psychology and education,' *Procedia—Social and Behavioral Sciences* 128: 275–80.

Iverson, J. M. and **S. Goldin-Meadow**. 2005. 'Gesture paves the way for language development,' *Psychological Science* 16/5: 367–71.

Jakonen, T. 2020. 'Professional embodiment: Walking, re-engagement of desk interactions, and provision of instruction during classroom rounds,' *Applied Linguistics* 41/2: 161–84.

Jun, S.-A. and **C., Fougeron** 1995. 'The accentual phrase and the prosodic structure of French' in K. Elenius and P. Branderud (eds.): *Proceedings of the 13th International Congress of Phonetic Science*s, Vol. 2. Stockholm, pp. 722–25.

Jun, S.-A. and **C. Fougeron**. 2000. 'A phonological model of French intonation' in A. Botinis (ed.): *Intonation: Analysis, Modelling and Technology*. Kluwer, pp. 209–42.

Juntunen, M.-L. 2016. 'The Dalcroze approach: Experiencing and knowing music through the embodied exploration' in C. R. Abril and B. Gault (eds): *Approaches to Teaching General Music: Methods, Issues, and Viewpoints*. Oxford University Press, pp. 141–67.

Kääntä, L. 2015. 'The multimodal organisation of teacher-led classroom interaction' in C. J. Jenks and P. Seedhouse (eds): *International Perspectives on ELT Classroom Interaction*. Palgrave Macmillan, pp. 64–83.

Kang, O. 2010. 'Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness,' *System* 38/2: 301–15.

Kang, O., **D. Rubin**, and **L. Pickering**. 2010. 'Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English,' *The Modern Language Journal* 94/4: 554–66.

Keily, H. 2019. 'Actions as a basis for online embodied concepts,' *Languages* 4/1: 16.

Kelly, S., **A. Bailey**, and **Y. Hirata**. 2017. 'Metaphoric gestures facilitate perception of intonation more than length in auditory judgments of non-native phonemic contrasts,' *Collabra: Psychology* 3/1: 7.

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

Klomjit, S. 2013. 'Using oral reading to self to improve oral fluency of English language learners,' Doctoral dissertation, University of Arkansas (ScholarWorks@UARK).

Kreiner, H. and **Z. Eviatar**. 2014. 'The missing link in the embodiment of syntax: Prosody,' *Brain and Language* 137: 91–102.

Lee, J., **J. Jang**, and **L. Plonsky**. 2015. 'The effectiveness of second language pronunciation instruction: A meta-analysis,' *Applied Linguistics* 36/3: 345–66.

Levis, J. and **L. Pickering**. 2004. 'Teaching intonation in discourse using speech visualization technology,' *System* 32: 505–24.

Li, P., **F. Baills**, and **P. Prieto**. 2020. 'Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts,' *Studies in Second Language Acquisition* 42/5: 1015–39.

Li, P., **X. Xi**, **F. Baills**, and **P. Prieto**. 2021. 'Training non-native aspirated plosives with hand gestures: Learners' gesture performance matters,' *Language, Cognition and Neuroscience* 36/10: 1313–28.

Liu, Y.-T. and **W.-T. Tseng**. 2019. 'Optimal implementation setting for computerized visualization cues in assisting L2 intonation production,' *System* 87: 102145.

Llanes-Coromina, J., **P. Prieto**, and **P. L. Rohrer**. 2018. 'Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task' in K. Klessa, J. Bachan, A. Wagner, M. Karpiñski, and D. Sledziñski (eds.): *Proceedings of the 9th International Conference on Speech Prosody*. Vientiane: International Speech Communication Association, pp. 498–502.

Loehr, D. P. D. 2012. 'Temporal, structural, and pragmatic synchrony between intonation and gesture,' *LabPhon12* 3: 71–89.

Lomotey, C. F. 2013. 'Does first language prosodic transfer affect second language prosody?,' *The Journal of the Acoustical Society of America* 134/5: 4246.

Ludke, K. M. 2018. 'Singing and arts activities in support of foreign language learning: An exploratory study,' *Innovation in Language Learning and Teaching* 12/4: 371–86.

Macedonia, M. 2014. 'Bringing back the body into the mind: Gestures enhance word learning in foreign language,' *Frontiers in Psychology* 5: 1467.

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

Macedonia, M. and **W. Klimesch**. 2014. 'Long-term effects of gestures on memory for foreign language words trained in the classroom,' *Mind, Brain, and Education* 8/2: 74–88.

MacPherson, I. R. 1975. *Spanish Phonology: Descriptive and Historical*. Manchester University Press.

Madan, C. R. and **A. Singhal**. 2012. 'Using actions to enhance memory: Effects of enactment, gestures, and exercise on human memory,' *Frontiers in Psychology* 3: 507.

Majlesi, A. R. 2015. 'Matching gestures—Teachers' repetitions of students' gestures in second language learning classrooms,' *Journal of Pragmatics* 76: 30–45.

Martins, C, and J.-J., Mabilat 2003. Conversations Pratiques De L'oral. Cornelsen.

Mascaró, J. 1976. 'Catalan phonology and the phonological cycle', Doctoral dissertation, Massachusetts Institute of Technology (DSpace@MIT).

Matsumoto, Y. and **A. M. Dobs**. 2017. 'Pedagogical gestures as interactional resources for teaching and learning tense and aspect in the ESL grammar classroom,' *Language Learning* 67/1: 7–42.

Mister, B., **A. Baker**, and **H. Chen**. 2021. 'Punching through the barrier: Using gesture to activate productive oral vocabulary,' *Journal of Second Language Pronunciation* 7/2: 265–90.

Morett, L. M. 2018. 'In hand and in mind: Effects of gesture production and viewing on second language word learning,' *Applied Psycholinguistics* 39/2: 355–81.

Morett, L. M. and **L.-Y. Chang**. 2015. 'Emphasising sound and meaning: Pitch gestures enhance Mandarin lexical tone acquisition,' *Language, Cognition and Neuroscience* 30/3: 347–53.

Mori, J. and **A. Hasegawa**. 2009. 'Doing being a foreign language learner in a classroom: Embodiment of cognitive states as social events,' *IRAL—International Review of Applied Linguistics in Language Teaching* 47/1: 65–94.

Munro, M. 1995. 'Nonsegmental factors in foreign accent: Rating of foreign speech,' *Studies in Second Language Acquisition* 17: 17–33.

Munro, M. J. and **T. M. Derwing**. 2015. 'A prospectus for pronunciation research in the 21st century,' *Journal of Second Language Pronunciation* 1/1: 11–42.

Nakatsukasa, K. 2016. 'Efficacy of recasts and gestures on the acquisition of locative prepositions,' *Studies in Second Language Acquisition* 38/4: 771–99.

Nibert, H. J. 2000. Phonetic and phonological evidence for intermediate phrasing in Spanish intonation, Doctoral dissertation, University of Illinois at Urbana—Champaign (IDEALS, the Illinois Digital Environment for Access to Learning and Scholarship).

Prieto, P., **J. Borràs-Comes**, **T. Cabré**, **V. Crespo-Sendra**, **I. Mascaró**, **P. Roseano**, **R. Sichel-Bazin**, and **M. M. Vanrell**. 2015. 'Intonational phonology of Catalan and its dialectal varieties' in S. Frota and P. Prieto (eds.): *Intonation in Romance*. Oxford University Press, pp. 9–62.

Quinn-Allen, L. 1995. 'The effects of emblematic gestures on the development and access of mental representations of French expressions,' *The Modern Language Journal* 79/4: 521–9.

Ramírez Verdugo, D. 2006. 'A study of intonation awareness and learning in non-native speakers of English,' *Language Awareness* 15/3: 141–59.

Renard, R. 2002. *Apprentissage D'une Langue Étrangère Seconde 2. La Phonétique Verbo-Tonale*. De Boeck Université.

Rieser, J. J., A. E. Garing, and M. F. Young. 1994. 'Imagery, Action, and Young Children's Spatial Orientation: It's Not Being There That Counts, It's What One Has in Mind,' Child Development 655.

Riquelme Gil, I., J. Roca De Larios, and Y. Coyle Balibrea. 2020. 'The effect of repeated reading on the pronunciation of young EFL learners,' Porta Linguarum Revista Interuniversitaria De Didáctica De Las Lenguas Extranjeras 27: 7–19.

Rizzolatti, G, and L. Craighero. 2004. 'The mirror-neuron system,' Annual Review of Neuroscience 27: 169–92.

Rueschemeyer, S.-A., O. Lindemann, D. Van Rooij, W. Van Dam, and H. Bekkering. 2010. 'Effects of intentional motor actions on embodied language processing,' Experimental Psychology 574260–6.

Saito, K, and R. Lyster. 2012. 'Effects of Form-Focused Instruction and Corrective Feedback on L2 Pronunciation Development of f /ɹ/ by Japanese Learners of English,' Language Learning 62, 2, 593–633.

Saito, K. and **L. Plonsky**. 2019. 'Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis,' *Language Learning* 69/3: 652–708.

Saito, Y. and **K. Saito**. 2017. 'Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: The case of inexperienced Japanese EFL learners,' *Language Teaching Research* 21/5: 589–608.

Saito, K., **P. Trofimovich**, and **T. Isaacs**. 2016. 'Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels,' *Applied Psycholinguistics* 37/2: 217–40.

Shapiro, L. and **S. A. Stolz**. 2019. 'Embodied cognition and its significance for education,' *Theory and Research in Education* 17/1: 19–39.

Siakaluk, P. D., **P. M. Pexman**, **L. Aguilera**, **W. J. Owen**, and **C. R. Sears**. 2008. 'Evidence for the activation of sensorimotor information during visual word recognition: The body-object interaction effect,' *Cognition* 106/1: 433–43.

Smith, C. P., **B. King**, and **J. Hoyte**. 2014. 'Learning angles through movement: Critical actions for developing understanding in an embodied activity,' *The Journal of Mathematical Behavior* 36: 95–108.

Smotrova, T. 2014. 'Instructional functions of speech and gesture in the L2 classroom', Doctoral dissertation, Pennsylvania State University (Pennstate Electronic Theses and Dissertations for Graduate School).

Sullivan, J. V. 2018. 'Learning and embodied cognition: A review and proposal,' *Psychology Learning & Teaching* 17/2: 128–43.

Suzukida, Y. and **K. Saito**. 2019. 'Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle,' *Language Teaching Research* 25: 431–50.

Taguchi, E., M. Takayasu-Maass, and G. Gorsuch. 2004. 'Developing reading fluency in EFL: How assisted repeated readingand extensive reading affect fluency development,' Reading in a Foreign Language 16, 2, 70–96.

Tanner, M. W. and **M. M. Landon**. 2009. 'The effects of computer-assisted pronunciation readings on ESL learners' use of pausing, stress, intonation, and overall comprehensibility,' *Language Learning and Technology* 13/3: 51–65.

Tellier, M. 2008. 'Dire avec des gestes,' *Le Français Dans Le Monde, Recherche et Application* 44.

Thomson, R. I. and **T. M. Derwing**. 2015. 'The effectiveness of L2 pronunciation instruction: A narrative review,' *Applied Linguistics* 36/3: 326–44.

Thompson, A. A. and **W. A. Renandya**. 2020. 'Use of gesture for correcting pronunciation errors,' *Teflin Journal—A Publication on the Teaching and Learning of English* 31/2: 342–59.

Trofimovich, P. and **W. Baker**. 2006. 'Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech,' *Studies in Second Language Acquisition* 28/1: 1–30.

Trofimovich, P. and **T. Isaacs**. 2012. 'Disentangling accent from comprehensibility,' *Bilingualism: Language and Cognition* 15: 905–16.

Ueyama, M. 2000. *Prosodic Transfer: An Acoustic Study of L2 English vs. L2 Japanese*. University of California.

Vaissière, J. 1991. 'Rhythm, accentuation and final lengthening in French' in J. Sundberg, L. Nord, and R. Carlson. (eds.): *Music, Language, Speech and Brain*. Macmillan Press, pp. 108–20.

Wang, X. 2020. 'Segmental versus suprasegmental: Which one is more important to teach?,' *RELC Journal* 2020: 397–412.

Wang, W., and **S. Loewen**. 2016. 'Nonverbal behavior and corrective feedback in nine ESL university-level classrooms,' *Language Teaching Research* 20/4: 459–78.

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

Wilson, M. and **K. Emmorey**. 1997. 'A visuospatial "phonological loop" in working memory: Evidence from American Sign Language,' *Memory & Cognition* 25/3: 313–20.

Wilson, R. A. and **L. Foglia**. 2017. 'Embodied cognition' in E. N. Zalta (ed.): *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab. pp. 319–25.

Yang, J. and **H. Shu**. 2016. 'Involvement of the motor system in comprehension of non-literal action language: A meta-analysis study,' *Brain Topography* 29/1: 94–107.

Yuan, C., **S. González-Fuente**, **F. Baills**, and **P. Prieto**. 2019. 'Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers,' *Studies in Second Language Acquisition* 41/1: 5–32.

Zhang, Y., **F. Baills**, and **P. Prieto**. 2020. 'Hand-clapping to the rhythm of newly learned words improves L2 pronunciation: Evidence from training Chinese adolescents with French words,' *Language Teaching Research* 24/5: 666–89.

Zhang, R. and **Z. Yuan**. 2020. 'Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation,' *Studies in Second Language Acquisition* 42/4: 905–18.

NOTES ON CONTRIBUTOR

Florence Baills owns a Ph.D. in Linguistics and is member of the 'Prosodic Studies Group' at the Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona. She owns a Master degree in Theoretical and Applied Linguistics from the same university. She has dedicated many years to the teaching of French as a Foreign Language and is interested in doing research that can help language teachers and learners to better achieve their goals. Her current work explores the role of prosodic, gesture, and music training on L2 pronunciation. *Address for correspondence*: Universitat Pompeu Fabra, C/Roc Boronat 138, Barcelona 08018, Spain. <florence.baills@upf.edu>

Charlotte Alazard-Guiu is an Associate Professor at the University of Toulouse 2, France, and she is affiliated to the Octogone Lordat laboratory. She is an expert practitioner of the verbotonal pronunciation teaching method and in her research, she addresses the question of the impact of prosody and multimodality on the acquisition of pronunciation in L2. She currently leads the INGPRO project on the role of corrective hand gestures on the acquisition of L2 pronunciation (both at a segmental and a suprasegmental level) through the use of a custom pronunciation training platform.

Pilar Prieto is an ICREA Research Professor at the Department of Translation and Language Sciences at Universitat Pompeu Fabra (UPF), Barcelona, Catalunya. After obtaining her doctoral degree in Romance Linguistics at the University of Illinois at Urbana–Champaign, she worked as a postdoctoral fellow at Bell Laboratories (Murray Hill, NJ), where she continued working on the linguistic meaning of prosody across languages. Since 2008 she coordinates the 'Prosodic Studies Group' at the Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona.