



Brief training with rhythmic beat gestures benefits L2 pronunciation

Daria Gluhareva

Màster: Lingüística Teòrica i Aplicada

Edició: 2014-2015

Directors: Dra. Pilar Prieto

Any de defensa: 2015

Col·lecció: Treballs de fi de màster

Programa oficial de postgrau
"Comunicació lingüística i mediació multilingüe"

Departament de Traducció i Ciències del Llenguatge

Acknowledgements

I would first like to express my sincere gratitude to Dr. Pilar Prieto, for her thoughtful feedback and guidance throughout this project. It would have been impossible to do this without her knowledge and support.

I am very grateful to Maya Raghavan for her help with recording, data collection, and the study as a whole. I would also like to thank the following GrEP members for their generous help on various aspects of the project: Joan Borràs-Comes and Olga Kushch for their assistance with statistical analyses, and Santiago González for his help with the online survey.

Additionally, thank you to my classmates in the MLTA program, for the many interesting and inspiring discussions that we have had, both in and out of class. I am also grateful to the many Pompeu Fabra students who have participated in this study, for their patience and cooperation.

And finally, a huge thank you to my friends and family for their support throughout this project:

To my raters (Yael Green, Rob McGillis, Vaughan Ollier, Craig Ryan, and Brian Salazar)—thank you for devoting so much of your time to this...I owe you all many dinners in the future.

To David Lechuga and Barrett Hedges, for their help in the pilot stage of this study, as well as for the many laughs throughout the year.

To my mom, for constantly offering to help.

And to Craig, for always being there for me.

Abstract

This study investigates the impact of beat gesture observation on the acquisition of native-like speech patterns in English by examining the effect of a brief training with or without beat gestures on participants' ratings of accentedness. In a within-subjects design, twenty Catalan-dominant undergraduate students learning English as a Foreign Language at an upper-intermediate level were shown a training video in which a native speaker of English gave short responses to 12 prompts. For half of the responses, the instructor used co-speech rhythmic beat gestures, while in the other half, she did not gesture at all. Participants' answers to the prompts were recorded before and after training and evaluated by five native speaker judges. The results demonstrated that while beat gesture training did not significantly improve the participants' accentedness ratings on the entire set of items, their performance on *initially difficult* items improved significantly as a result of beat gesture training.

Keywords: Second language acquisition, rhythm, suprasegmentals, accentedness, English language

Introduction

1.1. Segmental and suprasegmental influences on L2 pronunciation

In the field of second language acquisition, much has been written about possible ways to optimize learners' phonological abilities in their L2. Even though most classroom pronunciation training tends to center around segmental instruction (that is, one focused solely on individual sounds), several studies have highlighted the importance of suprasegmental instruction for improving learners' overall fluency and comprehensibility and reducing their foreign accent. In one of the earliest comparison studies on the subject, Derwing, Munro, and Wiebe (1998) compared the effectiveness of three conceptions of pronunciation pedagogy in improving ESL speakers' ratings of comprehensibility, accentedness, and fluency: segmental (instruction focused on individual sound contrasts), global¹ (which was said to address elements such as speaking rate, intonation, rhythm, projection, and word and sentence stress), and no pronunciation-specific instruction. The exact methods used for each learner group were determined by the respective instructors and participants' speech was sampled before and after an 11-week ESL course. While both the segmental and global training groups improved in comprehensibility and accentedness on a sentence-pronunciation task, only the global group showed significant improvement when spontaneous speech production was assessed, suggesting the importance of prosodic instruction in allowing learners to transfer their training to natural speech.

Similarly, Derwing and Rossiter (2003) compared three instructional methods: segmental (focusing on individual phonemes and discrimination of minimal pairs), global (emphasizing suprasegmentals such as word and sentence stress, intonation, rhythm, projection, and speech rate) and no specific pronunciation instruction, as a control. Participants in the experimental

¹ Several studies have used *global instruction* as an umbrella term to encompass a wide range of suprasegmental components of L2 speech.

groups received 20 hours of pronunciation instruction over the course of 12 weeks. Participants' pre- and post-training recordings were evaluated by five native speaker judges (professional ESL teachers), whose ratings showed that only the Global instruction improved significantly in terms of their comprehensibility and fluency. None of the three groups improved significantly on accentedness. Interestingly, a detailed error analysis showed that the Segmental group made significant gains in phonological accuracy, but this did not transfer over to their overall ratings.

More recently, Gordon, Darcy, and Ewert (2013) tested the effects of explicit phonetic instruction in the classroom for three groups of ESL learners, comparing segmental training (i.e., vowel distinctions) to suprasegmental instruction (focusing on stress, rhythm, linking, and reductions), as well as a group that did not receive any explicit pronunciation instruction. The instruction was given to the learners for three weeks, three days per week, and 25 minutes each day. The researchers found that only the group that received suprasegmental training significantly improved from pretest to posttest in their overall comprehensibility.

Behrman (2014) also compared the effects of segmental and prosody training on reducing speakers' foreign accent. Segmental training focused on the articulation of consonants, while prosodic training focused on four prosodic utterance levels: rise-fall pitch in one-word utterances, rising, falling, and rise-fall intonation in three-word utterances, informational and yes/no questions, and prosodic rhythm of longer utterances. The training in this case involved explicitly drawing the participants' attention to the prosodic target in question, and then coaching them in the actual production of the target. The study's approach differed from the previous literature because it used individualized instruction, and all participants underwent *both* segmental and prosodic training. Each participant received a minimum of five sessions per each type of instruction, and accuracy was assessed following each session. The results showed that a combination of both types of instruction produced the most successful outcomes in English

learners, and that neither type of training (segmental or prosodic) resulted in improvement in the other domain, suggesting the interconnectedness of these two aspects of speech, at least in terms of accent.

The aforementioned studies have provided strong evidence for the importance of a global (suprasegmental) approach in improving certain key aspects of L2 learners' speech. When segmental and suprasegmental features are compared on how they contribute to nonnative speakers' measures of foreign language competence, suprasegmental deviance tends to contribute more to accentedness, comprehensibility, and intelligibility than segmental deviance (e.g., Anderson-Hsieh, Johnson, and Koehler, 1992, Field, 2005, Edmunds, 2010, and Ulbrich, 2013). In a series of studies by Munro and Derwing (1995, 1997, and 1999), the researchers analyzed the relative contribution of segmental and suprasegmental features to native speakers' judgments of accentedness, intelligibility, and comprehensibility. As a whole, the results of the studies demonstrated that prosodic errors contributed more to L1 listeners' perception of accentedness and comprehensibility; ratings of intelligibility, on the other hand, were influenced more by deviations on the segmental level of speech (for example, by the incorrect usage of individual sounds).

Along this line of research, Kang (2010) analyzed the speech of 11 international teaching assistants using native speaker ratings, discovering that certain suprasegmental features explained much of the variance in ratings: pitch range, word stress measures, and mean length of pauses accounted for 41% of variance in accentedness ratings, while speech rate variables alone explained 35% of the variance in comprehensibility. Additionally, van Maastricht, Krahmer, and Swerts (2015) found that L1 speakers of Dutch were able to distinguish recordings of native speakers from non-native ones based on prosodic cues alone, even when segmental deviance was controlled for. Crucially, the researchers emphasized the importance of using recordings of

(semi-) spontaneous speech, instead of samples obtained from read-aloud tasks, as most previous studies on the topic have done². It has been previously shown that read speech differs significantly from spontaneous speech in terms of its prosodic characteristics (e.g., Swerts, Strangert & Heldner, 1996, Laan, 1997, and De Ruiter, 2015). Thus, recording participants' semi-spontaneous speech allowed the researchers to more closely approximate prosodic features that occur in real-life communication.

Despite the apparent importance of suprasegmental instruction, there exists little concrete evidence showing the superiority of one suprasegmental training method over another, as the previous studies have trained participants in a mixture of features using various teaching methods and techniques. Most suprasegmental training has involved explicit instruction and analysis of prosodic targets, and not much has been cited in terms of concrete, empirically-tested exercises or tasks that promote suprasegmental improvement in an L2. One of the main motivations behind this study stems from a desire to fill this gap in the research. Thus, in order to gain a better understanding of the phenomenon at hand, it seems important to address the different components of suprasegmental competence separately, notwithstanding their irrefutable interdependence. The present study will focus on the development of rhythm and the prosodic components that closely relate to it.

There is evidence pointing to the fact that above all other suprasegmental components, rhythm plays a particularly essential role in the processing of speech, with particular emphasis placed on the rhythm of L2 speakers (e.g., Adams, 1979; Faber, 1986; Anderson-Hsieh, Johnson & Koehler (1992); Tajima, Port & Dalby, 1997). White & Mattys (2007b) found that all three previously-identified rhythm metrics were significantly correlated with native English speakers'

² Non-spontaneous speech (obtained by means of read-aloud or repetition tasks) has been used by Anderson-Hsieh, Johnson & Koehler (1992), Derwing, Munro & Wiebe (1998), Munro & Derwing (1999), Field (2005), and Gordon, Darcy & Ewert (2013), among others.

ratings of foreign accent in L2 speech; one of these metrics in particular, VarcoV, was found to be an effective predictor of accent ratings, especially when coupled with speech rate. The rhythm of L2 speech, in other words, was shown to play a significant role in the extent to which a speaker is judged as being native.

Despite this, there does not seem to be a consensus in the field about how to explicitly teach L2 rhythm, and the subject is often neglected by teachers (Faber, 1986). Various pedagogical methods have been proposed-- for instance, Graham's *Jazz Chants*, which presents students with short, poem-like structures that enhance the rhythm of natural speech, was used by Derwing, Munro, and Wiebe (1998) and Derwing and Rossiter (2003). In this method, students are instructed to tap out beats with their fingers, which may aid them in noticing the target rhythm and following along with it. Besides this method, very little has been written about utilizing gesture as a tool for L2 rhythm training. Additionally, we have yet to see an independent assessment of the efficacy of these techniques, or an identification of the most crucial components of L2 suprasegmental training.

1.2. The role of gesture

In the literature on gesture, there is a growing body of evidence that shows the potential benefit of using gesture as a tool for language learning. Gullberg (2006) provides a comprehensive summary of reasons for further investigating the connection between gesture and second language acquisition, highlighting the fact that hand gestures may provide learners with additional input to aid comprehension and overall acquisition.

As evidence of this, Kelly et al. (2009) found that presenting novel Japanese words with congruent (matching) iconic hand gestures (for instance, presenting the Japanese word for *drink* while also mimicking the action of drinking) proved to be beneficial in helping participants to later recall the target vocabulary, as opposed to when words were presented with incongruent

gestures (e.g., presenting the word for *drink* while performing a *washing* motion with the hands), or no gestures at all. These findings lead the researchers to conclude that speech-congruent iconic gestures have a beneficial effect on learning foreign language vocabulary; additionally, it was found that these gestures do more than simply capturing the learner's attention (they are "more than mere hand waving"), because presenting words with *incongruent* gestures (non-matching) seemed to have a detrimental effect on later recall. Tellier (2008) found similar results in young children—gestures had a significant positive influence on French children's memorization of novel English words.

The gestures used by the aforementioned studies fall into the (relatively well-studied) category of iconic gestures-- ones that convey some semantic information about the word in question. The present study involves a different type of hand gesture-- beat gestures, which are usually characterized as up-and-down movements of the hand that convey temporal information, marking the prosody and rhythm of speech (McNeil, 1992). Hirata and Kelly (2010) investigated the potential effect of beats on learning a Japanese (non-native) sound contrast and found that participants in the speech-gesture condition did not perceive the contrast any better than those in the speech-alone condition. Similar mixed results were found by Hirata et al. (2014) regarding the effect of beat gestures on segmental-level learning—participants who were trained using hand gestures did show improvement in identifying the target vowel length contrast, but the researchers pointed out that the highest gains were seen in the audio-mouth (no gesture) training condition in Hirata and Kelly (2010).

Little is known about the effect of beat gestures on the acquisition of suprasegmental elements of language, although previous research has suggested a closely synchronous relationship between gesture and speech prominence in natural interactions (Loehr, 2012; Wagner et al., 2014). Biau and Soto-Faraco (2013) found that beat gestures play a significant

role in helping the listener regulate the parsing of a stream of speech, as well as focus his or her attention on the most relevant aspects of the information being conveyed. Additionally, further highlighting the role of beat gestures in native language processing, Kraemer and Swerts (2007) found that seeing a manual beat gesture on a word resulted in increased prominence perception of that word.

The effect of beat gestures on second language processing and acquisition remains underexplored. On the one hand, recent research has suggested that beat gestures do assist in word memorization: in So et al. (2012), beat gestures were shown to aid native word recall in adults, but not in children. More recently, Igualada et al. (2015) showed that children can, in fact, benefit from beat gestures in word memorization when the words are presented in a relevant context and serial sequential effects are controlled for. However, the effect that beat gestures may have on prosody acquisition is not as clear-- while McCafferty (2006) suggested a relationship between beat gestures and emerging second language prosody, no empirical study to our knowledge has systematically investigated the potential effect of gesture on the acquisition of suprasegmental elements, namely its potential role in teaching the rhythm of a second language. Based on the previous work highlighting the important role that beat gestures play in native language processing, it does not seem unreasonable to suggest that beats can be used as an effective aid in the processing, and eventual acquisition, of suprasegmental elements in a second language. Because little is known about the exact mechanisms behind the acquisition of L2 suprasegmentals in general (and rhythm, in particular), approaching the topic from the angle of beat gestures as a learning tool seems to be a promising place to start. Investigating the potential benefit of beat gestures in the acquisition of L2 rhythm may provide both researchers and language teachers a useful tool for aiding L2 learners in attaining comprehensible, less-accented speech in their second language.

1.3. Rhythm in Catalan and English

The present study will investigate the effects of short rhythm training on Catalan-dominant speakers learning English as a foreign language. Early work on linguistic rhythm established a binary system in which languages are classified as either stress-timed or syllable-timed (see Abercrombie, 1967, among others). Within this framework, so-called syllable-timed languages (for instance, Spanish and French) are characterized as having syllables of equal duration. Alternatively, in stress-timed languages such as English or Dutch, syllables have different lengths, yet the interval of time between *stressed* syllables is said to remain constant. While in more recent research, this categorical division has been questioned, if not refuted altogether, the two terms remain useful for discussing general properties of rhythm across languages.

Prieto et al. (2012) analyzed rhythmic patterns in Catalan, Spanish, and English, concluding that Catalan, which has often been cited as existing in an “intermediate” area between stress- and syllable-timed rhythm, demonstrates patterns more similar to those of Spanish. Despite the fact that Catalan possesses a more complex syllable structure than Spanish and demonstrates vowel reduction (qualities that have traditionally been cited as characteristics of stress-timed languages), a detailed analysis of seven rhythm metrics revealed that it is significantly different from English on all measures, and similar to Spanish in all indices except for two, leading the researchers to conclude that Catalan generally follows the pattern of syllable-timed languages.

For the purposes of this study, it is important to note how the language background of the participants affects their acquisition of English rhythm. Catalan and Spanish display similar rhythmic properties, while English seems to place itself firmly on the opposite side of the supposed continuum between stress- and syllable-timed languages. This marked difference

between the two classes of rhythm may pose an additional challenge for Catalan/Spanish bilinguals in acquiring a native-like level of spoken English.

1.4. The goal of the study

The present study aims to investigate the potential benefit that rhythm training (and more specifically, training with beat gestures) may have on the development of less-accented speech in a second language.

2. Method

2.1. Overall method

The study utilizes a training paradigm, before and after which participants' speech is recorded and assessed by native speakers. The training items are presented in one of two conditions (beat or no beat), and the training outcomes for these two conditions will be compared.

2.2. Participants

Participants ($N = 20$; 14 females and 6 males) were recruited from a larger pool of undergraduate students studying either Translation and Interpreting or Applied Languages at Universitat Pompeu Fabra in Barcelona. Their age ranged from 18 to 33 ($M = 19.3$; $SD = 3.31$). Participants were Catalan-dominant, and also bilingual in Spanish, reporting a mean daily usage of Catalan of 68.25% ($SD = 22.2\%$), with an upper-intermediate level of English³. All of the participants had submitted written consent prior to their participation in the study. The participants were paid 5 euros for participating in experiment, which took approximately 30 minutes.

2.3. Training Materials

³ Students in the Translation and Interpreting and Applied Languages degrees at UPF are required to have at least a B1 level of English (according to the Common European Framework of Reference for Languages) prior to beginning the program. The majority (18) of the participants in the present study reported having a B2 or C1 level of English; one participant reported having a B1 and another estimated his English level as C2.

Participants were given 12 items (prompts) for the pre- and post-training assessments. Each of the prompts consisted of an image of an everyday situation that the participants may encounter while living abroad in an English-speaking country, as well as a short set of instructions describing the situation. For example, in one of the items, the participants were shown a photograph of a group of tourists holding a map and speaking to a passerby. The image was presented with the following instructions: “You are trying to find Central Park. You ask a stranger for directions” (see Appendix A for the entire set of prompts). The items encompassed a range of difficulty levels, and were later divided into two groups based on their initial difficulty.

For each of the prompts, two training recordings were made that corresponded to the two conditions of the study (Beat and No Beat). The instructor in the training video was a native speaker of American English. In the beat recording, the instructor said the target phrase while also producing rhythmic beat gestures that marked the relevant prosodic components of the utterance (see Appendix B for a transcript of the training recordings, including a representation of the syllables that received beat gestures). The recordings were evaluated by four native speakers of English, who judged them based on how natural the renditions appear. Each of the recordings received a naturalness score of at least a four out of five. Conversely, in the recordings for the No Beat condition, the instructor’s hands remained in a neutral position. In order to match the prosody of the Beat recording, the rhythm and overall speech of the No Beat recording was exaggerated, as well. Thus, this condition also constituted as rhythmic training.

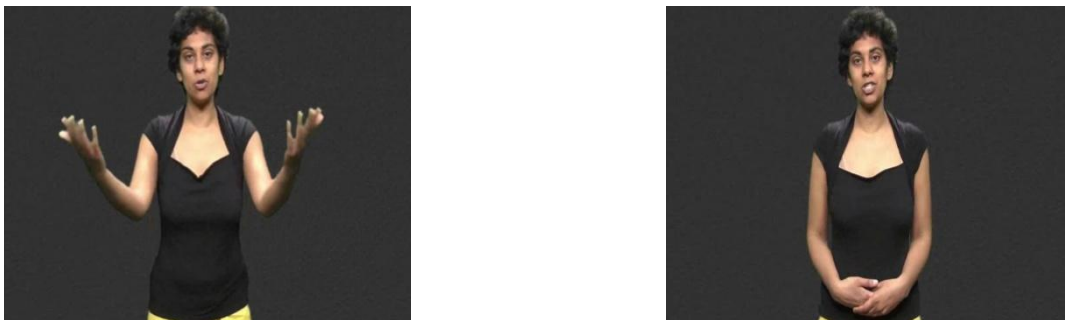


Fig. 1: Still images from the training videos: Beat condition (left) and No Beat condition (left)

2.4. Procedure

In the pre- and post-training tasks, in order to elicit more natural-sounding speech and to avoid allowing the participants to read off the screen while producing their response, the caption (instructions) were taken away on the Recording slide for each prompt. The participant then recorded a short, 1-3 sentence response to each of the test prompts, putting themselves in the role of the speaker in question. Each of the images included a blank speech bubble to mark the speaker. The participants were first shown an example response to a prompt in order to give

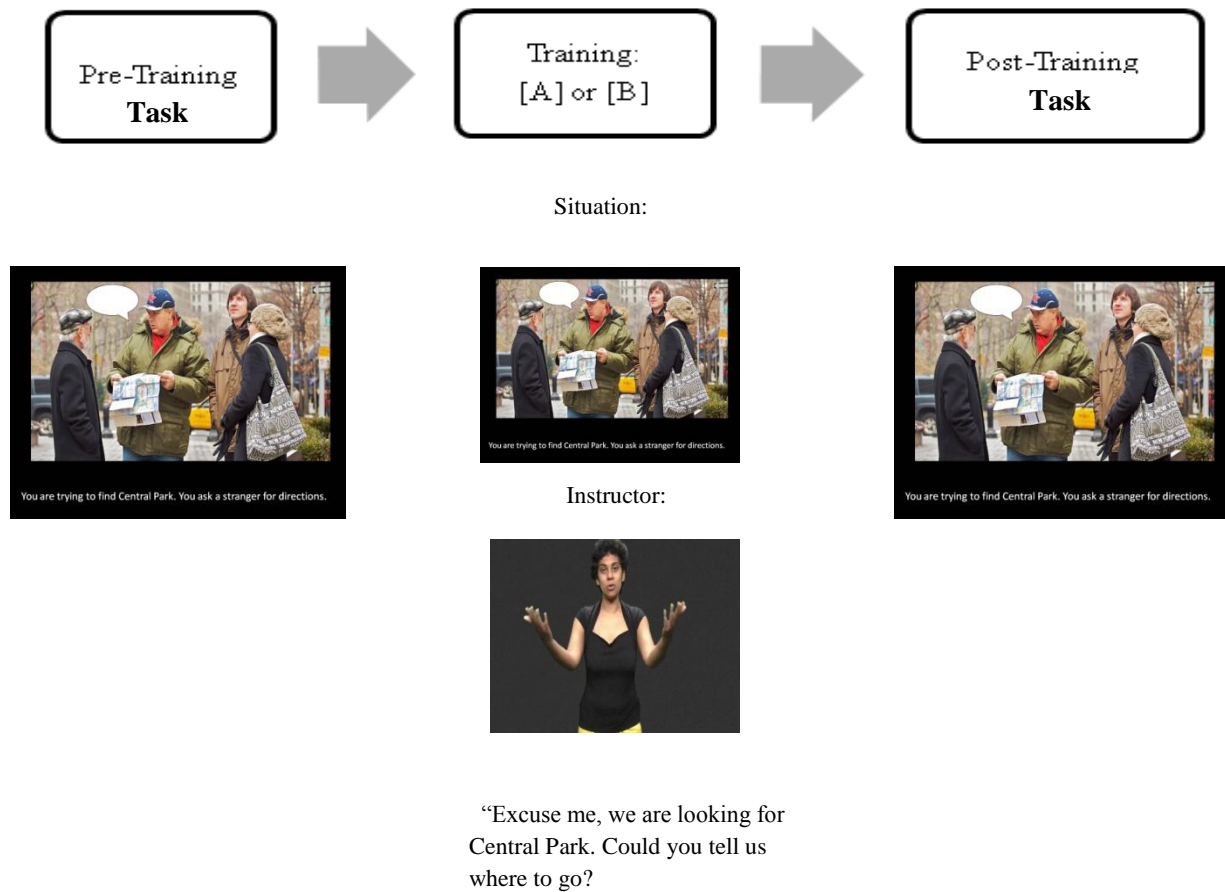


Fig. 2: Overall experimental procedure

them an idea of the length and complexity of the target responses. All testing was audio recorded with consent of the participants. The participants were tested and trained individually using a laptop computer. During the pre- and post-training assessments, the experimenter left the room, after verifying that the participant understood the instructions for the task. Following the training phase, the participants were given a five minute break, after which they were given the post-training assessment.

2.5. Training conditions

The goal of this study was to test the effect of observed beat gestures on the acquisition of rhythm using a within-subjects design. Thus, the training items were split into two groups, Beat (B) and No Beat (NB), in order to investigate the potential benefit of presenting training items with beat gestures. During the training phase, the items in the Beat group were presented with beat gestures produced by the instructor, while those in the No Beat group were given by the instructor while her hands were in a consistently neutral position. Participants were split into two conditions (A and B), each having a corresponding training video, which differed from each other in the following way: participants in condition A saw items 1-6 presented with beat gestures, and items 7-12 presented without them, while those in condition B saw 7-12 with beat gestures, and 1-6 without. Thus, each participant saw six items presented with beats and six without beats. Items were presented in the same order in both of the training videos, alternating between items presented with beats and those with no beats. Each recording was played three times in the training video. The duration of the entire training video was approximately seven minutes.


2.6. Speech Ratings


2.6.1. Procedure

The participants' recordings from both the pre-training and post-training sessions were rated by five native speakers of American English. Each rater evaluated a total of 480 participant recordings (20 participants x 12 items x 2 recordings for each item), split into 240 pairs (each pair consisted of a participant's pre- and post-training recordings for a specific item). All rating was performed via a three-part online survey-- each survey consisted of 80 pairs of recordings, and the raters were given a 24-hour break between performing each part. The raters reported that each part of the survey took approximately an hour to complete.

Prior to performing the ratings, the raters were reminded to evaluate the recordings based on the speaker's pronunciation and overall comprehensibility, instead of the content or grammar that his or her utterance conveyed. Each page of the survey presented the raters with a pair of recordings, after which the raters were asked to make a direct comparison between the two and indicate which recording sounded more native-like. They were then asked to evaluate each of the two recordings on a 7-point accentedness scale, from "1" (native/no accent) to "7" (very strong foreign accent). Accentedness was chosen as the target measurement because, as highlighted by van Maastricht, Krahmer & Swerts (2015), native listeners are very quick to mark nonnative speech as accented, while ratings of intelligibility and comprehensibility are not as extreme (in other words, heavily-accented speech may still be rated as relatively intelligible). The pre- and post-training recordings were presented in pairs because, as noted by Avello, Mora & Pérez-Vidal (2012), framing the accentedness question as a paired comparison is a more sensitive measure of slight changes in pronunciation from one testing time to another. The order in which the pre- and post- training recordings were shown was randomized. Refer to Figure 3 for a sample page from the rating survey.

Please listen carefully to the two clips.

Clip 1


Clip 2


3. Which clip sounds more native-like? *

Clip 1
 Clip 2

4. How native does CLIP 1 sound to you, on a scale of 1 to 7? *

Native/No foreign accent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very strong foreign accent
	1	2	3	4	5	6	7	

5. How native does CLIP 2 sound to you, on a scale of 1 to 7? *

Native/No foreign accent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very strong foreign accent
	1	2	3	4	5	6	7	

Fig. 3: Sample page from the online rating survey

2.6.2. Raters

The participants' recordings were rated by five native speakers of American English, three male and two female, aged from 18 to 28 ($M=24.2$; $SD=3.77$). The raters were all currently residing in the United States and had no previous training in linguistics or in teaching English as a Second Language. At the time of data collection, all of the raters reported having normal hearing. The raters reported having no significant contact with Peninsular Spanish- or Catalan-accented English.

2.6.3. Inter-rater reliability

Inter-rater reliability was assessed using an intra-class correlation (ICC) analysis for each pre- and post-training test item, and then obtaining an aggregate mean of the results. This yielded a Cronbach's Alpha score of .64, which is slightly lower than the generally-accepted measure of .7 (Larson-Hall, 2010). It is likely that this was caused by the relatively small number of raters

involved, considering that this measure of inter-rater reliability is attenuated when the number of judges is small (LeBreton & Senter, 2008). Nevertheless, the fact that the obtained value approaches that of the customary benchmark, even when using a small group of raters, suggests that there is a certain degree of cohesion in the raters' evaluations of the participant recordings. Therefore, all of the raters' scores were combined to produce a mean rating for each recording.

3. Results

3.1. Overall improvement from pre- to post-training

In order to assess whether training had an overall effect on the participants' accentedness ratings, a paired samples t-test was conducted to compare pre-training and post-training ratings across all items. There was a significant difference between the pre-training scores ($M = 4.66$, $SD = .56$) and the post-training scores ($M = 4.48$, $SD = .54$); $t(19) = 2.697$, $p = .014$, indicating that the participants' speech was perceived as being significantly less accented after they had undergone the training phase of the experiment.

3.2. The effect of beat gestures

An "improvement" variable was created for each participant's performance on each of the items by calculating the difference between their pre- and post-training score, with positive improvement value indicating less-accented post-training scores. An independent samples t-test was conducted to examine the effect of training with beat gestures versus no-gesture training on participants' improvement from pre-training to post-training. The test did not show a significant difference between in improvement between items that were presented with beat gestures ($M = .202$, $SD = .698$) and those presented without beat gestures ($M = .150$, $SD = .642$); $t(238) = .597$, $p = .551$.

While the initial test did not demonstrate a significant effect of beat gestures on accent improvement, it did suggest a positive trend in the expected direction. In order to further

investigate the question, the training items were subdivided into two categories based on their initial difficulty for the participants, which was assessed according to the average score that each item received in the pre-training. The six items that received the highest (most accented) initial scores (in other words, those that the participants seemed to have struggled with most prior to training) were coded as “difficult” in the Difficulty category, while the six with the lowest (least accented) scores were coded as “easy.” Table 1 shows the mean pre-training score for each of the items and their corresponding difficulty level.

Item	Mean Pre-training score (accentedness)	Assigned Difficulty (1 = easy; 2 = difficult)
Central Park	4.30	1
Time	4.35	1
Introduction	4.50	1
Pizza	4.56	1
Steak	4.56	1
Pharmacy	4.65	1
Shirt	4.71	2
Taxi	4.78	2
Necklace	4.84	2
Luggage	4.84	2
Professor	4.86	2
Apartment	5.00	2

Table 1: List of items, along with their mean pre-training score and the corresponding difficulty categories.

For the following analysis, a Generalized Linear Mixed Model (GLMM) ANOVA was conducted in IBM SPSS Statistics 19.0 (IBM Corporation, 2010). The GLMM analysis was conducted with improvement as the dependent variable, condition (beat, no beat) and difficulty (easy, difficult) as fixed factors, and item and subject as crossed random factors. As previously indicated, the difference in improvement between the beat condition ($M = .219$, $SD = .476$) and the no beat condition ($M = .129$, $SD = .476$) was not significant: $F(1,236) = 1.320$, $p = .252$. There was also no significant effect of difficulty on improvement-- the test revealed that

participants' improvement on easy items ($M = .132$, $SD = .476$) and difficult items ($M = .216$, $SD = .476$) did not differ significantly: $F(1,236) = .1145$, $p = .286$. However, a highly significant interaction was found between the two factors: $F(1,236) = 12.536$, $p < .001$, indicating that high-difficulty items that were presented with beat gestures showed the highest levels of improvement in the post-training scores. A follow-up pairwise contrast showed that within the set of difficult items, there was a significant difference in improvement between items that were presented with beat gestures ($M = .401$, $SD = .480$) and those presented without gestures ($M = .031$, $SD = .479$); $t(236) = 3.26$, $p = .001$. Within the easier set of items, those that were shown with beat gestures in the training ($M = .037$, $SD = .479$) demonstrated less improvement than items shown without gestures ($M = .226$, $SD = .480$), although this difference was not found to be significant: $t(236) = 1.747$, $p = .082$. Figure 4 summarizes these results.

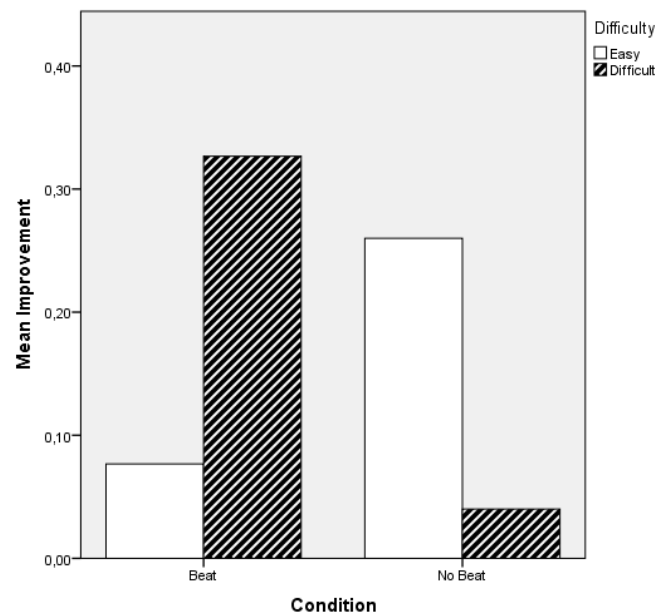


Fig. 4: Mean improvement as a function of condition and difficulty.

4. Discussion and Conclusions

This study examined whether observing beat gestures has an effect on the development of native-like speech (and more specifically, native-like rhythm) by nonnative speakers of English using a pre- and post-test design. As a whole, rhythm training, in both the speech (no beat) and beat conditions, served to make the participants' speech significantly less accented in the post-training task. Training with beat gestures had a significant beneficial effect on outcome (less-accented speech) when it was used with items that were initially more difficult for participants to produce at a native-like level-- items that tended to involve higher-level situations and require longer, more complex sentences. The results indicate that when, during training, these (comparatively) difficult items are presented with beat gestures, participants produce them in a more native-like manner in the post-training, compared to when the items are shown without any accompanying gestures.

While the role of beat gestures in improving overall pronunciation and suprasegmentals in L2 learners has been underexplored, this result of the present study is in line with the previous literature on the relationship between gesture and prosody in one's first language (Krahmer and Swerts, 2007, Biau and Soto-Faraco, 2013, among others). In the present study, while both conditions involved rhythm training, training with beat gestures resulted in significantly better outcomes for difficult items than no-beat (speech only) training. Perhaps it is the case that beat gestures perform a comparable prominence-enhancing function in perception when witnessed by L2 speakers as they do when the viewers in question are native speakers of the target language (as shown by Krahmer & Swerts, 2007). In this case, seeing beat gestures would serve as an additional source of linguistic information for L2 speakers, helping them to perceive key prosodic aspects of the target phrase, and later facilitating their language production. In order to be able to draw broad conclusions regarding this phenomenon, further research is needed in L2

gesture perception and its possible effect on subsequent speech production. Additionally, it is important to note that the key factor in this study was the presence of beat gestures in training—while the no-gesture condition still constitutes as a sort of rhythm-based prosodic instruction, the post-training assessment showed that these items differed significantly from those presented with beat gestures.

Our results also show that for lower-difficulty items, no-gesture training proved to be more successful (although not significantly so) than training with beat gestures. While this finding may initially seem to contradict the previous one, upon further examination, there are several possible reasons behind this result. Items that were quite easy for the participants to produce in the pre-training task (as indicated by their comparatively less-accented scores) may have already been at ceiling level-performance for these speakers. The items in this “easy” category tended to be very common phrases that language learners are exposed to from the very beginning of their studies (asking for directions, ordering food, introducing yourself, etc). Participants are likely to have heard these phrases countless times throughout their English education, and even if they do not necessarily produce them in a perfectly native-like manner, they are likely to do so naturally and without much second-guessing. A short amount of exposure to these phrases with accompanying beat gestures may not be enough to “override” the participants’ previous knowledge of prosodic prominence in these phrases, even if this knowledge may be somewhat erroneous. Therefore, while participants’ performance on the easy items benefited as a result of speech-only training, the presence of beat gestures did not add to the positive outcome (unlike the outcome of difficult items, which benefited significantly from beat gestures).

The findings of this study provide additional support for the importance of an explicit, global approach to L2 pronunciation instruction. As a whole, participants improved significantly

on their ratings of accentedness from pre- to post-training, after been exposed to each item only three times in the training phase of the experiment. This further highlights the potential benefits of targeted, effective pronunciation instruction, and suggests that the systematic use of gesture in instruction may further enhance acquisition of crucial aspects of competence in a second language. Additionally, the innovative method of the present study provides support for the use of natural materials, as well as for spontaneous (non-read) methods of eliciting speech from L2 speakers.

Finally, it would be interesting to continue investigating the benefit of beat gestures for L2 learners, including seeing whether gesture production results in higher language gains than gesture observation. Several studies in recent years have demonstrated that in the case of iconic gestures, producing them facilitates learning mental tasks more than simply observing them (see Goldin-Meadow 2013, among others). It would be valuable to explore whether a similar effect occurs in second language acquisition and if, in fact, participants show higher gains in accent improvement if they are instructed to actually imitate the experimenter and produce beat gestures themselves, rather than only observe them.

References

- Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.
- Adams, C. (1979). *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The Relationship between Native Speaker Judgments of Nonnative Pronunciation and Deviance in Segmentals, Prosody, and Syllable Structure. *Language Learning*, 42(4), 529–555.
- Avello, P., Mora, J. C., & Pérez-Vidal, C. (2012). Perception of FA by Non-Native Listeners in a Study Abroad Context. *Research in Language*, 10(01), 63-78.
- Behrman, A. (2014). Segmental and Prosodic Approaches to Accent Management. *American Journal of Speech-Language Pathology*, 23, 546–561.
- Biau, E. & Soto-Faraco, S. (2013). Beat gestures modulate auditory integration in speech perception. *Brain and Language*, 124, 143-152.
- Darcy, I., Ewert, D. & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL Teachers' pronunciation "toolbox". In J. Levis & K. Lavelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference*, Sept. 2011. (pp. 93-108). Ames, IA: Iowa State University.
- De Ruyter, L.E. (2015). Intonation status marking in spontaneous vs. read speech in story-telling tasks-Evidence from intonation analysis using GToBI. *Journal of Phonetics*, 48, 29-44.
- Derwing, T.M., & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility. *Studies in second language acquisition*, 19, 1-16.
- Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48(3), 393–410.
- Derwing, T.M., & Rossiter, M.J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning*, 13(1), 1-17.
- Edmunds, P. (2010). ESL speakers' production of English lexical stress: The effect of variation in acoustic correlates on perceived intelligibility and nativeness. Ph.D. dissertation, The University of New Mexico, Albuquerque.

- Faber, D. (1986). Teaching the Rhythms of English: A New Theoretical Base. *IRAL: International Review of Applied Linguistics in Language Teaching*, 24(3), 205–216.
- Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL quarterly*, 39, 399-423.
- Goldin-Meadow, S, Levine, S. L., Zinchenko, E., Yip, T.K-Y, Hemani, N., & Factor, L. Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*. 2012, 15(6), 876-884.
- Gordon, J., Darcy, I., & Ewert, D. (2013). Pronunciation Teaching and Learning: Phonetic Instruction in the L2 Classroom. In J. Levis & K. LeVelle (Ed.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 194–206). Ames, IA.
- Graham, C. (1978). *Jazz chants: Rhythms of American English for students of English as a second language*. New York: Oxford University Press.
- Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (Homage a Adam Kendon). *IRAL - International Review of Applied Linguistics in Language Teaching*, 44(2), 103–124.
- Hirata, Y., Kelly, S. D., Huang, J., & Manansala, M. (2014). Effects of hand gestures on auditory learning of second-language vowel length contrasts. *Journal of Speech, Language, and Hearing Research*, 57, 2090-2101.
- Igualada, A. Esteve-Gibert, N., & Prieto, P. (2015). Cognitive effects of beat gestures in pre-school children in a word recall task. *Child Language Symposium 2015*. University of Warwick: Coventry (United Kindom), July 20-21.
- Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2), 301–315.
- Kelly, S. D., Hirata, Y., Manansala, M., & Huang, J. (2014). Exploring the role of hand gestures in learning novel phoneme contrasts and vocabulary in a second language. *Frontiers in Psychology*, 5, 673.
- Kelly, S. D., McDevitt, T., & Esch, M. (2009). Brief training with co-speech gesture lends a hand to word learning in a foreign language. *Language and Cognitive Processes*, 24(2), 313-334.

- Krahmer, E., & Swerts, M. (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.
- Laan, G. (1997). The contribution of intonation, segmental duration, and special features to the perception of a spontaneous and real speaking style. *Speech Communication*, 22, 43-65.
- Larson-Hall (2010). *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York: Routledge.
- LeBreton, J.M. & Senter, J.M. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11(4), 815-852.
- Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, 3(1), 71–89.
- Mccafferty, S.G. (2006). Gesture and the materialization of second language prosody. *IRAL - International Review of Applied Linguistics in Language Teaching*, 44(2), 197–209.
- McNeill, D. (1992). *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press.
- Munro, M.J. & Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
- Prieto, P., Vanrell, M. D. M., Astruc, L., Payne, E., & Post, B. (2012). Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication*, 54(6), 681–702.
- So, W.C., Sim, C., & Low, W.S. (2012). Mnemonic effect of iconic gesture and beat gesture in adults and children: Is meaning in gesture important for memory recall? *Language and Cognitive Processes*, 5, 665-681.
- Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. In *Proceedings of the Fourth International Conference on Spoken Language*, 96, 1501-1504.
- Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25, 1-24.
- Tellier, M. (2008). The effect of gestures on second language memorization by young children. *Gesture*, 8, 219-235.

Ulbrich, C. (2013). German pitches in English: Production and perception of cross-varietal differences in L2. *Bilingualism: Language and Cognition*, 16, 397-419.

Van Maastricht, L., Kraemer, E., Swerts, M. (2015). Native speaker perceptions of (non-)native prominence patterns: The effect of divergence in pitch accent distributions on accentedness, comprehensibility, intelligibility, and nativeness. *Bilingualism: Language and Cognition* (in press).

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication*, 57, 209–232.

White, L., & Mattys, S. L. (2007). Rhythmic typology and variation in first and second languages. In P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 237–257). Amsterdam: John Benjamins.

Appendix A: Materials for the pre-/post-training task

1.

Instructions

Test your English survival skills!


You are going to see a series of situations you may encounter while living in an English-speaking country. Record how you would express yourself in each situation. Each response should be 1-2 sentences long.

2.

First, you will hear an example

3.


Example
You will see...



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.


4.

Example
You will see...



You will say... 🗣️
"I would like to open a student bank account. Could you tell me what documents I need to provide?"

5.



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

6.



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

7.



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

8.



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

9.



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

10.



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

11.



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.

12.



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.

13.



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.

14.



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

15.



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.

16.



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.

Appendix B
Training video transcript⁴

1. ExCUSE me, what TIME is it?
2. ExCUSE me, we are looking for Central PARK. Could you TELL us where to GO?
3. HI, I'd like to place an ORder for deLivery. Two large Pizzas with CHEESE and peppeROni.
4. SORRY, what did the professor just SAY? I couldn't HEAR him.
5. How much is this NECKlace? Can I get it for five DOLLars?
6. HI, I'm MAAya. It's GREAT to meet you.
7. My LUggage is LOST. Could you HELP me?
8. I'd like to get a STEAK with FRENCH fries, and a glass of red WINE, please.
9. I'm looking for this SHIRT in a bigger SIZE. Could you check and SEE if you have it in the BACK?
10. Can you TAKE me to the AIRport? As fast as you CAN please. I'm LATE for my flight.
11. Does this aPARTment get a lot of LIGHT in the mornings?
12. I have a sore THROAT and a FEver. Could you presCRIbe something for me?

⁴ Full beats are marked with capital letters and intermediate beats are underlined. Emphasis was placed on getting video recordings that appeared natural; therefore, not all stressed syllables were marked with beat gestures.