This is the post print of the article. Please cite it in this form: Ozakin, A.S., Xi, X., Li, P. & Prieto, P. (2022). Thanks or tanks: A brief multisensory training with tactile cues facilitates the pronunciation of non-native English interdental consonants in a discourse reading task. *Language Learning and Development*. doi: 10.1080/15475441.2022.2107522

Thanks or tanks: A brief multisensory training with tactile cues facilitates the pronunciation of non-native English interdental consonants in a discourse reading task

Alev Senem Ozakin [a]*, Xiaotong Xi [a], Peng Li [a], and Pilar Prieto [b, a]

[a] *Department of Translation and Language Sciences, Universitat Pompeu Fabra, Barcelona, Spain;* [b] *Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain*

Correspondence concerning this article should be addressed to Alev Senem Ozakin, E-mail: senemozakin@gmail.com. Depending on the journal's preference, co-authors can be informed about the process as well.

Alev Senem Ozakin, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain. Research interests: Audiovisual processing, multimodal pronunciation training, and the role of gestures in second language acquisition.

Xiaotong Xi, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain. Research interests: Second language segment learning, multimodal training, and hand gestures. E-mail: xiaotong.xi@upf.edu

Peng Li, Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain. Research interests: Gestural studies, second language pronunciation, and language contact between Japanese and Chinese. E-mail: peng.li@upf.edu

Pilar Prieto, Institució Catalana de Recerca i Estudis Avançats, and Department of Translation and Language Sciences, Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain. Research interests: The role of prosody and co-speech gestures in human communication. E-mail: pilar.prieto@upf.edu

Thanks or tanks: A brief multisensory training with tactile cues facilitates the pronunciation of non-native English interdental consonants in a discourse reading task

The present study investigates whether training second language pronunciation with tactile cues facilitates the production of non-native sounds involving accessible articulatory features. In a between-subjects experiment with a pretest-training-posttest design, 50 Turkish learners of English received audiovisual training on a set of target words and sentences containing two English interdental fricatives, /θ/ and /ð/, in one of two conditions, tactile and non-tactile. The tactile condition involved self-touching the tongue as it protruded during pronunciation of the two target sounds. Participants' pronunciation performance was assessed through a word-imitation task, a sentence-imitation task, and a discourse reading task. Results showed that while both training conditions helped learners to improve their pronunciation performance in all three tasks, the tactile condition triggered greater improvements in the discourse reading task. These results extend previous findings on the benefits of tactile input for speech perception and suggest the efficacy of multisensory training paradigms for improving second language pronunciation.

Introduction

Multisensory teaching, an innovative method that integrates several sensory modalities, including verbal instruction, visual modeling, and visual-tactile cues, has been shown to promote learning in educational settings. Particularly in language learning, multisensory approaches are argued to be beneficial (Minogue & Jones, 2006) because learners may learn better when the material is perceived through multiple sensory modes (e.g., visual, auditory, kinesthetic, and tactile). In what follows, we will briefly summarize previous research on language training paradigms using different multisensory cues with a focus on second language (L2) pronunciation.

Effects of Articulatory Training with Visual Cues

Previous studies have confirmed the effectiveness for improving L2 sound pronunciation accuracy of articulatory training based on phonetic explanations. When learners receive an explicit explanation of the articulatory postures and movements of the target sounds, they are more likely to obtain gains in their L2 speech production (Cartford & Pisoni, 1970; Saloranta et al., 2015).

In addition to verbal explanation, providing visual feedback of articulation can promote L2 sound production. This visual articulation feedback can be either direct or indirect (Kartushina et al., 2015). Direct feedback offers dynamic and immediate visualization of the articulatory information via various techniques (Gick et al., 2008). Ultrasound imaging, for example, allows learners to see their own articulatory movements in real time, which in turn helps them to control their tongue movements more effectively and thus improve their output (Ouni, 2014). Indirect feedback, on the other hand, typically provides visual information derived from acoustic analyses of the learner's speech (e.g., spectrograms), which learners can then use to try progressively match their pronunciation to target model speech.

Effects of Audiovisual Training on L2 Pronunciation

Providing visual information from a speaker's face can also enhance L2 pronunciation learning. The consensus is that human beings perceive natural speech in a multisensory fashion that combines auditory and visual sources. Several empirical studies have demonstrated that access to multiple sensory dimensions can promote phonological learning. Because observing the instructor's articulatory movements helps learners correctly identify L2 consonants with salient articulatory movements (Hazan et al., 2006), training with audiovisual input boosts learning (Hazan et al., 2005). Moreover, having

access to the instructor's articulatory movement can also help the production of L2 sounds with less salient articulatory movements, such as nasal vowels (Inceoglu, 2016).

Effects of Hand Gestures in L2 Pronunciation Training

Recent studies on articulatory training have attempted to involve sensorimotor cues to the target segmental features, such as hand gestures. Training with hand gestures, coupled with visual and auditory input, may lead to better phonemic awareness and help learners produce non-native sounds more accurately (Amand & Touhami, 2016). However, hand gestures used for L2 pronunciation training must adequately encode the phonetic features being trained (Xi et al., 2020), and learners must be able to perform the target gestures appropriately during training (Li et al., 2021). Furthermore, the type of gesture also plays a role in successful learning. For instance, Hoetjes and van Maastricht (2020) trained Dutch speakers to produce the Spanish /θ/ and /u/ with pointing gestures or gestures mimicking the articulatory features. However, only pointing gestures helped the production of both phonemes. Gestures mimicking the target articulatory features favored the learning of /u/ but hindered the learning of /θ/. Briefly, when using hand gestures to promote the production of non-native sounds, choosing the correct type of hand gesture is crucial in this training paradigm.

Effects of Tactile Information on L2 Pronunciation Training

In L2 pronunciation teaching practice, researchers have proposed a set of teaching strategies using tactile information. For example, the "Butterfly" technique asks learners to tap their shoulders when uttering a stressed syllable and the opposed elbow for an unstressed syllable (Burri & Baker, 2016); the "Touchinami" technique uses sweeping hand movements, including touch, to mimic intonational patterns; the "Tai Chi" technique asks learners to hold a ball and stretch their arms to learn stressed syllables;

and the "Rhythm Fight Club" instructs learners to perform boxing-like movements to mimic syllable stress (Acton et al., 2013; Teaman & Acton, 2013). A recent qualitative study showed that students found haptic techniques to be both highly engaging and beneficial, suggesting that the incorporation of touch, movement, and systematic hand gestures might be of great utility to language teachers (Burri & Baker, 2019).

Despite these proposals, contradictory findings have been reported in empirical studies. Bara et al. (2004) found that adding tactile information to the shape of letters can strengthen the link between orthographic and phonological representations and thus can help children learn alphabetic principles. Gick and Derrick (2009) found that listeners' perceptions of aspirated plosives were more accurate when auditory stimuli were accompanied by aero-tactile cues than when they were exposed to auditory stimuli alone. More recently, Cibelli (2020) used tactile information to cue laryngeal control (e.g., using hands to feel the vocal fold vibration for voicing Hindi /d/ and airburst for the aspirated Hindi /tʰ/) in combination with visual information (e.g., a sagittal section image of the mouth) showing tongue position (e.g., the Hindi retroflex /ʈ/). They found that this training paradigm was effective in improving L2 pronunciation. By contrast, Esteve-Gibert et al. (2019) reported that adding tactile information had no significant effect on L2 perception. They trained Catalan/Spanish bilingual children to learn the English /æ/–/ʌ/ vowel contrast with tactile information by watching themselves in a mirror and touching their own lips to feel the differences in lip shape or without such information. However, the tactile training implied no significant benefits when it came to distinguishing between the target vowel pair.

The Current Study

As previous studies do not show full consensus on the role of tactile cues in L2 pronunciation learning, more evidence is needed. In particular, relatively few studies have

directly employed tactile information to cue articulatory features such as lip and tongue movements. The current study therefore complements the existing literature on articulatory training by cueing the tongue position via self-touch. We address the following research question: Does adding tactile cues to multisensory training enhance learners' production of non-native sounds involving visible articulatory movements?

To answer this question, we trained Turkish learners of English to produce non-native interdental fricatives /θ/ and /ð/. It is well known that Turkish learners of English tend to replace the interdental /θ, ð/ with the closest Turkish counterparts /t, d/, pronouncing 'thanks' as 'tanks' (Ercan, 2018). Since interdental sounds are produced with salient articulatory movements (i.e., a protruded tongue), a more robust integration of tactile information in the form of self-touching might well improve learners' production of these sounds. Based on the Embodied Cognition framework, we hypothesized that adding tactile cues would improve Turkish learners' pronunciation accuracy of the English interdental fricatives /θ/ and /ð/.

Embodied Cognition holds that mind and body play a joint role in human cognitive processing (Foglia & Wilson, 2013; Wellsby & Pexman, 2014), and it thus has important implications for education (Shapiro & Stolz, 2019). A growing number of studies on Embodied Cognition have provided evidence that the body plays a role in language processing (Wilson, 2002) and language learning (Nathan, 2021). In educational settings, perceiving through multiple sensory-motor channels (e.g., kinesthetic, visual, and auditory) may improve learning by activating the cognitive system and reducing cognitive demands. For example, Goldin-Meadow et al. (2001) showed that gesturing can help children recollect more resources from their working memory and lighten their cognitive load during oral explanation. Similarly, Ping and Goldin-Meadow (2010) observed that the cognitive benefit is greater when novice learners produce meaningful

gestures than when they do not gesture, and that this sort of gesturing saves cognitive resources even when the referred objects are not physically present at the time.

In line with this, we hypothesize that exposing the learner to another sensory layer of information will facilitate phonological learning. Thus, including sensory-motor actions, in this case touching with the hand, coupled with audiovisual training, will help reduce learners' cognitive load and therefore boost their pronunciation of non-native segmental phonemes.

Methodologically, we aimed to assess L2 pronunciation through the use of two complementary tasks, namely imitation and discourse reading. Although imitation tasks have been used frequently for the evaluation of L2 pronunciation patterns, they have been shown to not necessarily reflect learner's abilities to produce difficult sound contrasts as shown by word-reading tasks (Llompart & Reinisch, 2019). Thus, following Saito and Plonsky (2019), in order to provide a more comprehensive and reliable assessment of the acquisition of L2 pronunciation patterns, we included a discourse reading task, which is a complex task involving more natural and non-imitative behaviors, in addition to two imitation tasks.

Methods

The experiment consisted of a between-subjects one-session multisensory training session with a pretest-posttest design. We designed and conducted an online controlled experiment, and participants received the links to the experiment by e-mail. They were trained for approximately 18 minutes with 12 English words and 12 sentences containing the interdental fricative sounds /θ/ and /ð/ under one of the two conditions, namely the Tactile (T) condition and the Non-Tactile (NT) condition. Whereas in the T condition participants watched a male instructor pronouncing the target words and sentences and modeling the self-touch gesture (touching the tip of his tongue) whenever he produced

the interdental consonants, in the NT condition they watched the same instructor pronouncing the same target words and sentences but without performing any self-touch gestures. For both conditions, participants were asked to imitate the instructor, either by repeating the target words and sentences with (T) or without (NT) tactile cues.

Participants

Fifty monolingual Turkish undergraduate and graduate students (31 females, 19 males, M_{age} = 24.2 years, 18–29 years old) with an elementary or intermediate English proficiency level were recruited from Turkey. They took an online English proficiency test consisting of 25 questions (https://www.cambridgeenglish.org/test-your-english/for-schools/). Participants' test scores on the proficiency test were collected before the experiment. Then they were assigned to either the T or NT group based on their proficiency test scores in such a way that the distribution of English proficiency levels was balanced across the two groups.

Once recruited, participants completed a short online questionnaire requesting information about their age, gender, linguistic background, musical experience, history of formal English language experience, and extra-classroom exposure to English. The musical experience questions were adapted from Li et al. (2020) and coded according to the method described therein. All the participants signed a consent form giving permission to use their experimental data for the purposes of this study.

Table 1 summarizes all the individual factors obtained from the questionnaire across the two groups (T vs. NT); none of the measures showed a significant difference between groups (all p >.05).

[t]Table 1 near here[/t]

Materials

This section describes the audiovisual and audio materials used for the training session and pre- and post-tests. All the audiovisual materials were recorded with a Canon EOS 2000d professional portable digital camera and were edited with *Adobe Premiere Pro 2020* software. All questionnaires and audio or audiovisual materials were uploaded to the online platform Alchemer (https://www.alchemer.com/) to prepare the two versions of the experimental survey.

Audiovisual materials for the familiarization phase

For the familiarization phase, two short 4-minute audiovisual sequences were created for each condition to explain and illustrate the main articulatory feature of the English interdental fricatives, that is, the position of the tongue. At the end of these short videos, participants were shown two sample training trials in order to familiarize themselves with the experimental procedure. These sample trials featured the two target interdentals in words or phrases that did not appear in the actual experimental materials. In the familiarization video for the T condition, the instructor produced the target words and touched the tip of his tongue whenever he pronounced an interdental while written instructions instructed the participant to repeat the utterances and mimic the self-touch. In the video for the NT condition, the instructor produced the same speech stimuli without any tactile information and participants were instructed to simply imitate the speech.

Audiovisual materials for the training session

The training video consisted of two blocks, one in which the English interdental fricatives were trained at the word level and another in which the target phonemes were trained at the sentence level. Each video clip for word stimuli lasted approximately 2 seconds, whereas for sentence stimuli lasted 5 seconds.

For the first training block, a set of 12 high-frequency monosyllabic English words containing the target phonemes were selected as the training stimuli. The target phonemes were distributed equally in two different positions, namely word-initial and word-final. Eight of these words contained the voiceless interdental fricative phoneme /θ/, whereas the remaining four included the voiced interdental fricative phoneme /ð/. The number of voiceless tokens was deliberately made higher to reflect the fact that in English, though voiced phoneme /ð/ has a higher frequency in functional words, the frequency of /θ/ is greater overall (219 vs. 66 out of 9,174,650 tokens, Gilner & Morales, 2010).

For the second block, another set of 12 monosyllabic and disyllabic high-frequency English words containing the two target phonemes in word-initial or word-final positions were selected and embedded into sentences as the training stimuli. Again, eight words contained voiceless /θ/, whereas the remaining four included voiced /ð/. In order to guarantee a salient pronunciation of the target words, they were placed in prominent prosodic positions within the utterance, that is, in either sentence-initial or sentence-final positions. The sentence length was between four and five words.

A male native American-English-speaking instructor was video-recorded while producing the target words and sentences for the two training conditions, i.e., with and without tactile cues. The tactile cue consisted of the instructor placing his right index finger on the tip of his protruded tongue whenever he pronounced an interdental fricative. This is illustrated in Figure 1.

[t]Figure 1 near here[/t]

For both blocks, the instructor was video-recorded as follows. First, the instructor faced the camera while producing the training stimuli without the touch cues, and then he looked sideways while producing the same stimuli again, either with the touch cues (for

the T condition) or without (for the NT condition). This sideways angle allowed viewers to see the touching procedure and/or the tongue clearly.

The videos in the two conditions were edited to compose a video for each trial which included three repetitions of the target words or sentences, with the following temporal sequence. First, the word or sentence containing the target phoneme appeared as English text on the screen with the target phonemes highlighted in yellow; then the video clip with the instructor pronouncing the stimulus was played; this was followed by text in Turkish instructing the participant to repeat the stimulus. The following video clip showed the instructor looking sideways and repeating the stimulus, either performing touch cues (T condition) or not (NT condition); this was again followed by text instructing the participant to either repeat the speech while imitating the tactile cue (T condition) or repeat the speech only (NT condition). Finally, the last two steps were repeated, which enabled each training stimulus to be produced three times in one trial. All the trials appeared twice in different orders within each block. Thus, each target word or sentence was presented six times. The first (word-level) block lasted 7 minutes and 54 seconds, and the second (sentence-level) block lasted 9 minutes and 42 seconds. Figures 2 and 3 show the still frames of one trial from the training videos.

[t]Figure 2 near here[/t]

[t]Figure 3 near here[/t]

Materials for pre- and posttest tasks

The auditory stimuli for the word and sentence-imitation tasks consisted of 12 words and 12 sentences containing the target phonemes. In both imitation tasks, half of the stimuli were identical to those used in the training session, while the other half were not. In the word-imitation task, eight words contained /θ/ while four contained /ð/. Half of the words

occurred word-initially and the other half word-finally. In the sentence-imitation task, again, there were eight /θ/ and four /ð/, while eight target phonemes occurred word-initially and four word-finally. Exactly the same set of stimuli for the word and sentence-imitation tasks were used in the pretest and posttest.

Audio clips were digitally recorded with professional equipment and featured the same native English speaker that produced the training materials. All words and sentences were recorded twice at a normal speech rate, then the recording that sounded clearer and more natural-sounding of each pair was selected for the final 24 audio files.

The short story for the discourse reading task was *Theo the Thief*, which was taken from the website Teaching Cave (https://www.teachingcave.com), a site that provides teaching materials for EFL teachers. The ten-sentence story has 12 words that contain interdental phonemes, some of them being repeated more than once, for a total of 18 instances (e.g., *Theo, thinks, then, thief*). Each sentence contains a maximum of three words with the target phonemes.

Procedure

As noted above, after recruitment participants were assigned to the T or NT groups according to their Cambridge proficiency test scores. Due to the outbreak of COVID19, the entire experiment was carried out remotely, with participants gaining access to the experiment hosted on the online platform Alchemer through a link sent by e-mail. For each task, the instructions were provided in Turkish so as to avoid any potential misunderstanding. The order of the presentation of the questions was automatically randomized within each test by the software. The entire 35-minute session was self-recorded by participants, each participant having been informed about how to use the self-video recording tool (https://webcamera.io/tr/) to ensure that they performed the tasks properly. When they had finished, they sent their recording to the first author.

A schematic diagram of the experimental procedure can be seen in Figure 4.

[t]Figure 4 near here[/t]

Participants took the pretest, which consisted of word-imitation (2 min), sentence-imitation (3 min), and a discourse reading task (1 min). This was followed by a familiarization phase in which participants watched a 4-minute video introducing the main articulatory phonetic features of the English interdental fricatives, specifically illustrating the tongue position and the distinctive air friction that occurs during articulation. After this short video, participants undertook the training session that corresponded to their group and which consisted in both cases of two blocks (isolated words and sentences), lasting 18 minutes in total. Finally, participants were asked to perform the posttest, which consisted of the same tasks as the pretest (6 minutes). Altogether, the full experiment lasted about 35 minutes.

Pretest and posttest

In the word- and sentence-imitation tasks, participants were instructed to repeat 12 words and 12 sentences, respectively. The stimulus for repetition was exclusively auditory, with no written text being presented. After hearing the audio file once, participants were asked to repeat the word or sentence they had heard and then to confirm that they had done so by mouse-clicking on a circle below the text.

In the discourse reading task, participants were asked to read aloud the written text once at a normal pace. After they finished reading, they clicked on a circle below the text to confirm that they had completed the task.

Training session

After completing the pretest, participants watched the training videos. In the NT condition

participants watched the instructor pronounce the words and sentences, whereas in the T condition, participants watched the instructor performing the tactile cues whenever he pronounced the target interdental phonemes. In both conditions, participants were asked to replicate the words that they heard, and in the T group, they were also asked to replicate the actions they had seen.

Data Coding

To evaluate the L2 speech at pre-and post-test, we recruited six native American English speakers to perceptually rate participants' output. The decision to conduct perceptual ratings using native speakers, as opposed to acoustic analyses, was based on two reasons. First, the mispronunciation patterns of the target phonemes can vary depending on participants. Turkish learners of English tend to pronounce the /θ, ð/ as /t, d/, but they occasionally realize them as /s, z, f, v/. Therefore, a speech assessment based on acoustic cues could not provide a straightforward analysis. Second, since the goal of the current study was to assess the improvement in pronunciation quality from pretest to posttest, we felt that native speakers' impressions would provide a more uniform measure (i.e., how accurate the pronunciation sounded regardless of the specific acoustic features they relied on to give the score).

A total of 2,400 video-recorded clips (50 participants × 12 imitation items × 2 tasks × 2 tests) were obtained from the word- and sentence-imitation tasks, and a total of 100 clips (50 participants × 2 tests) were obtained from the discourse reading task. The audio tracks of these clips were separated from the video material and then rated on pronunciation accuracy by the six native speakers (3 females, M_{age} = 35.5 years). Before performing the ratings, all raters were trained simultaneously in a 45-minute session with speech samples to become familiar with the evaluation system and practice applying it. In the rating session for each task, the raters rated the audio clips in random order. That

is, they did not know which testing session and training condition each audio clip they heard was taken from.

For the imitation tasks, raters were asked to listen to each audio clip and assess the segmental accuracy of the interdental fricatives of each item using a 5-point Likert scale (5 'very accurate' to 1 'not accurate'). Raters were asked not to pay attention to the suprasegmental features but rather focus on the target phonemes. The mispronunciation of the non-target phonemes was not penalized. For instance, if the original stimulus word was "throw" /θroʊ/ but the participant produced /θroʊn/ (throne), the production still received a high rating because the target phoneme /θ/ had been produced accurately.[1]

For the discourse reading task, in order to limit the length of the rating process and avoid fatigue caused by monotony, raters were instructed to give a general accuracy score on a 5-point Likert scale based on the pronunciation of all the target phonemes in each clip overall rather than on each target phoneme individually. This score depended directly on the proportion of correctly pronounced target phonemes (i.e., 5 = all tokens correct, 4 = most tokens correct, 3 = half of the tokens correct, 2 = most tokens incorrect, 1 = all incorrect).

Interrater reliability across the six raters for the three outcome measures was checked by using the Intraclass Correlation Coefficient (ICC) test. ICC estimates and their 95% confidence intervals were calculated using the *irr* package version 0.84.1 (Gamer et al., 2019). The ICC for segmental accuracy in both word-imitation (ICC = 0.78, 95% CI [0.77, 0.80]) and sentence-imitation (ICC = 0.77, 95% CI [0.75, 0.79]) tasks was good, and was excellent in the discourse reading task (ICC = 0.86, 95% CI [0.78, 0.90]).

[1] In the few cases (N = 11) where speakers self-corrected, we took the corrected utterance as the target item to evaluate.

Statistical Analyses

A set of Linear Mixed Models (LMM) were performed for each of the three response variables (segmental accuracy scores from the word-imitation, sentence-imitation, and discourse reading tasks) with *glmm TMB* package version 1.0.2.1 (Brooks et al., 2017) of R (version 4.0.2 in R Studio). For the three LMMs, the response variable (segment accuracy) was obtained by calculating the average of the six ratings of each item pronounced by each participant, as suggested by Nagle (2018) for modeling rating data. Based on our research questions, all three LMMs included contrast-coded fixed factors for Condition (-.5 = NT, .5 = T) and Test (-.5 = pretest, .5 = posttest), as well as their interaction. A set of initial LMM analyses involved the main effects of familiarity (trained vs. untrained), position (initial vs. final), and voicing (voiced vs. voiceless), as well as the interaction of each of them with Condition × Test. Since no significant 3-way interaction was found, these three factors were excluded from the analysis to nest the models to our research question. See Appendix A for the reports of the models.

As for the random structures, a set of LMMs with different random effect structures were modeled and compared to find the best-fitting model for our dataset. The models were built from the one with the most complex random effects structure to a marginal model with a random intercept for participants. Then only those models reporting no convergence or singular fitting problems were included in the model comparison that was performed using the *performance* package version 0.4.8 (Lüdecke et al., 2019). Thus, the random effect structures of the best fitting models for the three outcome measures were as follows: (a) for the word-imitation task, a by-participant and a by-item random slope for test, and random intercepts for participant and item; (b) for the sentence-imitation task, a by-participant random slope for test, a by-item random slope for condition and test, and random intercepts for participant and item; and (c) for

the discourse reading task, a random intercept for participant. Correlations between random effects were also included in model building. Significance tests for fixed effects were then performed with Type II Wald Chi-squared tests using the *car* package (Fox & Weisberg, 2019). Sequential Bonferroni corrections were applied to the post-hoc pairwise comparisons using *emmeans* package version 1.5.1 (Lenth et al., 2020). Effect size with Cohen's *d* was calculated. We followed the benchmarks proposed by Plonsky and Oswald (2014) for L2 research, which considered the effect size as small for a *d* value around 0.60, medium for a *d* value around 1.00, and large for a *d* value around 1.40.

Results

Word-Imitation Task

Figure 5 shows the mean segment accuracy scores obtained by participants in the word-imitation task across conditions (NT and T) and tests (pretest and posttest). Results (see Table 2) revealed a significant main effect of test, a main effect of condition, and a significant two-way interaction between Condition × Test. This indicates that (a) participants' segment accuracy score in word-imitation task differed significantly from pretest to posttest; (b) segment accuracy score differed between conditions; and (c) the difference in scores from pre- to posttest was different between the two conditions. Post-hoc analyses of the main effect of condition revealed that the participants from the NT condition had a significantly higher segment accuracy score than participants in the T condition ($d = 0.51$, $p = .001$, 95% CI [0.20, 0.82]). Post-hoc results of the main effect of test showed that the segment production of all participants significantly improved after the training session ($d = 0.62$, $p < .001$, 95% CI [0.46, 0.79]). For the 2-way interaction of Condition × Test, the results can be assessed in two ways. The improvement from pretest to posttest was significantly different between the T condition ($d = 0.81$, $p < .001$,

95% CI [0.59, 1.03]) and the NT condition (d = 0.44, p < .001, 95% CI [0.22, 0.66]). At the same time, the NT condition obtained a significantly higher score than the T condition at both pretest (d = 0.70, p < .001, 95% CI [0.29, 1.10]) and posttest (d = 0.33, p = .016, 95% CI [0.06, 0.60]).

[t]Table 2 near here[/t]

[t]Figure 5 near here[/t]

Sentence-Imitation Task

Figure 6 shows the mean segment accuracy scores obtained in the sentence-imitation task across conditions (NT and T) and tests (pretest and posttest). Results (Table 3) revealed a significant main effect of test, suggesting that participants' segmental performance at sentence level was statistically different from pre- to posttest. Post-hoc analyses indicated that there was a significant improvement from pretest to posttest regardless of training condition (d = 0.58, p < .001, 95% CI [0.35, 0.80]). Yet, the two-way interaction of Condition × Test was not found to be significant, which suggests that the training with tactile information did not yield more improvement for segment accuracy at sentence level compared to training without tactile information.

[t]Table 3 near here[/t]

[t]Figure 6 near here[/t]

Discourse Reading Task

Figure 7 shows the mean segment accuracy scores obtained in the discourse reading task across condition (NT and T) and test (pretest and posttest). Results (Table 4) revealed a significant main effect of test, indicating that participants' performance differed from

pretest to posttest. A significant 2-way interaction between Condition × Test was found, revealing that the performance of the participants differed significantly between pre-and post-test across conditions. Post-hoc results showed that the segment accuracy score was not statistically different between the two conditions either at pretest (d = 0.41, p = .410, 95% CI [-0.59, 1.42]) or at posttest (d = 0.37, p = .462, 95% CI [-1.38, 0.63]) and the score improved significantly from pre- to posttest in both the NT (d = 1.40, p < .001, 95% CI [0.80, 1.99]) and the T condition (d = 2.19, p < .001, 95% CI [1.54, 2.83]). Though both training conditions revealed large effect sizes, the significant 2-way interaction suggests that the T condition yielded a significantly larger improvement (effect size) than the NT condition.

[t]Table 4 near here[/t]

[t]Figure 7 near here[/t]

Discussion and Conclusions

The present study examined the effectiveness of a short 18-minute multisensory training session with visual and tactile cues on the acquisition of the English interdental fricative /θ, ð/. The pronunciation gains were assessed through three complementary tasks, namely a word-imitation task, a sentence-imitation task, and a discourse reading task. Moreover, assessing segment accuracy through a set of controlled and less controlled tasks allowed us to measure pronunciation gains in a comprehensive fashion (Saito & Plonsky, 2019).

The results of the study revealed a contrast between the participants' performance in the imitation tasks and their performance in the discourse reading task. First, we did not observe clear beneficial effects of adding tactile information in the word- and sentence-imitation tasks. The unbalanced pretest score in the word-imitation task makes it impossible to discuss the implication of this task further. As for the sentence-imitation

task, adding tactile information was not more helpful for language learners. Taken together, adding tactile information was not proven to significantly improve L2 speech imitation compared to not having it.

The results of the segmental accuracy scores in the discourse reading task revealed a larger improvement in the T group than in the NT group. Crucially, the two groups did not differ significantly from each other at pretest in their segmental accuracy scores, suggesting that the larger improvement observed in the T group can be attributed exclusively to the tactile training. These results thus clearly support our hypothesis that tactile training would yield greater improvement in the production of interdental fricatives /θ, ð/ than training without tactile cues. The mechanism that may account for the current findings is the perception-production relationship in L2 speech learning. According to the Speech Learning Model (SLM, Flege, 1995, 2003), learners must first detect the phonetic differences between similar sounds to correctly produce them. Our training paradigm, based on embodied cognition, provides a multimodal perceptual cue (i.e., tactile information) to enhance the detectability of the perceptual differences between two sets of phonemes (e.g., /t-d/ vs. /θ-ð/). Therefore, learners in the T group may have had more complementary channels (tactile + audiovisual) to help them perceive the target phonetic features compared to the NT group (audiovisual).

Notably, in the discourse reading task we observed a larger effect size from pretest to posttest than in the two imitation tasks. Task difficulty might account for this difference. Specifically, since the imitation tasks provided participants with a native model speech, participants obtained relatively higher pretest scores in this task than in the discourse reading task. Thus, the relatively low scores on the pretest might have given participants more room to improve in the discourse task.

All in all, the results of our study show a positive joint effect of visual, auditory, and tactile input on pronunciation learning. These results confirm and expand previous knowledge of how L2 learners integrate relevant tactile information in the production of L2 sounds.

The main novelty of the current study is that it shows that articulatory training with tactile cues, in combination with audiovisual information, can facilitate the acquisition of articulatory features related to tongue position. While previous studies have confirmed the effectiveness of articulatory training based on phonetic explanations in improving L2 phoneme production (Cartford & Pisoni, 1970; Cibelli, 2020; Gick et al., 2008; Ouni, 2014; Saloranta et al., 2015), the current study provides positive evidence in favor of integrating haptic approaches to articulatory training on L2 phonemes which involve visible articulatory movements.

Our results contrast with the null results obtained by Esteve-Gibert et al. (2019), where tactile training did not improve children's perception of L2 English /æ/-/ʌ/ contrast. This difference may be due to methodological reasons. First, the distinction between /æ/ and /ʌ/ is not only due to lip shape but also due to tongue position, reflected by the significantly different F2 values that characterize the two sounds (Mora & Fullana, 2007). Therefore, lip touch may not provide full information on articulatory movements. By contrast, in our study, the main articulatory difference between /θ-ð/ and /t-d/ lies in the tongue position. This difference is highlighted for learners by our "tongue touch" strategy. Second, Esteve-Gibert and colleagues (2019) used an indirect training paradigm in which participants were not provided with visual input from the native speakers. Rather, participants touched their own lips while looking in a mirror. This indirect training paradigm leaves them to rely on their own non-native articulatory movements. In addition,

since they did not report the production data for the imitation task, we cannot make a direct comparison between our imitation data and theirs.

Interestingly, our findings confirm and expand on Llompart and Reinisch's (2019) work, which suggest that imitation is modulated by perceptual skills but does not necessarily reflect productive knowledge. Our study showed that tactile training showed more benefits than non-tactile training in the discourse-reading task, but not in the imitation tasks. It may be that even though imitation abilities are related to learners' phonological representation of L2 sound contrasts, participants found it more difficult to draw on those representations when called upon to produce non-native sounds in the discourse reading task. Here, adding tactile information may have reinforced their awareness of the target phonological categories, thus leading to improvement in pronunciation in read-aloud settings. Moreover, imitation tasks are not as demanding as discourse reading task because participants only imitated what they had heard. This can be verified by the significantly higher segment accuracy scores obtained at pretest in the word-imitation and sentence-imitation tasks compared to those in the discourse reading task. Finally, the participants' individual speech-imitation abilities may have influenced their improvement in the imitation tasks independently of training. This suggests that future studies may need to control for general imitation abilities when assessing the results from imitation tasks.

Our findings also have practical implications for language teachers. Activating embodiment and touch in classroom settings will boost L2 learners' phonological awareness and thus facilitate their ability to accurately hear and produce novel phonemes. Importantly, in the present study, the positive role of multisensory training was mainly documented in the discourse reading task, suggesting that L2 teachers should consider

implementing broader sets of training materials that include not only words and short chunks but also more complex sequences and whole discourse.

The present study has several limitations. First, since the T group both observed the tactile gesture from the instructor and at the same time imitated it while producing the training items, it is conceivable that the combination of both viewing and enacting the tactile information jointly boosted pronunciation learning. Therefore, future investigations might want to test the effects of observing tactile information only. Second, participants' general imitation abilities were not assessed. If, for example, speech imitation abilities had been treated as a control measure across groups, the results for imitation could have been assessed in a more robust way. Future studies should assess individual differences in phonological knowledge to generate more homogenous between-subject groups for the study. Third, it would have been beneficial if any long-lasting gains were tested. Future studies should take this aspect into consideration as well.

In conclusion, the present study has shown evidence that a short 18-minute multisensory training session with tactile information (in this case, self-touching the tip of the tongue when producing interdental consonants) can facilitate the pronunciation of novel phonological features. We believe that the results of the current study lay the ground for further experimental testing of multisensory training not only in the L2 pronunciation classroom but possibly also in clinical contexts.

Acknowledgments

We appreciate the assistance of Dr. Ayşen Değer and Dr. Seda Altıner in the recruitment of students to conduct the pilot task and the experiment. We extend our gratitude to the external members of the Universitat Pompeu Fabra TFM–MLTA committee, as well as Dr. Sílvia Perpiñán, and Dr. Núria Esteve-Gibert. We would like to thank all participants for their effort and time. Also, we would like to thank Patrick Louis Rohrer for his time and help in generating the ideal training materials for the research. Finally, many thanks to all members of *Grup d'Estudis de*

Prosòdia for providing additional information that helped us better structure the literature review section.

Disclosure Statement

The authors have no (financial) conflict of interest in the subject matter or materials discussed in the manuscript.

Funding

This work was supported by the Spanish Ministerio de Ciencia, Innovación y Universidades (MCIU), Agencia Estatal de Investigación (AEI) and Fondo Europeo de Desarrollo Regional (FEDER) [grant number PGC2018-097007-B-l00], and Catalan government's Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR) [grant number 2017 SGR_971]. The second author is supported for her PhD study by the Secretaria d'Universitats i Recerca de la Generalitat de Catalunya and the European Social Fund under the Grant for the recruitment of early-stage research staff [grant number 2020FI_B 00237].

References

Acton, W., Baker, A., Burri, M., & Teaman, B. (2013). Preliminaries to haptic-integrated pronunciation instruction. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 234–244). Iowa State University.

Amand, M., & Touhami, Z. (2016). Teaching the pronunciation of sentence final and word boundary stops to French learners of English: Distracted imitation versus audio-visual explanations. *Research in Language*, *14*(4), 377–388. https://doi.org/10.1515/rela-2016-0020

Bara, F., Gentaz, E., Colé, P., & Sprenger-Charolles, L. (2004). The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive Development*, *19*(3), 433–449. https://doi.org/10.1016/j.cogdev.2004.05.003

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Machler, M., & Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400.

Burri, M., & Baker, A. (2016). Teaching rhythm and rhythm grouping: The butterfly technique. *English Australia Journal*, *31*(2), 72–77.

Burri, M., & Baker, A. (2019). "I never imagined" pronunciation as "such an interesting thing": Student teacher perception of innovative practices. *International Journal of Applied Linguistics (United Kingdom)*, *29*(1), 95–108. https://doi.org/10.1111/ijal.12247

Cartford, J. C., & Pisoni, D. E. (1970). *Autitory vs. articulatory training in exotic sounds*.

Cibelli, E. (2020). Training Non-Native Consonant Production with Perceptual and Articulatory Cues. *Phonetica*, *77*(1), 1–28. https://doi.org/10.1159/000495728

Ercan, H. (2018). Pronunciation problems of the Turkish EFL learners in Northern Cyprus. *International Online Journal of Education and Teaching (IOJET)*, *5*(4), 877–893.

Esteve-Gibert, N., Del Mar Suárez, M., Vasylets, O., & Serrano, R. (2019). Children's use of tactile input when acquiring non-native phonological contrasts. *Presentation at XIV International Symposium of Psycholinguistics*.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). York Press,.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In N. O. Schiller & A. S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319–355). Mouton de Gruyter.

Foglia, L., & Wilson, R. A. (2013). Embodied cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(3), 319–325. https://doi.org/10.1002/wcs.1226

Fox, J., & Weisberg, S. (2019). *An {R} Companion to Applied Regression* (Third). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/

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

Gick, B., Bernhardt, B., Bacsfalvi, P., & Wilson, I. (2008). Ultrasound imaging applications in second language acquisition. In J. G. H. Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 309–322). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.36.15gic

Gick, B., & Derrick, D. (2009). Aero-tactile integration in speech perception. *Nature*, *462*(7272), 502–504. https://doi.org/10.1038/nature08572

Gilner, L., & Morales, F. (2010). Functional Load: Transcription and Analysis of the 10,000 Most Frequent Words in Spoken English. *The Buckingham Journal of Language and Linguistics*, *3*, 135–162. https://doi.org/10.5750/bjll.v3i0.27

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing Lightens the Load. *Psychological Science*, *12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Hazan, V., Sennema, A., Faulkner, A., Ortega-Llebaria, M., Iba, M., & Chung, H. (2006). The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America*. https://doi.org/10.1121/1.2166611

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*. https://doi.org/10.1016/j.specom.2005.04.007

Hoetjes, M., & van Maastricht, L. (2020). Using gesture to facilitate L2 phoneme acquisition : The importance of gesture and phoneme complexity. *Frontiers in Psychology*, *11*(03178), 1–16. https://doi.org/10.3389/fpsyg.2020.575032

Inceoglu, S. (2016). Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics*, *37*(5), 1175–1199. https://doi.org/10.1017/S0142716415000533

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America*, *138*(2), 817–832. https://doi.org/10.1121/1.4926561

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *Emmeans: Estimated marginal means, Aka Least-Squares means*. R package 1.5.1. https://cran.r-project.org/package=emmeans

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

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

Llompart, M., & Reinisch, E. (2019). Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech*, *62*(3), 594–622. https://doi.org/10.1177/0023830918803978

Lüdecke, D., Makowski, D., Waggoner, P., & Patil, I. (2019). *Performance: Assessment of regression models performance*. R package 1.5.1.

Minogue, J., & Jones, M. G. (2006). Haptics in Education: Exploring an Untapped Sensory Modality. *Review of Educational Research*, *76*(3), 317–348. https://doi.org/10.3102/00346543076003317

Mora, J. C., & Fullana, N. (2007). Production and perception of English /iː/-/ɪ/ and /æ/-/ʌ/ in a formal setting: investigating the effects of experience and starting age. In J. Trouvain (Ed.), *Proceedings of the 16th International Congress of Phonetic Sciences* (Issue August, pp. 1613–1616). Saarbrücken Univ. des Saarlandes. http://www.icphs2007.de/conference/Papers/1594/1594.pdf

Nagle, C. (2018). Pronunciation in Second Language Learning and Teaching. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 82–105). Iowa State University. https://apling.engl.iastate.edu/alt-content/uploads/2015/05/PSLLT_5th_Proceedings_2013.pdf

Nathan, M. J. (2021). *Foundations of Embodied Learning: A Paradigm for Education*. Routledge.

Ouni, S. (2014). Tongue control and its implication in pronunciation training. *Computer Assisted Language Learning*, *27*(5), 439–453.

Ping, R., & Goldin-Meadow, S. (2010). Gesturing saves cognitive resources when talking about nonpresent objects. *Cognitive Science*, *34*(4), 602–619. https://doi.org/10.1111/j.1551-6709.2010.01102.x

Plonsky, L., & Oswald, F. L. (2014). How Big Is "Big"? Interpreting Effect Sizes in L2 Research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Saito, K., & Plonsky, L. (2019). Effects of Second Language Pronunciation Teaching Revisited: A Proposed Measurement Framework and Meta-Analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Saloranta, A., Tamminen, H., Alku, P., & Peltola, M. S. (2015). Learning of a non-native vowel through instructed production training. *Proceedings of the 18th International Congress of Phonetic Sciences*, 1–5. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0235.pdf

Shapiro, L., & Stolz, S. A. (2019). Embodied cognition and its significance for education. *Theory and Research in Education*, *17*(1), 19–39. https://doi.org/10.1177/1477878518822149

Teaman, B. D. ., & Acton, W. R. . (2013). Haptic (movement and touch for better) pronunciation. In N. Sonda & A. Krause (Eds.), *JALT2012 Conference Proceedings* (pp. 402–409). Japan Association for Language Teaching.

Wellsby, M., & Pexman, P. M. (2014). Developing embodied cognition: Insights from children's concepts and language processing. *Frontiers in Psychology*, *5*(MAY), 1–10. https://doi.org/10.3389/fpsyg.2014.00506

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625–636. https://doi.org/10.3758/BF03196322

Xi, X., Li, P., Baills, F., & Prieto, P. (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. https://doi.org/10.1044/2020_JSLHR-20-00084

Appendices

A. Linguistic and Musical Experience Questionnaire

Linguistic Experience

1. What is your native language?
2. Apart from Turkish, which language(s) do you speak?
3. How long have you been learning English?
4. Have you ever had phonology or pronunciation training beforehand?

Musical Experience

1. How many years of musical education have you received?
2. Do you play any musical instrument? If yes, how many instrument(s) do you play?
3. How often do you listen to music?
    a. Every day
    b. 5-6 days per week
    c. 3-4 days per week
    d. 1-2 days per week
    e. Occasionally
    f. Never
4. How often do you sing?
    a. Every day
    b. 5-6 days per week
    c. 3-4 days per week
    d. 1-2 days per week
    e. Occasionally

f. Never

B. English target words used for the training session (Block 1)

Word	IPA	Position	Word	IPA	Position
think	[θ]	word-initial	Ear**th**	[θ]	word-final
thank	[θ]	word-initial	Mou**th**	[θ]	word-final
throw	[θ]	word-initial	**Th**ose	[ð]	word-initial
thief	[θ]	word-initial	**Th**em	[ð]	word-initial
clo**th**	[θ]	word-final	Brea**th**e	[ð]	word-final
ma**th**	[θ]	word-final	Ba**th**e	[ð]	word-final

C. English target words and sentences used for the training session (Block 2)

Words	IPA	Position	Target Sentences
thumb	[θ]	word-initial	I hurt my **th**umb!
theory	[θ]	word-initial	I like your **th**eory.
thin	[θ]	word-initial	He is very **th**in.
thick	[θ]	word-initial	My book is **th**ick.
mon**th**	[θ]	word-final	April is my favorite mon**th**.
sou**th**	[θ]	word-final	He is moving sou**th**.
pa**th**	[θ]	word-final	What a nice pa**th**!
bo**th**	[θ]	word-final	Bo**th** my parents are doctors.
this	[ð]	word-initial	**Th**is is her room.
then	[ð]	word-initial	She was working back **th**en.
there	[ð]	word-initial	I want to go **th**ere.

these	[ð]	word-Initial	**Th**ese are my classmates.

D. Stimuli used for the Word-Imitation Task

Word	IPA	Position	Training	Word	IPA	Position	Training
throw	[θ]	word-initial	trained	eigh**th**	[θ]	word-final	Untrained
thief	[θ]	word-initial	trained	you**th**	[θ]	word-final	Untrained
thick	[θ]	word-initial	untrained	**th**em	[ð]	word-initial	Trained
thermos	[θ]	word-initial	untrained	**th**ey	[ð]	word-initial	Untrained
ear**th**	[θ]	word-final	trained	ba**th**e	[ð]	word-final	Trained
mou**th**	[θ]	word-final	trained	wri**th**e	[ð]	word-final	Untrained

E. Stimuli used for the Sentence-Imitation Task

Word	IPA	Position	training	Embedded Sentences
theory	[θ]	word-initial	trained	1.I like your **th**eory.
thumb	[θ]	word-initial	trained	2.I hurt my **th**umb!
thing	[θ]	word-initial	untrained	3.What a nice **th**ing!
theme	[θ]	word-initial	untrained	4.Did he say '**th**eme'?
sou**th**	[θ]	word-final	trained	5.He is moving sou**th**.
mon**th**	[θ]	word-final	trained	6.April is my favorite mon**th**.
fai**th**	[θ]	word-final	untrained	7.Don't lose your fai**th**!
loa**th**	[θ]	word-final	untrained	8.Can you say 'loa**th**'?
this	[ð]	word-initial	trained	9.**Th**is is her room.
then	[ð]	word-initial	trained	10.She was working back **th**en.
their	[ð]	word-initial	untrained	11.**Th**eir house is huge.

that	[ð]	word-initial	untrained	12.**Th**at boy is my friend.

F. Stimuli used for the Discourse Reading Task

Theo the Thief

Theo is a **th**ief. He **th**inks **th**at he can steal whatever he wants and get away wi**th** it. He is tall, **th**in and has a gold front too**th** and **th**ick black hair. Yesterday I saw him walking down a pa**th** wi**th** a ba**th**. The rain was teeming down and I could even hear **th**under. Suddenly, the ba**th** began to fill with water. It got heavier and heavier and **th**en it fell and trapped **Th**eo under it. Just **th**en, a policeman came from nowhere. "You're under arrest!" said the policeman. "No," said **Th**eo, "I'm under a ba**th**!"

Table 1. Means, standard deviations (SD), and median values (Mdn) of age, English proficiency test scores, musical experience, time devoted to studying English, and other exposure to English for each group (T vs NT), as well as Mann-Whitney *U* test results.

	T			NT				
	M	*SD*	*Mdn*	*M*	*SD*	*Mdn*	*U*	*p*
Age	23.72	5.55	22.00	24.72	4.99	26.00	252.50	.242
English proficiency test	15.04	4.49	14.00	15.16	4.45	14.00	309.50	.953
Musical experience	10.28	4.87	9.00	9.56	3.04	9.00	314.00	.977
Age of onset learning	11.08	4.08	10.00	10.08	3.29	10.00	325.00	.806
Years of learning	8.56	3.04	10.00	9.40	3.80	9.00	274.50	.465
Months of extracurricular English courses	10.84	23.54	0.00	1.20	2.06	0.00	396.00	.067
Months of study abroad	0.76	2.17	0.00	2.04	4.33	0.00	261.00	.182
Books read	2.16	4.71	0.00	2.04	4.42	0.00	323.50	.814
Movies in English watched per week	8.64	10.23	4.50	7.26	6.99	5.00	295.50	.740

Table 2. Results from the Linear Mixed Models of segment accuracy scores of the word-imitation task

	Fixed effects		Random effects	
			By participant	By item
	χ^2	*p*	*SD*	*SD*
Intercept	-	-	0.36	0.23
Condition	5.59	.018	-	-
Test	57.99	<.001	0.24	0.07
Condition × Test	5.83	.016	-	-

Table 3. Results from the Linear Mixed Models of segment accuracy scores of the sentence-imitation task

	Fixed effects		Random effects	
			By participant	By item
	χ^2	*p*	*SD*	*SD*
Intercept	-	-	0.35	0.14
Condition	0.35	.554	-	0.11
Test	25.19	<.001	0.24	0.2
Condition × Test	0.4	.526	-	-

Table 4. Results from the Linear Mixed Models of segment accuracy scores of the discourse reading task

	Fixed effects		Random effects	
			By participant	By item
	χ^2	*p*	*SD*	*SD*
Intercept	-	-	0.66	-
Condition	0.002	.962	-	-
Test	80.27	<.001	-	-
Condition × Test	3.89	.048	-	-

Figure 1

Figure 2

Figure 3

Figure 4

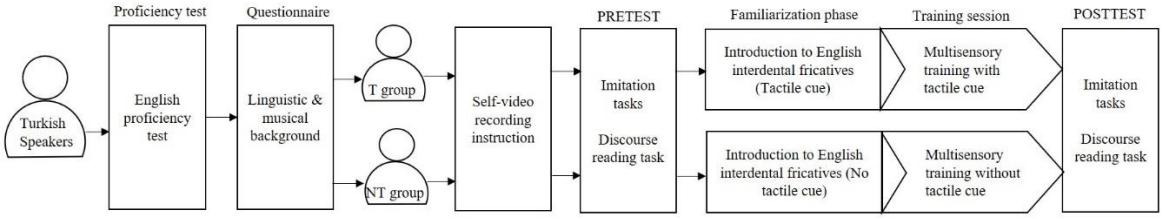


Figure 5

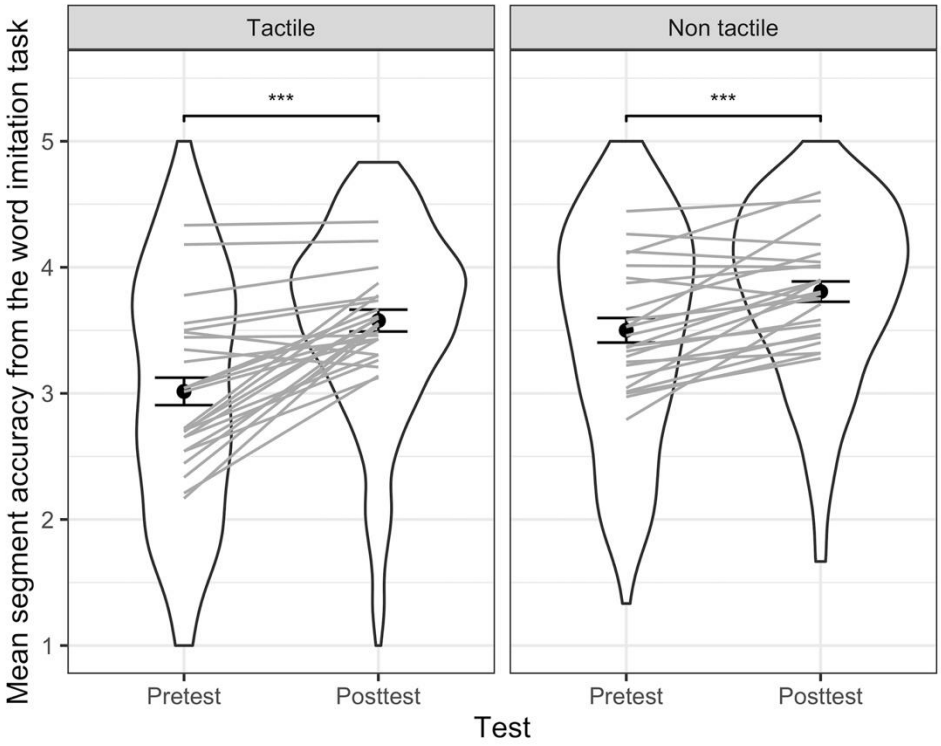


Figure 6

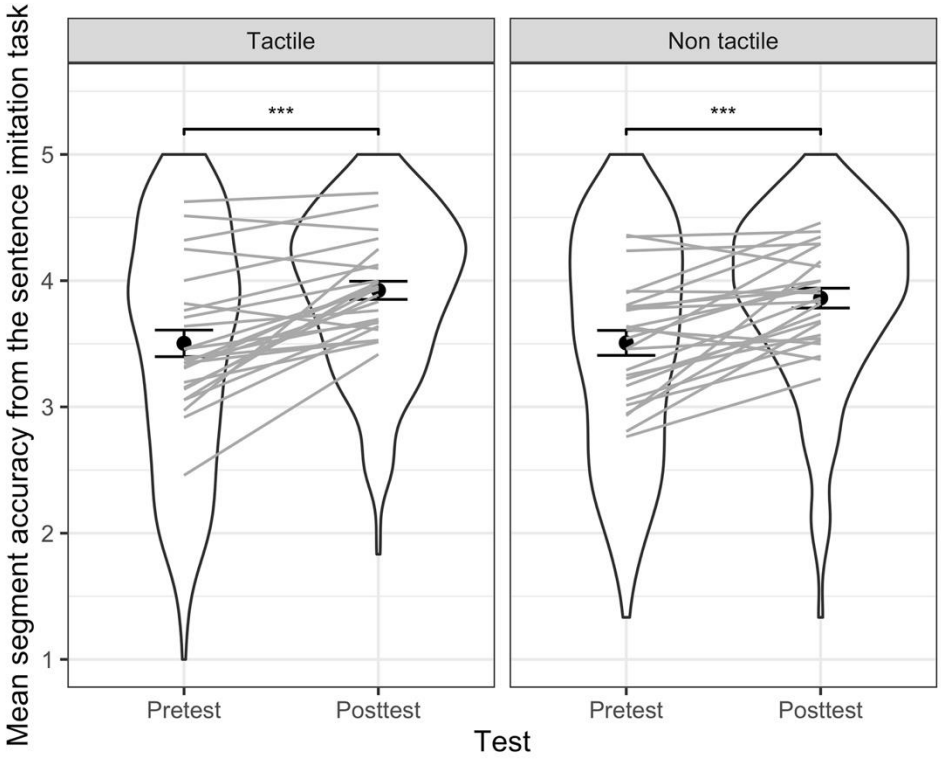
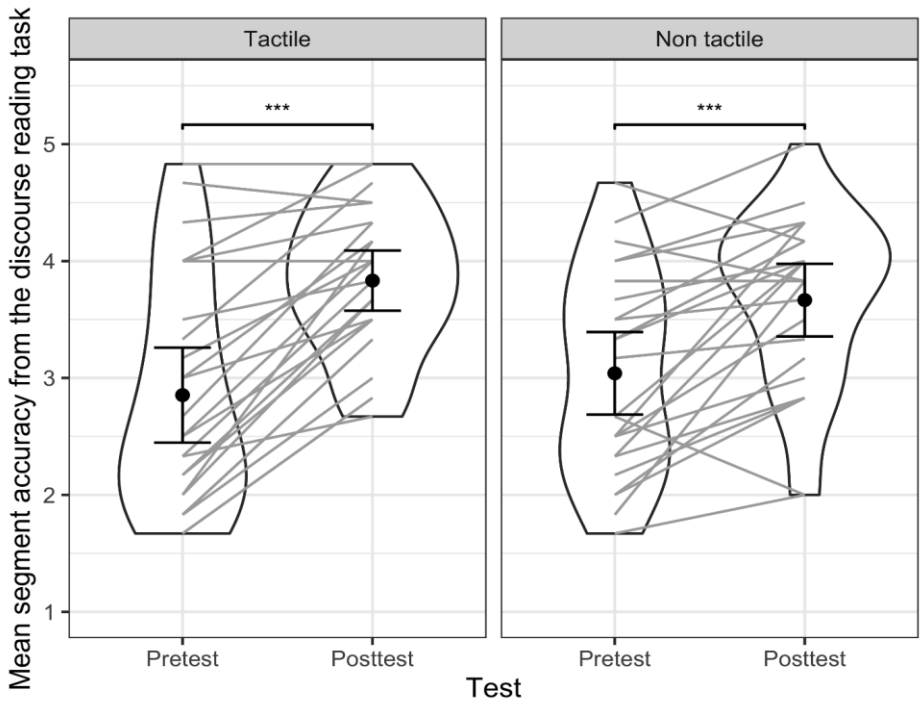


Figure 7



Captions in a list

Figure 1. Side views of instructor during articulation of /θ/ or /ð/. Left: while executing the tongue-touch gesture. Right: without the tongue-touch gesture.

Figure 2. Stills from training materials showing instructor as he utters one of the two phonemes being taught, in this case /θ/. Upper row: T (tactile) condition. Lower row: NT (non-tactile) condition.

Figure 3. Stills from training materials showing instructor as he utters one of the two phonemes being taught, in this case /ð/. Upper row: T (tactile) condition. Lower row: NT (non-tactile) condition.

Figure 4. Schematic diagram of the experimental procedure.

Figure 5. Violin plot of segment accuracy scores in the word imitation task across condition (T and NT) and test (pretest and posttest). The dots show the mean value, the error bars indicate 95% CI, the light lines show individual changes in mean score from pretest to posttest for each of the participant, and the asterisks mark significant contrasts.

Figure 6. Violin plot of segment accuracy scores in the sentence imitation task across condition (T and NT) and test (pretest and posttest). The dots show the mean value, the error bars indicate 95% CI, the light lines show individual changes in mean score from pretest to posttest for each of the participant, and the asterisks mark significant contrasts.

Figure 7. Violin plot of segment accuracy scores in the discourse reading task across condition (T and NT) and test (pretest and posttest). The dots show the mean value, the error bars indicate 95% CI, the light lines show individual changes in mean score from pretest to posttest for each of the participant, and the asterisks mark significant contrasts.