Article

LANGUAGE
TEACHING
RESEARCH

Language Teaching Research
2017, Vol. 21(5) 609–631

DOI: 10.1177/1362168816651463
journals.sagepub.com/home/ltr
SAGE

Training with rhythmic beat gestures benefits L2 pronunciation in discourse-demanding situations

Daria Gluhareva
Universitat Pompeu Fabra, Spain

Pilar Prieto
ICREA-Universitat Pompeu Fabra, Spain

Abstract

Recent research has shown that beat gestures (hand gestures that co-occur with speech in spontaneous discourse) are temporally integrated with prosodic prominence and that they help word memorization and discourse comprehension. However, little is known about the potential beneficial effects of beat gestures in second language (L2) pronunciation learning. 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-participants, pre-/post-test design, participants (undergraduate students learning English as a foreign language) watched a training video in which an L2 instructor gave spontaneous responses to discourse prompts. The prompts belonged to one of two categories (easy and difficult), and were presented by the instructor either with or without accompanying beat gestures. Participants' own answers to the prompts were recorded before and after training and evaluated by five native speaker judges. The results of the comparison between the participants' pre-training and post-training speech samples demonstrated that beat gesture training significantly improved the participants' accentedness ratings on the set of difficult (more discourse-demanding) items. The results of the study support the role of beat gestures as highlighters of rhythmic information and have implications for pronunciation instruction practices.

Keywords

accentedness, English language, pronunciation training, rhythm, second language acquisition, suprasegmentals

Corresponding author:
Daria Gluhareva, Department of Translation and Language Sciences, Universitat Pompeu Fabra, C/Roc Boronat, 138, Barcelona, 08018, Spain.
Email: dasha.gluhareva@gmail.com

I Introduction

1 Segmental and suprasegmental influences on second language (L2) pronunciation

In the field of second language acquisition, much has been written about possible ways to optimize learners' phonological abilities in their second language (L2). Even though most classroom pronunciation training tends to center around segmental instruction (that is, one focused solely on specific 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 (English as a second language) speakers' ratings of comprehensibility, accentedness, and fluency: segmental (instruction focused on individual sound contrasts), global[1] (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 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 group 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 instruction significantly improved in their overall comprehensibility from pre-training to post-training.

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 native listeners' perception of accentedness, comprehensibility, and intelligibility[2] than segmental deviance (e.g. Anderson-Hsieh, Johnson, and Koehler, 1992; Edmunds, 2010; Field, 2005; Ulbrich, 2013). In a series of studies by Munro and Derwing (1995, 1999; Derwing & Munro, 1997), the researchers analysed 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 first language (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) analysed 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 done.[3] It has been previously shown that read speech differs significantly from spontaneous speech in terms of its prosodic characteristics (e.g. De Ruiter, 2015; Laan, 1997; Swerts, Strangert & Heldner, 1996). Thus, recording participants' semi-spontaneous speech allowed the researchers to more closely approximate prosodic features that occur in real-life communication.

While suprasegmental training has proven to be successful in improving second language learners' overall fluency and comprehensibility, almost no work has tested the efficacy of specific training paradigms. 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 value of training the rhythmic component in a second language (and the prosodic components that closely relate to it) through the use of beat gestures, which will act as visual enhancers of this foreign language rhythm.

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; Anderson-Hsieh, Johnson & Koehler, 1992; Faber, 1986; Tajima, Port & Dalby, 1997). White & Mattys (2007) found that, in the case of three previously-identified rhythm metrics (VarcoV, nPVI-V, and %V), all were significantly correlated with native English speakers' ratings of foreign accent in L2 speech. One of these metrics in particular, VarcoV,[4] 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* (Graham, 1978), 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.

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, as the presence of iconic 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 rhythmic up-and-down movements of the hand that are associated with prosodic prominence in speech (McNeill, 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 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).

Recent research has also suggested that beat gestures do assist in first language word memorization: in So et al. (2012), beat gestures were shown to aid native word recall in adults, but not in children. 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. Recently, Kusch et al. (2015) found that beat gestures produced naturally (those that coincide with a focal pitch accent in speech) enhance the acquisition of novel words in a second language, while beat gestures that are not accompanied by prosodic prominence do not have a beneficial effect on learning.

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. Further highlighting the role of beat gestures in native language processing, Krahmer and Swerts (2007) found that seeing a manual beat gesture on a word resulted in increased prominence perception of that word. Additionally, Wang and Chu (2013) showed that beat gestures enhance speech comprehension, while other non-beat like hand movements do not.

The effect of beat gestures on second language processing and acquisition remains underexplored. The effect that beat gestures may have on second language prosody acquisition is not 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.

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, among others, Abercrombie, 1967). 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) analysed 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, even when syllable structure is controlled for, Catalan is significantly different from English on all measures. Specifically, the vocalic variability measures nPVI-V, DV, and VarcoV were shown to be robust tools for discrimination between the two languages, leading the researchers to conclude that there are important differences in durational patterns between English, on the one hand, and Catalan/Spanish, on the other. Thus, although Catalan's phonological properties may suggest that it is an intermediate language, its actual rhythm patterns are much more closely aligned with those of syllable-timed languages. These authors claim that the rhythmic class distinctions between syllable-timed (e.g. Spanish, Catalan) and stress-timed languages (e.g. English) finely correlate with differences in the way these languages instantiate two

prosodic timing processes, namely, the durational marking of prosodic heads, and pre-final lengthening at prosodic boundaries.

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.

4 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.

II Method

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 pre- and post-training outcomes for these two conditions will be compared.

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 a university 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 English.[5] 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.

3 Materials

a Prompts. Participants were given 12 items (prompts) for the pre- and post-training assessments (see Appendix 2). 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' (for the

entire set of prompts, see Appendix 1). The items encompassed a range of difficulty levels, and were later divided into two groups based on their initial difficulty.

b Training videos. For each of the prompts, two training recordings were made that corresponded to the two conditions of the study (Beat and No Beat); see Figures 1 and 2. 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 (for a transcript of the training recordings, including a representation of the syllables that received beat gestures when presented in the Beat condition, see Appendix 2). In all 12 of the Beat training videos, all of the nuclear pitch accents received full beat gestures. Other stressed syllables were also marked, using intermediate beat gestures; it is worth mentioning, however, that not all stressed syllables received beat gestures, as this would have appeared unnatural. In recording the training materials, emphasis was placed on obtaining videos that replicated naturally-occurring co-speech gestures as much as possible; thus, the instructor only placed beat gestures on words that carry the most semantic (and consequently, prosodic) weight. The beat gestures consisted of simple up-and-down or back-and-forth motions of the hands. Special attention was paid to avoiding gestures that could be interpreted as iconic (in other words, those that carry any sort of semantic information; see Figure 1).

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

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 ($M = 4.21$, $SD = .57$). Conversely, in the recordings for the No Beat condition, the instructor's hands remained in a neutral position. During the recording process, special attention was paid to ensuring that the prosodic patterns of each Beat video matched those of its No Beat counterpart. The instructor was trained to produce the utterances with the same prosody in both of the conditions. Table 1 shows that the acoustic measures of duration and pitch range over critical words and critical syllables were comparable across both conditions. The recordings in each pair were also equal to each other in terms of the number of prosodic phrases that they contained. Thus, it is important to emphasize that the training materials for the Beat and No Beat conditions differed from each other on the presence of gestures alone, allowing us to separate it from other possible factors. See also Table 2.

Table 1. Mean acoustic measures for corresponding pairs of Beat and No Beat utterances.

Measure	With gestures (Beat)		Without gestures (No Beat)		F	Sig.
	Mean	SD	Mean	SD		
Critical word (ms)	.509	.769	.514	.738	.028	.870
Accented syllable (ms)	.355	.774	.364	.671	.098	.757
Pitch accent range (Hz)	285.260	56.929	267.348	43.598	.749	.396
Pitch range (Hz)	378.893	39.706	354.348	27.153	3.124	.091

Table 2. List of items, along with their mean pre-training score and the corresponding difficulty categories.

Item	Mean Pre-training score (accentedness)	SD	Assigned Difficulty*
Central Park	4.30	.94	1
Time	4.35	.74	1
Introduction	4.50	1.02	1
Pizza	4.56	.90	1
Steak	4.56	.86	1
Pharmacy	4.65	.73	1
Shirt	4.71	.69	2
Taxi	4.78	.70	2
Necklace	4.84	.64	2
Luggage	4.84	,70	2
Professor	4.86	.94	2
Apartment	5.00	.76	2

Notes. *1 = easy; 2 = difficult.

4 Procedure

In the pre- and post-training tasks, for each of the prompts, the participants were first shown an Instruction slide. 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 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 5-minute break, after which they were given the post-training assessment; see Figure 2.

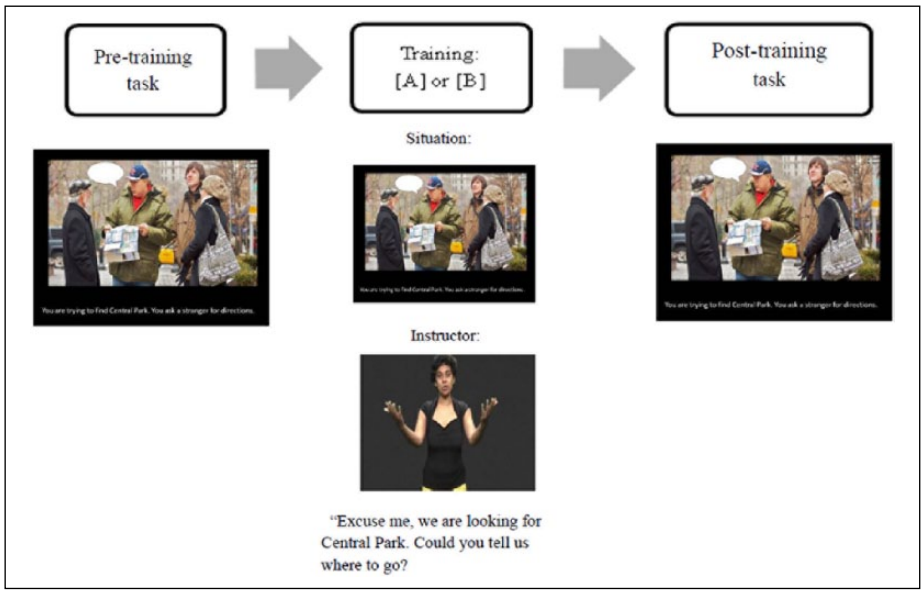


Figure 2. Overall experimental procedure.

Because the aim of the present study was to elicit (semi-)natural speech instead of imitations of the target phrases, participants were discouraged from memorizing the training utterances word for word. While watching the training video, they were not aware of the contents of the post-training assessment (in other words, the participants did not know that they would be tested on the same phrases). The authors found that, in general, these measures prevented the participants from attempting to memorize the target phrases, which allowed for both their pre- and post-test recordings to sound equally natural.

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-participants 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. Further, these Beat and No Beat groups contained an equal amount of two types of test items, distinguished by their level of difficulty: each group contained six so-called 'easy' items and six 'difficult' items. Level of difficulty was determined by the expected complexity/length of the triggered response, as well as how common/familiar the situation may be for a speaker of English as a foreign language. For instance, a prompt that simply required the participant to introduce himself or herself was classified as easy, and one that required the participant to ask about the condition of an apartment was classified as difficult.

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 7 minutes.

6 Speech ratings

a Raters. The participants' recordings from both the pre-training and post-training sessions 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 USA 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.

b Procedure. Each rater evaluated a total of 480 participant recordings (20 participants × 12 items × 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 60 minutes 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). Additionally, a comprehensive measurement such as accentedness was chosen in lieu of asking the raters to evaluate more specific characteristics of the speech samples (e.g. stress patterns, intonation, etc.) because we aimed to assess the participants' pronunciation based on the global impression that it produces in native speaker listeners who are not necessarily trained in linguistics or language teaching, as a way to more closely approximate 'real-world' conditions for learners of English as a foreign language.

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

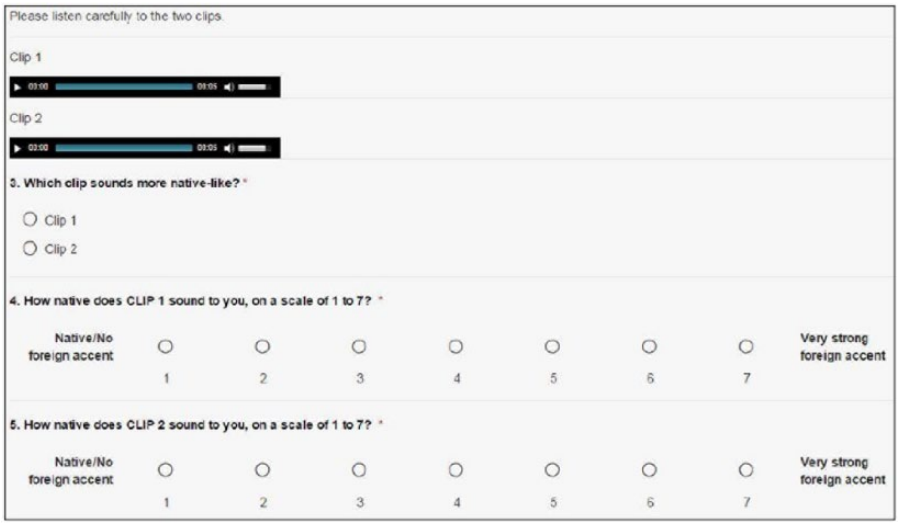


Figure 3. Sample page from the online rating survey.

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. Thus, the raters were not aware which of the recordings in each pair corresponded to the pre- and post-training assessment (for a sample page from the rating survey, see Figure 3).

c *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.

III Results

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 mean 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. Further, Cohen's effect size value (d = .603) suggested a medium-sized treatment effect. This indicates that the participants' speech was perceived as being significantly less accented after they had undergone the training phase of the experiment, independently of the training condition (e.g. Beat and No Beat).

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 values indicating less-accented post-training scores. Additionally, the efficacy of using item difficulty level as a fixed factor in the analysis was assessed using the mean pre-training scores that participants obtained on each of the items. The pre-training scores corresponded with the initial distinction: items initially classified as easy received lower (less-accented) scores in the pre-training than the more difficult set of items. Table 1 shows the mean pre-training score for each of the items, as well as their corresponding difficulty level.

This separation between easy and difficult items was later confirmed by another measure: the average length and complexity that the participants produced in response to each of the prompts. Responses for the two groups of items differed significantly in terms of their length: responses to items in the so-called difficult group were significantly longer in terms of sentences per discourse (M = 1.88) than responses to items in the easy group (M = 1.63); $t(238)$ = 2.625, p = .009. Similarly, items in the difficult group elicited answers of significantly more words per discourse (M = 16.97) than those in the easy group (M = 13.55); $t(238)$ = 4.438, $p < .001$. Thus, we decided to view items in the difficult group as being more 'discourse-demanding' than those in the easy group, because they required the participant to produce longer, more complex responses.

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 participant as crossed random factors. We found no main effect of condition: $F(1,236)$ = 1.320, p = .252) nor difficulty $F(1,236)$ = .1.145, p = .286) on improvement. 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.

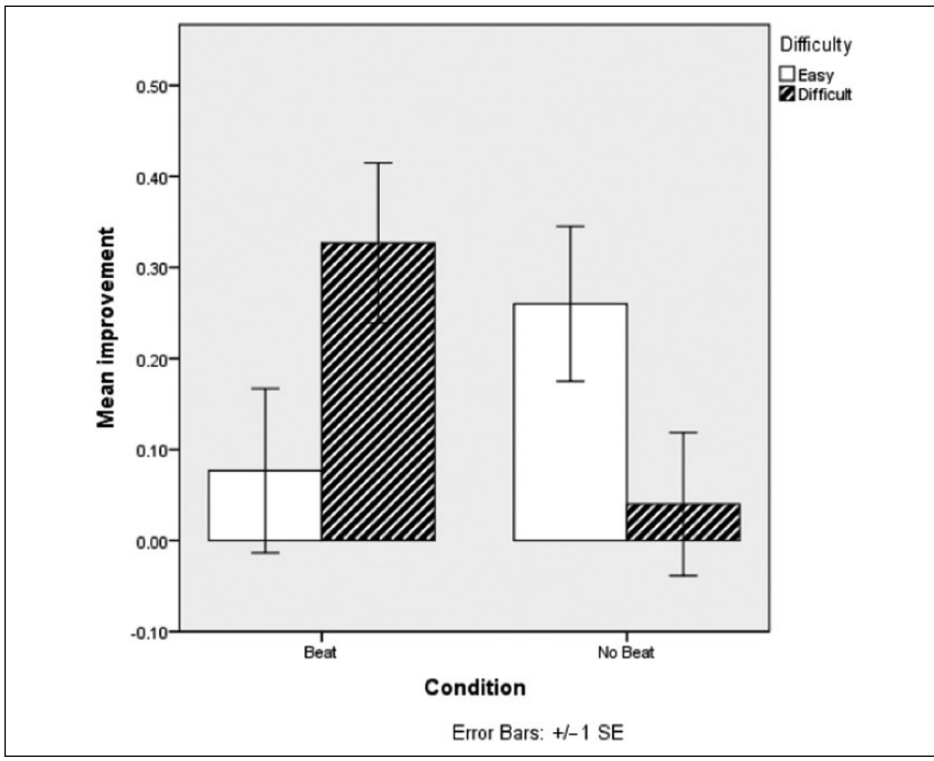


Figure 4. Mean improvement as a function of condition and difficulty.

IV Discussion and conclusions

1 Limitations

Because the present study involved a very short-term, exploratory pronunciation training design, some limitations need to be acknowledged. First, because the participants' post-training recordings were taken only 10 minutes after they were shown the training video, no long-term effects of the training were assessed, thereby not allowing us to evaluate the extent to which the participants will retain the benefits of the training over an extended period of time. Additionally, we exposed our participants to a small set of prompts, thus limiting the scope of the training significantly. Finally, it is important to recognize that the present training paradigm involved a relatively homogenous group of participants in terms of age, language background, and level of English; more work is needed to assess how the use of beat gestures may affect different groups of learners.

2 Conclusions

This study examined whether observing rhythmic 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 triggered 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.

Even though the improvement in accentedness ratings observed in this experiment may not seem substantial, it is important to note that these changes occurred after only a very short (7-minute) training session. The majority of pronunciation training studies have tended to take place over an extended period of time and be comprised of multiple training sessions, thereby making it somewhat difficult to compare the results of the present study to previously-investigated training methods. Additionally, actual numerical improvement was not reported by several of the aforementioned pronunciation training studies. It is, however, possible to draw some parallels: the highest improvement in the present study (.4 on an accentedness scale of 1 to 7), observed in the high difficulty-beat training group, is comparable to, for instance, results obtained by Gordon, Darcy, and Ewert (2013), who found that participants in their suprasegmental group improved by .6 points on a comprehensibility scale of 1 to 9. Thus, while our results seem to show only a moderate change from pre- to post-training assessment in terms of actual accentedness score, this improvement should still be considered meaningful within the field of pronunciation research (especially considering the moderate-to-high effect size of the treatment).

While the role of beat gestures in improving overall pronunciation and suprasegmentals in L2 learners has been underexplored, the result of the present study is in line with the previous literature on the relationship between gesture and prosody in one's first language (see, amongst others, Biau and Soto-Faraco, 2013; Krahmer and Swerts, 2007). 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 enhance their perception of key prosodic aspects of the target language, 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 while the key factor in this study was the presence vs. absence of rhythmic beat gestures in training, the no-gesture condition still constitutes a sort of rhythm-based prosodic instruction which triggers clear L2 pronunciation gains. However, the post-training assessment showed that training without gestures is not as beneficial (at least for higher-difficulty items) as training with beat gestures.

The results of the present study stand in contrast to the findings of Hirata and Kelly (2010) and Hirata et al. (2014), which, taken together, do not indicate a beneficial effect of beat gestures on learning in an L2. A possible explanation for this incongruity may stem from the fact that while Hirata and Kelly assessed segmental-level improvement, the present study targeted a different level of language learning, namely, suprasegmental (rhythmic) improvement. Perhaps it is the case that, because of their inherent rhythmic properties (synchronization with speech prominence), observing beat gestures impacts rhythm perception (and subsequent production) much more than segmental perception. While this is certainly an interesting possibility, more research is needed to confirm it.

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 corroborated by their comparatively less-accented scores) may have already been at high levels of 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 off-target. 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).

Another possible explanation for the somewhat-unexpected finding may have to do with the length of the responses elicited by the two categories of prompts. In general, the difficult prompts required longer responses than the easy ones, so it is entirely possible that, because the raters had more language output to work with, their judgments of the responses to the difficult prompts were more sensitive to (relatively minor) changes in accentedness. This, in turn, would mean that the evaluations of the responses to the difficult items are more valid for the objectives at hand than those of the responses for the easy items. While both explanations are entirely plausible, more research would be needed to disentangle the issue.

The findings of this study may 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 having been exposed to each item only three times in the training phase of the experiment. While it is not possible to decisively prove the advantage of such an approach without comparing our participants to a control group, there is little doubt that the obtained results are promising, given the very brief nature of the training involved. The study further highlights the potential benefits of targeted 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 used in 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, among others, Goldin-Meadow et al., 2012). 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.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by grant MINECO-FFI2013-31995 from the Ministerio de Educación, Cultura y Deporte of Spain and grant 2014SGR-925 from the Generalitat de Catalunya.

Notes

1. Several studies have used the term 'global instruction' as an umbrella term to encompass a wide range of suprasegmental components of L2 speech.
2. As defined by Munro and Derwing (1999), accentedness refers to the degree of foreign accent as perceived by the listener, comprehensibility is the listeners' perceived difficulty in understanding the utterance, and intelligibility reflects the degree to which the utterance is understood.
3. Non-spontaneous speech (obtained by means of read-aloud or repetition tasks) has been used by, among others, Anderson-Hsieh, Johnson & Koehler, 1992; Derwing, Munro & Wiebe, 1998; Field, 2005; Gordon, Darcy & Ewert, 2013; Munro & Derwing, 1999.
4. VarcoV is a rate-normalized measure of vocalic variability, calculated as the standard deviation of vocalic interval duration divided by the mean vocalic interval duration and multiplied by 100. As demonstrated by Prieto et al. (2012), this metric seems to be purely dependent on prosodic factors and is independent from phonotactic factors (namely, syllable structure).
5. Students in the Translation and Interpreting and Applied Languages degrees at the university 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.

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*, 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*, 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.

De Ruiter, 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.M., & Rossiter, M.J. (2003). The Effects of Pronunciation Instruction on the Accuracy, Fluency, and Complexity of L2 Accented Speech. *Applied Language Learning*, *13*, 1–17.

Derwing, T., Munro, M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, *48*, 393–410.

Edmunds, P. (2010). ESL speakers' production of English lexical stress: The effect of variation in acoustic correlates on perceived intelligibility and nativeness. Unpublished PhD dissertation, The University of New Mexico, Albuquerque, NM, USA.

Faber, D. (1986). Teaching the rhythms of English: A new theoretical base. *IRAL: International Review of Applied Linguistics in Language Teaching*, *24*, 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. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, *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 (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 194–206). Ames, IA: Iowa State University.

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 (Hommage à Adam Kendon). *IRAL: International Review of Applied Linguistics in Language Teaching*, *44*, 103–124.

Hirata, Y., & Kelly, S. (2010) Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research*, *53*, 298–310.

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.

IBM Corporation. (2010). *IBM SPSS Statistics for Windows*, Version 19.0. Armonk, NY: IBM Corporation.

Igualada, A., Esteve-Gilbert, N., & Prieto, P. (2015). Cognitive effects of beat gestures in preschool children in a word recall task. Paper presented at the Child Language Symposium 2015, University of Warwick, Coventry, UK, 20–21 July.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, *38*, 301–315.

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*, 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*, 396–414.

Kusch, O., Igualada, A., & Prieto, P. (2015). Beat gestures favor second language novel word acquisition only when they go together with prosodic prominence in speech. Paper presented at International Conference 'Prominence in Language', University of Cologne, Germany, 15–17 June.

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*, 815–852.

Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*, *3*, 71–89.

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

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, *45*, 73–97.

Munro, M.J., & Derwing, T.M. (1999). 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*, 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., Krahmer, 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.

Wang, L., & Chu, M. (2013). The role of beat gesture and pitch accent in semantic processing: an ERP study. *Neuropsychologia*, *51*, 2847–2855.

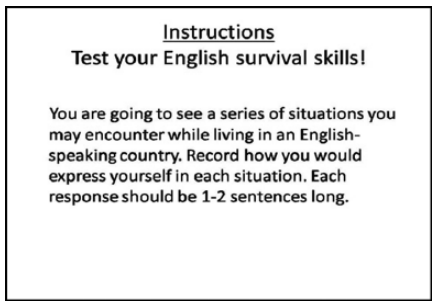
White, L., & Mattys, S.L. (2007). Rhythmic typology and variation in first and second languages. In: P. Prieto, J. Mascaró, & M.-J. Solé (Eds.), *Segmental and prosodic issues in Romance phonology* (pp. 237–257). Amsterdam: John Benjamins.

Appendix 1

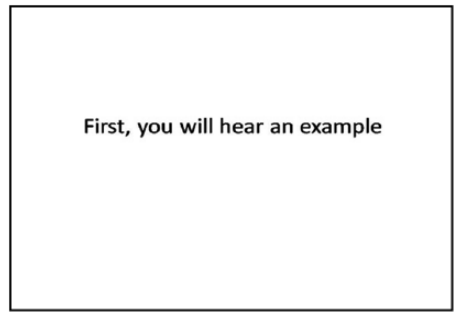
Materials for the pre-/post-training task

('difficult' items are marked with an asterisk)

1.



2.



3.

4.

5.

6.

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 2

Training video transcript

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 MAya. 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?

Notes

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.