This is the preprint of the article. Please cite it in this form: Li, P., Baills, F., Baqué, L., & Prieto, P. (in press). The effectiveness of the embodied prosodic training in L2 accentedness and vowel accuracy. *Second Language Research*. 10.1177/02676583221124075

The effectiveness of embodied prosodic training in L2 accentedness and vowel accuracy

Peng Li[1], Florence Baills[1], Lorraine Baqué[2], and Pilar Prieto[3,1]

[1] Department of Translation and Language Sciences, Universitat Pompeu Fabra

[2] Department of French and Romance Languages, Universitat Autònoma de Barcelona

[3] Institució Catalana de Recerca i Estudis Avançats

Author Note

This work was funded by the Ministerio de Ciencia, Innovación y Universidades, Agencia Estatal de Investigación and European Regional Development Fund under Multimodal language learning (MLL): Prosodic and Gestural Integration in Pragmatic and Phonological Development [PGC2018-097007-B-I00]. This work is also supported by the Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) under Pla de Recerca de Catalunya [2017 SGR_971], and the Ministerio de Economía y Competitividad under Cognipros project [FFI2013-40419-P]. The second author acknowledges a grant from the Department of Translation and Language Sciences at University Pompeu Fabra.

The authors thank Ms. Cécile Devart, Ms. Delphine Fauvet, and Ms. Gaëlle Saragnet for their interpretation of the dialogues, Dr. Élisabeth Miche (UPF) for her kind cooperation during the experiment, and all the participants at the Department of Translation and Language Sciences (UPF) in this study.

Correspondence concerning this article should be addressed to Peng Li, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain. E-mail: peng.li@upf.edu

Abstract

This study explores the effects of embodied prosodic training on the production of non-native French front rounded vowels (i.e., /y, ø, œ/) and the overall pronunciation proficiency. Fifty-seven Catalan learners of French practiced pronunciation in one of two conditions: one group observed hand gestures embodying prosodic features of the sentences they were listening to, while the other group did not see any such gestures. The learning outcome was assessed in a pretest, posttest, and delayed posttest through a dialogue-reading task and a sentence imitation task in terms of accentedness, comprehensibility, and fluency scores and through formant analysis of participant-produced target vowels. The results showed that compared to non-embodied training, embodied prosodic training yielded continuous improvement in accentedness in both tasks and improved the F2 values of French front rounded vowels (more fronted). As for comprehensibility and fluency scores, both groups showed similar levels of significant improvement. This study highlights the interaction between prosodic and segmental features of speech by showing that training with embodied prosodic features benefitted accentedness and the production accuracy of non-native vowels.

1 Introduction

The importance of pronunciation instruction in second language (L2) teaching and learning has received increasing attention (Lee et al., 2015; Saito and Plonsky, 2019). Recent studies suggest that both prosody (e.g., intonation, rhythm, etc.) and segments (e.g., vowels and consonants) should be included in training programs (Lee et al., 2015) and that teachers should take advantage of the relationship between the two (Wang, 2020; Zielinski, 2015). This is in accordance with the studies suggesting that prosodic (Anderson-Hsieh et al., 1992; Trofimovich and Baker, 2006) and segmental accuracy (Isaacs and Trofimovich, 2012; Saito et al., 2016, 2017) is important for reducing accentedness. In addition, prosodic accuracy and fluency influence comprehensibility (Crowther et al., 2016; Isaacs and Trofimovich, 2012; Saito et al., 2016, 2017), while only segments with high functional load seem to affect comprehensibility (Munro and Derwing, 2006).

Notwithstanding, most research on L2 pronunciation has focused on specific sounds to improve learners' pronunciation proficiency (Lambacher et al., 2005; Lord, 2005, 2008; Ozakin et al., 2022; Saito and Munro, 2014; Xi et al., 2020), little is known about the potential effects of prosody-based pronunciation training and embodied prosodic training on the production of non-native sounds. The main goal of the present study is to empirically assess whether embodied prosodic training (i.e., instructors use hand gestures to highlight the L2 prosody) can help improve learners' overall pronunciation proficiency and the pronunciation accuracy of front rounded vowels in an L2.

1.1 Prosodic training and L2 pronunciation instruction

Previous research has compared the effects of prosodic pronunciation instruction (i.e., focusing on intonation, rhythm, etc.) to segmental pronunciation instruction (i.e., focusing on specific vowels and consonants) in improving L2 learners' pronunciation proficiency. Focusing on L2 prosody can generally improve learners' L2 speech production

(Gordon and Darcy, 2016), especially the comprehensibility and fluency in free speech (Derwing et al., 1998), enhance intonation accuracy (Saito and Saito, 2017), and trigger delayed positive effects on comprehensibility in spontaneous speech (Zhang and Yuan, 2020).

However, to our knowledge, only a handful of empirical studies have explored the potential effects of prosodic training on improving non-native sounds, with mixed results. Focusing first on positive results, training in prosody yielded greater gains in overall pronunciation and segmental accuracy than training in individual speech sounds (Missaglia, 2007). Likewise, the gains in prosody obtained from intonation training were generalized to segmental accuracy (Hardison, 2004). Moreover, prosodic training could help improve vowel accuracy in L2 speech production (Saito and Saito, 2017), which seemed to stem from the gains in accurately reproducing rhythmic structures. By contrast, some research failed to find beneficial effects of prosodic training on L2 segmental accuracy, despite its effects on improving L2 speech comprehensibility (Gordon and Darcy, 2016).

There are several reasons to hypothesize that prosody may help the pronunciation of individual sounds. First, prosodic and segmental structures are two integrated and interdependent components in speech production. For instance, the Sonority Expansion hypothesis (Beckman et al., 1992) holds that speech prominence can make vowels more open, while the Hyperarticulation hypothesis (de Jong, 1995) claims that despite openness, speech prominence may even affect the lip roundedness and backness of the vowels. In addition, a non-prominent position, in general, may compress the duration and formant frequencies of a speech sound (Walker, 2011: 16). Therefore, previous studies concerning L2 pronunciation training (Derwing et al., 1998; Gordon and Darcy, 2016; Missaglia, 2007; Saito and Saito, 2017) took advantage of the interaction between prosody and segments.

In a broader perspective, prosody is regarded as a bootstrapping component in language learning. For instance, the Prosodic Bootstrapping hypothesis postulates that prosodic features (e.g., rhythm, tempo, pitch) may bootstrap syntactic and lexical features in early first language acquisition (Christophe et al., 1997, 2008). Recent evidence has shown that the bootstrapping effects of rhythmic training could improve the L2 imitation abilities (Campfield and Murphy, 2014), which had already been applied to speech therapy (Bedore and Leonard, 1995). Given the interdependence between prosody and segments, we hypothesize that prosody can bootstrap the pronunciation of challenging L2 speech sounds.

1.2 Embodied approaches to training L2 pronunciation

Embodied Cognition captures the relation between mind and body (Ionescu and Vasc, 2014), which holds that the body is tightly involved in human cognitive processes (Barsalou, 2008) and may therefore have an impact on learning and education (Shapiro and Stolz, 2019). The cognitive offloading theory holds that people tend to offload their cognitive load onto the environment or the body to reduce the occupation of working memory or attention abilities (Risko and Gilbert, 2016). These theories thus have important implications for education in that "embodiment offers either a causal route to more effective learning or a diagnostic tool for measuring conceptual understanding" (Shapiro and Stolz, 2019: 30).

Like many areas of learning where the Embodied Cognition paradigm has been extensively applied, L2 teachers frequently use embodied strategies in their classrooms (Smotrova, 2017). A growing body of empirical research has shed light on the positive role of embodied training in L2 pronunciation. For example, beat gestures highlighting speech prominence can improve L2 pronunciation proficiency (Kushch, 2018). Hand gestures tracing pitch contours in space performed over the nuclear-accented syllables could favor the pronunciation of L2 intonational patterns (Yuan et al., 2019). Moreover, illustrating vowel length by short and long horizontal hand sweep gestures could also improve the pronunciation of the L2 vowel length (Li, Baills, et al., 2020). In a recent study, Baills et al. (2022) showed that gestures encoding sentence-level prosodic features accompanied by the oral repetition of logatomes (i.e., a series of identical nonsense CV syllables that maintain prosodic structure intact) improved learners' accentedness and pronunciation of suprasegmental features in a reading task compared to merely repeating the sentences orally.

In practice, researchers have proposed various types of teaching techniques involving embodiment; among them, the so-called verbotonal method (Guberina, 2008; Intravaia, 2000) has drawn much attention. It encourages the combination of prosody and body movements like hand gestures for phonetic corrections for both speech sounds and prosody. For example, to trigger a more target-like pronunciation of the French front rounded /y/ for Spanish speakers (who often pronounce it as /u/), teachers may place the /y/ in a rising intonation contour (Renard, 2002), embody the rise with an upward hand gesture (Billières, 2002), and put it in various prosodic positions and meaningful discourses (Wlomainck, 2002).

Nevertheless, only a few studies have empirically assessed the role of the verbotonal method in actual teaching practice, and they have yielded inconclusive findings. The verbotonal method could improve learners' L2 fluency in oral reading (Alazard et al., 2010), and this method might particularly benefit beginning learners' L2 speech fluency (Alazard, 2013). However, in a more recent study (Alazard-Guiu et al., 2018), the verbotonal method could not outperform the traditional focus-on-form training in improving segmental accuracy. Given the limited sample size (eight participants) in the last study, one might expect more conclusive results in experiments with larger populations.

1.3 The present study

To assess the value of embodied prosodic training on L2 pronunciation, especially its effects at the segmental level, the present study investigated whether Catalan learners of French could benefit from embodied prosodic training in improving their overall pronunciation and the production accuracy of the non-native front rounded vowels.

Catalan learners face clear challenges in learning the French pronunciation at both segmental and prosodic levels. At the segmental level, French has ten oral vowels /i, e, ɛ, y, ø, œ, a, u, o, ɔ/ (Hannahs, 2007), where the front rounded /y, ø, œ/ contrast with the back rounded /u, o, ɔ/ (Darcy et al., 2012). However, /y, ø, œ/ are not part of the phonological inventory of Central Catalan, which has only seven vowels /i, e, ɛ, a, u, o, ɔ/ (Wheeler, 2005). Based on the observation that learners of French whose native languages do not have front rounded vowels tend to assimilate them to their back counterparts (see Darcy et al., 2012; Levy and Law, 2010 for English speakers; Racine and Detey, 2019 for Spanish speakers; Hannahs, 2007 for a review), we hypothesize that Catalan speakers may also produce the front /y, ø, œ/ as back /u, o, ɔ/ in their L2 French speech. Regarding prosody, French stress is assigned at the phrase level (i.e., Accentual Phrase, or AP), marked by a phrase-initial optional high tone and a phrase-final obligatory high tone. This implies that stress is a demarcation of the AP rather than the word (Fougeron and Jun, 2002; Jun and Fougeron, 2000). Contrastingly, Catalan does not show evidence for AP, whereas the intermediate phrase generally consists of more than one prosodic word (Prieto et al., 2015). Therefore, these differences in prosody may influence Catalan learners' French pronunciation.

Hence, we addressed two main research questions in the present study, as follows:

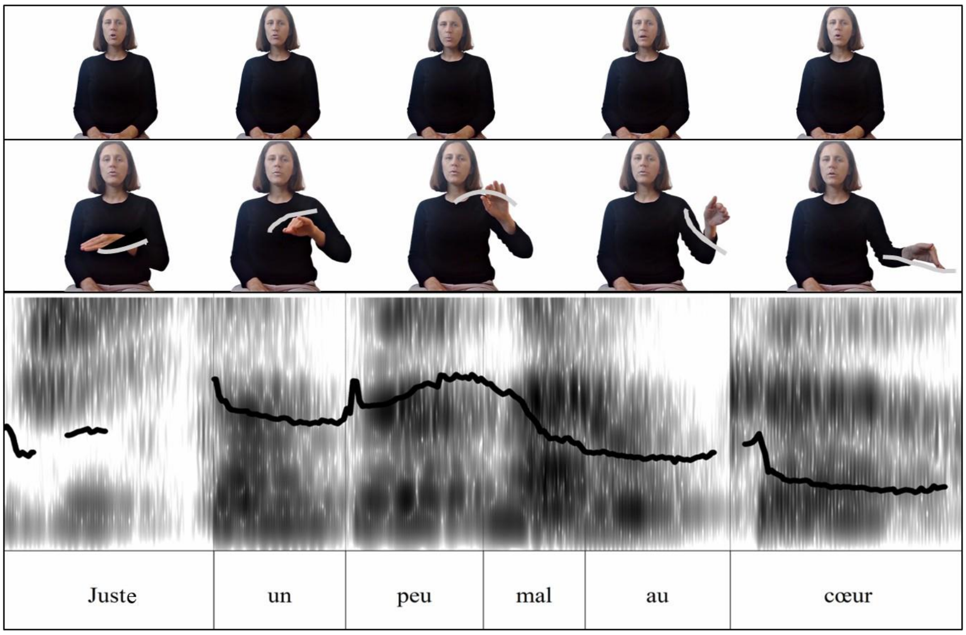
RQ1: To what extent does the embodied prosodic training improve Catalan learners' overall pronunciation proficiency of French? To globally assess the quality of learners' L2 speech (henceforth, overall pronunciation proficiency), we will adopt three measures: accentedness, comprehensibility, and fluency, as suggested by previous research (Munro and Derwing, 2015).

RQ2: To what extent does the embodied prosodic training improve Catalan learners' pronunciation accuracy of French front rounded vowels /y, ø, œ/?

For the embodied prosodic training, we used hand gestures to mimic French prosodic features on the sentence level. Figure 1 illustrates one of the instructor's hand movements illustrating the prosodic structure of the sentence *J**u**ste un p**eu** mal au c**œu**r*. 'Just a bit sick at heart.' The up and down hand movements show pitch peaks and valleys and indicate speech prominence, thus highlighting rhythmic and intonational structures. The rightmost images show how the instructor prolonged her horizontal hand movement to show the phrase-final lengthening patterns. Note that the target front rounded vowels (boldface in the sample sentence above) appeared in different positions in the utterance.

Figure 1

The sequence of images illustrating the instructor's hand gestures while producing a sample sentence in the non-embodied (upper panel) or embodied condition (lower panel)



2 Methods

2.1 Participants

Two classes of Catalan-speaking students were recruited from the BA program in Translation and Applied Languages at a public university in Spring 2020. They were in their first and second year of study. The pronunciation training sessions were incorporated into their French classes. Fifty-seven of the students (53 females, 18–46 years old, M_{age} = 19.89 years, *SD* = 3.63) voluntarily signed a written consent which allowed the researchers to collect and analyze the audio and video recordings obtained during the experiment[1]. The 57 participants reported using Catalan for their daily verbal communication an average of 62.81% of the time (*SD* = 29.67) and were thus considered Catalan-dominant speakers.

According to the teaching syllabus, students had to practice their French-speaking skills in two parallel groups after each week's main lecture. In each class, students were divided into two groups according to the alphabetic order of their names. This teaching design provided two groups of participants for a between-subject experiment. In each class, one group of students received embodied prosodic training (*n* = 28; first-year = 9; second-year = 19, henceforth "embodied condition"), while the other group received non-embodied training (*n* = 29; first-year = 7; second-year = 22, henceforth "non-embodied condition").

Since the participants were not necessarily beginning learners of French when they were admitted to the BA program, we surveyed their French language learning background in terms of age of onset of learning, years of formal learning, months of study abroad in a French-speaking country, and months of extracurricular French courses. In addition, they were asked to self-assess their French proficiency from 1 (A1) to 6 (C2) according to the Common European Framework of Reference for Languages. Their answers are summarized in Table 1. Although participants displayed large individual differences within group, no statistical differences were found between groups (all $p > .05$). Therefore, the two training groups were comparable in French learning experience and proficiency level. See also Appendix A for individual information of each participant.

Table 1.

[1] Although all the students had to take part in the training sessions and complete the tasks in pretest, posttest and delayed posttest as their class assignments, the researchers only had access to the data of the 57 participants who signed the written consents.

Means, Medians, Standard Deviations, and Mann-Whitney-Wilcox Test Results of Individual Differences in French Learning Experience and Self-Assessed Proficiency Across Groups

	Non-embodied			Embodied			*W*	*p*
	M	*Mdn*	*SD*	*M*	*Mdn*	*SD*		
Age of onset of learning	13.72	13	3.06	14.11	13	2.99	413	.910
Formal learning (years)	5.52	6	3.01	5.18	6	2.57	381	.687
Study abroad (months)	0.45	0	1.24	0.71	0	1.43	473.5	.142
Extracurricular courses (months)	1.97	3	3.09	1.61	1.5	2.85	385	.721
Self-assessed proficiency	3.07	3	1.00	3.25	3	0.93	434	.637

2.2 Materials

The experiment was a between-subject training study with a pretest/posttest/delayed posttest paradigm. In this section, we describe the materials created for the experiment, including the audio-visual stimuli for the training sessions as well as the auditory and textual stimuli for the three tests.

2.2.1 Stimuli for the pronunciation training sessions

(a) Training dialogues

The training materials were adapted from a French pronunciation textbook which provided a series of dialogues (Martinie and Wachs, 2006). For this experiment, we selected three dialogues designed to train /y, ø, œ/, and modified the content to increase the frequency of the target vowels (see Appendix B).

Table 2 summarizes the prosodic positions of the three target vowels in the three training dialogues. More than half (57%) of the target vowels were pitch-accented, and of those, most (59%) carried a high tone or were in a rising intonation. It should be noted that since many functional words contain /y/ (e.g., *tu* ‘you’, *du* ‘of the’, *une* ‘a’, etc.), the frequency of /y/ is inevitably higher than that of /ø/ and /œ/ and functional words are often unaccented.

Table 2

Count and Proportion of the Target Front Rounded Vowels separated by Prosodic Patterns in the Training Dialogues

	/y/	/ø/	/œ/
Prosodic Pattern	*n* (%)	*n* (%)	*n* (%)
Unaccented	28 (53%)	10 (50%)	0 (0%)
Accented	25 (47%)	10 (50%)	16 (100%)
High/rising	14 (56%)	7 (70%)	9 (56%)
Low/falling	11 (44%)	3 (30%)	7 (44%)

(b) Audio-visual training stimuli

The audio-visual stimuli consisted of two parts: the enactments of the three dialogues and the sentence-by-sentence training clips with or without gestures.

To create the enactments of the three dialogues, four female amateur actors (all native French speakers) performed the dialogues in pairs. Each performance lasted around 45 seconds.

For the sentence-by-sentence training clips, two experienced female teachers of French with native French proficiency were video-recorded, producing each of the sentences of the three training dialogues. Before recording, the two instructors watched the dialogue enactments as many times as necessary and imitated the actors' speech. They were filmed against a white wall, with their face and the upper half of their body visible. For the non-embodied condition, the instructors produced the sentences naturally without any body movements other than those strictly related to oral articulation. For the embodied condition, they produced each sentence along with hand gestures to visually illustrate the prosodic information of the sentence (see Figure 1 for an example) while keeping the rest of the body still. To accurately trace the pitch track by hand movements, they viewed pitch contour curves generated by Praat (Boersma and Weenink, 2020). The gesture accuracy was checked by comparing them to the visual intonation patterns of each sentence in Praat.

Regarding the non-gestural movements, the instructors were asked to perform as similarly as possible across the two conditions. All sentences were filmed four times for each condition. Then the authors compared the instructors' non-gestural performance (such as lip movements) across conditions (embodied vs. non-embodied) and selected the most similar clips with and without gestures for each sentence. A total of 128 video clips were selected, 64 for each condition.

In addition, to avoid any potential differences in speech across the two conditions, we added the audio tracks of the gesture videos to the corresponding no-gesture videos and removed the original audio recording. In order to ensure that the resulting video clips were equally natural across conditions, three native French speakers rated all the video clips (N = 128). For each video clip, they gave a naturalness rating from 1 (very unnatural) to 5 (very natural) based on how well they thought the lip movements and the sound were temporally aligned. The results showed a high mean naturalness rating for the items in the two conditions ($M_{\text{Non-embodied}}$ = 4.96, SD =0.12; M_{Embodied} = 4.94, SD = 0.13). A Mann-Whitney-Wilcox test showed no statistical difference (W = 1862, p = .142) in the naturalness scores between the two conditions. Therefore, we concluded that the video clips used for the two training conditions had equally good quality.

Eventually, we created six training videos (3 dialogues × 2 conditions). Importantly, each instructor was assigned a separate role for each dialogue so that the two instructors spoke in alternating turns.

2.2.2 Pretest and posttest materials

For the dialogue-reading task, in addition to the three dialogues that had been trained, we selected a fourth untrained dialogue from the same textbook (Martinie and Wachs, 2006). Like the trained dialogues, the untrained dialogue was adapted to increase the frequency of the target front rounded vowels (see Appendix B, dialogue 4).

The stimuli for the sentence imitation task were 15 sentences selected from the dialogue-reading task, twelve from the trained dialogues, and three from the untrained dialogue (see Appendix C). The four actors who performed the dialogues recorded the sentences. Each sentence was read by the person who said it in the video. The second author (a native French speaker) read the three untrained sentences.

2.3 Procedure

Due to the impact of the COVID-19 pandemic, all spring 2020 courses were conducted online, so testing and training for this study also took place online. While the training sessions were conducted online under the teachers' supervision for speaking practice, participants did the testing sessions individually as course homework.

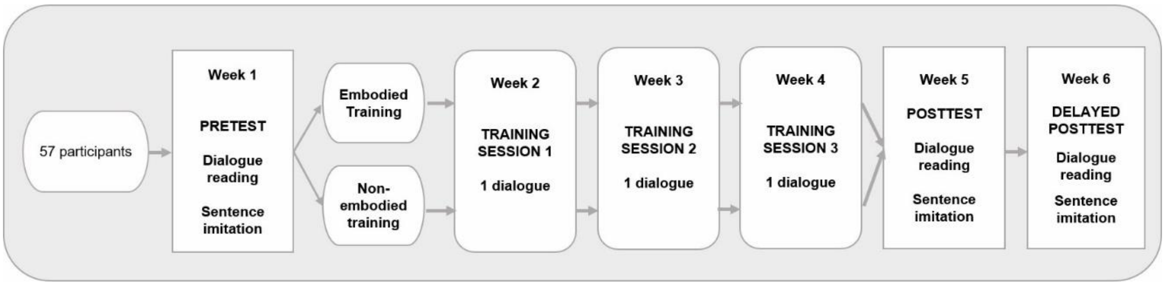
The experiment lasted six weeks. In the first week, the participants performed the pretest tasks: dialogue-reading and sentence imitation tasks. In the dialogue-reading task, participants had to read aloud the text of the dialogues presented online using Alchemer software (https://www.alchemer.com/). In the sentence imitation task, participants listened to each of the 15 French sentences once and imitated each of them immediately afterward. The presentation of the sentences was randomized automatically. In both tasks the participant's voice was automatically recorded by an online camera (http://webcamera.io).

From the second week to the fourth week, participants received three sessions of audio-visual training, one session per week. Each training session followed the same procedure. First, the seminar teacher explained the words that might be unfamiliar to the participants in the training materials. Then participants watched the video and performed the training. Specifically, they first watched the enactment of the dialogue to be trained (45 s) and then read the dialogue script aloud by themselves (90 s). Following this, the dialogue was trained sentence by sentence, in the order they appeared in the dialogue. Each sentence was trained in two blocks. In the first block participants first saw the sentence written in French (5s); then, one of the two instructors read the sentence twice with or without gestures, depending on the condition. After that, a black screen displaying “Repeat that once” appeared (5 s), allowing the participants to repeat the sentence once. In the second block, however, the instructor uttered each sentence only once, while the rest of the procedure remained the same as that of the first block. After the audio-visual training, participants were asked to watch the dialogue enactment a second time (45 s) and then read aloud the script of the dialogue again (90 s). In total, each training session lasted around 15 minutes.

In the fifth and sixth weeks, the participants took the posttest and delayed posttest, respectively, in which they repeated the dialogue-reading and sentence imitation tasks as in the pretest. The procedure of the three tests was the same. The participants completed the three tests by themselves. After each test, they had to upload the video recording files to a Google Drive folder, to which only the researcher, the course lecturer, and the participant themself had access. Note that the interval between two training sessions and/or tests was one week. Figure 2 summarizes the main procedure of the experiment.

Figure 2

Experimental procedure



2.4 Data Coding

In order to ensure that the audio files were adequate for subsequent acoustic analyses, we digitally adjusted the recordings. First, we extracted the audio tracks from the video recordings and saved each soundtrack as a monophonic .wav audio file[2]. After that, for each

[2] Regarding the properties of the recordings, all the original files had a 16-bit resolution, but the sample rate varied. While 12 participants used a 44.1 kHz sample rate, the rest used 48 kHz. The number of 44.1-kHz files was similar between the two groups (n= 5 in the embodied; n = 7 in the non-

audio file, we used Audacity to identify a short period of the typical background noise and applied noise reduction throughout the file. Finally, the digitally modified audio files were used for formant analyses.

We obtained 684 recordings (4 dialogues × 3 tests × 57 participants) from the dialogue-reading and 2,565 recordings (15 sentences × 3 tests × 57 participants) from the sentence imitation task. We did perceptual rating and acoustic analyses to assess all the recordings. In what follows, we will report the details of the two analyses.

2.4.1 Perceptual rating on overall pronunciation proficiency

Three native French-speaking teachers (aged 29–39 years, M_{age} = 35 years) rated all the recordings. The three raters had taught French in a Catalan-speaking city for at least three years by the time of recruitment. On a five-point Likert scale (1 = not at all; 5 = completely), raters' self-evaluation of familiarity with Catalan-accented French yielded an average of 4.6 and self-reported knowledge in French phonology an average of 4. The result suggests that they were very sensitive to Catalan-accented French and had enough linguistic knowledge to make reliable judgments.

Before the rating task, all raters received a 45-minute training session to familiarize themselves with the evaluation system. For the dialogue-reading task, they rated accentedness, comprehensibility, and fluency, while for the sentence imitation task, they only rated accentedness because the sentences were very short (around 2 s). Raters had to listen to each recording and rate each measure on a scale from 1 to 9. Accentedness measured "the difference in pronunciation as compared with the native speakers" (1 = "very strong foreign accent", 9 = "no foreign accent at all"); comprehensibility was defined as "the degree of

---

embodied). We therefore digitally converted the sample rate of all files to 44.1 kHz. We acknowledge that since the videos were recorded in a .mp4 format, the audio quality was inevitably compressed. Based on a set of arguments and following Sanker et al. (2021), we consider the variation in recording, including the compression patterns in the audio files, would not affect our formant analyses. See the "Discussion" section for more information.

difficulty in understanding the speech" (1 = "incomprehensible", 9 = "completely comprehensible"); while fluency referred to the "fluidity of speech" (1 = "disfluent", 9 = "fluent") (Munro and Derwing, 2015).

The ratings were performed online in 21 one-hour batches, with each batch containing 57 dialogues or 285 sentences. All the recordings were presented to the raters randomly through the online software Alchemer. The raters completed one batch per day over 21 days.

Inter-rater reliability was checked by a series of two-way mixed, consistency, average-measures (k = 3) Intra-Class Correlation (ICC) analyses (Hallgren, 2012). The ICC and their 95% confidence intervals were calculated using the *irr* package version 0.84.1 (Gamer et al., 2019) in R version 4.0.2 (R Core Team, 2014). The three raters showed a good level of agreement in the rating of accentedness (ICC = 0.81, 95% CI [0.79, 0.84]), comprehensibility (ICC = 0.83, 95% CI [0.80, 0.84]), and fluency (ICC = 0.78, 95% CI [0.75, 0.81]) for the dialogue-reading task, as well as a good level of agreement in the accentedness rating (ICC = 0.81, 95% CI [0.79, 0.82]) for the sentence imitation task (see Koo and Li, 2016 for the interpretation of ICC scores). Therefore, the ratings of the three raters of each measure were averaged per each item for each participant in order to create the scores for accentedness, comprehensibility, and fluency to be analyzed.

2.4.2 Acoustic analyses on the target vowels

In the field of acoustic phonetics, formant analysis is often employed to capture vowel quality differences, and the first two formant frequencies (i.e., F1 and F2) are often used to describe the vowels in terms of tongue height (mouth aperture) and tongue frontness/backness (tongue position) (Johnson, 2011). In order to minimize gender influence on formant frequencies, we only included female participants (N = 53). This did not substantially reduce our sample size because there were only two male participants in each condition. In order to ensure that the acoustic analyses were comparable between the two tasks, the first author annotated the same front rounded vowels that the female participants produced in both the dialogue-reading task and the sentence imitation task using Praat software. Accordingly, a total of 6,042 tokens (19 tokens × 53 participants × 3 tests × 2 tasks) were annotated. After annotation, the mean F1 and F2 values of each token were extracted from Praat. The acoustic data were then transformed from Hertz to Bark to normalize the individual differences in vocal tract length using the following formula: $Bark = 7\ln\{(Hz/650) + [(Hz/650)^2 + 1]^{1/2}\}$ (Traunmüller, 1990, see Gordon and Darcy, 2016; Saito and Munro, 2014 for similar decisions).

2.5 Statistical analyses

We applied 16 Generalized Linear Mixed Models to the following outcome measures using the *glmmTMB* package, version 1.0.2.1 (Brooks et al., 2017) in R, version 4.0.2. For the dialogue-reading task, we used three averaged perceptual rating scores (accentedness, comprehensibility, and fluency) and six acoustic measures (the Bark normalized F1 and F2 values of /y, ø, œ/); for the sentence imitation task, we used the accentedness score and six acoustic measures. For all models, the fixed factors were condition (two levels: non-embodied and embodied), test (three levels: pretest, posttest, and delayed posttest), and their interaction. Other possible fixed factors were training (trained items vs. untrained items), participants' French proficiency (high vs. low), position of the target vowels (AP-medial vs. AP-final), and the pitch of the target vowels (unaccented vs. rising vs. falling). However, a series of analyses revealed that none of these factors showed interactions with Test × Condition, which means that these factors did not affect the training effects. Moreover, the training and test materials were not designed to test the abovementioned effects. Therefore, we excluded all these factors so that the models were nested into our research questions. Finally, we added two random intercepts for each model: participant and item, to account for the variance caused by individual differences and item difficulties. Significance was determined by Type II Wald *chi*-squared tests using the *car* package, version 3.0.9 (Fox and Weisberg, 2019). The post-hoc analyses were a series of Bonferroni pairwise comparisons performed with the *emmeans* package, version 1.4.8 (Lenth et al., 2020). The effect size was assessed using Cohen's *d* in the post-hoc analyses (small: $d \geq 0.2$; medium: $d \geq 0.5$; large: $d \geq 0.8$, see Cohen, 1988).

3 Results

In what follows, we report the statistical results according to the two research questions.

3.1 RQ1: To what extent does the embodied prosodic training improve Catalan learners' overall pronunciation proficiency of French?

3.1.1 Dialogue-reading task

The descriptive data for the three perceptual measures across conditions and tests are displayed in Table 3.

Table 3

Means and Standard Deviations of the Accentedness, Comprehensibility, and Fluency Scores Across Conditions and Tests for the Dialogue-Reading Task.

	Non-embodied		Embodied	
	M	*SD*	*M*	*SD*
Accentedness				
Pretest	4.41	1.39	4.53	1.35
Posttest	4.64	1.32	4.85	1.17
Delayed posttest	4.70	1.34	5.07	1.18
Comprehensibility				
Pretest	5.33	1.43	5.44	1.45
Posttest	5.80	1.41	5.82	1.21
Delayed posttest	5.92	1.32	5.92	1.29
Fluency				
Pretest	5.18	1.19	5.32	1.61
Posttest	5.94	1.31	6.09	1.28
Delayed posttest	6.08	1.23	6.27	1.21

In what follows, we will report the results of the statistical analyses for each dependent variable separately.

- Analysis of accentedness scores revealed a significant two-way interaction of Test × Condition ($\chi^2(2) = 6.41$, $p = .041$), which indicates that the accentedness score varied across condition and test. The post-hoc results are as follows. For the non-embodied group, there was a significant improvement from pretest to posttest ($\Delta = 0.23$, $SE = 0.07$, $t(675) = 3.27$, $p = .003$, $d = 0.43$) and this improvement was maintained at delayed posttest, indicated by a significant improvement from pretest to delayed posttest ($\Delta = 0.29$, $SE = 0.07$, $t(675) = 4.04$, $p < .001$, $d = 0.53$). By contrast, the accentedness score progressively improved in the embodied group across the three tests, indicated by a significant improvement from pretest to posttest ($\Delta = 0.32$, $SE = 0.07$, $t(675) = 4.38$, $p < .001$, $d = 0.59$), and from posttest to delayed posttest ($\Delta = 0.22$, $SE = 0.07$, $t(675) = 3.07$, $p = .007$, $d = 0.41$).
- Analysis of comprehensibility scores revealed a significant main effect of test ($\chi^2(2) = 101.58$, $p < .001$), but no interaction of Test × Condition ($\chi^2(2) = 0.97$, $p = .616$) was found. Post-hoc analyses confirmed that the comprehensibility score of all the participants improved from pretest to posttest ($\Delta = 0.43$, $SE = 0.06$, $t(675) = 7.59$, $p < .001$, $d = 0.71$) and from pretest to delayed posttest ($\Delta = 0.53$, $SE = 0.06$, $t(675) = 9.51$, $p < .001$, $d = 0.89$). Other comparisons, however, did not reveal significant results.
- Similarly, analysis of fluency scores only revealed a significant main effect of test ($\chi^2(2) = 235.91$, $p < .001$). The post-hoc analysis found that participants showed a continuous improvement from pretest to posttest ($\Delta = 0.77$, $SE = 0.06$, $t(675) = 11.93$, $p < .001$, $d = 1.12$) and from posttest to delayed posttest ($\Delta = 0.16$, $SE = 0.06$, $t(675) = 2.42$, $p = .048$, $d = 0.23$), regardless of training method.

3.1.2 Sentence imitation task

The descriptive data for accentedness scores across conditions and tests are summarized in Table 4. Recall that we did not measure comprehensibility and fluency for the sentence imitation task.

Table 4

Means and Standard Deviations of the Accentedness Score Across Conditions and Tests in the Sentence Imitation Task

	Non-embodied		Embodied	
	M	*SD*	*M*	*SD*
Pretest	4.80	1.49	4.65	1.32
Posttest	5.02	1.49	5.02	1.24
Delayed posttest	5.11	1.49	5.22	1.19

Analysis of accentedness scores in the sentence imitation task revealed a significant two-way interaction of Test × Condition ($\chi^2(2) = 9.02$, $p = .011$). The post-hoc analysis showed similar patterns to those of the dialogue-reading task. Specifically, the non-embodied group showed a significant improvement from pretest to posttest ($\Delta = 0.22$, $SE = 0.06$, $t(2556) = 3.55$, $p = .001$, $d = 0.24$) and maintained this improvement at delayed posttest, indicated by a significant improvement from pretest to delayed posttest ($\Delta = 0.31$, $SE = 0.06$, $t(2556) = 5.08$, $p < .001$, $d = 0.34$). By contrast, the improvement observed in the embodied group was continuous, namely, a significant improvement from pretest to posttest ($\Delta = 0.37$, $SE = 0.06$, $t(2556) = 5.95$, $p < .001$, $d = 0.41$) followed by a further improvement from posttest to delayed posttest ($\Delta = 0.20$, $SE = 0.06$, $t(2556) = 3.23$, $p = .004$, $d = 0.22$).

3.2 RQ2: To what extent does the embodied prosodic training improve Catalan learners' pronunciation accuracy of French front rounded vowels /y, ø, œ/?

This section will mainly report the acoustic results of the three front rounded vowels /y, ø, œ/. We will report the statistical analyses of the F1 and F2 values from the dialogue-reading task followed by the sentence imitation task.

3.2.1 Dialogue-reading task

(a) Vowel height (F1)

First, although a significant main effect of test was found for /y/ ($\chi^2(2) = 9.88$, $p = .007$) and /œ/ ($\chi^2(2) = 22.54$, $p < .001$) , test and condition did not show any significant interaction for the F1 (Bark) in any of the three vowels (/y/: $\chi^2(2) = 0.55$, $p = .761$; /ø/: $\chi^2(2) = 1.37$, $p = .504$; /œ/: $\chi^2(2) = 4.23$, $p = .121$). Therefore, the changes in F1 across tests did not significantly differ by condition. In both Catalan and French, the front vowels show a three-way contrast in vowel height, which can be measured by F1. If the learners could maintain the three-way contrast in the target front rounded vowels throughout the three tests, the F1 would not be a key acoustic measure to assess the pronunciation accuracy of the target vowels. We thus checked whether participants could distinguish the three degrees of vowel height by further analyzing the F1 data. To this end, we additionally ran three Generalized Linear Mixed Models for F1 with vowel (three levels: /y, ø, œ/) being the fixed effect and participant and item as random intercepts in each of the three tests. The results revealed a significant main effect of vowel on the F1 value (Bark) in all three tests (pretest: $\chi^2(2) = 94.43$, $p < .001$; posttest: $\chi^2(2) = 92.46$, $p < .001$; delayed posttest: $\chi^2(2) = 96.79$, $p < .001$). Post-hoc analyses revealed that there was always a clear distinction in terms of F1 between the three target vowels ($F1_{/œ/} > F1_{/ø/} > F1_{/y/}$, all $p < .05$, see Table 5 for descriptive data and Table D1 for post-hoc results). This indicates that participants could produce the three target vowels using three different levels of vowel height throughout the whole experiment.

Table 5

Means (Standard Deviations) of the Bark Normalized First and Second Formant Frequencies of the Three Front Rounded Vowels in the Dialogue-Reading Task Across Conditions and Tests

	F1 (Bark)		F2 (Bark)	
	Non-embodied	Embodied	Non-embodied	Embodied

/y/				
Pretest	4.20 (0.81)	4.15 (0.65)	11.48 (1.96)	11.30 (2.32)
Posttest	4.27 (0.79)	4.28 (0.73)	11.39 (2.07)	11.75 (1.98)
Delayed posttest	4.32 (0.72)	4.32 (0.89)	11.48 (1.80)	11.95 (1.76)
/ø/				
Pretest	4.75 (0.78)	4.88 (0.89)	11.50 (1.72)	11.40 (1.92)
Posttest	4.88 (0.78)	4.92 (0.87)	11.09 (1.81)	11.65 (1.64)
Delayed posttest	4.89 (0.82)	4.91 (0.77)	11.25 (1.62)	11.75 (1.53)
/œ/				
Pretest	5.64 (0.77)	5.78 (0.92)	11.02 (1.43)	10.89 (1.49)
Posttest	5.89 (0.74)	5.89 (0.81)	10.55 (1.58)	11.29 (1.29)
Delayed posttest	5.96 (0.89)	5.90 (0.83)	10.96 (1.33)	11.31 (1.33)

(b) Vowel frontness/backness (F2)

The second formant frequency (F2) is related to vowel frontness (the higher the F2, the more front the vowel). Because front rounded vowels do not exist in the learners' first language (Catalan), an increase in F2 would reflect a more target-like pronunciation in posttest sequences. The analyses of all the three vowels revealed a significant interaction of Test × Condition (/y/: $\chi^2(2) = 12.00$, $p = .002$; /ø/: $\chi^2(2) = 14.38$, $p = .001$; /œ/: $\chi^2(2) = 24.46$, $p < .001$), which suggests that the change in F2 values across tests differed by condition. In what follows, we report the significant changes revealed by post-hoc comparisons.

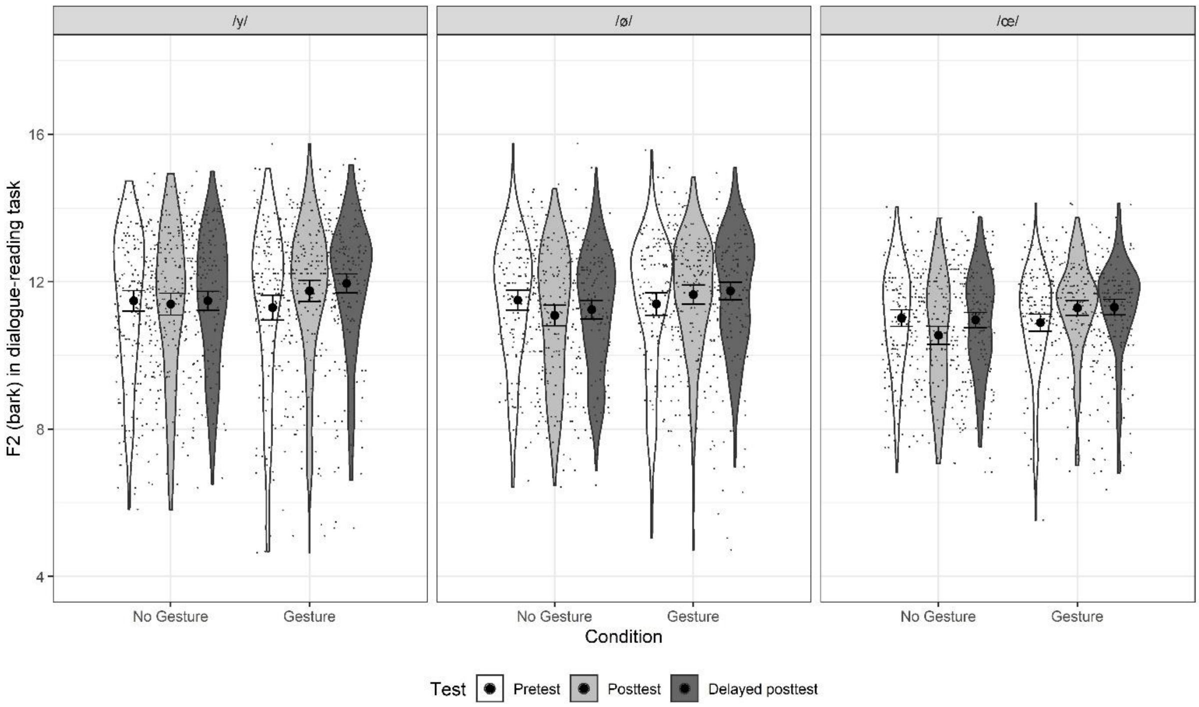
- For /y/, the F2 of the embodied group increased from pretest to posttest ($\Delta = 0.45$ Bark, $SE = 0.15$, $t(1104) = 3.12$, $p = .006$, $d = 0.33$) and from pretest to delayed posttest ($\Delta = 0.65$ Bark, $SE = 0.15$, $t(1104) = 4.52$, $p < .001$, $d = 0.47$). However, no significant difference was found for the non-embodied group in any of the three tests.
- For /ø/, the F2 of the embodied group showed a significant increase from pretest to delayed posttest ($\Delta = 0.36$ Bark, $SE = 0.14$, $t(945) = 4.52$, $p = .033$, $d = 0.29$), while the non-embodied group significantly decreased the F2 value from pretest to posttest ($\Delta = -0.42$ Bark, $SE = 0.14$, $t(945) = -3.03$, $p = .008$, $d = -0.34$).

- For /œ/, the F2 of the embodied group again showed a significant increase from pretest to posttest (Δ = 0.40 Bark, *SE* = 0.13, *t*(945) = 3.21, *p* = .004, *d* = 0.36) and from pretest to delayed posttest (Δ = 0.42 Bark, *SE* = 0.13, *t*(945) = 3.35, *p* = .003, *d* = 0.38). By contrast, in the non-embodied group, there was a significant decrease in F2 from pretest to posttest (Δ = -0.47 Bark, *SE* = 0.12, *t*(945) = -3.78, *p* = .001, *d* = -0.42), although this decrease was adjusted at delayed posttest by a significant improvement from posttest (Δ = 0.42 Bark, *SE* = 0.12, *t*(945) = 3.36, *p* = .002, *d* = 0.37).

Briefly, in the embodied group, all three vowels, except for /ø/, showed an enhanced F2 at posttest (i.e., one week after training) and maintained this enhancement at delayed posttest (i.e., two weeks after training). This result reflected a fronting effect of tongue position when participants produced the three target vowels. However, the non-embodied group did not show such a progressive pattern. Figure 3 visually plots the results.

Figure 3

Mean Bark Normalized Second Formant Frequencies “F2 (Bark)” of the Three Front Rounded Vowels Across Conditions and Tests in the Dialogue-Reading Task



Note. The larger dots indicate the mean values. The error bars mark the 95% confidence intervals.

3.2.2 Sentence imitation task

The acoustic analyses of the sentence imitation task followed the same procedure as that of the dialogue-reading task.

(a) Vowel height (F1)

As in the dialogue-reading task, F1 analyses found no significant two-way interaction of Test × Condition for any of the three vowels (/y/: $\chi^2(2) = 1.26$, $p = .532$; /ø/: $\chi^2(2) = 0.67$, $p = .714$; /œ/: $\chi^2(2) = 5.98$, $p = .050$). Yet a significant main effect of test was found for the three vowels (/y/: $\chi^2(2) = 10.26$, $p = .006$; /ø/: $\chi^2(2) = 13.46$, $p = .001$; /œ/: $\chi^2(2) = 12.48$, $p = .002$). In order to check if the participants distinguished the three levels of vowel height (F1 Bark), the same models were run as those applied to the dialogue-reading task. Again, for all three tests, there was a significant main effect of vowel (pretest: $\chi^2(2) = 58.74$, $p < .001$; posttest: $\chi^2(2) = 59.64$, $p < .001$; delayed posttest: $\chi^2(2) = 67.61$, $p < .001$), and post-hoc pairwise comparisons showed that the F1 value was significantly different between the three target vowels in all the three tests (all $p < .05$, see Table D2), with /y/ producing the lowest figures, and /œ/ producing the highest (see Table 6).

Table 6

Means (Standard Deviations) of the Bark Normalized First and Second Formant Frequencies of the Three Front Rounded Vowels in the Sentence Imitation Task Across Condition and Test

	F1 (Bark)		F2 (Bark)	
	Non-embodied	Embodied	Non-embodied	Embodied
/y/				
Pretest	4.08 (0.75)	4.05 (0.78)	11.69 (1.83)	11.54 (2.23)
Posttest	4.27 (0.88)	4.13 (0.77)	11.67 (1.94)	12.02 (1.57)
Delayed posttest	4.21 (0.76)	4.12 (0.74)	11.74 (1.47)	12.08 (1.43)
/ø/				
Pretest	4.71 (0.88)	4.79 (0.86)	11.32 (1.71)	11.28 (1.99)
Posttest	4.88 (0.81)	4.98 (0.99)	11.33 (1.62)	11.69 (1.45)
Delayed posttest	4.80 (0.83)	4.82 (0.80)	11.17 (1.61)	11.91 (1.38)
/œ/				
Pretest	5.69 (1.00)	5.82 (0.98)	10.80 (1.32)	10.74 (1.59)
Posttest	5.97 (0.91)	5.91 (0.80)	10.85 (1.22)	11.34 (1.07)
Delayed posttest	5.96 (0.82)	5.83 (0.70)	10.86 (1.13)	11.45 (1.06)

(b) Vowel frontness/backness (F2)

The analyses of the F2 (Bark) for all three vowels revealed a significant two-way interaction of Test × Condition (/y/: $\chi^2(2) = 7.60$, $p = .022$; /ø/: $\chi^2(2) = 15.22$, $p < .001$; /œ/: $\chi^2(2) = 19.62$, $p < .001$). This again suggests that the F2 value of the target vowels in the sentence imitation task varied across condition and test. Significant contrasts revealed by post-hoc comparisons are reported as follows.

- For /y/, the embodied group revealed a significant improvement in F2 from pretest to posttest ($\Delta = 0.48$ Bark, $SE = 0.15$, $t(1104) = 3.26$, $p = .004$, $d = 0.34$) and from pretest to delayed posttest ($\Delta = 0.54$ Bark, $SE = 0.15$, $t(1104) = 3.66$, $p = .001$, $d = 0.38$). Yet no significant contrast was observed in the non-embodied group between any of the three tests.
- For /ø/, the embodied group revealed a progressive pattern similar to that of /y/, with the pretest outperformed by the posttest ($\Delta = 0.41$ Bark, $SE = 0.14$, $t(945) =$

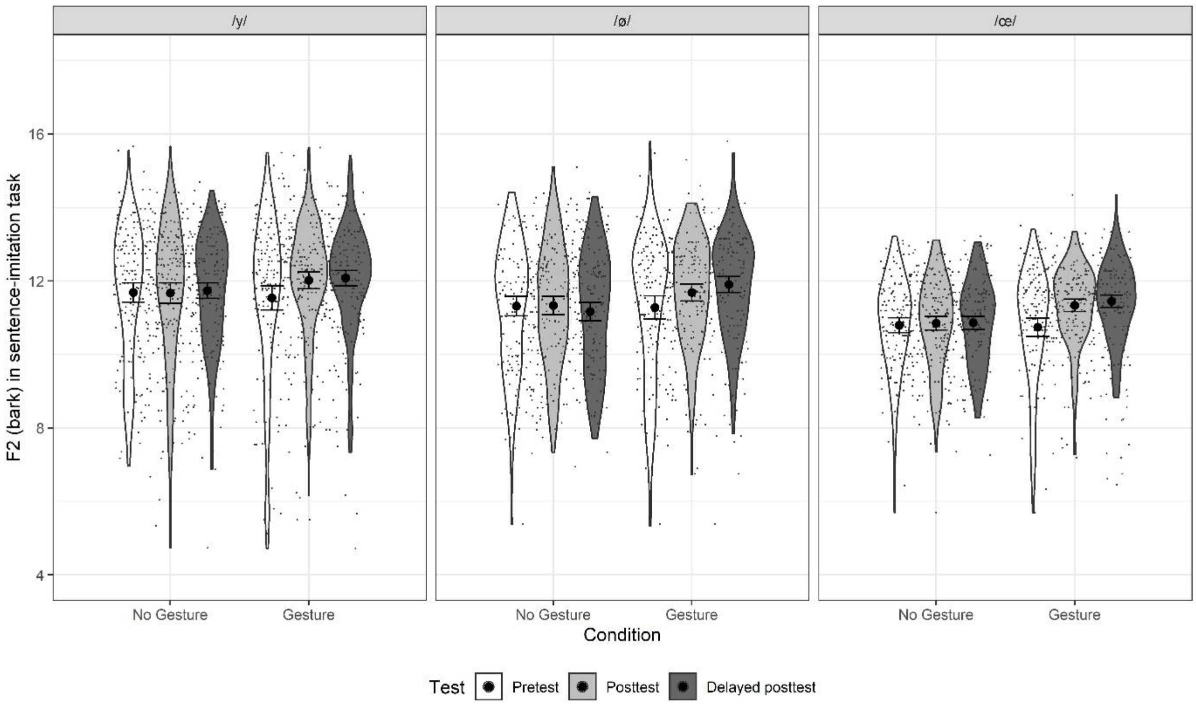
2.88, p = .012, d = 0.33) and delayed posttest (Δ = 0.63 Bark, SE = 0.14, t(945) = 4.43, p < .001, d = 0.50), while no significant change was shown in the non-embodied group over the three tests.

- For /œ/, again, the embodied group significantly improved their F2 from pretest to posttest (Δ = 0.59 Bark, SE = 0.11, t(945) = 5.36, p < .001, d = 0.61) and from pretest to delayed posttest (Δ = 0.71 Bark, SE = 0.11, t(945) = 6.36, p < .001, d = 0.72). The non-embodied group did not show any significant improvement or worsening between the three tests.

In summarizing, the analyses on the F2 of the target vowels in the sentence imitation task revealed a similar pattern as the one reported for the dialogue-reading task. That is, while participants in the embodied group could front their tongue when producing the front rounded vowels one week after training and maintained this effect after two weeks, this was not the case for the participants in the non-embodied group. Table 6 shows the descriptive data and Figure 4 visually plots the results.

Figure 4

The Bark Normalized Second Formant Frequency “F2 (Bark)” of the Three Front Rounded Vowels Across Conditions and Tests in the Sentence Imitation Task



Note. The larger dots indicate the mean values. The error bars mark the 95% confidence intervals.

4 Discussion

The present study assessed the effects of a three-session embodied prosodic training program using a listen-and-repeat paradigm to improve both overall pronunciation proficiency and segmental accuracy in an L2. We will organize the discussion in accordance with our two research questions.

4.1 Effects of embodied prosodic training on overall pronunciation proficiency

The first research question was whether embodying prosodic features through hand gestures would improve Catalan learners' overall French pronunciation proficiency (as measured by accentedness, comprehensibility, and fluency) more than non-embodied training. The results showed that the gains obtained by embodied prosodic training were larger than those achieved by non-embodied training for reading and imitation tasks.

Regarding accentedness, the additional improvement from posttest (one week after training) to delayed posttest (two weeks after training) in both tasks in the embodied group suggested that hand gestures that highlight prosodic properties may help maintain the training effects. Therefore, our study supports the results of previous studies showing positive effects of embodied prosodic training techniques on pronunciation (Kushch, 2018; Llanes-Coromina et al., 2018) and provides further evidence that gestures can play a role in the maintenance of these training effects.

By contrast, comprehensibility and fluency measures obtained in the discourse-reading task did not show significant two-way interactions of Test × Condition. That is, the two training methods improved the participants' performance similarly. With respect to comprehensibility, both groups significantly improved their comprehensibility score at posttest and maintained this improvement at delayed posttest with a large effect size (d = 0.89). As for fluency, the participants in both groups showed a continuous improvement across the three tests (d = 1.35). Taken together, these results show that both embodied and non-embodied training paradigms were beneficial in improving the learners' L2 speech comprehensibility and fluency.

Why is it the case that embodied prosodic training could not outperform non-embodied training in comprehensibility and fluency? In our view, this is because accentedness is mainly related to pronunciation factors (Saito et al., 2016, 2017; Trofimovich and Isaacs, 2012), while comprehensibility is also determined by other factors like grammar and lexis (Crowther et al., 2016; Isaacs and Trofimovich, 2012). Moreover, the two tasks in the present study were at a controlled speech level so that the content was predictable to the raters. It would thus be interesting for a future study to assess whether the role of gestures in improving L2 comprehensibility and fluency could also be noticeable in spontaneous speech. Finally, it might well be that a three-session training program was insufficient to improve these measures, given that in other research a longer training period (eight weeks and two hours per session) with the verbotonal method revealed a positive effect on fluency (Alazard, 2013). Future studies are needed to assess the effects of embodied training over a longer period.

4.2 Effects of embodied prosodic training on the production of non-native vowels

The second research question was whether training with hand gestures mimicking L2 prosodic features would benefit Catalan learners' pronunciation of non-native vowels. Results of the formant analyses of the three front rounded vowels /y, ø, œ/ confirmed that participants in the embodied prosodic training condition generally improved their pronunciation accuracy from pretest to posttest or delayed posttest in both reading and imitation tasks. This was shown by the expected F2 increase across tests, indicating a shift from back to front rounded vowels in both tasks. By contrast, participants in the non-embodied condition did not display such an improvement between the three tests in either of the tasks. Furthermore, in the dialogue-reading task, /ø/ and /œ/ even showed a temporary decrease (from pretest to posttest) in F2. These results indicate that embodied prosodic training led to a more target-like pronunciation of *front* rounded vowels by helping participants move the place of articulation forward.

These findings are thus in line with the previous studies (Hardison, 2004; Missaglia, 2007; Saito and Saito, 2017), showing that training prosody can trigger pronunciation gains at the segmental level. The underlying mechanism may be based on several components. First, highlighting the prosodic structure of target sentences could serve as an integrative strategy for learning the target sounds, given that they were placed in various prosodic positions. Our results thus corroborate the hypotheses that prosodic and segmental features are interdependent in phonological learning and that prosodic features can act as a scaffold for improving articulatory production. In addition, the visualization of the target prosodic structure through hand gestures may enhance the bootstrapping function of prosody for learning L2 vowels. This follows the Embodied Cognition paradigm, in which cognition is viewed to be grounded in sensory-motor processes (Ionescu and Vasc, 2014). In relation to this, gestures have been shown to save cognitive resources and reduce the cognitive load during language processing (Cook et al., 2012; Goldin-Meadow et al., 2001). Thus, observing hand movements mimicking prosodic structures during pronunciation training may bootstrap phonological processes.

Moreover, participants in the embodied condition improved accentedness and vowel accuracy at posttest in the dialogue-reading and sentence imitation tasks. Note that the difficulty level was not the same across tasks since imitation can provide both lexical and acoustic-phonological information while only lexical pathways are available to participants in a reading task. Therefore, the consistency of these results across tasks shows that the positive effects of embodied prosodic training can be generalized from imitation tasks to more challenging reading tasks.

4.3 Limitations

A potential limitation of the current design is that we asked participants to view the gestures without performing them. Since learners' gesture performance accuracy may significantly influence training effects (Li et al., 2021; Li, Xi, et al., 2020), we did not encourage the participants to imitate the gestures encoding complex prosodic features. This design leaves it an open question whether learners would gain more if they had performed the target gestures after the instructors. Furthermore, as pointed out by an anonymous reviewer, participants may tend to observe either articulatory or gestural movements, which may impact learning outcomes. However, the current study did not assess which visual cues participants attended. Future studies might want to fill up this gap by using eye-tracking techniques.

Finally, we acknowledge that online recordings may induce deviances and trigger errors in formant analyses (Freeman and De Decker, 2021; Ge et al., 2021; Sanker et al., 2021). Sanker et al. (2021) provided suggestions for online experiments in linguistic fieldwork, phonetics, and other studies involving speech data collection. Our methodology conformed to Sanker et al.'s (2021) main recommendations in that (a) the participants did the recordings on their own devices and directly uploaded them onto online storage services (Google Drive) to avoid further compression; (b) the recording platform was the same across participants and sessions, and (c) our sample size was reasonable for group-level comparisons which avoided individual-level analyses. However, it was not possible to (a) make uniform participants' computer setups, (b) install external microphones for each of them, or (c) generate non-compressed recordings from the free online recording service. Even though these factors might trigger random deviances or errors, two facts back up the validity of our acoustic analyses. First, for speech data collected online via video recording, Sanker et al. (2021)'s study showed that the formant values (i.e., F1 and F2) did not significantly differ by recording conditions (Tables S34-35 in Sanker et al., 2021). Second, since each participant performed the recordings on their own computer, the recording condition for a single participant was consistent across three tests. Therefore, our online experiment setting does not seem to affect the validity of the formant analyses largely. Future studies may seek to establish more stringent protocols for online data collection, specifically for L2 speech research.

5 Conclusion

The present study showed that hand gestures mimicking prosodic features could not only enhance learners' overall pronunciation proficiency but also improve the accuracy of French front rounded vowels. These results expanded our knowledge about the effects of prosodic embodiment in L2 acquisition. In practice, the current results have several implications for L2 instruction practices. First, teaching L2 pronunciation in a meaningful context may achieve more robust training effects. Second, hand gestures can be incorporated into classroom teaching to cue various L2 phonological features (e.g., sounds or prosody). Third, instead of training individual sounds or specific prosodic features separately, teachers can embed the target sounds into various prosodic contexts to obtain a global improvement.

All in all, despite some limitations, the current study provided empirical evidence to support the role of embodiment in L2 pronunciation acquisition and demonstrated that hand gestures imitating L2 melodic and rhythmic features are an effective tool in teaching practice as they improve the overall pronunciation proficiency and the production accuracy of L2 sounds.

References

Alazard-Guiu C, Santiago F and Mairano P (2018) L'incidence de la correction phonétique sur l'acquisition des voyelles en langue étrangère : étude de cas d'anglophones apprenant le français. In: *Journées d'Études sur la Parole, JEP 2018*, 2018, pp. 116–124. Laboratoire Parole et Langage. DOI: 10.21437/jep.2018-14 . hal-01779110 HAL.

Alazard C (2013) *Rôle de la prosodie dans la fluence en lecture oralisée chez des apprenants de Français Langue Étrangère*. Université Toulouse le Mirail. DOI: 10.13140/RG.2.2.24924.23686.

Alazard C, Astésano C and Billières M (2010) The implicit prosody hypothesis applied to foreign language learning: From oral abilities to reading skills. In: *Proceedings of the 5th International Conference on Speech Prosody*, Chicago, IL, 2010, pp. 1–4.

Anderson-Hsieh J, Johnson R and Koehler K (1992) The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning* 42(4): 529–555. DOI: 10.1111/j.1467-1770.1992.tb01043.x.

Baills F, Alazard-Guiu C and Prieto P (2022) Embodied prosodic training enhances L2 read speech pronunciation. *Applied Linguistics*. DOI: http://doi.org/10.1093/applin/amac010.

Barsalou LW (2008) Grounded cognition. *Annual Review of Psychology* 59: 617–645. DOI: 10.1146/annurev.psych.59.103006.093639.

Beckman M, Edwards J and Fletcher J (1992) Prosodic structure and tempo in a sonority model of articulatory dynamics. In: Docherty GJ and Ladd DR (eds) *Papers in Laboratory Phonology II: Gesture*, *Segment*, *Prosody*, pp. 68–89. DOI: 10.1017/cbo9780511519918.004.

Bedore LM and Leonard LB (1995) Prosodic and syntactic bootstrapping and their clinical applications. *American Journal of Speech-Language Pathology* 4(1): 66–72. DOI: 10.1044/1058-0360.0401.66.

Billières M (2002) Le corps en phonétique corrective. In: Renard R (ed.) *Apprentissage d'une Langue Étrangère/Seconde Vol.2 La Phonétique Verbo-Tonale*. Bruxelle, Belgique: De Boeck Université, pp. 38–70.

Boersma P and Weenink D (2020) Praat: doing phonetics by computer [computer program]. 6.1.16. Available at: http://www.praat.org.

Brooks ME, Kristensen K, van Benthem KJ, et al. (2017) glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal* 9(2). Technische Universitaet Wien: 378–400.

Campfield DE and Murphy VA (2014) Elicited imitation in search of the influence of linguistic rhythm on child L2 acquisition. *System* 42(1): 207–219. DOI: 10.1016/j.system.2013.12.002.

Christophe A, Guasti T, Nespor M, et al. (1997) Reflections on phonological bootstrapping: Its role for lexical and syntactic acquisition. *Language and Cognitive Processes* 12(5–6): 585–612. DOI: 10.1080/016909697386637.

Christophe A, Millotte S, Bernal S, et al. (2008) Bootstrapping lexical and syntactic acquisition. *Language and Speech* 51(1–2): 61–75. DOI: 10.1177/00238309080510010501.

Cohen J (1988) *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New York, NY: Lawrence Erlbaum Associates.

Cook SW, Yip TKY and Goldin-Meadow S (2012) Gestures, but not meaningless movements, lighten working memory load when explaining math. *Language and Cognitive Processes* 27(4): 594–610. DOI: 10.1080/01690965.2011.567074.

Crowther D, Trofimovich P and Isaacs T (2016) Linguistic dimensions of second language accent and comprehensibility. *Journal of Second Language Pronunciation* 2(2): 160–182. DOI: 10.1075/jslp.2.2.02cro.

Darcy I, Dekydtspotter L, Sprouse RA, et al. (2012) Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English- L2 French acquisition. *Second Language Research* 28(1): 5–40. DOI: 10.1177/0267658311423455.

de Jong KJ (1995) The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *The Journal of Acoustic Society of America* 97(1): 491–504.

Derwing TM, Munro MJ and Wiebe G (1998) Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48(3): 393–410. DOI: 10.1111/0023-8333.00047.

Fougeron C and Jun S-A (2002) Realizations of accentual phrase in French intonation. *Probus* 14(2002): 147–172. DOI: 10.1515/prbs.2002.002.

Fox J and Weisberg S (2019) *An {R} Companion to Applied Regression*. Third. Thousand Oaks {CA}: Sage. Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/.

Freeman V and De Decker P (2021) Remote sociophonetic data collection: Vowels and nasalization over video conferencing apps. *The Journal of the Acoustical Society of America* 149(2): 1211–1223. DOI: 10.1121/10.0003529.

Gamer M, Lemon J, Fellows I, et al. (2019) irr: Various Coefficients of Interrater Reliability and Agreement version 0.84.1[software]. Available at: https://cran.r-project.org/package=irr.

Ge C, Xiong Y and Mok P (2021) How reliable are phonetic data collected remotely? Comparison of recording devices and environments on acoustic measurements. In: *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2021, pp. 3984–3988. DOI: 10.21437/Interspeech.2021-1122.

Goldin-Meadow S, Nusbaum H, Kelly SD, et al. (2001) Explaining math: Gesturing lightens the load. *Psychological Science* 12(6): 516–522. DOI: 10.1111/1467-9280.00395.

Gordon J and Darcy I (2016) The development of comprehensible speech in L2 learners. *Journal of Second Language Pronunciation* 2(1): 56–92. DOI: 10.1075/jslp.2.1.03gor.

Guberina P (2008) *Retrospección* (J Murilloed. & tran.). Mons: Éditions du CIPA - Asociación Española Verbotonal.

Hallgren KA (2012) Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology* 8(1): 23–34. DOI: 10.20982/tqmp.08.1.p023.

Hannahs SJ (2007) French phonology and L2 acquisition. In: Dalila A (ed.) *French Applied Linguistics*. Amsterdam, The Neetherlands: John Benjamins, pp. 50–74.

Hardison DM (2004) Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology* 8(1): 34–52. DOI: http://dx.doi.org/10125/25228.

Intravaia P (2000) *Formation Des Professeurs de Langue En Phonétique Corrective. Le Système Verbo-Tonal*. Paris, France: Didier Érudition.

Ionescu T and Vasc D (2014) Embodied Cognition: Challenges for psychology and education. *Procedia - Social and Behavioral Sciences* 128: 275–280. DOI: 10.1016/j.sbspro.2014.03.156.

Isaacs T and Trofimovich P (2012) Deconstructing comprehensibility. *Studies in Second Language Acquisition* 34(3): 475–505. DOI: 10.1017/S0272263112000150.

Johnson K (2011) *Acoustic and Auditory Phonetics*. 3rd ed. Malden, MA: Blackwell Publishing.

Jun S-A and Fougeron C (2000) A phonological model of French intonation. In: Botinis A (ed.) *Intonation: Analysis, Modeling and Technology*. Dordrecht: Kluwer Academic Publishers, pp. 209–242. DOI: 10.1007/978-94-011-4317-2_10.

Koo TK and Li MY (2016) A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15(2): 155–163. DOI: 10.1016/j.jcm.2016.02.012.

Kushch O (2018) *Beat gestures and prosodic prominence: Impact on learning*. Universitat Pompeu Fabra (Spain).

Lambacher SG, Martens WL, Kakehi K, et al. (2005) The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics* 26(2): 227–247. DOI: 10.1017.S0142716405050150.

Lee J, Jang J and Plonsky L (2015) The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics* 36(3): 345–366. DOI: 10.1093/applin/amu040.

Lenth R, Singmann H, Love J, et al. (2020) Emmeans: Estimated marginal means, Aka Least-Squares means. R package 1.5.1. Available at: https://cran.r-project.org/package=emmeans.

Levy ES and Law FF (2010) Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America* 128(3): 1290–1305. DOI: 10.1121/1.3466879.

Li P, Xi X, Baills F, et al. (2020) Appropriately performing hand gestures cueing phonetic features facilitates simultaneous speech imitation in an L2. In: *Proceedings of the 7th Gesture and Speech in Interaction GESPIN 7*, 2020.

Li P, Baills F and Prieto P (2020) Observing and producing durational hand gestures facilitates the pronunciation of novel vowel-length contrasts. *Studies in Second Language Acquisition* 42(5): 1015–1039. DOI: 10.1017/S0272263120000054.

Li P, Xi X, Baills F, et al. (2021) Training non-native aspirated plosives with hand gestures: learners' gesture performance matters. *Language, Cognition and Neuroscience* 36(10): 1313–1328. DOI: 10.1080/23273798.2021.1937663.

Lord G (2005) (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania* 88(3): 557. DOI: 10.2307/20063159.

Lord G (2008) Podcasting communities and second language pronunciation. *Foreign Language Annals* 41(2): 364–379. DOI: 10.1111/j.1944-9720.2008.tb03297.x.

Martinie B and Wachs S (2006) *Phonétique En Dialogues*. Paris, France: CLE International SEJER.

Missaglia F (2007) Prosodic training for adult Italian learners of German: the Contrastive Prosody Method. In: Trouvain J and Gut U (eds) *Non-Native Prosody: Phonetic*

Description and Teaching Practice. Berlin, Germany: De Gruyter Mouton. DOI: 10.1515/9783110198751.2.237.

Munro MJ and Derwing TM (2006) The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34(4): 520–531. DOI: 10.1016/j.system.2006.09.004.

Munro MJ and Derwing TM (2015) A prospectus for pronunciation research in the 21st century. *Journal of Second Language Pronunciation* 1(1): 11–42. DOI: 10.1075/jslp.1.1.01mun.

Ozakin AS, Xi X, Li P, et al. (2022) Thanks or tanks: Training with tactile cues improves learners' accuracy of English interdental consonants in an oral reading task. *Language Learning and Development*. DOI: 10.1080/15475441.2002.2107522.

Prieto P, Borràs-Comes J, Cabré T, et al. (2015) Intonational phonology of Catalan and its dialectal varieties. In: Frota S and Prieto P (eds) *Intonation in Romance*. Oxford: Oxford University Press, pp. 9–62. DOI: 10.1093/acprof:oso/9780199685332.003.0002.

R Core Team (2014) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.r-project.org/.

Racine I and Detey S (2019) Production of French close rounded vowels by Spanish learners. In: Gibson M and Gil J (eds) *Romance Phonetics and Phonology*. New York, NY: Oxford University Press, pp. 381–394. DOI: 10.1093/oso/9780198739401.003.0019.

Renard R (2002) Une phonétique immergée. In: Renard R (ed.) *Apprentissage d'une Langue Étrangère/Seconde Vol.2 La Phonétique Verbo-Tonale*. Bruxelle, Belgique: De Boeck Université, pp. 12–24.

Risko EF and Gilbert SJ (2016) Cognitive offloading. *Trends in Cognitive Sciences* 20(9): 676–688. DOI: 10.1016/j.tics.2016.07.002.

Saito K and Munro MJ (2014) The early phase of /ɹ/ production development in adult Japanese learners of English. *Language and Speech* 57(4): 451–469. DOI: 10.1177/0023830913513206.

Saito K and Plonsky L (2019) Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning* 69(3): 652–708. DOI: 10.1111/lang.12345.

Saito K, Trofimovich P and Isaacs T (2016) Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics* 37(2): 217–240. DOI: 10.1017/S0142716414000502.

Saito K, Trofimovich P and Isaacs T (2017) Using listener judgments to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization Study. *Applied Linguistics* 38(4): 439–462. DOI: 10.1093/applin/amv047.

Saito Y and Saito K (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. DOI: 10.1177/1362168816643111.

Sanker C, Babinski S, Burns R, et al. (2021) (Don't) try this at home! The effects of recording devices and software on phonetic analysis. *Language* 97(4): 360–382. DOI: 10.1353/lan.2021.0079.

Shapiro L and Stolz SA (2019) Embodied cognition and its significance for education. *Theory and Research in Education* 17(1): 19–39. DOI: 10.1177/1477878518822149.

Smotrova T (2017) Making pronunciation visible: Gesture in teaching pronunciation. *TESOL Quarterly* 51(1): 59–89. DOI: 10.1002/tesq.276.

Traunmüller H (1990) Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America* 88(1): 97–100. DOI: 10.1121/1.399849.

Trofimovich P and Baker W (2006) Learning second language suprasegmentals: Effects of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28(1): 1–30. Available at: http://eprints.whiterose.ac.uk/72874/.

Trofimovich P and Isaacs T (2012) Disentangling accent from comprehensibility. *Bilingualism* 15(4): 905–916. DOI: 10.1017/S1366728912000168.

Walker R (2011) *Vowel Patterns in Language*. Cambridge, UK: Cambridge University Press.

Wang X (2020) Segmental versus suprasegmental: Which one is more important to teach? *RELC Journal*: 397–412. DOI: 10.1177/0033688220925926.

Wheeler MW (2005) *The Phonology of Catalan*. New York, NY: Oxford University Press.

Wlomainck P (2002) Le travail du rythme et de l'intonation dans l'apprentissage d'une langue étrangère. In: Renard R (ed.) *Apprentissage d'une Langue Étrangère/Seconde Vol.2 La Phonétique Verbo-Tonale*. Bruxelle, Belgique: De Boeck Université, pp. 156–161.

Xi X, Li P, Baills F, et al. (2020) Hand gestures facilitate the acquisition of novel phonemic contrasts when they appropriately mimic target phonetic features. *Journal of Speech, Language, and Hearing Research* 63(11): 3571–3585. DOI: 10.1044/2020_JSLHR-20-00084.

Yuan C, González-Fuente S, Baills F, et al. (2019) Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition* 41(1): 5–32. DOI: 10.1017/S0272263117000316.

Zhang R and Yuan ZM (2020) Examining the effects of explicit pronunciation instruction on the development of L2 pronunciation. *Studies in Second Language Acquisition* 42(4): 905–918. DOI: 10.1017/S0272263120000121.

Zielinski B (2015) The segmental/suprasegmental debate. In: Reed M and Levis JM (eds) *The Handbook of English Pronunciation*. Chichester, UK: Wiley-Blackwell, pp. 397–412.

Appendix A. Individual Learning Experience and Self-Assessed Proficiency of French Summarized by Groups and Courses

Non-embodied condition						Embodied condition					
ID	AOL	YoL	SA	ExC	Prof.	ID	AOL	YoL	SA	ExC	Prof.
First-year class											
48	18	2	0	0	A2	65	12	6	0	0	B1
51	17	2	0	0	A2	73	11	5	0	9.3	B2
52	17	1	0	2.8	A1	78	16	2	0	1	A2
56	13	6	0	6.8	B2	80	19	0.5	0	0	A2
61	10	9	0	3	B2	82	13	5	0	0	B1
62	20	0.5	0	0	A1	86	12	6	0	0	B2
64	14	7	0	0	B1	89	16	2	0	0	A2
						90	20	1	0	0	A2
						91	21	1	0	0.5	A2
Second-year class											
24	10	8	0	0	B2	2	16	6	5	9	B2
25	16	3	0	5.5	B1	3	12	6	0	0	B1
26	12	8	0	9	B2	4	11	8	0	0	C1
27	13	6	0	0	B2	5	18	3	0	0	A2
28	14	6	0	3	B1	7	13	7	0.8	0	B1
29	17	1	4	1	B1	8	17	2	0	0.5	B1
30	16	3	0	2.3	A2	9	15	4	0.8	4	B2
31	12	7	0	0.8	A2	10	12	7	3	1.5	B2
32	12	7	0	12	B1	11	11	7	4	0	B2
33	14	6	1	3	B1	13	10	10	0.5	0	B1
34	15	5	0	0	A2	14	14	5	0	2.3	B1
35	12	7	0	0	B2	15	15	5	0	3	C1
36	6	14	0	0	B2	16	12	8	1	9	B2
37	19	2	0	0	A2	17	13	7	0	0.3	B1
38	12	8	0	6	B2	18	12	8	1	0	B2
39	16	4	4	0	B2	19	11	8	4	0.8	B2
40	13	7	0	0	B1	20	17	3	0	3	A2
41	14	6	4	0.3	A2	21	14	6	0	1	B1
42	10	8	0	0.5	B2	22	12	7	0	0	B2
43	12	5	0	0	B2						
44	12	4	0	0	B2						
45	12	8	0	1.5	B2						

Note. ID = Randomly assigned code to anonymize the participants; AOL= Age of onset of learning; YoL = Years of formal learning; SA = Months of study abroad in a French-speaking country; ExC = Months of extracurricular French courses taken; Prof. = Self-assessed proficiency level.

Appendix B. Dialogues Used in the Training Sessions and for the Dialogue-Reading Task

1. Dialogue for training session 1

Tu ne vas pas t'amuser avec cet hurluberlu **!**

Muriel : Eh ! Sal**u**t L**u**cie !

Lucie : Sal**u**t M**u**riel. Mais ch**u**t ! Ne parle pas si fort. T**u** es folle ! Où vas-t**u** ?

Muriel : Faire un tour dans la r**u**e. Et toi, pourquoi es-t**u** descend**u**e ?

Lucie : Parce que j'aime la r**u**e, c'est tout.

Muriel : C'est tout ? T**u** es s**û**re ?

Lucie : Oui, je suis s**û**re. Pourquoi ?

Muriel : Je te trouve pl**u**tôt triste. Allez, viens ! On sort **u**ne min**u**te.

Lucie : T**u** as v**u** tous ces nuages ? Il va pl**eu**voir, c'est s**û**r. On n'a pas **eu** de pluie depuis le déb**u**t d**u** mois de mars...

Muriel : Pourquoi t**u** es si triste, L**u**cie ? T**u** n'aimes pas ces vacances ?

Lucie : M**u**riel, c'est trop d**u**r. L**u**c ne me parle pl**u**s depuis sept jours...

Muriel : T**u** sais, c'est un vrai sauvage, ce L**u**c. T**u** ne vas pas t'am**u**ser avec cet h**u**rl**u**berl**u** !

(Adapted from Martinie and Wachs, 2006, p. 15)

2. Dialogue for training session 2

Tu as eu peur ?

Chloé : Je ne trouve pas la ceint**u**re de séc**u**rité.

Eugénie : Normal ! Cette voit**u**re est de quatre-vingt-d**eu**x, il n'y avait pas de ceint**u**re à l'époque.

Chloé : Ah bon ? Mais c'est très danger**eu**x !

Eugénie : Danger**eu**x ? Pas d**u** tout !

Chloé : Mon **œ**il… À quelle h**eu**re est le rendez-vous chez le coiff**eu**r ?

Eugénie : À 13 h**eu**res. Mais ne t'inquiète pas, c'est j**u**ste à côté. Profite des paysages !

Chloé : Oui, c'est très beau, mais ... je suis un p**eu** malade en voit**u**re.

Eugénie : T**u** n'as pas à avoir p**eu**r : je suis le meill**eu**r chauff**eu**r de toute la M**eu**se !

Chloé : Je vois le vide, quelle horr**eu**r… Au secours !

Eugénie : On est arrivés. T**u** as **eu** p**eu**r ?

Chloé : Non, non. J**u**ste un p**eu** mal au c**œu**r... Merci **Eu**génie, heim… Mais, je ne v**eu**x pas te retenir : je rentrerai s**eu**le…

(Adapted from Martinie and Wachs, 2006, p. 22)

3. Dialogue for training session 3

Une gentille vagabonde

Marie : Bonjour, généré**eu**se demoiselle ! Vous avez bien **u**ne petite pièce pour **u**ne pauvre miséré**eu**se !

La voyageuse : **Eu**h voyons voir... Ah, mais quel malh**eu**r ! Je ne trouve pl**u**s mon portef**eu**ille. Regardez : la poche de ma veste est vide !

Marie : La chance n'est pas en ma fav**eu**r !

La voyageuse : Ah, je suis bien anxi**eu**se maintenant : j'ai un train de banli**eu**e à sept h**eu**res, et je n'ai pas de billet ! Comment vais-je trouver d**eu**x **eu**ros ?

Marie : Un mouchoir, d**u** tabac, un vi**eu**x bout de ficelle ... Ah, voilà vos d**eu**x **eu**ros !

La voyageuse : Non, je ne v**eu**x pas ab**u**ser de votre amabilité.

Marie : Si, si ! Prenez ! C'est de bon c**œu**r.

La voyageuse : Alors j'accepte bien volontiers. Merci beaucoup. Au revoir !

(Adapted from Martinie and Wachs, 2006, p. 44)

4. Untrained item used for the dialogue-reading task

Je travaille, moi !

La jeune fille : Je ne p**eu**x pas passer ! Soyez gentil, monsi**eu**r, dégagez le passage, par pitié !

Le livreur : Dégager le passage ? S**û**rement pas. Je travaille, moi ! D'aill**eu**rs, c'est vous qui gênez. À vous de bouger votre voit**u**re.

La jeune fille : Je n'en crois ni mes y**eu**x, ni mes oreilles ! C'est l'h**eu**re de déj**eu**ner, et je suis déjà en retard... Que malh**eu**r !

Le livreur : C'est bien dommage. Bon, allez, entend**u** ! Je suis prêt à vous arranger. Mais vous m'aidez à décharger ces cartons !

La jeune fille : Oh ! Espèce de m**u**fle ! Que vous êtes mal élevé !

Le livreur : Des inj**u**res, à présent ? J**u**sque-là, je suis resté gentil. Mais pl**u**s question que je parte. Voyez-vous, je p**eu**x patienter toute la journée avec mon journal !

(Adapted from Martinie and Wachs, 2006, p.48)

Note. The target phonemes are in boldface.

Appendix C. Sentences Included in the Sentence Imitation Task

Sentences containing /y/

1. Faire un tour dans la r**u**e.

2. Pourquoi es-t**u** descend**u**e ?

3. C'est tout ? T**u** es s**û**re ?

4. Je te trouve pl**u**tôt triste.

5. (Untrained) À vous de bouger votre voit**u**re.

Sentences containing /ø/

6. Ah bon ? Mais, c'est très danger**eu**x.

7. Je ne v**eu**x pas te retenir.

8. Comment vais-je trouver d**eu**x **eu**ros ?

9. Je suis bien anxi**eu**se maintenant.

10. (Untrained) Je n'en crois, ni mes y**eu**x, ni mes oreilles.

Sentences containing /œ/

11. Je rentrai s**eu**le.

12. Je suis le meill**eu**r chauff**eu**r

13. Si, si, prenez. C'est de bon c**œu**r.

14. La chance n'est pas en ma fav**eu**r.

15. (Untrained) D'aill**eu**rs, c'est vous qui gênez.

Note. The target phonemes are in boldface.

Appendix D. Pairwise Comparisons of F1 (Bark) of the Three Target Front Rounded Vowels by Tests in Dialogue-Reading and Sentence Imitation Tasks

Table 1

Pairwise Comparisons of F1 (Bark) of the Target Front Rounded Vowels by Tests in the Dialogue-Reading Task

	Estimate	*SE*	*df*	*t* ratio	*p* value	Cohen's *d*
Pretest						
/y/-/ø/	-0.64	0.16	1001	-4.04	<.001	-0.97
/y/-/œ/	-1.53	0.16	1001	-9.71	<.001	-2.32
/ø/-/œ/	-0.90	0.16	1001	-5.46	<.001	-1.36
Posttest						
/y/-/ø/	-0.63	0.17	1001	-3.72	.001	-0.99
/y/-/œ/	-1.61	0.17	1001	-9.58	<.001	-2.54
/ø/-/œ/	-0.99	0.18	1001	-5.65	<.001	-1.56
Delayed posttest						
/y/-/ø/	-0.58	0.17	1001	-3.54	.001	-0.86
/y/-/œ/	-1.61	0.17	1001	-9.78	<.001	-2.38
/ø/-/œ/	-1.03	0.17	1001	-6.01	<.001	-1.51

Table 2

Pairwise Comparisons of F1 (Bark) of the Target Front Rounded Vowels by Tests in the Sentence imitation task

	Estimate	*SE*	*df*	*t* ratio	*p* value	Cohen's *d*
Pretest						
/y/-/ø/	-0.69	0.22	1001	-3.10	.006	-1.02
/y/-/œ/	-1.69	0.22	1001	-7.65	<.001	-2.52
/ø/-/œ/	-1.01	0.23	1001	-4.38	<.001	-1.50
Posttest						
/y/-/ø/	-0.73	0.23	1001	-3.22	.004	-1.07
/y/-/œ/	-1.74	0.23	1001	-7.71	<.001	-2.55
/ø/-/œ/	-1.01	0.23	1001	-4.33	.001	-1.49
Delayed posttest						

/y/-/ø/	-0.64	0.21	1001	-3.05	.007	-1.07
/y/-/œ/	-1.73	0.21	1001	-8.18	<.001	-2.87
/ø/-/œ/	-1.09	0.22	1001	-4.95	<.001	-1.80