ASJU, 2023, 57 (1-2), 805-823
https://doi.org/10.1387/asju.25982

ISSN 0582-6152 – eISSN 2444-2992

Training ESL students to reproduce beat gestures in discourse leads to L2 pronunciation improvements[1]

Pilar Prieto*
ICREA, Institució Catalana de Recerca i Estudis Avançats,
Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra

Carmen Pérez-Vidal
Departament de Traducció i Ciències del Llenguatge, Universitat Pompeu Fabra

Olga Kushch
Facultad de Lenguas y Educación, Universidad Nebrija

Joan Borràs-Comes
Laboratori de Fonètica, Universitat de Barcelona

Daria Gluhareva
Department of Linguistics, Simon Fraser University

ABSTRACT: The main goal of the present study is to assess whether training foreign language students to reproduce natural beat gestures in discourse can trigger pronunciation gains. A total of 18 young adult Catalan learners of English with an intermediate proficiency level participated in a 15-minute discourse-based pronunciation training session. Participants were randomly as-

[1] We would like to thank Joan Carles Mora, Núria Esteve-Gibert and Alfonso Igualada for useful comments on an earlier version of this article. We are grateful to the students from Universitat Pompeu Fabra who participated in the experiment, and especially to Judith Llanes-Coromina for her assistance with the data collection. We also thank Irene Cañada for helping in the assessment of the gesture production training videos by participants. This study has been supported by the Ministerio de Ciencia, Innovación y Universidades [PID2021-123823NB-I00] and by the Agència de Gestió d'Ajuts Universitaris i de Recerca of the Generalitat de Catalunya [2021 SGR 00922].

**** Corresponding author:*** Pilar Prieto. Departament de Traducció i Ciències del Llenguatge. Despatx 53.720, Campus de Comunicació, ICREA - Universitat Pompeu Fabra. Poblenou, Roc Boronat, n.º 138 (08018, Barcelona). – Pilar.Prieto@upf.edu – https://orcid.org/0000-0001-8175-1081

How to cite: Prieto, Pilar; Kushch, Olga; Borràs-Comes, Joan; Gluhareva, Daria; Pérez-Vidal, Carmen (2023). «Training ESL students to reproduce beat gestures in discourse leads to L2 pronunciation improvements», *ASJU*, 57 (1-2), 805-823. (https://doi.org/10.1387/asju.25982).

Received: 2022-11-12; Accepted: 2023-03-24.

ISSN 0582-6152 - eISSN 2444-2992 / © UPV/EHU Press

signed to two groups. While one group was asked to simply repeat the instructor's multimodal responses to discourse prompts by focusing on speech, the other group was asked to repeat the utterances together with the natural beat gestures that the instructor was using. Before and after training, participants were recorded producing a discourse completion task and their speech was rated for accentedness. Results showed that participants who accompanied their verbal repetition with beat gestures during training significantly reduced their accentedness scores more than those who were asked to only repeat the utterances without reproducing the beat gestures. These results support recent findings that show the value of embodied prosodic training for pronunciation instruction.

KEYWORDS: Beat gestures; L2 pronunciation instruction; L2 training.

1. Introduction

1.1. The role of suprasegmental training for second language (L2) pronunciation

In the field of second language acquisition (SLA), pronunciation instruction has typically focused on finding ways to improve segmental and phonemic aspects of speech (see Derwing & Munro's 2015 book for a review). In the last few decades, however, a new line of research has highlighted the importance of suprasegmental instruction for reducing learners' overall accentedness and improving their comprehensibility. In a recent review article assessing evidence-based design principles for the teaching of pronunciation, Colantoni and colleagues (2021) emphasize the value of focusing on prosodic features for improving all dimensions of L2 speech, including intelligibility and comprehensibility. More specifically, a number of classroom studies have demonstrated that teaching suprasegmental components of a L2 can help learners improve their overall fluency and comprehensibility (e.g., Derwing, Munro & Wiebe 1998; Derwing & Rossiter 2003). For example, Derwing *et al.* (1998) confirmed that, after an 11-week English as a Second Language (ESL) course, learners who had received suprasegmental-based pronunciation instruction (also called *prosodic* instruction) achieved significantly better pronunciation scores in spontaneously-produced speech than learners who were exposed to only segmental pronunciation instruction, or those who received no pronunciation-specific instruction at all. While suprasegmental-based pronunciation instruction attended to speaking rate, intonation, rhythm and stress at the word and sentence levels (that is, to suprasegmentals), segmental pronunciation instruction was centered on training individual phonemes and performing discrimination exercises using minimal pairs. Similarly, Derwing & Rossiter (2003) compared the outcomes of three instructional methods, namely segmental, suprasegmental-based, and no explicitly focused pronunciation instruction, the latter serving as a control condition. In the two experimental groups, participants were exposed to 20 hours of pronunciation training over the course of 12 weeks, after which pre-and post-training recordings of the participants were evaluated by five judges who were native English speakers and ESL experts. Though none of the groups showed significant improvements in accentedness ratings, only the suprasegmental-based instruction group improved significantly in terms of comprehensibility and fluency.

While the aforementioned studies provide evidence for the importance of suprasegmental instruction in L2 pronunciation teaching, more research is needed to strengthen the existing findings and to assess the effectiveness of rhythm-based training proposals. In the context of ESL, an interesting rhythm-based pronunciation instruction scheme is Graham's (1978) "jazz chants" approach, which combines rhythmic repetition of English sentences or phrases with the use of kinesthetic actions such as clapping or finger-tapping. Yet to our knowledge little academic work has focused on comparing different suprasegmental training approaches. The large-scale survey of English pronunciation instruction practices across Europe (Henderson *et al.* 2015) confirmed that little emphasis is placed on teaching suprasegmental elements of the language.

In the present study, we focus on the value of using beat gestures, which are hand or arm movements used by speakers to highlight the rhythmic structure of their utterances, for L2 pronunciation learning. Though research has detected a close relationship between beat gestures and emerging L2 prosody (McCafferty 2006), there is still little research on whether these gestures can be used for L2 rhythm training and to promote improvements in L2 pronunciation. More specifically, to our knowledge no previous studies have investigated the potential beneficial effect of including the observation and/or production of beat gestures in L2 pronunciation instruction.

1.2. The role of hand gestures in L2 pronunciation learning

In the context of L2 pronunciation learning, different types of hand gestures have been shown to be helpful (see Baills 2022, for a recent review; see also Kushch *et al.* 2018 for an assessment of the role of hand gestures on second language novel word learning.. For example, one group of studies has explored the potential benefits of so-called pitch gestures (hand gestures that mimic or visually represent the tonal movements of speech), produced by the instructor, on the learning of L2 tones and intonation, with positive results (e.g., Yuan *et al.* 2017; Morett & Chang 2015; Baills *et al.* 2018). However, little is currently known about the value of other types of gestures, such as gestures.

Beat gestures have been typically associated with prosodically prominent positions in discourse and have been shown to reinforce the viewer/listener's perceptions of prominence (Krahmer & Swerts 2007). Beat gestures may thus be useful to highlight the rhythmic patterns of an L2, leading to improvements in a learner's performance in terms of prosody. To our knowledge, only two studies have investigated the value of rhythmic beat gestures in discourse. In one, a within-subject study, Gluhareva & Prieto (2017) asked 20 Catalan learners of English to watch an English instructor produce a set of responses framed in a discourse situation. Half of the utterances were accompanied by beat gestures, while the others were produced without. When tested using the same contextual prompts, participants who had been exposed to the beat gesture condition were rated as less accented than those who had not. To further explore the benefits of producing beat gestures, Llanes-Coromina *et al.* (2018) encouraged adolescent low-intermediate-level Catalan learners of English to intentionally produce beat gestures during an oral reading task. The authors found that these participants obtained greater improvement in terms of accentedness, com-

https://doi.org/10.1387/asju.25982

prehensibility, and fluency in an oral reading task compared to participants who had not been instructed to move their hands while reading. Importantly, however, the first of these studies was based exclusively on the observation of beat gestures by participants, and the second involved having participants produce beat gestures, suggesting that further work comparing the effectiveness of these two modalities would be of interest.

1.3. Effects of self-performing *vs.* observing gestures

In the gesture literature, there is substantial evidence that for general learning processes producing gestures is more effective in some contexts than merely observing them (Beilock & Goldin-Meadow 2010; Goldin-Meadow 2014; Goldin-Meadow, Cook & Mitchell 2009). For example, Beilock & Goldin-Meadow (2010) carried out two experiments that involved solving and explaining the Tower of Hanoi task with gestures. Gesturing during the task had beneficial effects on later speech performance, presumably because gesturing helped participants to change their thought processes by adding action information to their mental representations of the task. The results support the hypothesis that producing gestures can change the gesturer's mental representations and in the process contribute positively to task solving.

The value of involving the sensorimotor system during learning is grounded in the embodied cognition paradigm (e.g., Barsalou 2008), which claims that the physical body plays a key role in shaping our cognition. Some studies have yielded evidence that language and body movements are supported by the same neural substrates (Glenberg & Kaschak 2002; Pulvermüller *et al.* 2005). The cognitive system uses the body as an external informational structure that supports internal representations (Barsalou *et al.* 2003; Niedenthal *et al.* 2005). From this perspective, the production of gestures is considered an important form of embodiment in language, closely linked to representations in memory.

Embodied cognition has important implications for education. According to a recent review of the research by Jusslin *et al.* (2022), the notion of embodiment might well have applications in language learning and teaching because of its potential to enhance learner attention and creativity, suggesting that it is a field ripe for exploration in future research. In the context of vocabulary learning, for example, neurophysiological studies provide evidence that self-performing a gesture when learning verbal information helps a learner to construct sensorimotor networks that represent and store the words, whether in their first (Masumoto *et al.* 2006) or second language (Macedonia *et al.* 2011).

To our knowledge, only two studies have assessed the value of both observing and producing gestures in the context of phonological learning. In the first, Vilà-Giménez & Prieto (2020) found that children trained to produce beat gestures while retelling a narrative showed greater fluency in their speech than children who were simply asked to retell the story without any instructions regarding the use of gestures. In the second study, Bailly *et al.* (2018) found that for learners of Mandarin Chinese both observing and producing gestures visually depicting tones favored tone discrimination as well as word identification and recognition. In light of these previ-

ous results, we hypothesize that the potential benefits of producing beat gestures for phonetic training will be greater than merely observing them.

1.4. Goal of the study

The goal of the present study is to investigate whether participants achieve greater gains in approximating native-like pronunciation of an L2 if they are instructed to observe an instructor using beat gestures to mark prosodic prominences while speaking and subsequently imitate what they see by producing beat gestures themselves, in comparison to only repeating the instructor's utterances without imitating her gestures. First, following Gluhareva & Prieto's (2017) study, we hypothesize that the active use of visible and natural beat gestures working together with prosody can facilitate the learning of L2 pronunciation, measured in terms of accentedness. Second, following the embodied cognition paradigm, we hypothesize that producing such gestures will be more beneficial for pronunciation improvement than only observing them being performed by a native-speaking instructor.

2. Methods

The study consists of a between-subject training paradigm with a pre-test/post-test design.

2.1. Participants

Eighteen native Catalan-Spanish bilinguals (14 female and 4 male) (mean age = 21.5 years, SD = 3.327) volunteered to participate in the study. All were first-year students in the Universitat Pompeu Fabra's Faculty of Translation and Language Sciences. First, after being briefed on the goals and nature of the experiment, participants provided written informed consent for their data to be used anonymously for research purposes. They then completed a language questionnaire. All reported having an upper-intermediate B2[2] level of English. Participants also reported using Catalan (as opposed to Spanish) for an average of 75.7% (SD = 8.5) of their daily communication needs. Participants received five euros as remuneration for their participation in the experiment, which lasted approximately 30 minutes in total.

2.2. Materials

Pre-test and post-test discourse prompts. Each of the prompts consisted of a photo depicting an everyday situation which the participants might face if they lived abroad in an English-speaking country, accompanied by a short description of the situation and instructions indicating the speech act they were expected to perform. For example, one of the prompts showed a group of tourists in New York City trying to

[2] Prior to admission in any degree program related to translating or applied linguistics at the Universitat Pompeu Fabra, students are required to have a command of English at least equivalent to a B2 level according to the Common European Framework of Reference for Languages.

https://doi.org/10.1387/asju.25982

find their way with the help of a map. The accompanying text said: "You are trying to find Central Park. You ask a local person for directions". Some of the discourse prompts are provided in Appendix 1. The natural discourse situations expressed in the prompts are based on a conception of SLA which is strongly based on the communicative approach (see Pérez-Vidal 2009 for a review). Ten such discourse prompt situations were used for the pre-test, adapted from Gluhareva & Prieto (2017). Exactly the same prompts were used for the post-test that followed the training session, but ten completely new items were also included. Two further prompts were created for use in order to familiarize participants with the training procedures.

Training videos

The same discourse prompts were used in the training session, but in this case they were also accompanied by a set of video clips showing a native speaker of American English giving appropriate responses to the prompt situations while using beat gestures to mark the main prosodic component of her speech. For this study we adopted the set of training videos used in Gluhareva & Prieto (2017), as presented in Appendix 1. The instructors' beat gestures consisted of simple up-down open-hand palm-up movements (see Figure 1). In all ten training videos, all of the nuclear pitch accents received full beat gestures, while some non-nuclear stressed syllables were marked with less forceful beat gestures. However, not all stressed syllables were accompanied by beat gestures, because as pointed out by Gluhareva & Prieto (2017), this would have appeared unnatural; thus, the instructor only executed beat gestures when uttering the words with the heaviest semantic weight (the script used by the speaker is available in Appendix 2, with prosodic prominences indicated).

Figure 1

Still image from a training video showing the instructor making a palm-up beat gesture

2.3. Experimental setup and procedure

Participants were tested and trained individually using a laptop computer at Universitat Pompeu Fabra's Language Laboratory. First, participants were randomly assigned to one of two groups, each consisting of nine members, which we will hereafter denote the 'Beat Observation' group and 'Beat Production' group. The experiment consisted of three elements in sequence, a pre-test, a training session, and a post-test. The experimental design is schematically represented in Figure 2.

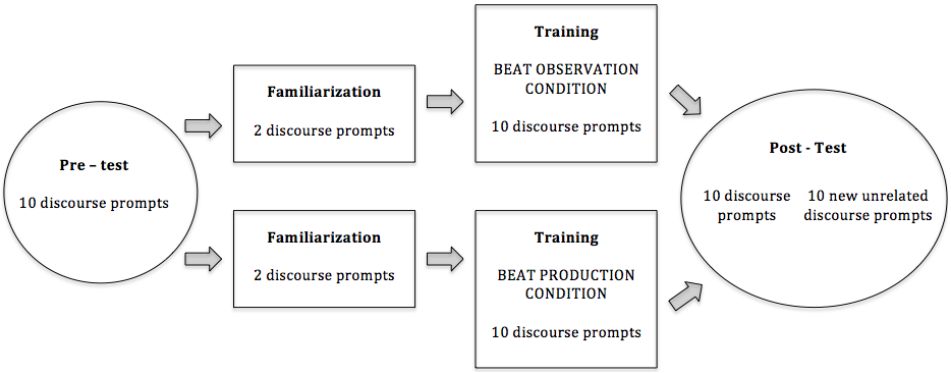


Figure 2

Overall experimental procedure

Pre-test

Participants carried out the pre-test alone in a quiet classroom using a laptop with a PowerPoint presentation which they controlled by clicking a mouse button. In each slide of the presentation, participants saw an image depicting a situation and simultaneously read the short description and instructions telling them what speech act they were to perform (e.g., ask for directions, introduce themselves, ask for the time, etc.) prompted by a blank speech bubble. In order to elicit natural speech and to avoid having the participants read off the screen while producing the responses, a black screen then appeared and the participants' response to each situation was audio recorded, using a digital voice recorder. The pre-test took roughly five minutes.

Training session

After completing the pre-test, participants were asked to start the training individually. The instructions differed depending on the experimental group to which they had been assigned. In order to familiarize participants with the training procedure, the experimenter presented two discourse prompts similar to those they had seen in the pre-test, but now each followed by videos of the native speaker modeling the appropriate response. The goal of this familiarization activity, which was not recorded, was to allow the experimenter to confirm that the participants fully under-

https://doi.org/10.1387/asju.25982

stood what was expected of them. After verifying that the participants understood the instructions for the task, the experimenter left the room because it was felt that the presence of the experimenter might inhibit participant performance. The participants proceeded to view the training video showing the native speaker modeling responses to the ten discourse prompts participants had seen in the pre-test. Prior to each modeling, they were shown the text description and instructions, as well as the photo with the speech bubble. At no time were they shown the text being spoken by the speaker in the video.

Throughout the training session, which took about 15 minutes altogether, participants were video-recorded with a Nikon d7000 camera, which was set up facing the participant at a distance of about two meters. A total of 18 video recordings were obtained from the training session (one per participant). These video recordings were not intended to provide data for analysis; rather, they were viewed after the session merely to ensure that participants had performed as expected. This was of particular concern in the case of participants belonging to the Beat Production group. In order to check that they had produced beat gestures that were appropriate and performed naturally with respect to form and rhythmic pattern, the video recordings were assessed by a research assistant not otherwise involved in the study. The rater judged participants' gesture performance using a Likert scale (1-bad performance, 5-good performance). The gestures produced by all nine participants received moderately high scores (M = 3.73, SD = 0.83), validating the inclusion of their pre-test and post-test data in the subsequent analysis. Figure 3 displays still images of the participants taken during the training session, with participants in the Beat Observation training group in the top row and participants in the Beat Production training group in the bottom row. The full training procedure took about 20 minutes.

Figure 3

Still images taken during the training session
(top row Beat Observation, bottom row Beat Production)

Post-test

Following the training session, the participants were given a five-minute break. They then proceeded to carry out the post-test, which was identical in procedure to the pre-test, but which included ten new discourse prompt items. Because participants had been exposed to the pre-test items not only in the pre-test but during the training session, including completely new prompts allowed us to control for any learning effect and detect any overall improvement in accentedness. As in the pre-test, all participant responses were audio-recorded. It was this audio data that was used to test our research hypothesis.

Ratings

Accentedness (i.e., deviation from native speaker pronunciation) was chosen as the target measure of listeners' global speech perception, in line with previous work by Gluhareva & Prieto (2017) and Llanes-Coromina *et al.* (2018). It was also thought that, for this data, accentedness may be a more sensitive measure than comprehensibility, given that the latter may be subject to ceiling effects when listeners assess short phrases produced by intermediate-advanced L2 speakers. All of the participants' spoken output from both the pre-test and post-test phases was rated for degree of accentedness by five native speakers of American English, four females and one male (mean age = 26; SD = 2.3), all residents of Barcelona. At the time of the assessment, all raters reported having normal hearing. Before they began, the raters received a brief training course on the rating procedure (following Gluhareva & Prieto 2017). Each rater evaluated a total of 540 recordings (18 participants × [10 pre-test recordings + 20 post-test recordings]) on a nine-point accentedness scale, from '1' (Native/No foreign accent) to '9' (Very strong foreign accent). The recordings were embedded in an online survey and appeared in random order (see Figure 4). Given the amount of time required to rate all the recordings, the full survey was broken into five parts. The raters reported that each part of the survey took them approximately 60 minutes to complete, for a total of 5 hours.

https://doi.org/10.1387/asju.25982

Figure 4

Sample item page from the online rating survey

Inter-rater reliability

Inter-rater reliability in the assessment of Accentedness across the five raters was assessed using an intra-class correlation (ICC) analysis for each pre-test and post-test. This yielded a Cronbach's Alpha score of .73, which surpasses the generally accepted measure of .7, thus indicating good inter-rater reliability (Larson-Hall 2010).

3. Results

A total of 2,700 tokens (5 raters · 3 tests · 10 situational prompts · 18 participants) were submitted to a Generalized Linear Mixed Model (GLMM) using IBM SPSS Statistics 24. The dependent variable was Rating (continuous). The fixed factors were Condition (two levels: Beat Production, Beat Observation), Test (three levels: pre-test, post-test trained item, post-test untrained item), as well as their interaction. A random intercept was set for Rater, with a random slope defined for both Participant and Situation.

The GLMM revealed significant results for the three fixed effects. The main effect of Condition ($F(1, 2694) = 7.935$, $p = .005$) indicated that the Beat Production condition received lower accentedness ratings than the Beat Observation condition ($\beta = .119$, $p = .005$). The main effect of Test ($F(2, 2694) = 12.661$, $p < .001$) indicated that perceived accentedness was lower in post-test trained items compared with either pre-test ($\beta = .523$, $p < .001$) or post-test new items ($\beta = .480$, $p < .001$), whereas no statistical difference was found between pre-test and post-test new items ($\beta = .043$, $p = .263$).

Finally, the interaction between CONDITION and TEST ($F(2, 2694) = 6.675$, $p = .001$) can be interpreted in two different ways. The first interpretation is that there was a significant effect of CONDITION, such that the Beat Production condition received lower accentedness ratings than the Beat Observation condition in post-test trained items ($\beta = .204$, $p = .001$), but not in pre-test ($\beta = .104$, $p = .106$) or post-test new items ($\beta = .049$, $p = .355$). The second interpretation is that though the effects of CONDITION were similar in each TEST, there was a size effect difference. The accentedness difference between pre-test and post-test was favorable to the Beat Production group. The accentedness distance between pre-test and post-test old items was stronger in the Beat Production condition ($\beta = .573$, $p < .001$) than in the Beat Observation condition ($\beta = .473$, $p = .003$); and similarly, the difference between post-test old items and new items was larger for the Beat Production condition ($\beta = .558$, $p < .001$) than for the Beat Observation condition ($\beta = .402$, $p = .003$); meanwhile, no significant differences were found between the pre-test and the post-test new items either in the Beat Production condition ($\beta = .016$, $p = .681$) or in the Beat Observation condition ($\beta = .071$, $p = .241$). Figure 5 shows the mean accentedness ratings obtained for the pre-test, post-test trained items and post-test untrained items as a function of the Beat Production and Beat Observation conditions.

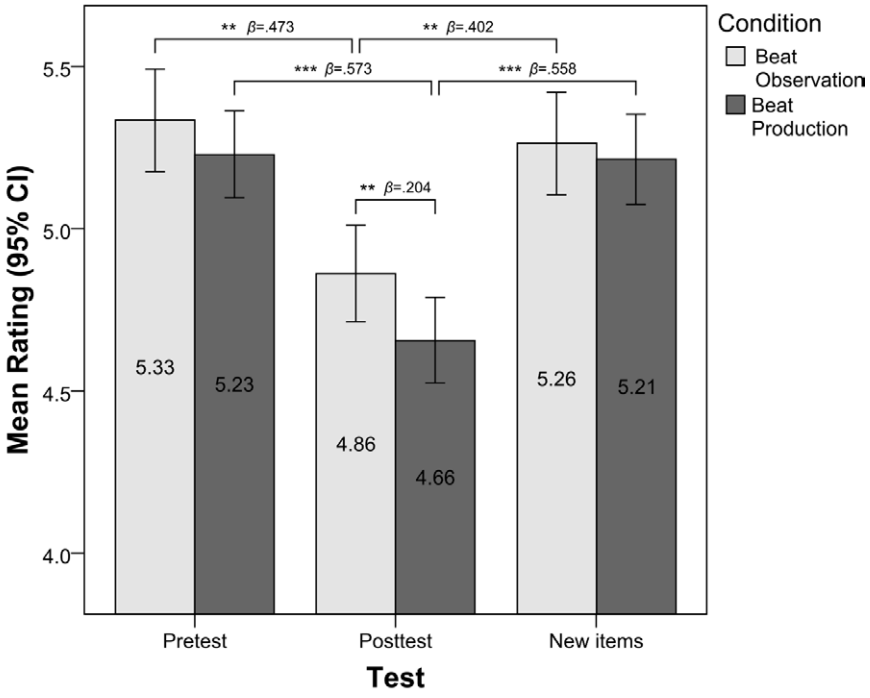


Figure 5

Mean accentedness rating scores obtained from the pre-test, post-test trained items, and post-test new items) as a function of the beat production and beat observation conditions. Note that lower scores indicate less accented speech

https://doi.org/10.1387/asju.25982

On the whole, the results confirm that the group of participants who were trained to accompany their utterances with beat gestures significantly reduced their accentedness when uttering trained items (post-test old items) more than the group of participants who were trained to merely observe the gestures.

4. Discussion and conclusions

The main goal of the present study was to investigate whether training participants to imitate the beat gestures produced by an English instructor would lead to higher pronunciation gains than training participants to merely observe these beat gestures. To this end, 18 young adult Catalan learners of English were exposed to a short training video in which a native speaker of English produced short utterances, accompanying her speech with rhythmic gestures synchronized with prosodic prominences. Participants in the experiment were randomly assigned to two groups, one of which was instructed to repeat the utterances after the instructor and imitate her use of beat gestures while the other group was instructed to simply repeat her utterances. Our analysis of participants' non-native accentedness before and after training showed that while both training methods reduced accentedness, being trained to produce beat gestures resulted in significantly better outcomes in pronunciation, at least for utterances with which the participants had been trained, than being trained to merely observe the performance of beat gestures.

The results of the present study complement previous studies on the potential utility of hand gestures for pronunciation instruction. It is important to emphasize that the beneficial effect of producing beat gestures was obtained after a relatively short 15-minute training session. This pattern of results is similar to that reported for the use of pitch gestures (gestures that mimic tone and other intonational movements in speech). Various studies have demonstrated that a relatively short training session involving pitch gestures can facilitate the production of tones by learners of a tonal language (Morett & Chang 2015; Yuan *et al.* 2017; Baills *et al.* 2018). By contrast, most other studies on L2 pronunciation learning (Munro & Derwing 1995; Derwing & Munro 1997; Gordon *et al.* 2013; and others) have utilized more lengthy training designs and included multiple training sessions (but see Gluhareva & Prieto 2017).

We believe that the beneficial role we report here of the use of beat gestures in natural discourse to improve pronunciation is related to the fact that these gestures highlight in a visual way the suprasegmental prosodic properties of the target language. Note that we have seen that actively producing gestures has a greater effect than merely observing them, findings which are in line with the embodied cognition paradigm and with research that demonstrates that the production of gestures facilitates learning mental tasks of various sorts more than the observation of gestures (see, among others, Goldin-Meadow *et al.* 2012; Goldin-Meadow 2014; Goldin-Meadow *et al.* 2009; Masumoto *et al.* 2006; Macedonia *et al.* 2011).

Some limitations of this study need to be acknowledged. First, a sample size of 18 participants is rather small and thus precludes any broad conclusions. Second, the study involved a training session lasting only 15 minutes, and post-test recordings of participants were made only five minutes later, so that we are in no position to make any claims with regard to the long-term retention of training effects. Also, the signifi-

cant beneficial effect of producing beat gestures we detected was not transferred to new non-trained items that were included in the post-test, thus no generalization of the training effect was observed. Finally, it is also important to recognize that the present training paradigm involved a relatively homogenous group of participants in terms of age, first language, and level of English; more research will be needed to assess how the use of beat gestures might affect different profiles and combinations of learners.

All in all, the findings from the present study seem to suggest that actively using rhythmic hand gestures to highlight the suprasegmental elements of a foreign language is a valuable technique that can prove useful in the context of second language pronunciation instruction. In general, embodied methodologies that involve the active production or imitation of body actions may be more effective for pronunciation learning than more passive classroom practices that are limited to observation and speech. In this regard, further research should continue to assess the value of embodied prosodic-based training for improving pronunciation.

References

Bailles, Florence. 2022. *Using the hands to embody prosody boosts phonological learning in a foreign language*. Barcelona: Universitat Pompeu Fabra doctoral dissertation. http://hdl.handle.net/10803/673777 (28/12/2023).

Bailles, Florence, Nerea Suárez-González, Santiago González-Fuente & Pilar Prieto. 2018. Observing and producing pitch gestures facilitates the learning of Mandarin Chinese tones and words. *Studies in Second Language Acquisition* 41(1). 35-58.

Barsalou, Lawrence W. 2008. Grounded cognition. *Annual Review of Psychology* 59. 617-645.

Barsalou, Lawrence W., Kyle W. Simmons, Aron K. Barbey & Cristine D. Wilson. 2003. Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences 7*(2). 84-91.

Beilock, Seian L. & Susan Goldin-Meadow. 2010. Gesture changes thought by grounding it in action. *Psychological Science* 21(11). 1605-1610.

Colantoni, Laura, Paola Escudero, Victoria Marrero-Aguiar & Jeffrey Steele. 2021. Evidence-based design principles for Spanish pronunciation teaching. *Frontiers in Communication 6*. Article 639889.

Derwing, Tracey M. & Murray J. Munro. 1997. Accent, intelligibility, and comprehensibility. *Studies in Second Language Acquisition* 19(1). 1-16.

Derwing, Tracey M. & Murray J. Munro. 2015. *Pronunciation Fundamentals: Evidence-based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins

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

Derwing, Tracey M., Murray J. Munro & Grace Wiebe. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48(3). 393-410.

Glenberg, Arthur M. & Michael P. Kaschak. 2002. Grounding language in action. *Psychonomic Bulletin and Review* 9. 558-565.

Gluhareva, Daria & Pilar Prieto. 2017. Training with rhythmic beat gestures favors L2 pronunciation in discourse-demanding situations. *Language Teaching Research* 21(5). 609-631.

https://doi.org/10.1387/asju.25982

Goldin-Meadow, Susan. 2014. Widening the lens: What the manual modality reveals about language, learning and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369(1651). Article 20130295.

Goldin-Meadow, Susan, Susan Wagner Cook & Zachary A. Mitchell. 2009. Gesturing gives children new ideas about math. *Psychological Science* 20(3). 267-272.

Goldin-Meadow, Susan, Susan Levine, Elena Zinchenko, Terina Yip, Naureen Hemani & Laiah Factor. 2012. Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science* 15(6). 876-884.

Gordon, Joshua, Isabelle Darcy & Doreen Ewert. 2013. Pronunciation teaching and learning: Phonetic instruction in the L2 classroom. In John Levis & Kimberly LeVelle (eds.), *Pronunciation and assessment: Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference 2012*, 194-206. Ames, IA: Iowa State University.

Graham, Carolyn. 1978. *Jazz chants: Rhythms of American English for students of English as a second language*. Oxford: Oxford University Press.

Henderson, Alice, Lesley Curnick, Dan Frost, Alexander Kautzsch, Anastazija Kirkova-Naskova, Elina D. Levey & Ewa Waniek-Klimczak. 2015. The English Pronunciation Teaching in Europe Survey: Factors inside and outside the classroom. In Jose A. Mompean, & Jonás Fouz-González (eds.), *Investigating English Pronunciation. Trends and Directions*, 260-291. London: Palgrave Macmillan. https://doi.org/10.1057/9781137509437

Jusslin, Sofia, Kaisa Korpinen, Niina Lilja, Rose Martin, Jahanna Lehtinen-Schnabel & Eeva Anttila. 2022. Embodied learning and teaching approaches in language education: A mixed studies review. *Educational Research Review* 37. Article 100480.

Krahmer, Emiel & Marc Swerts. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language* 57(3). 396-414.

Kushch, Olga, Alfonso Igualada & Pilar Prieto. 2018. Prominence in speech and gesture favor second language novel word learning. *Language, Cognition and Neuroscience 33*(8). 992-1004.

Larson-Hall, Jenifer. 2010. *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York: Routlege.

Llanes-Coromina, Judith, Pilar Prieto & Patrick L. Rohrer. 2018. Brief training with rhythmic beat gestures helps L2 pronunciation in a reading aloud task. In Katarzyna Klessa, Jolanta Bachan, Agnieszka Wagner, Maciej Karpiński & Daniel Śledziński (eds.), *Proceedings of the 9th Speech Prosody Conference*, 498-502. Poznań: International Speech Communication Association.

Macedonia, Manuela, Karsten Müller & Angela D. Friederici. 2011. The impact of iconic gestures on foreign language word learning and its neural substrate. *Human Brain Mapping* 32(6). 982-998.

Masumoto, Kouhei, Masahiko Yamaguchi, Kouichi Sutani, Satoru Tsuneto, Ayako Fujita & Mitsuo Tonoike. 2006. Reactivation of physical motor information in the memory of actionevents. *Brain Research* 1101(1). 102-109.

McCafferty, Steven G. 2006. Gesture and the materialization of second language prosody. *IRAL*: *International Review of Applied Linguistics in Language Teaching* 44(2). 197-209.

Morett, Laura M. & Li-Yun Chang. 2015. Emphasising sound and meaning: pitch gestures enhance Mandarin lexical tone acquisition. *Language, Cognition and Neuroscience* 98(3). 773-794.

Munro, Murray J. & Tracey M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1). 73-97.

Niedenthal, Paula M., Lawrence W. Barsalou, Piotr Winkielman, Silvia Krauth-Gruber & François Ric. 2005. Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review* 9(3). 184-211.

Pérez-Vidal, Carmen. 2009. The integration of content and language in the classroom: A European approach to education (The second time around). In Emma Dafouz, & Michele Guerrini (eds.), *CLIL Across Educational levels: Experiences from primary, secondary and tertiary contexts*, 25-40. Madrid: Santillana & Richmond Publishing.

Pulvermüller, Friedemann, Olaf Hauk, Vadim V. Nikulin & Risto J. Ilmoniemi. 2005. Functional links between motor and language systems. *European Journal of Neuroscience* 21(3). 793-797.

Vilà-Giménez, Ingrid & Pilar Prieto. 2020. Encouraging kids to beat: Children's beat gesture production boosts their narrative performance. *Developmental Science* 23(6). Article e12967.

Yuan, Chenjie, Santiago González-Fuente, Florence Baills & Pilar Prieto. 2017. Observing pitch gestures favors the learning of Spanish intonation by Mandarin speakers. *Studies in Second Language Acquisition* 41(1). 5-32.

Appendix 1

Discourse prompt materials

Familiarization items

You are at your classmate's birthday party. You meet her boyfriend for the first time and introduce yourself to him.

You are trying to find Central Park. You ask a stranger for directions.

https://doi.org/10.1387/asju.25982

Experimental items (used in pre-test, post-test, and training session)

You are in a restaurant and would like to order a steak with French fries and a glass of reed wine

You are in the metro and would like to ask a stranger for the time.

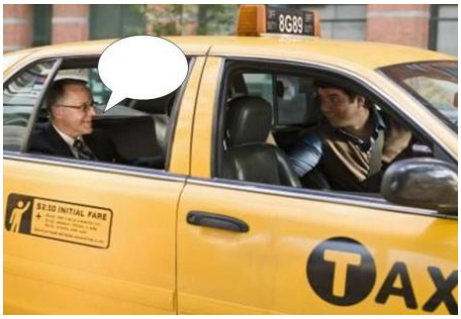
You are at the market. You want to ask the price of the necklace and ask if you can get it for $5.

You arrive at the airport in New York. You realize that your luggage is lost. You as, an airport employee for help.

You are at the pharmacy. You would like to tell the pharmacist that you have a sore throat and a fever and ask her to prescribe something for you.

You are trying to find an apartment in your new city. You want to ask the agent if this apartment gets a lot of light in the mornings.

You call a pizzeria. You would like to place an order for delivery—two large pizzas with cheese and pepperoni.



You get into a taxi. You would like to ask the driver to take you to the airport as fast as he can, because you are running late for your flight.

You are in a lecture at the university. You didn't hear what the professor just said and would like to ask your friend to repeat it for you.

You are in a clothing store. You would like to tell the clerk that you are looking for this shirt in a bigger size, and ask her if they have it in the back of the store.

Items used only in the post-test

You are at your university's Student Administration Centre. You would like to tell the secretary that you applied for a new student ID card last week and ask her if it is ready yet.

You go to the computer repair shop. You would like to tell the technician that your computer has been running slowly and ask him to figure out what the problem is.

https://doi.org/10.1387/asju.25982

You go to the phone store. You ask the clerk to show you their newest phone model.

You are at the bank. You would like to ask the teller how to apply for a new student bank account and what documents you need to provide

You go to the cinema. You want to ask the clerk if there is a discount for students.

You new roommate has been leaving her dirty dishes in the sink. You want to ask her to clean up after herself.

You go to your professor's office. You want to ask him why you got question number 7 wrong on the test.

You see your classmate at the library. You ask her if she wants to study for the Economics test together.

The roof in your apartment is leaking. You call the repairman and ask him when is the earliest he can come.

You go to a gym in your new neighborhood. You ask the employee how much a new membership costs, and if the gyms is open late on Sundays.

Appendix 2

Orthographic transcription of familiarization and training videos. The association of beat gestures to the target syllables is indicated by capital letters (highly prominent beats) or by underlined text (less prominent beats). Because video-recorded performances should appear plausibly naturalistic, not all stressed syllables were associated with gestures.

Familiarization items

— HI, *I'm MAya.* It's GREAT to *meet* you.
— ExCUSE *me*, we're *loo*king for Central PARK. Could you TELL us *where* to GO?

Training items

1. HI, *I'd like* to *place* an ORder for deLIvery. Two l*arge* PIzzas *with* CHEESE and peppeROni.
2. SORRY, *what* did the professor just SAY? I couldn't HEAR him.
3. How *much* is this NECKlace? Can I get it for *five* DOllars?
4. *Ex*CUSE me, *what* TIME is it?
5. *My* LUggage is LOST. *Could you* HELP *me*?
6. I'd *like* to get a STEAK with FRENCH fries, and a *glass* of *red* WINE, please.
7. I'm *lookin*g for this SHIRT in a *bigge*r SIZE. Could you *check* and SEE if you *have* it in the BACK?
8. Can you TAKE *me* to the AIRport? As *fast* as you CAN *please*. I'm LATE for my *flight.*
9. Does *this* aPARTment get a *lot of* LIGHT in the mornings?
10. I have a *sore* THROAT and a FEver. Could you presCRIbe *some*thing for *me*?

https://doi.org/10.1387/asju.25982