The development of comprehensible speech in L2 learners

A classroom study on the effects of short-term pronunciation instruction

Joshua Gordon and Isabelle Darcy
Universidad Nacional, Sede Región Brunca, Costa Rica / Indiana University, Bloomington, USA

Developing comprehensible speech is an important goal for L2 learners. At present, there is clear evidence indicating that pronunciation instruction can help develop comprehensibility compared to no instruction at all (see Thomson & Derwing, 2015, for a review). However, it is unclear whether rapid improvements can be obtained through explicit pronunciation instruction. This study investigated the effects of explicit and nonexplicit instruction in the development of comprehensible speech in ESL learners. Three groups ($n = 12$) received about four hours of pronunciation instruction over three weeks using the same teaching sequence. Groups differed in the type of explicit instruction received: two experimental groups received explicit instruction either on suprasegmental features or four vowel sounds. The same content was presented to the third group without explicit instruction. Speech samples collected from all L2 learners before and after treatment revealed an effect of explicit instruction on comprehensibility: the group instructed in suprasegmentals was rated as more comprehensible. No significant improvement was seen in the nonexplicit group. These results suggest that focusing on suprasegmental aspects seems to be most effective for comprehensibility when time is limited, and argue for a major role for explicit phonetic instruction in developing enhanced comprehensibility (e.g., Derwing, Munro, & Wiebe, 1998; Munro, 1995).

Keywords: explicit phonetic instruction, pronunciation instruction, suprasegmentals, comprehensibility

In today's world of increasing mobility, millions of people are nonnative speakers of languages they use daily. Efficient oral communication skills in a second

Journal of Second Language Pronunciation 2:1 (2016), **56–92**. DOI 10.1075/jslp.2.1.03gor
ISSN 2215–1931 / E-ISSN 2215–194X

language (L2) have become crucial. Developing comprehensible and intelligible speech in an L2 is particularly crucial in settings where learners need to integrate into society personally and professionally, such as in the case of English as a second language (ESL) learners who live in an English-speaking environment (Derwing, 2008). A lack of intelligible pronunciation can result in comprehension difficulties. Both often preclude L2 learners from obtaining stable jobs and contribute to social isolation and educational inequality (Derwing, Thomson, & Munro, 2006; Yates, 2011; Zielinski, 2012). However, it is well known that nonnative pronunciation patterns can be very hard to modify even when one's proficiency in the L2 is considered high. As seen in the case of many international physicians in the health-care system or teaching assistants at universities, accented L2 speech is independent from education levels and intelligence. Generally speaking, pronunciation (which we term "phonological skills" to encompass both speaking and listening) is the area of language with the largest individual variation in performance, compared to, for example, grammar or vocabulary.

What is it specifically that makes an L2 speaker more or less intelligible? Research has shown that different but related concepts such as intelligibility (i.e., the extent to which a speaker's utterance is actually understood) and comprehensibility (i.e., a listener's estimation of difficulty in understanding an utterance produced by an L2 speaker) do not necessarily correlate with degree of foreign accent (see Derwing & Munro, 1997; Munro & Derwing, 1995); thus, heavily accented L2 speech can still be highly comprehensible and intelligible. As a result, and because native-like pronunciation patterns are unlikely to be attained by most L2 learners (Abrahamson & Hyltenstam, 2009), developing intelligible and comprehensible speech has become a more realistic goal for pronunciation teaching (see Levis, 2005).

On the whole, it is accepted that more accurate pronunciation patterns for segmental and suprasegmental features, along with higher fluency, contribute to intelligible and comprehensible speech (see Goodwin, 2014, for a review). For instance, nontarget segmental realizations have been shown to influence the perceived degree of foreign accent, comprehensibility, and intelligibility (e.g. Magen, 1998; Munro, 1993; Schoonmaker-Gates, 2012; Tajima, Port, & Dalby, 1997; Zielinski, 2008). Similarly, misplaced or missing prominence (Hahn, 2004), incorrect word stress (Field, 2005), and inappropriate syllable timing (insufficient differentiation in syllable duration between stressed and unstressed syllables) (Setter, 2006; Tajima, Port & Dalby, 1997) negatively affect comprehensibility. For fluency, Kang (2010) showed that too many pauses, or pauses that are too long, are particularly detrimental to intelligibility and comprehensibility, as were speaking too slowly or too fast (see also Munro & Derwing, 2001).

Knowing what is detrimental to comprehensibility and intelligibility is important. Yet, given the time constraints that often characterize pronunciation instruction in classroom settings, there is a need to make pronunciation instruction as efficient as possible, and to determine what specific aspects of nonnative pronunciation are more or less detrimental to intelligibility and comprehensibility.

Research generally supports suprasegmentals' key role in shaping the degree of perceived foreign accent, intelligibility, and comprehensibility (e.g. Anderson-Hsieh, Johnson, & Koehler, 1992; Derwing & Munro, 1997; Kang, Rubin, & Pickering, 2010; Munro, 1995; Munro & Derwing, 1995; Trofimovich & Baker, 2006; Wennerstrom, 2000). For instance, Munro and Derwing (1995) and Derwing and Munro (1997) analyzed specific accent features in L2 speech samples and correlated them with accentedness, comprehensibility, and intelligibility ratings. They found that prosodic error scores — as opposed to phonemic errors — contributed more strongly to ratings of stronger accentedness and lower comprehensibility, whereas intelligibility ratings in turn were influenced by both. However, researchers agree on a balanced approach for pedagogy encompassing segmentals, suprasegmentals, and fluency, given the importance of all three aspects of language to comprehensibility and intelligibility (Derwing, Munro & Wiebe, 1998; Morley, 1991).

If the goal for L2 learners is comfortable, intelligible, and comprehensible speech, how can they attain it? Some may argue for the possibility that learners will simply "pick up" accurate and intelligible pronunciation patterns with more exposure. Indeed, although at the outset the naturalistic acquisition of an L2 phonological system is, to a large extent, shaped by interference or transfer from first language (L1) phonological knowledge during processing, this is generally expected to diminish over time. For both segmental (i.e., vowels, consonants, and their language-specific combinations) and suprasegmental (e.g. stress, rhythm, intonation) dimensions of phonology, the L1 phonological system influences the perception and production of nonnative/L2 speech (e.g. Darcy, et al., 2012a; Darcy, Ramus, Christophe, Kinzler & Dupoux, 2009; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997; Guion, 2005; McAllister, Flege, & Piske, 2002; Munro, 1993; Pallier, Bosch, & Sebastian-Gallés, 1997; Werker & Tees, 1984; for a review, see Sebastian-Gallés, 2005).

The L1-L2 interference effects during L2 phonological acquisition can be modulated by external factors, and this may hint at the possibility that learners might pick up better phonological skills on their own. In fact, studies often conclude that an increased amount of exposure to an L2 or more L2 use positively influences the accuracy with which the sounds of the L2 are perceived and produced (e.g. Darcy, Peperkamp, & Dupoux, 2007; Flege, Bohn & Jang, 1997; Flege, Frieda & Nozawa, 1997; Guion, Flege & Loftin, 2000; Levy & Strange, 2008; Munro & Derwing, 2008; Purcell & Suter, 1980; Trofimovich & Baker, 2006). However, it is not fully clear

how much exposure to the language is necessary to reach comfortable phonological accuracy. Best and Tyler (2007) suggest that in the case of phonetic properties, little additional perceptual benefit seems to accrue from experience past the initial learning period of 6–12 months of residence in a new language environment for most late learners. (Here, length of residence, LOR, is an indirect measure of experience). If this is the case, it is possible that the learners in studies that failed to find LOR effects on phonological processing had already reached their personal ceiling performance. As a result, more L2 exposure would not trigger further development (see also Darcy et al., 2012a; Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Levy & Strange, 2008, for the difficult /u/-/y/ contrast, or also Pallier et al., 1997). However, it is also possible that such a plateau in phonological development only ends after several years of sustained and intensive exposure (see also Han, 2004; VanPatten, 1988). For instance, in studies that found effects of experience beyond the 6–12 months identified by Best and Tyler, "experienced" groups often had very extended LORs (e.g., 7 years or longer in Flege et al., 1997).

Given the long time apparently necessary for adult L2 learners to accumulate enough exposure to show improvements in naturalistic phonological acquisition, and given the obvious possibility that learners do not just pick up accurate pronunciation patterns (Grant, 2014; Zielinski, 2012), at least within a reasonable amount of time (see Piske, MacKay, & Flege, 2001, p. 197), a long-standing question of interest has been the extent to which exposing learners to L2 speech through instruction or laboratory training studies can help speed up this process, and in particular, what time-frame for instruction and what methodology will yield the best results.

Different studies have examined whether learner training is effective in improving perception and production of an L2. Studies using high variability training paradigms have generally shown that in controlled laboratory conditions, training can cause L2 learners to improve their perception and production of segmentals (e.g. Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1997) and suprasegmentals (e.g. Wang, Spence, Jongman & Sereno, 1999; Wang, Jongman & Sereno, 2003). The findings of such training studies are important because they show that knowledge gained through perceptual training can be transferred to the production domain, and also because they suggest that a specific treatment designed to direct learners' attention toward specific features of the L2 can also facilitate acquisition. However, training studies often require many hours of exposure to sounds (see also Nishi & Kewley-Port, 2007), and their pedagogical relevance remains unclear, given the time constraints that often characterize classroom L2 instruction. One important goal, therefore, is to know whether gains in comprehensibility can be obtained in less time and in more ecologically valid contexts, such as in a regular, instructed classroom context, in order to maximize instruction a efficiency.

Different studies have focused on the effects of instruction on the acquisition of L2 phonological features in the classroom (e.g., Derwing, Munro, & Wiebe, 1998; Elliott, 1997; Lord, 2005; Missaglia, 1999; Perlmutter, 1989; see Thomson & Derwing, 2015, for a review). Using an acoustic analysis of speech samples collected before and after instruction, Lord (2005) showed that explicit phonetic instruction increased the accuracy with which learners articulated nine Spanish phonemes. Although an L1-Spanish control group was used as a comparison baseline (as opposed to another group of L2 learners with a different instructional focus), this study suggested that explicit phonetic instruction might contribute to learners' pronunciation accuracy, perhaps by making them aware of features in the L2 input that otherwise are difficult to notice (see Pennington & Ellis, 2000). A greater qualitative language awareness (according to Benson & Lor's [1999] framework) was also suggested by Kennedy and Trofimovich (2010) as one possible factor linked to greater comprehensibility. The positive effects of explicit phonetic instruction documented by Lord's study took place in a group of learners of Spanish studying the language's phonetics for a semester. In addition, in contrast to classroom-based studies like those of Derwing et al. (1998) or Elliot (1997) reviewed below, Lord's explicit instruction took place as part of a Spanish phonetics course and not in a more ecologically diverse L2 class.

In a classroom-based study, Derwing et al. (1998) contrasted two experimental instruction groups: group 1 ("segmental") was trained on segmental accuracy (i.e., narrow, word level), whereas group 2 ("global") was trained on general speaking habits and prosodic factors (e.g., broad, discourse level). A third group received no pronunciation specific instruction (control group). Speech samples taken at the beginning and the end of the 10-week instructional period — in the form of short sentences and extemporaneous narratives — were presented to L1-English listeners. Although both groups improved in comprehensibility and accentedness on a sentence repetition task, only the global group improved in comprehensibility and fluency in the narrative task, suggesting that comprehensibility is likely to improve more generally as a result of instruction than accentedness is. Overall, given the larger and more generalized improvement in comprehensibility obtained by the global group on extemporaneous productions, these results again suggest that suprasegmentals are important in L2 pronunciation teaching focusing on intelligibility or comprehensibility. However, Derwing and colleagues also clarify that these results do not call for an abandonment of segmental instruction because segmental errors that cause communication breakdowns can potentially be repaired if L2 learners are aware of relevant differences.

Other studies have also evidenced the positive results of explicit pronunciation instruction. Elliot (1997) investigated the production of different sounds in learners of Spanish at an American university during one semester. In this study,

an experimental group that received explicit phonetic instruction showed overall improvement in the production of 19 Spanish sounds over a control group. This classroom-based study demonstrated that explicit phonetic instruction can be effective when incorporated into the regular curriculum of the class.

In a more recent study to investigate the effects of explicit phonetic instruction, Kissling (2013) analyzed the acquisition of different Spanish consonants in L1-English learners of Spanish. The participants were grouped to receive either explicit phonetic instruction with input, practice, and feedback, or a similar treatment without the explicit phonetic instruction component. This pretest-posttest experimental design study included a total of three weeks of treatment, and the participants took a delayed posttest 3 weeks after treatment. Through an acoustic analysis, the results of this investigation showed that learners in both groups improved their pronunciation of Spanish consonants equally, which Kissling presented as evidence that it is input, practice, and feedback that facilitate improvement in pronunciation, as opposed to explicit phonetic instruction itself.

Although these studies have highlighted the positive role of explicit phonetic instruction, they have mostly focused on the study of segments only, and not necessarily on more global, suprasegmental aspects of speech that could also be beneficial for learners to improve their comprehensibility. Additionally, some of these studies were conducted in the context of semester-long courses, either specifically dedicated to phonetic instruction, or embedded in regular language courses (one exception is Kissling, 2013, which lasted only a few weeks). Therefore, it is necessary to investigate the effects of short-term explicit pronunciation instruction that goes beyond the study of just different phonemes and which therefore could provide efficient tools for learners to improve their comprehensibility if incorporated little by little in class.

One problem of pronunciation instruction is that many teachers are resistant to implementing it due to a lack of time. In many teaching contexts, such as ESL programs, the intensive nature of the curriculum makes teachers hesitant to incorporate pronunciation instruction in their regular classes, and teachers have to be selective as to what specific aspects to teach. In other cases, pronunciation instruction is left to a single course that students can take as an elective, often reserved for advanced-level learners. One recent solution to this problem is to make pronunciation instruction a curricular component that is integrated into every lesson (see Darcy, Ewert, & Lidster, 2012; Sicola & Darcy, 2015). The rationale behind this call is that it has strong potential to enable more teachers to provide pronunciation instruction, even if for short periods of time, which could be beneficial for learners in the long run. However, it remains essential to verify to what extent improvement can be obtained in a short-time frame, through pronunciation instruction components that are embedded in regular English language classroom

instruction, which is a less controlled setting than the previously-mentioned laboratory and phonetics course-based studies have examined.

Our goal in this study is to evaluate the potential for improvement that a modest amount of pronunciation instruction can achieve. This, if found, might prove useful for encouraging teachers to provide more pronunciation instruction even if they are not teaching in dedicated pronunciation classes.

The current study

This study presents the results of a short classroom intervention that investigated the effects of short-term explicit pronunciation instruction in suprasegmental features and four vowel sounds on the comprehensibility of ESL learners' production. Its purpose is to investigate to what extent embedding a short pronunciation component into a regular ESL class — that is, by incorporating explicit pronunciation instruction little by little into the regular class as opposed to doing it in a separate pronunciation class — can help learners achieve more comprehensible speech. Although there is limited time in many intensive ESL programs, incorporating pronunciation instruction in short amounts into class could help learners develop comprehensible speech in their regular classes — an important part of communicative competence — instead of having to address their pronunciation needs in a separate course. For this study, we specifically compared improvement in groups of ESL learners receiving explicit pronunciation instruction to a group of learners receiving nonexplicit pronunciation instruction (rather than no pronunciation instruction at all). Carried out in an L2 classroom context with intact classes in an intensive ESL program at a large public university in the American Midwest, this study also extends comparable previous studies (e.g., Derwing, et al., 1997; 1998) by examining learners who for the most part had recently arrived in the L2 environment, and by implementing a shorter, 3-week instruction period — as opposed to a 12-week treatment for learners whose LOR was above 10 years (Derwing et al., 1997), and a 10-week treatment for learners with 2 years LOR (Derwing et al., 1998). The study was motivated by the following research questions:

1. Does a short-term explicit pronunciation instruction component incorporated in intact speaking classes improve comprehensibility ratings for L2 learners of English as compared to nonexplicit instruction?
2. If so, does a short-term pronunciation instruction component in suprasegmental features yield larger comprehensibility increases than instruction in four vowel sounds?

Method

The general procedure used in this study followed a pretest-posttest design. Pronunciation instruction was conducted in three level-six speaking (i.e., oral communication skills) classes in a seven-level intensive ESL program. Level-six students are high-intermediate learners according to the institutional levels of this program: beginning to low-intermediate learners are placed in levels 1 through 3, intermediate learners are in level 4, high-intermediate students belong to levels 5 and 6, and advanced learners are placed in level 7. The participants in this study had average TOEFL scores of 499, 514, and 485 respectively in the paper-based version of the test. In this program, learners typically receive 5 hours of instruction per day, five days a week for six weeks, and final exams are taken in the seventh week of each session. For this study, learners received pronunciation instruction for three weeks (the third, fourth, and fifth weeks of the session) and were audio recorded individually before and after receiving this instruction (during the second and sixth weeks of the session). In addition to the three groups of ESL learners, a group of 10 L1-English speakers also recorded speech samples that were used as a baseline for comparison purposes. For uniformity, these participants were undergraduate students from the same university and residents from the same Midwestern region. The ESL learners' speech samples were mixed with L1-English speaker baseline samples, and comprehensibility ratings of all those samples were obtained from L1-English raters to assess improvement in phonological skills.

Pronunciation instruction

We used three intact ESL classes for this study, and we assigned a specific experimental treatment to each class: the first group (referred to as the *suprasegmental group* hereafter) was originally composed of 12 participants whose L1s were Arabic, Turkish, Korean, and Japanese. The second experimental group (the *vowel group*) was originally composed of 8 participants, whose L1s were Arabic, Portuguese, French, Russian, Korean, and Japanese. Finally, the third group (the *nonexplicit group*) was composed of 10 participants who were L1 speakers of Arabic, Turkish, Korean, and French. It is important to stress that we used intact classes composed of students with a variety of L1 backgrounds, which is a common characteristic of ESL programs like this. Because the study was carried out in a classroom-based context in an intensive ESL program, it would have been impossible to randomly assign participants to groups with the three different conditions (see Mackey & Gass, 2005). Therefore, for convenience and due to the impossibility of randomly assigning participants to different groups, we decided to use intact classes for this study. Additionally, in our view it is important to examine to what extent the

principles and findings of laboratory research are applicable to actual classrooms in order for language teachers to have a psycholinguistically- and pedagogically-oriented model of what is possible to achieve in an actual class in terms of pronunciation instruction. Each class was taught by a different L1-American English teacher. The two instructors in charge of the explicit groups were senior teachers in the program with more than 15 years each of experience teaching ESL and EFL. The teacher in charge of the nonexplicit group was a graduate student in TESOL and had 3 years of experience teaching ESL.

The three learner groups received instruction for 3 weeks, 3 days per week (Monday, Wednesday, and Friday), for 25 minutes a day, either at the beginning or end of class. All students in the three groups attended their 50-minute communication classes 5 days per week from Monday to Friday. These communication classes focus mainly on the development of oral skills, and the students usually participate in discussions of current events, debates, role plays, and oral presentations as part of the instruction. The pronunciation component for the two explicit groups was designed following the presentation-practice-production sequence used commonly in many L2 instructional settings (Chamot, Barnhardt, El-Dinary, & Robbins, 1999). This approach was selected for explicit instruction because it allows teachers to introduce topics and concepts, guide the L2 learners in different tasks, and assess their performance in production. Both explicit groups received instruction following the same sequence of activities but with a focus on either suprasegmentals or vowels. The suprasegmental group received instruction on the perception and production of four aspects of English prosody: stress (word and sentence stress), rhythm (e.g., stressed and unstressed syllables, pauses), reductions (e.g., function words with vowels reduced to schwa), and linking (e.g., re-syllabification in CC-V sequences, intervocalic consonants in VC-V sequences, consonant lengthening in geminate consonants). The vowel group, in contrast, received instruction on four English vowel sounds: /i/, /ɪ/, /æ/, and /ɛ/. Such instruction included explicit explanations about the characteristics of each of these vowel sounds (e.g., tense and lax vowels, tongue position) as well as perception and production exercises that contrasted minimal pairs, both in individual words as well as in sentence contexts. These very specific vowels and suprasegmentals were selected for various reasons. First, they represent problems for the majority of L1 groups who come to study in the program (Arabic, Chinese, Japanese, Korean, and Turkish speakers), either in perception and production, or in the stress-timing nature of English (Korean: Darcy, Park, & Yang, 2015; Flege, 1995; Flege, Bohn, & Jang, 1997; Japanese: Strange et al., 1998; Arabic: Anani, 1989; Flege, 1995; Munro, 1993; Turkish: Bayraktaroğlu, 2008; Darcy & Krüger, 2012). Additionally, some common textbooks on pronunciation instruction and English phonology also point out the difficulties for L2 learners of English with these vowels and prosodic

aspects (see Avery & Ehrlich, 1992; Yavaş, 2006). Another reason is that in this program, instructors usually have to limit the content they can address due to time constraints and the intensive nature of these courses (i.e., each session only lasts 7 weeks), so they usually focus only on those aspects that are problematic for the majority of their learners. Additionally, although the number of vowels studied by the vowel group was limited in comparison to the prosodic aspects studied by the suprasegmental group, these four vowel sounds are also problematic for many learners because of their high functional load in different minimal pair combinations (see Brown, 1991), such as /æ/ and /ɛ/ (e.g., *pet-pat*, *bet-bat*, *wreck-rack*), /ɛ/ and /ɪ/ (e.g., *bet-bit*, *pet-pit*, *bed-bid*), or /i/ and /ɪ/ (e.g., *peel-pill*, *feel-fill*, *these-this*).

Instruction combined a bottom-up (detailed analysis of specific phonetic information on features of English) and top-down (activities developing fluency in pronunciation) skills approach to keep a balance between fluency and accuracy, given that both aspects affect each other in L2 oral production (see Celce-Murcia, 2001; Celce-Murcia, Brinton, Goodwin, & Griner, 2010; Hinkel, 2006; Lazaraton, 2001).

Before the study, the three collaborating teachers met with the researchers and received instructional guidelines to administer treatment. The researchers provided the materials and lesson plans for the treatments, and the teachers were in charge of their ultimate implementation so that they could deliver instruction in a way they felt comfortable with. The first author observed and audio recorded all classes in the three groups to verify treatment fidelity, and also to take note of factors potentially affecting the study. For example, attendance was recorded and data from students who were absent in class were not included in the final analyses. Figure 1 summarizes the characteristics of the instruction for each group.

In the experimental groups, at the beginning of a pronunciation lesson, the teachers introduced topics and concepts with explanations of the different vowels or suprasegmental features to be studied in each lesson using explicit phonetic information, audio samples, repetition, or visual aids (e.g., pictures, diagrams, phonetic symbols, articulatory charts, words and phrases with prominence markers — for instance, dots of different sizes placed above syllables — as well as capital and bold letters in texts) to develop awareness of L2 phonetic features (Avery & Ehrlich, 1992; Celce-Murcia et al., 2010). As for the nonexplicit group, the instructor announced practice in pronunciation but did not present any phonetic information, and simply asked the students to repeat words and sentences after her (taken in equal proportions from the materials used in the two experimental groups). Students' attention was never called to specific features, nor was the target of instruction specified for the nonexplicit group (Ellis et al., 2009). We also wanted to see the effects of repetition in learners without the explicit instruction component given to the other two experimental groups since it is a common technique used in pronunciation instruction by many teachers (see Baker, 2014).

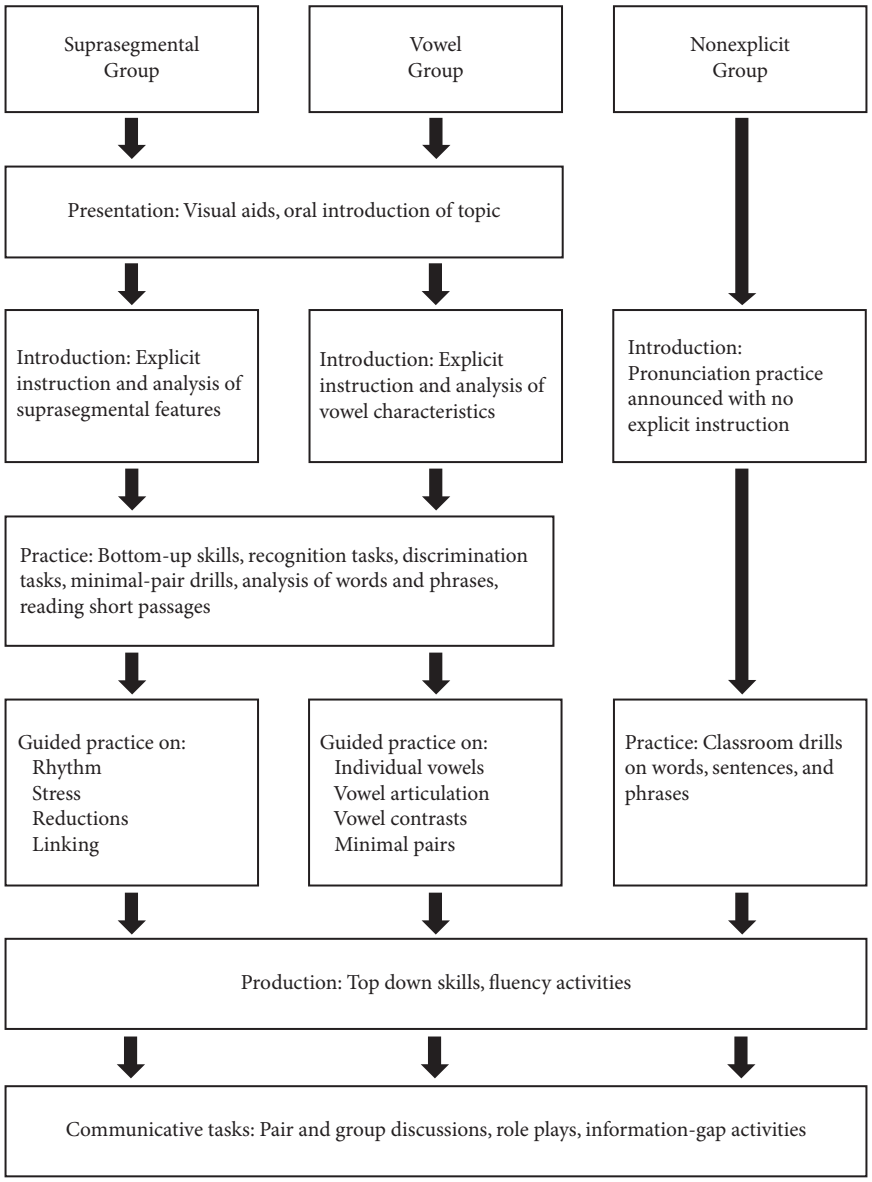

Figure 1. Pronunciation instruction treatment implemented in this study.

Thus, the students in the nonexplicit group only practiced repeating words and sentences without being given explicit phonetic information.

The topic introduction was followed by a guided practice stage. This part of the treatment included traditional classroom techniques for pronunciation instruction; for example, decontextualized recognition and discrimination tasks,

minimal-pair drills, reading of short passages and sentences out loud, and contextualized minimal-pair recognition and discrimination tasks (see Celce-Murcia et al., 2010). Visual and kinesthetic reinforcement was also used with the explicit groups (e.g., clapping while reading sentences, tapping out the rhythm of a passage with a pencil on a desk, pointing out the length of vowels through gestures, etc.).

The last part of each treatment session consisted of communicative tasks integrated into the content of communicatively-oriented lessons where learners put into practice the patterns learned with another classmate or in small groups for realistic, communicative purposes (see Celce-Murcia et al., 2010; Levis & Grant, 2003). The activities for this stage were carefully designed to incorporate the content used in the second stage. For example, the vowel group performed role plays using minimal pairs with the vowel sounds studied. Similarly, the suprasegmental group performed tasks in small groups (usually 3 participants) in which 1 participant monitored the appropriate use of stress and rhythm by the other two participants while they performed an information-gap activity.

The nonexplicit group worked with a combination of the same materials used in the two experimental groups. Participants were guided by the teacher on drilling activities (e.g., listen and repeat) with words, phrases, and sentences that were also part of the materials used by the other two groups. During the last part of the class, the participants performed the same communicative activities as the two experimental groups, but with the difference that there was no feedback from the teacher on the use and production of segmental or suprasegmental features (see Figure 1). In sum, the difference between explicit and nonexplicit instruction in the design targeted both instruction and feedback. In explicit instruction, participants' attention was directed explicitly to phonetic errors (as opposed to just meaning) and how these errors could lead to problems in communication. Both individual and group errors and their difficulties were explicitly stated and delineated, and students were explicitly told how to correct them. In the nonexplicit group, none of the above was the case. Figure 2 presents a taxonomy that explains the criteria that teachers in the experimental groups used to provide explicit instruction and feedback to students.[1]

Comprehensibility rating task

Speech samples

The original stimuli for the task consisted of 24 prompts and responses in the pretest plus 48 prompts and responses in the posttest. To check for test-retest effects from pretest to posttest, the design included a set of new sentences in the

1. We would like to thank Ryan Lidster for helping us operationalize this taxonomy.

Explicit vs Non-Explicit Instruction	
+/– Attention on error (and not meaning)	
+/– Statement of difficulty and error	Instruction: "Look, this is difficult, this is where people make mistakes." Feedback: "You've made a mistake."
+/– Delineation of the target and error	Instruction: "Look at this specific word and its pronunciation." Feedback: "You pronounced __ like __."
+/– Means of correction	Instruction: "This is how you can correct it." Feedback: "This is what you should do."

Figure 2. Explicit and nonexplicit instruction taxonomy used by the teachers in the experimental groups.

posttest — that is, it included the same 24 prompts and responses from the pretest plus 24 new ones. The L1-English speakers completed the task only once with all 48 sentences, since no systematic differences would be expected across multiple repetitions. All the sentences contained vowels and suprasegmental features studied by the two experimental groups. Each sentence contained at least one target vowel (i.e. all 48 sentences contained one word with one of the four vowel sounds studied, /i/, /ɪ/, /æ/, and /ɛ/), as well as function words such as articles, prepositions, and pronouns, which are generally reduced in regular speech. Sentences with these characteristics were included because vowel reduction is a key aspect in the production of stress timing and rhythm in English (Cutler, Wales, Cooper, & Janssen, 2007; Ling, Grabe, & Nolan, 2000; Trofimovich & Baker, 2006), which were aspects targeted in instruction in the suprasegmental group.

Data elicitation

Learners were recruited in the three groups during the first week of classes. All volunteered to participate in the pretest and posttest that were to take place in weeks 2 and 6 of the session. All participants were tested individually. They were seated in a sound-isolated recording booth equipped with a Sennheiser microphone, and their productions were recorded using a Roland external sound card UA25, at a sampling rate of 44.1 Hz with a 16 bit resolution, on a mono channel. The participants were presented with audio prompts, which were recorded by two L1 speakers of American English (one male and one female so that they could distinguish their voices) at a normal speed. These were played from a PC, and the participants heard them in the cabin booth through high quality headphones (Sennheiser HD515).

Both tests had the same design and consisted of producing sentences in a delayed sentence-repetition task (e.g., Guion, et al., 2000; Ratner, 2000; Trofimovich

& Baker, 2006). In this task, the participants hear a question (e.g. one prompt in a male voice) followed by a specific answer (e.g. a response in a female voice). This first question-answer pair is then followed by the first question again (the prompt), to which the participants have to provide the answer heard previously. The task was entirely auditory, and no written sentences were presented at any time. All participants completed a total of 10 prompts and responses as warm-ups before the task, and these did not include vocabulary present in the target prompts, nor any of the target vowel sounds. The 10 prompts and responses in the warm-up were recorded by two speakers different from the ones in the actual task. Although other speech data collection techniques may elicit more natural and authentic speech (e.g., a picture description narrative), this particular type of task was selected because it elicited the production of similar and fluent speech samples that could be compared across all groups and participants in the comprehensibility rating task. Additionally, we chose an auditory presentation format to avoid the use of reading, which can affect pronunciation patterns.

Screening

Since this study was carried out in an intact-classroom context, it was not exempt from common variables that affect language teaching daily, such as student absences or unforeseen changes in the lesson plan. Given the short period of time for treatment, speech samples from those participants who were absent from more than 1 treatment session (due to absences or late enrollment) were not included in the comprehensibility rating task. In addition, because of its exclusively auditory nature, the delayed-sentence repetition task used in both the pretest and posttest was demanding for some participants. Not all learners were able to successfully repeat the full set of 24 and 48 responses in the pretest and posttest, respectively. Given the nature of the task and the need to memorize the answer to repeat, there were cases of hesitations or misremembered words. We chose to only use sentences that the students were able to repeat continuously and without hesitations, unnatural pauses, or replacing of words with others not included in the original stimuli (e.g. *father* for *dad*). This was very important in order to ensure that ratings would be only given to the same, grammatically correct, sentences. Otherwise, different ratings may have been due to factors related to grammaticality, naturalness or other characteristics of the sentences unrelated to pronunciation. Therefore, in order to avoid unintentional confounds in the comprehensibility rating task, only data from the 4 participants in each group that correctly repeated all stimuli were included in the final analysis. In total 24 sentences per participant were retained for the rating task: 8 sentences from the pretest and 16 from the posttest (i.e., the same 8 sentences from the pretest plus 8 brand new sentences; see the stimuli used

in the final analysis in Appendix A). This left 12 learners as well as 4 L1-English participants — randomly selected out of the 10 recorded for the L1-English sample baseline — who produced a total of 384 responses (12 L2 learners × 24 = 288 and 4 L1-English speakers × 24 = 96; 288 + 96 = 384). These responses were presented to a group of listeners (L1-English graduate students), blocked by sentence, in randomized blocks (and with pseudo-random order of sentences within blocks) for the rating task.

All learners who were retained as speakers for the comprehensibility rating had resided in the United States between 5 months and 1 year — except 1 Korean learner who had arrived 1 week before the beginning of the program and 1 Japanese learner who had spent 6 years in the United States (see Table 1 below for details). The median LOR was 6.5 months in the suprasegmental group (range: 5–8), 10.5 months in the vowel group (range: 9–72), and 6 months in the nonexplicit group (range: 0.25–12). Groups varied in how long they had learned English. Average length of learning was 3.86 years in the suprasegmental group (range: 0.58–10, SD = 5.32; median: 1), 8.75 years in the segmental group (range: 4–21, SD = 8.18; median: 5), and 4.25 years in the nonexplicit group (range: 1–7, SD = 2.75; median 4.5).

Raters

A group of 12 L1-English speakers carried out a comprehensibility rating task. They were graduate students in language teaching or linguistics and had not participated in the data elicitation. None of them had taught any of the L2 learners in this study. Previous studies (e.g., Kennedy & Trofimovich, 2008; Winke & Gass, 2013) on intelligibility and comprehensibility have pointed out that it is problematic when experienced raters such as teachers evaluate learners' comprehensibility based on how well they understand them, or based on their familiarity with specific L2 accents — as opposed to other, inexperienced raters with limited exposure to L2 speech. However, in order to obtain in-depth subjective reports from raters about the specific phonological features that they thought enhanced or constrained comprehensibility in the speech production of these L2 learners, we chose raters with a background in linguistics and language teaching for this study. Thus, all the raters filled out a short written questionnaire right after the rating task to point out what pronunciation features they thought affected their ratings of these sentences (see Appendix B).

Rating task

The L1-English raters listened to sentences through high quality headphones and rated the comprehensibility of each sentence on a 9-point Likert scale, where 1 means *extremely easy to understand* and 9 means *impossible to understand* (Derwing & Munro, 1997; Munro & Derwing, 1995). The lower the ratings the more comprehensible the sentences were perceived to be. We selected this rating scale because similar 9-point Likert scales have been used successfully in other studies (e.g., Derwing & Munro, 1997; Derwing, Munro, & Wiebe, 1998; Kennedy & Trofimovich, 2008; Munro & Derwing, 1995), and have been shown to yield high inter-reliability ratings. Levis (2005; 2006) explains that in spite of their differences, the terms *intelligibility* and *comprehensibility* are related. Intelligibility, in the broad sense given by Levis (2006), refers to the listeners' ability to understand speech and "is not usually distinguished from closely related terms such as comprehensibility" (p. 252). However, for the purpose of this study, because raters rated the same sentences for all speakers, intelligibility could not be our principal measure as we would have had to ask raters to transcribe the exact same sentences. Thus, we used *comprehensibility* as the main goal to attain in pronunciation instruction.

To avoid fatigue, the 384 sentences were split into two lists with an equal number of sentences from the pretest and posttest, evenly distributed across the four speaker groups. Half the listeners were assigned to one list, the other half to the other list. Additionally, a set of 15 warm-up prompts and sentences (not included in the analysis) were used to familiarize listeners with the use of the scale. All raters performed the task individually on a PC in a computer lab with high-quality headphones. All the sentences were presented in randomized blocks of the same sentence in which the first sentence of each block was uttered by one of the L1 speakers. This was done to avoid effects of increased familiarity with voices that could make the listeners rate some sentences more leniently than others. Before the rating task, the raters received instructions from one of the researchers orally. The same instructions were also given in written form through the computer at the beginning of the task. The raters were instructed to listen to each sentence and then rate them based on how comprehensible each sentence sounded. They were told to use the entire scale and also reminded that comprehensibility referred to their estimation of difficulty in understanding a sentence in spite of a speaker's accent. They were given the choice to listen to each sentence more than once, but they were asked not to change the ratings once they were marked. None of the raters reported any particular difficulty with this task.

Results

Comprehensibility ratings

The interrater reliability coefficients (Cronbach's alpha) computed across all ratings given for each list were high (.92 and .92), which indicated very strong agreement (LeBreton & Senter, 2008). Figure 3 displays the average posttest rating in each group for repeated sentences (i.e., those used in both pretest and posttest) and new sentences used only in the posttest. This also allowed verification that the sentences used in both tests were not easier or harder than the new sentences used in the posttest only. Finding that participants were perceived as more comprehensible in the new sentences (i.e., for just "receiving lower ratings according to the scale used) could compromise attributing any improvement of performance in the posttest (compared to the pretest) to the kind of training received during treatment.

For the suprasegmental group, the mean rating for the repeated vs. new sentences was 3.58 [95% CI = 2.59; 4.55] and 3.62 [2.62; 4.58] respectively. For the vowel group, these were 4.99 [4.01; 5.97] and 4.60 [3.65; 5.60].[2] For the nonexplicit group, the mean for repeated sentences was 4.14 [3.18; 5.15], and for new sentences, it was 4.13 [3.17; 5.13]. To examine whether L2 learners had equal performance for repeated and new sentences, a linear mixed effects model was conducted on the

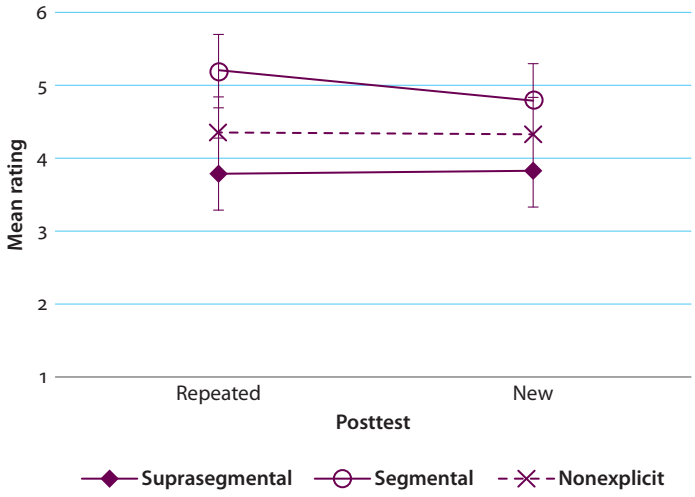

Figure 3. Average rating obtained in repeated and new sentences for each training group during posttest. Error bars enclose ± 1 standard error. (1 = extremely easy to understand; 9 = impossible to understand).

2. From here on, the mention of 95% Confidence Intervals (CI) will be omitted when presenting confidence intervals in the following format [lower; upper].

posttest ratings of the 3 L2-learner groups. It declared the factors *group* (suprasegmental, vowel, nonexplicit) and *sentence type* (repeated, new) as fixed effects, and *speakers*, *token*, and *listeners* as random effects. In this and all following analyses, the significance level for *p*-values was set at $\alpha = .05$. There was no effect of *sentence type* ($F_{1, 14.6} < 1$, $p > .1$) and no effect of *group* ($F_{2, 8.8} = 2.8$, $p > .1$) on the ratings, but most importantly, there was no interaction between the type of sentence and the group ($F_{2, 1097.1} = 2.5$, $p = .077$). No post hoc univariate comparisons reached significance. As further supported by the confidence intervals, which overlap for all new vs. repeated values, there was no difference in performance for repeated and new sentences for any of the groups.

These data thus allowed us to collapse the repeated and new sentences for the posttest ratings and to compare the ratings at the pretest and posttest. Two mean comprehensibility ratings were computed for each speaker: the first one across all pretest sentences, and the second one across all posttest sentences (the merged repeated and new sentences). The averaged ratings for each group at each time (pretest vs. posttest) are displayed in Figure 4.

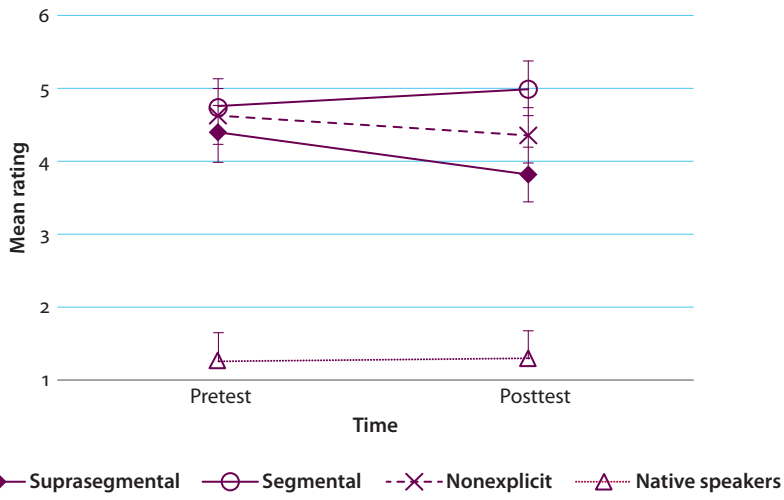

Figure 4. Average rating obtained in pretest and posttest for each group. Error bars enclose ± 1 standard error (1 = extremely easy to understand; 9 = impossible to understand).

Visually, the results are clear. Although L1-English speakers were consistently rated the most comprehensible — thereby verifying that the raters did not give random ratings — the L2 learners were similar in the pretest, but differed from each other in the posttest, which suggests that the differences observed in the posttest are likely a result of the specific treatment received.

For the suprasegmental group, the mean rating on the pretest was 4.16 [3.38; 4.95]; it was 4.51 [3.73; 5.30] for the vowel group, and 4.42 [3.64; 5.20] for the nonexplicit group. For the L1-English speakers, it was 1.06 [.28; 1.85]. As shown by the confidence intervals here as well, which overlap for all three learner groups, performance was similar on the pretest except for the native speakers. On the posttest, some differences emerge. For the suprasegmental group, the mean rating was 3.61 [2.84; 4.38]. For the vowel group, it was 4.80 [4.04; 5.57]. The CIs overlap a lot less in this case. For the nonexplicit group, the mean rating was 4.15 [3.38; 4.92]. And finally, for the native English group, it was 1.09 [.30; 1.87].

To corroborate these observations, a linear mixed effects model on the ratings was conducted in SPSS 22, declaring the factors *test* (pretest, posttest), and *group* (suprasegmental, vowel, nonexplicit, L1-English speakers) as fixed effects, and *speakers*, *token*, and *raters* as random effects. There was no main effect of *test* ($F_{1,\ 1371} = 2.3$, $p > .1$) on the ratings, but a significant effect of *group*, ($F_{3,\ 12.1} = 39.0$, $p < .001$), and a significant interaction of *group* and *test* ($F_{3,\ 2032} = 7.5$, $p < .001$) suggesting that the groups received different ratings at each time, and that this difference was modulated by the kind of treatment received. Clearly, however, the large effect of *group* is mainly due to the L1-English speakers. We restricted the same analysis to the L2-learner groups only. The mean rating of the learners for the pretest was 4.36 [3.58; 5.13]. Univariate tests (all Bonferroni-corrected) confirmed that the groups did not differ on the pretest ($F_{2,\ 10.7} = .3$, $p > .1$). At the posttest, the mean rating showed overall slightly more comprehensibility ($M = 4.19$ [3.42; 4.97]). This difference between pretest and posttest (the main effect of *test*), was now nonsignificant ($F_{1,\ 1526.3} = 3.2$, $p = .072$), and the main effect of *group* disappeared ($F_{2,\ 9} = 1.7$, $p > .1$). However, it is important to point out that the interaction remained significant ($F_{2,\ 1649.3} = 9.8$, $p < .001$), indicating that the performance at each test varied as a function of the treatment received.

Post hoc univariate analyses (all Bonferroni-corrected) of the effect of *test* on the ratings showed that the suprasegmental group improved between the pretest and the posttest: comprehensibility increased (i.e. the ratings decreased) by 0.54 points [95% CI for difference = 0.26; 0.82], a significant improvement ($F_{1,\ 1659} = 14.1$, $p < .001$). In contrast, the vowel group obtained higher ratings on the posttest, which indicated that their comprehensibility diminished by −0.31 points [95% CI for difference = −0.59; −0.02], also a significant difference ($F_{1,\ 1651} = 4.5$, $p = .033$). The slight improvement seen in the nonexplicit group was not significant (more comprehensible by 0.25 points [95% CI for difference = −0.03; 0.53], $F_{1,\ 1646} = 3.1$, $p = .078$). Unsurprisingly, the ratings for the L1-English speakers did not change significantly between the pretest and the posttest (a negligible difference

Table 1. Individual Variables (Group, L1, Length of Residence, Length of Learning) and Ratings at Time 1 and Time 2

Participant	Group	Rating time 1	Rating time 2	L1	LOR (in months)	Length of learning (in years)
G108	Suprasegmental	3.38	2.88	Turkish	7	1
G103	Suprasegmental	4.54	3.72	Arabic	5	10
G105	Suprasegmental	4.08	3.75	Arabic	6	unknown
G104	Suprasegmental	4.86	4.13	Turkish	8	0.7
G206	Segmental	4.02	3.41	Japanese	72	21
G205	Segmental	4.35	4.8	Korean	12	5
G207	Segmental	4.85	5.27	Korean	9	5
G208	Segmental	4.97	5.74	Japanese	9	4
G306	Nonexplicit	4.17	3.83	Turkish	7	7
G308	Nonexplicit	4.9	3.89	Arabic	12	1
G303	Nonexplicit	5.04	4.17	Korean	0.25	6
G307	Nonexplicit	3.77	4.61	Arabic	5	3

Note: LOR = Length of Residence

of −0.02 points, $F_{1,\ 1773} < 1$).[3] However, since a limited number of subjects were observed, and given their mixed backgrounds, it was important to ascertain that there were no extraordinary performances in some individuals that could have had a significant impact on the results of the group, or that the results in one group were not due to an overall higher length of residence (LOR), or to an L1 effect.[4] Table 1 presents the ratings for each individual together with demographic information.

Simple Pearson correlations were performed between the rating obtained at time 2 and the factors LOR (in months), and Length of Learning (in years). Neither of these approached significance (LOR $r = -.256$, $p = .2$; Length of learning $r = -.237$, $p = .2$), suggesting the absence of any systematic pattern in the data which could be due to one of these factors rather than to the type of instruction received. However, effects of L1 background are difficult to evaluate given the mixed ı and the nonoverlap across training groups. We return to this issue in the discussion.

3. As mentioned before, the L1-English speakers recorded all the 48 sentences only once. For the purpose of the analyses of comprehensibility ratings, we split their recordings into pretest and posttest sentences like in the L2-learner groups.

4. An anonymous reviewer rightly pointed out that LOR does not equate with length of exposure. While we certainly agree with this, we used LOR only as an indirect measure of exposure for participants immersed in a context like the one of this study.

To reiterate, the only reason why the language backgrounds, LOR and Length of Learning differed across groups is because we used intact classes in an ESL program. The study was classroom-based, and students with various L1 backgrounds and learning experiences enrol. For logistics in terms of design (see Mackey & Gass, 2005), and in order to keep our study as realistic as possible, we could not control or modify this factor, since this is the case in most ESL programs. However, even if we could not control for these important variables (L1, LOR, etc.) given the classroom-based nature of the study, we are confident that there is most likely no significant effect of LOR or L1 background on the observed changes in performance.

Vowel acoustic analysis

The target vowels /i/, /ɪ/, /æ/, and /ɛ/ of all productions were analyzed with Praat (Boersma & Weenink, 2015). Since the word pairs containing the vowels were different in the two posttests, it was necessary to analyze the acoustic properties of the vowels for the two posttests (repeated vs. new) separately, unlike the results for comprehensibility reported above in Figure 4. To obtain spectral measures, F1, F2 and F0 were extracted from the midpoint of the steady-state portion of the second formant of the vowel. The two dimensions of tongue height and tongue fronting, which play a role in the spectral differentiation of our four target vowels, were obtained on a Bark-converted scale to normalize for vocal tract differences across speakers. The vowels examined also differ in duration. For instance, the two high vowels are commonly distinguished by duration, such that the tense /i/ is usually longer than the lax /ɪ/. Similarly, the two mid vowels differ in duration, /æ/ often being longer than /ɛ/. To capture these differences, we examined the four vowels pairwise, allowing us to compute a duration ratio to express how clearly speakers temporally differentiate the vowel pairs (the two high vowels, and the two mid vowels). For each pair, a ratio of 1 indicates that both vowels are of equal duration. A ratio below 1 indicates that /ɪ/ and /ɛ/ are shorter than /i/ and /æ/ respectively (e.g., a ratio of .5 means that one vowel is half as long as the other). A ratio above 1 indicates that /ɪ/ and /ɛ/ are longer than /i/ and /æ/ respectively. The duration data were submitted to a series of nonparametric Mann-Whitney tests.

Examining the duration ratios indicates that L1-English speakers produced /æ/ and /ɛ/ with a mean ratio of .59, indicating that /ɛ/ is about half as long as its mid counterpart /æ/. For the high vowels, this ratio was .78, indicating a slightly smaller difference in duration between /i/ and /ɪ/, with /ɪ/ being somewhat shorter than /i/. For the L2 learners, at Time 1, duration ratios were higher than the L1 speakers' for the mid vowels, indicating that they were producing both /æ/, and /ɛ/ with roughly equal durations. For high vowels, their Time 1 values were close to those of L1-English speakers. Over time, we observed an overall improvement

such that all learner groups produced duration ratios that were closer to the L1 speakers' range at the end of the study. This development is shown in Figure 5, for mid vowels (upper panel, /æ/-/ɛ/) and for high vowels (lower panel, /i/-/ɪ/).

For analysis, Mann-Whitney tests compared each group to the L1 speakers, first for vowels in sentences from the pretest, then for those from the posttest (including both repeated and new words).[5] There was no significant difference in pretest or posttest for the high-vowel duration ratio between any learner group and the L1 speakers (all $p > .1$). For the mid-vowel ratios however, the *suprasegmental* and *vowel* groups (but not the *nonexplicit* group) differed significantly from the L1 speakers both on the pretest and on the posttest. At the time of the pretest, both suprasegmental and vowel groups had higher ratios that were slightly above 1, indicating that their two mid vowels were of roughly the same duration (Suprasegmental vs. L1-English: Mann-Whitney $U = 0$, $z = -2.7$, $p = .004$; Vowel vs. L1-English: Mann-Whitney $U = 0$, $z = -2.7$, $p = .004$). The nonexplicit group did not differ from the natives (both $p > .1$) in terms of duration ratio. At the time of the posttest, the vowel duration ratios were closer to that of L1 speakers, but were still significantly higher (Suprasegmental vs. L1-English: Mann-Whitney $U = 3$, $z = -3.04$, $p < .001$; Vowel vs. L1-English: Mann-Whitney $U = 11$, $z = -2.2$, $p = .028$).

The normalized Bark-converted measures for tongue height (B2-B1) and frontness (B1-B0) are plotted in Figure 6, with height on the Y-axis and frontness on the X-axis, but were not analyzed statistically. The data were plotted by obtaining a total of two values (frontness and height) for each subject (across all test words containing that vowel) at each time point, and for each vowel.

We observe the following pattern in terms of the spectral difference between vowels: The L1-English speakers show a very clear spectral differentiation between all four vowels, where /i/ is the most fronted and highest, followed by /ɪ/, /ɛ/ and /æ/, which is the least fronted and the lowest vowel. We also see that at the time of the pretest, the learner groups do not manage to successfully produce a spectral difference between the two high vowels, or between the two mid vowels. Substantial overlap in both dimensions is visible. This also confirms that choosing these four vowel contrasts was justified from a pedagogical point of view, and that these four vowels did pose problems for the learners in the vowel group at the time of the pretest. Yet over time, unlike the *suprasegmental* and *nonexplicit* groups, the *vowel* group progresses to a clearer separation of the four vowels. Of the three learner groups, this group obtains the vowel plot that is closest to the L1-English speaker one. In addition, we computed the Euclidean distance (Liljencrants & Lindblom,

5. All sentences in the pretest and the posttest (i.e., repeated sentences from the pretest plus new sentences participants had not encountered before) contained one target word with one of the target vowels studied by the vowel group.

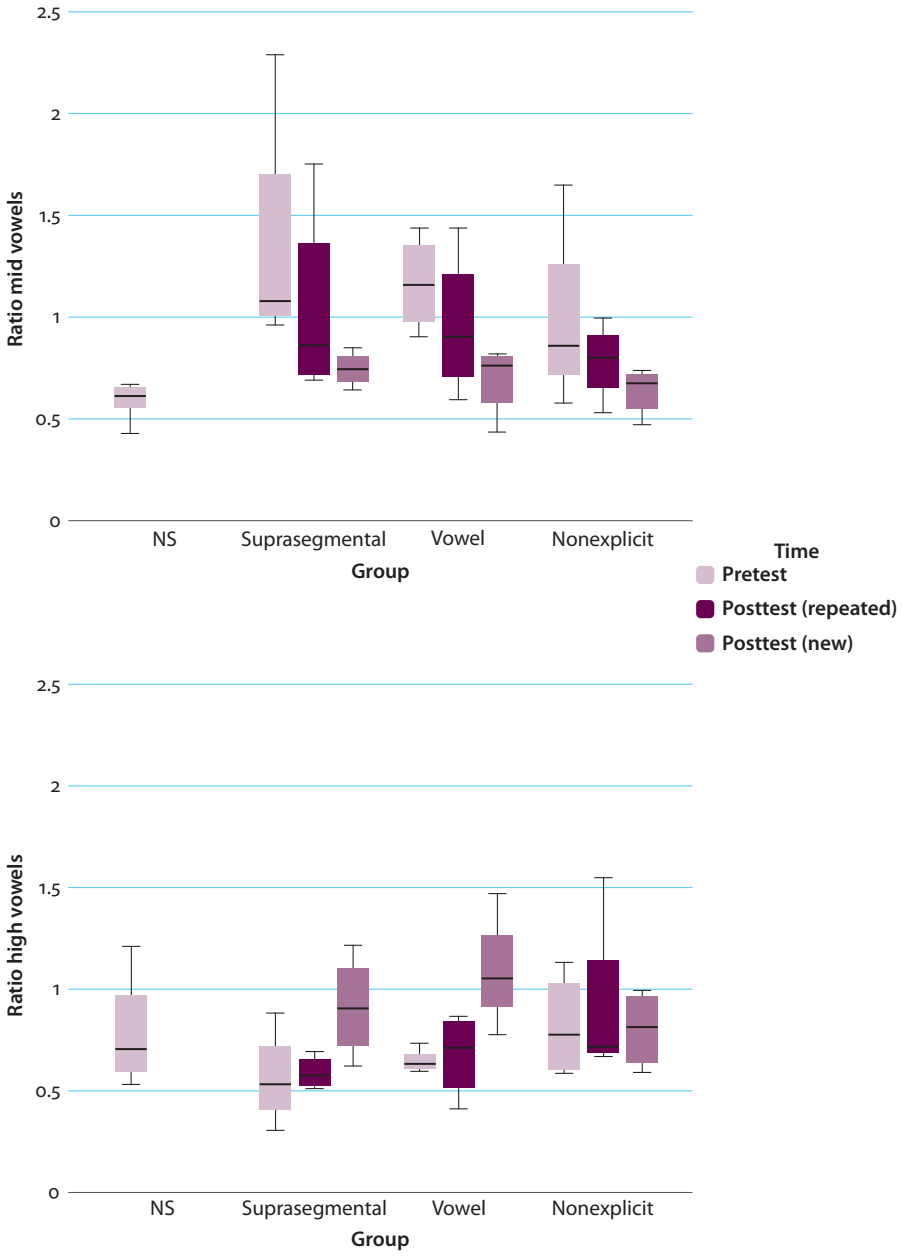

Figure 5. Boxplots showing the duration ratio of the two mid vowels (on top) and the two high vowels (below), by group and time.

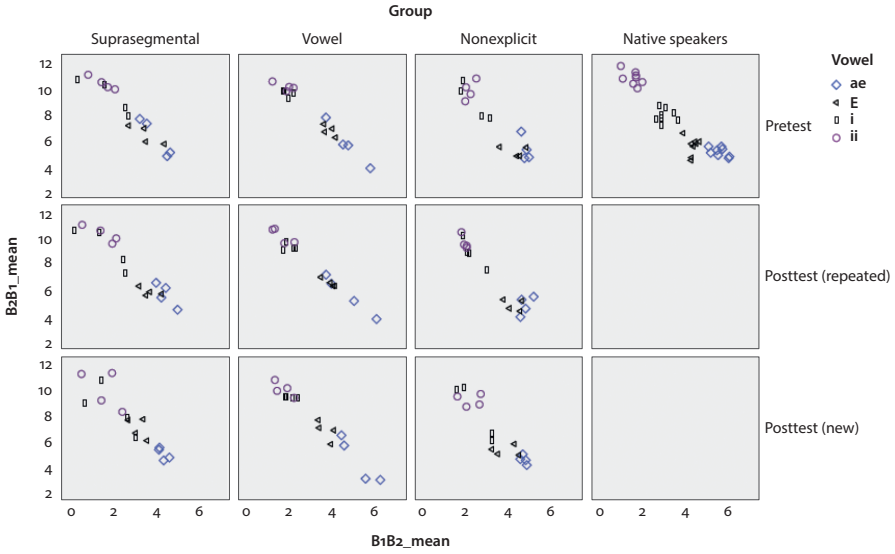

Figure 6. Scatterplot of frontness (B1-B0) and height (B2-B1) values for each participant at each time, for the four vowels. Note: "ae" = [æ]; "E" = [ɛ]; "ii" = [iː] and "i" = [ɪ].

1972) between the two vowels in each contrast (/ɪ/-/i/ and /ɛ/-/æ/), to compare how every individual speaker realizes the spectral distance between the vowels. We only included the pretest and the "repeated" posttest because these contained the same word pairs. Only the group trained on vowels increased their Euclidean Distance value between the pretest and the posttest, therefore being closer to the L1-English speakers. The other two groups did not. The mean Euclidean Distance (ED) of the L1-English speakers across the two vowel contrasts was 2.41 (*SD* = 1.16). Again, of the three learner groups, only the vowel group improved (i.e. increased) their ED. The mean ED on the pretest was 1.20 (*SD* = 1.28), and on the posttest, it was 1.49 (*SD* = 1.34). The two other groups showed a reduction in ED from the pretest to the posttest. For the suprasegmental group, the mean decreased from 1.72 (*SD* = 1.15) to 1.38 (*SD* = 0.97). For the nonexplicit group, the mean decreased from 1.36 (*SD* = 0.68) to 1.12 (*SD* = 0.59). Overall, there was a clear tendency for the vowel group to show some improvement in their realization of these vowel contrasts, whereas there was not for the two other groups.

Given the small sample size of our data, it is difficult to conduct meaningful statistical analyses on these values. In addition, given constraints on materials construction, words were not sufficiently well controlled for an in-depth acoustic analysis of this kind. Therefore, our findings are to be interpreted with caution in terms of statistical differences between the learners and the L1 speakers. However, the comparison among learner groups does suggest that the vowel-focused group also

improved most their pronunciation of the four vowels studied during treatment, as compared to the other groups.

Analysis of raters' comments

In general terms, the L1-English raters pointed out both segmental and suprasegmental aspects to explain their reasons for rating sentences as more or less comprehensible. Accurate pronunciation of very specific segments — almost demanding native-like renderings in the case of some raters — was mentioned as a contributing factor to improved comprehensibility in sentences. For example, comments like "individual segments produced in a targetlike (nativelike) way," "vowels and consonants produced nativelike and produced consistently," "unaspirated /p/ in 'peaches' was a significant factor because it changes the meaning" or "addition of consonants and vowels, if the speaker inserted sounds it was less comprehensible" suggest that there was a preference in some raters to have vowels and/or consonants articulated in a nativelike manner. Similarly, a rater stated that sentences sounded more comprehensible "if the important signifying words were pronounced clearly," and another one stated that "if the content words were entirely pronounced" those sentences were determined as more comprehensible, clearly pointing at the importance of accurate enunciation of individual vowels and consonants. However, this was not always the case, and other raters also pointed out how mispronunciation of individual segments did not necessarily affect global comprehensibility. Some raters may have tolerated mispronunciation of some errors more than others as long as their salience did not affect the whole comprehension of the sentence. As one rater stated, "there were certain nonnative features I was more accepting of than others, but I can't name others besides /l/ and /r/." Only one rater explicitly pointed out that mispronunciation of vowels did not necessarily affect the complete meaning of some sentences: "Interestingly, I don't think vowel quality played as big of a role for me, [o]nion is as intelligible as [ʌ]nion." Despite this isolated view, these comments suggest that clear realization of target vowels and consonants, especially in stressed syllables, continues to play an important role in determining comprehensibility or intelligibility (see also Jenner, 1989; Zielinski, 2008).

Clearly however, individual segments were not the sole determining factor underlying global comprehensibility ratings. The raters also mentioned that appropriate use of suprasegmental features and fluency were both related to better comprehensibility in some sentences. Whereas factors like "intonation and word stress," "speech rate more nativelike," or "more fluid speech" were mentioned as criteria for judging sentences as comprehensible, a more prominent role of prosody was also included as necessary to determine the degree of comprehensibility

of different sentences. For instance, raters' comments pointed out that "intonation, stress on words where natives would stress other words; less fluent, pausing between words or sounds," contributed to lower comprehensibility. Similarly, comments like "intonation and speech rhythm not used properly" were also part of the criteria used to rate sentences as less comprehensible. Comments like these are important because they indicate features the raters paid attention to during the rating task. During treatment, at least in the suprasegmental group, the learners were explicitly taught to stress the important words (i.e., content words) in sentences and reduce function words to maintain the stress-timing characteristic of English. Thus, it is possible that raters found sentences with those characteristics more natural and easier to comprehend as opposed to those in which L2 speakers pronounced all words with the same stress. In a related manner, fluency was mentioned as important for comprehensibility. "Lack of fluency, wrong pauses" were some of the comments given to point out lack of comprehensibility in sentences.

Overall, segmentals and suprasegmentals were mentioned to similar extents in raters' comments, suggesting that their importance in determining comprehensibility can be seen as comparable. Even if it is necessary to use caution when relating such explicit opinions to how listeners actually decide on their rating (Hayes-Harb & Hacking, 2011), all raters were language-teaching or linguistics graduate students whose coursework included instruction and practice in diagnosing pronunciation difficulties of L2 speakers.

General discussion

This study investigated whether a short-term explicit pronunciation instruction component yields comprehensibility gains for adult ESL learners when incorporated into regular instruction. The picture presented by the data is mixed and should be interpreted carefully. On the one hand, explicit phonetic instruction resulted in comprehensibility gains over the course of 3 weeks, but only for the group trained on suprasegmental features. The group trained in vowels seemed to present some gains in the production of the four vowels analyzed, but its comprehensibility for sentence production did not improve (it even decreased), as opposed to the group trained in suprasegmentals at the end of the study. On the other hand, comprehensibility did not change significantly for the nonexplicit group. In this sense, explicit phonetic instruction can benefit L2 learners when it is not restricted to vowel training only. These findings add to the growing evidence arguing in favor of explicit pronunciation instruction (Couper, 2003; Derwing, et al., 1998; Elliot, 1997; Lord, 2005). However, and as mentioned above, the group instructed on English vowels did not experience comprehensibility gains. The

sentences obtained from this group after the treatment were rated as significantly less comprehensible. Our second research question regarding whether instruction that focuses on suprasegmental features leads to larger comprehensibility gains — as opposed to a focus on segmental features only — can at first glance be answered affirmatively: broader-scope instruction that targets suprasegmentals yields rapid improvements in comprehensibility. However, as we argue below, inferring from this finding that instruction targeting segmentals is ineffective is not warranted. Overall, these findings confirm previous studies that demonstrate a relationship between the appropriate use of suprasegmentals and comprehensibility in L2 speech (e.g., Derwing & Munro, 1997; Kang et al., 2010; Munro, 1995; Munro & Derwing, 1995). In terms of instruction, our results align with studies by Derwing et al. (1998), who observed significant comprehensibility gains for those trained on suprasegmentals. In contrast, when explicit instruction focuses on segmental features, at least in the case of a limited number of specific vowels as in the current study, an increased attention to this specific feature (to the exclusion of others) might leave fewer attentional resources to allocate to complexity and fluency, and thus may slow down pronunciation improvements in the short term, decreasing comprehensibility (see Derwing et al., 1998; Schmidt, 2001).

Although this pattern of findings appears to suggest that explicit phonetic instruction on segmental features might be detrimental to comprehensibility, such a conclusion is not warranted for a number of reasons, and it is important to keep in mind that the comprehensibility gains obtained by the suprasegmental group — although remarkable for such a small sample and a short period of time — are not larger than half a point. It is possible that this group would have improved even more had instruction attended to segmental features (both vowels and consonants) as well. In fact, some studies remark that nontarget segmental realizations can seriously reduce intelligibility and comprehensibility (Zielinski, 2008). Additionally, as suggested by the raters in the open questionnaire given after the rating task, the production of more accurate segmentals was reported as playing a role in determining comprehensibility to an extent comparable to accurate suprasegmentals.

Second, a possible reason for the difference observed between the segmental and suprasegmental groups is the difference in scope of the pronunciation treatments. Whereas the vowel group focused on only four vowel sounds, the suprasegmental group was instructed in a range of more global prosodic phenomena, which was accompanied by practice on larger chunks of speech. It is possible that the vowel group's ratings would have improved if they had been trained on the entire segmental inventory of English, including both consonants and vowels, in a way similar to Derwing et al., (1998) or Derwing and Rossiter (2003), or on different segmentals that are perhaps more crucial to comprehensibility. Given the

limited focus of the segmental instruction, it is possible that learners in the vowel group focused their attention on the correct realization of these specific four vowels that were the target of instruction only, to the exclusion of the other dimensions. This may have made their speech sound less comprehensible because of pronunciation patterns affecting the other parts of the sentences that did not contain one of the target vowels. This possibility — rather than the focus of instruction per se — might be responsible for the lower comprehensibility ratings obtained by this group. The acoustic analysis seems to indicate that the L2 participants in the vowel group succeeded in modifying their vowels to approximate a more native-like spectral realization by the end of treatment (as opposed to the L2 participants in the other two groups), but the limited sample and the different L1 backgrounds of the participants presented a picture too complex to allow us to fully determine to what extent the production of these four vowels made a difference in the comprehensibility ratings. However, the ratings clearly indicate that the speech of participants in the vowel group was perceived as less comprehensible when compared to the group trained on suprasegmentals. After all, as pointed out by Thomson and Derwing (2015), measurable acoustic changes in vowels are not always perceptible to listeners, and in the end, it is what listeners perceive that matters for ratings.

A third possibility we cannot rule out is that segmental features need more time to develop in L2 learners than suprasegmental features. Whereas much L2 phonology research has focused on the development of L2 phonetic categories in speech perception and production (see Sebastian-Gallés, 2005; and Strange and Shafer, 2008, for reviews), comparatively less research has been conducted on the development of suprasegmental dimensions in L2 (e.g. Trofimovich & Baker, 2006). Even though suprasegmentals appear to follow the same developmental path as segmentals, it is unclear whether they respond to instruction in the same way. As a result, it is possible that improvements in suprasegmental domains take less time to emerge than improvements in segmental domains. When considering instruction duration, our study suggests that explicit phonetic instruction can yield rapid comprehensibility gains: focusing learners' attention to the key L2 phonological features that are necessary to improve L2 production on suprasegmental features of pronunciation appears to be most effective in the short-term, as in the current study (see Pennington and Ellis, 2000; Schmidt, 2001 for issues of attention to L2 features). It is important to remember that the L2 learners in this study were instructed for 3 weeks only, which is a much shorter time than 10 weeks as in the study by Derwing, et al. (1998).

Finally, it is also possible that the non-overlap in L1 backgrounds in the different groups may have interacted in unexpected ways with the training. Even though every effort was made to choose linguistic dimensions known to pose problems to speakers from these L1s (see background section), we cannot ascertain that all

dimensions are equally difficult for speakers of these L1 backgrounds, or that the groups reacted in similar ways to pronunciation training regardless of their L1. While clearly a limitation of this study which reduces the generalizability of the findings, this fact also highlights the difficulty of conducting studies in classroom contexts, and points to the need of more research in this area: obtaining converging data across varied L1 backgrounds would strengthen claims about the foci of pronunciation components.

Our findings also suggest that an approach favoring the development of both bottom-up and top-down skills through explicit instruction can be beneficial for learners. Interestingly, we observed that improvements were also generalized to new sentences that learners had not encountered before. Despite the fact that we did not include a delayed posttest to check for long-lasting effects (unlike Couper, 2006), the generalization obtained might be an indication that a short-term explicit pronunciation component could yield general comprehensibility gains. A better test of this possibility, however, would require a less "constrained" task than the delayed sentence-repetition task, likely including narratives or a picture description.

To conclude, we argue that an explicit pronunciation curricular component, integrated in oral communication classes — including both segmental and suprasegmental pronunciation features — can significantly improve comprehensibility, even in a short period of time. These results demonstrate that it is possible to effectively instruct L2 learners in pronunciation in real classroom contexts and in a short time frame, without the need for a long or intensive dedicated pronunciation course. That we observed measurable changes in participants after this short-term pronunciation module is promising, despite a small sample. Furthermore, the subjects who participated are, generally speaking, representative of the typical learners enrolled in classes in intensive ESL programs. Therefore, despite the sample being small, we are confident that our findings do in fact provide evidence of actual benefits of pedagogical intervention on speech production, and are likely replicable. However, it remains clear that more classroom pronunciation research with larger samples is needed. Taken together, our findings add support to the voice of a growing circle of researchers (e.g., Foote, Holtby, & Derwing, 2011; Hinkel, 2006; Isaacs, 2009; Levis, 1999; Sicola & Darcy, 2015; Zielinski & Yates, 2014, among others) advocating that integrating pronunciation instruction into the language classroom, even for small amounts of time every week — is possible and can make a positive difference for L2 learners.

Acknowledgments

We would like to thank our L2 participants and the three collaborating teachers Rebecca Mahan, Denise Shettle and Valerie Cross, without whom this research could not have been carried out. We also thank Ryan Lidster, Doreen Ewert, Kyle Swanson, the Second Language Psycholinguistics Lab members, the Department of Second Language Studies at Indiana University, as well as audiences at AAAL 2012, for insightful feedback.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning*, 59, 249–306. doi: 10.1111/j.1467-9922.2009.00507.x

Anani, M. (1989). Incorrect stress placement in the case of Arab learners of English. *IRAL : International Review of Applied Linguistics in Language Teaching*, 27, 15–21. doi: 10.1515/iral.1989.27.1.15

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. doi: 10.1111/j.1467-1770.1992.tb01043.x

Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.

Baker, A. (2014). Exploring teachers' knowledge of second language pronunciation techniques: Teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, 48, 136–163. doi: 10.1002/tesq.99

Bayraktaroğlu, S. (2008). Orthographic interference and the teaching of British pronunciation to Turkish learners. *The Journal of Language and Linguistic Studies*, 4, 107–143.

Benson, P., & Lor, W. (1999). Conceptions of language and language learning. *System*, 27, 459–472. doi: 10.1016/S0346-251X(99)00045-7

Best, C.T., & Tyler, M.D. (2007). Nonnative and second language speech perception: Commonalities and complementarities. In O.-S. Bohn & M.J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins. doi: 10.1075/lllt.17.07bes

Boersma, P., & Weenink, D. (2015). Praat: Doing phonetics by computer [Computer program] Version 5.4.06. Retrieved 7 March 2015 from <http://www.praat.org/>

Bradlow, A., Akahane-Yamada, R., Pisoni, D.B., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61, 977–985. doi: 10.3758/BF03206911

Brown, A. (1991). Functional load and the teaching of pronunciation. In A. Brown (Ed.), *Teaching English pronunciation: A book of readings* (pp. 211–224). London: Routledge.

Celce-Murcia, M. (Ed.). (2001). *Teaching English as a second or foreign language* (3rd ed.). Boston: Heinle & Heinle.

Celce-Murcia, M., Brinton, D.M., Goodwin, J.M., & Griner, B. (2010). *Teaching pronunciation: A coursebook and reference guide* (2nd ed.). New York, NY: Cambridge University Press.

Chamot, A., Barnhard, S., El-Dinary, P., & Robbins, J. (1999). *The learning strategies handbook*. London: Longman.

Couper, G. (2003). The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect*, 18, 53–70.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect*, 21, 46–66.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W.J. Barry (Eds.). *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 1913–1916). Saarbrücken.

Darcy, I., Dekydtspotter, L., Sprouse, R., Glover, J., Kaden, C., McGuire, M., & Scott, J.H.G. (2012a). Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English-L2 French acquisition. *Second Language Research*, 28, 5–40. doi:10.1177/0267658311423455

Darcy, I., Ewert, D., & Lidster, R. (2012b). Bringing pronunciation instruction back into the classroom: An ESL Teachers' pronunciation "toolbox". In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference* (pp. 93–108). Ames, IA: Iowa State University.

Darcy, I., & Krüger, F. (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics*, 40, 568–581. doi:10.1016/j.wocn.2012.05.001

Darcy, I., Park, H., & Yang, C.-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, 40, 63–72. doi:10.1016/j.lindif.2015.04.005

Darcy, I., Peperkamp, S., & Dupoux, E. (2007). Bilinguals play by the rules. Perceptual compensation for assimilation in late L2-learners. In J. Cole & J.I. Hualde (Eds.), *Laboratory Phonology 9* (pp. 411–442). Berlin: Mouton de Gruyter.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Kügler, C. Féry, & R. van de Vijver (Eds.), *Variation and gradience in phonetics and phonology* (pp. 265–309). Berlin: Mouton de Gruyter.

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.M. (2008). Curriculum issues in teaching pronunciation to second language learners. In J.G. Hansen Edwards & M.L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 347–371). Amsterdam: John Benjamins. doi:10.1075/sibil.36.17der

Derwing, T.M., & Munro, M.J. (1997). Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition*, 19, 1–16. doi:10.1017/S0272263197001010

Derwing, T.M., Munro, M.J., & Wiebe, G. (1997). Pronunciation instruction for 'fossilized' learners: Can it help? *Applied Language Learning*, 8, 185–203.

Derwing, T.M., Munro, M.J., & Wiebe, G.E. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning*, 48, 393–410. doi:10.1111/0023-8333.00047

Derwing, T.M., Thomson, R.I., & Munro, M.J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, 34, 183–193. doi:10.1016/j.system.2006.01.005

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology-Human Perception and Performance*, 25, 1568–1578. doi:10.1037/0096-1523.25.6.1568

Dupoux, E., Pallier, C., Sebastián-Gallés, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, 36, 406–421. doi:10.1006/jmla.1996.2500

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, 106, 682–706. doi:10.1016/j.cognition.2007.04.001

Elliot, A.R. (1997). On the teaching and acquisition of pronunciation within a communicative approach. *Hispania*, 80, 95–108. doi:10.2307/345983

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning, testing, and teaching*. Bristol, UK: Multilingual Matters.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly*, 39, 399–423. doi:10.2307/3588487

Flege, J.E. (1995). Second language speech learning. Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience. Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Flege, J.E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470. doi:10.1006/jpho.1997.0052

Flege, J.E., Frieda, E.M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, 25, 169–186. doi:10.1006/jpho.1996.0040

Foote, J.A., Holtby, A., & Derwing, T.M. (2011). Survey of pronunciation teaching in adult ESL programs in Canada, 2010. *TESL Canada Journal*, 29, 1–22

Goodwin, J. (2014). Teaching Pronunciation. In M. Celce-Murcia, D.M. Brinton, & M.A. Snow (Eds.), *Teaching English as a second or foreign language (4th ed.)* (pp. 136–152). Boston: Heinle & Heinle.

Grant, L. (Ed.). (2014). *Pronunciation myths: Applying second language research to classroom teaching*. Ann Arbor, MI: University of Michigan Press.

Guion, S.G. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition*, 27, 503–533. doi:10.1017/S0272263105050230

Guion, S.G., Flege, J.E., & Loftin, J.D. (2000). The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics*, 28, 27–42. doi:10.1006/jpho.2000.0104

Guion, S.G., Flege, J.E., Liu, S.H., & Yeni-Komshian, G.H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, 21, 205–228. doi:10.1017/S0142716400002034

Hahn, L.D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38, 201–223. doi:10.2307/3588378

Han, Z. (2004). *Fossilization in adult second language acquisition*. Bristol, UK: Multilingual Matters.

Hayes-Harb, R., & Hacking, J. (2011, October) Variability in Listeners' Bases for Accentedness Judgments of L2 Speech. Poster presented to the *Second Language Research Forum*, Ames, IA.

Hinkel, E. (2006). Current perspectives on teaching the four skills. *TESOL Quarterly*, 40, 109–131. doi:10.2307/40264513

Isaacs, T. (2009). Integrating form and meaning in L2 pronunciation instruction. *TESL Canada Journal*, 27, 1–12.

Jenner, B. (1989). Teaching pronunciation: The common core. *Speak Out!* 4, 2–4.

Kang, O. (2010). Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38, 301–315. doi: 10.1016/j.system.2010.01.005

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal*, 94, 554–566. doi: 10.1111/j.1540-4781.2010.01091.x

Kennedy, S., & Trofimovich, P. (2008). Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review*, 64, 459–489. doi: 10.3138/cmlr.64.3.459

Kennedy, S., & Trofimovich, P. (2010). Language awareness and second language pronunciation: A classroom study. *Language Awareness*, 19, 171–185. doi: 10.1080/09658416.2010.486439

Kissling, E.M. (2013). Teaching pronunciation: Is explicit phonetic instruction beneficial for FL learners? *Modern Language Journal*, 97, 720–744. doi: 10.1111/j.1540-4781.2013.12029.x

Lazaranton, A. (2001). Teaching oral skills. In M. Celce-Murcia (Ed.), *Teaching English as a second or foreign language (3rd ed*., pp. 103–115). Boston, MA: Heinle & Heinle.

LeBreton, J.M., & Senter, J.L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods*, 11, 815–852. doi: 10.1177/1094428106296642

Levis, J.M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal*, 12, 13–19.

Levis, J.M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33, 37–63. doi: 10.2307/3588190

Levis, J.M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39, 369–377. doi: 10.2307/3588485

Levis, J.M. (2006). Pronunciation and assessment of spoken language. In R. Hughes (Ed.), *Spoken English, TESOL and applied linguistics* (pp. 245–270). Houndmills: Palgrave Macmillan.

Levy, E.S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics*, 36, 141–157. doi: 10.1016/j.wocn.2007.03.001

Liljencrants, J., & Lindblom, B. (1972) Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, 48, 839–862. doi: 10.2307/411991

Ling, L.E.E., Grabe, E., & Nolan, F. (2000). Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. *Language & Speech*, 43, 377–401. doi: 10.1177/00238309000430040301

Lord, G. (2005). (How) Can we teach foreign language pronunciation? On the effects of a Spanish phonetics course. *Hispania*, 88, 557–567. doi: 10.2307/20063159

Mackey, A., & Gass, S.M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum Associates.

Magen, H.S. (1998). The perception of foreign-accented speech. *Journal of Phonetics*, 26, 381–400. doi: 10.1006/jpho.1998.0081

McAllister, R., Flege, J.E., & Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30, 229–258. doi: 10.1006/jpho.2002.0174

Missaglia, F. (1999). Contrastive prosody in SLA: An empirical study with Italian learners of German. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 551–554).San Francisco, CA.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481–520. doi: 10.2307/3586981

Munro, M.J., & Derwing, T.M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, 58, 479–502. doi: 10.1111/j.1467-9922.2008.00448.x

Munro, M.J. (1995). Nonsegmental factors in foreign accent: Ratings of filtered speech. *Studies in Second Language Acquisition*, 17, 17–34. doi: 10.1017/S0272263100013735

Munro, M.J., & Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73–97. doi: 10.1111/j.1467-1770.1995.tb00963.x

Munro, M.J., & Derwing, T.M. (2001). Modelling perceptions of the comprehensibility and accentedness of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, 23, 451–468.

Munro, M.J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36, 39–66.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language & Hearing Research*, 50, 1496–1509. doi: 10.1044/1092-4388(2007/103)

Pallier, C., Bosch, L., & Sebastian-Gallés, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, 64, B9–B17. doi: 10.1016/S0010-0277(97)00030-9

Pennington, M.C., & Ellis, N.C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *Modern Language Journal*, 84, 372–389. doi: 10.1111/0026-7902.00075

Perlmutter, M. (1989). Intelligibility rating of L2 speech pre- and postintervention. *Perceptual and Motor Skills*, 68, 515–521. doi: 10.2466/pms.1989.68.2.515

Piske, T., MacKay, I.R.A., & Flege, J.E. (2001). Factors affecting the degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191–215. doi: 10.1006/jpho.2001.0134

Purcell, E.T., & Suter, R.W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30, 271–287. doi: 10.1111/j.1467-1770.1980.tb00319.x

Ratner, N.B. (2000). Elicited imitation and other methods for the analysis of trade-offs between speech and language skills in children. In L. Menn & N.B Ratner (Eds.), *Methods for studying language production* (pp. 291–312). Mahwah, NJ: Lawrence Erlbaum Associates.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). New York, NY: Cambridge University Press. doi: 10.1017/CBO9781139524780.003

Schoonmaker-Gates, E. (2012). *Perception of foreign accent in Spanish by native and nonnative listeners*. Unpublished doctoral dissertation, Indiana University.

Sebastian-Gallés, N. (2005). Cross-language speech perception. In D.B. Pisoni & R.E. Remez (Eds.), *The handbook of speech perception* (pp. 546–566). Oxford: Blackwell. doi: 10.1002/9780470757024.ch22

Setter, J. (2006). Speech rhythm in World Englishes: The case of Hong Kong. *TESOL Quarterly, 40(4)*, 763-782

Sicola, L., & Darcy, I. (2015). Integrating pronunciation into the language classroom. In M. Reed & J.M. Levis (Eds.), *The handbook of English pronunciation* (pp. 471–487). Hoboken, NJ: John Wiley & Sons.

Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S.A., Nishi, K., & Jenkins, J.J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26, 311–344. doi: 10.1006/jpho.1998.0078

Strange, W., & Shafer, V. (2008). Speech perception in second language learners. The re-education of selective perception. In J.G. Hansen Edwards & M.L. Zampini (Eds.), *Phonology and*

second language acquisition (pp. 153–191). Amsterdam: John Benjamins. doi: 10.1075/sibil.36.09str

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. *Journal of Phonetics*, 25, 1–24. doi: 10.1006/jpho.1996.0031

Thomson, R.I., & Derwing, T.M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36, 326–344. doi: 10.1093/applin/amu076

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28, 1–30. doi: 10.1017/S0272263106060013

VanPatten, B. (1988). How juries get hung: Problems with the evidence for a focus on form in teaching. *Language Learning*, 38, 243–260. doi: 10.1111/j.1467-1770.1988.tb00411.x

Wang, W., Jongman, A., & Sereno, J.A. (2003). Acoustic and perceptual evaluations of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113, 1033–1043. doi: 10.1121/1.1531176

Wang, W., Spence, M.M., Jongman, A., & Sereno, J.A. (1999). Training American listeners to perceive Mandaring tones. *Journal of the Acoustical Society of America*, 106, 3649–3658. doi: 10.1121/1.428217

Wennerstrom, A. (2000). The role of intonation in second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 102–127). Ann Arbor: University of Michigan Press.

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7, 49–63. doi: 10.1016/S0163-6383(84)80022-3

Winke, P., & Gass, S. (2013). The influence of second language experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly*, 47, 762–789. doi: 10.1002/tesq.73

Yates, L. (2011). Language, interaction and social inclusion in early settlement. *International Journal of Bilingual Education and Bilingualism*, 14, 457–471. doi: 10.1080/13670050.2011.573068

Yavaş, M. (2006). *Applied English phonology*. Oxford: Blackwell.

Zielinski, B. (2012). The social impact of pronunciation difficulties: Confidence and willingness to speak. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 3rd pronunciation in second language learning and teaching conference* (pp. 18-26). Ames, IA: Iowa State University.

Zielinski, B. (2008). The listener: No longer the silent partner in reduced intelligibility. *System*, 36, 69–84. doi: 10.1016/j.system.2007.11.004

Zielinski, B., & Yates, L. (2014). Myth: Pronunciation instruction is not appropriate for beginner-level learners. In L. Grant (Ed.), *Pronunciation myths* (pp. 56–79). Ann Arbor, MI: University of Michigan Press.

Appendix A. Delayed-Sentence Response Task Stimuli

Pretest Stimuli	
Prompt	**Response**
Did you guys do anything exciting on Sunday?	No! We watched some bad movies.
Did you get something from the store to fix dinner?	Yes, I bought some cabbage and onions.
Hey, what is that annoying sound outside?	That's the bell from school.
So, how do you want to decorate the room for the party?	We want red balloons only.
So, did you have fun in New Orleans?	Yes, I collected beads during parades.
Can you recognize any animals in the picture?	Only a sheep and a cow.
What do you usually buy for snacks?	I always buy chips with salsa.
So, what was the book you ordered?	The big book of short stories.
Posttest Stimuli (the posttest also included the stimuli from the pretest)	
Prompt	**Response**
Why isn't your father home yet?	My dad usually works late.
Have you seen Paul around?	He was in the lab working.
So, how did your second assignment go?	I did better this time.
Are you still hungry?	My belly is full now.
Would you like some fruit with your order?	I want a peach and some grapes.
You look tired, did you have a good day?	I felt sleepy all day long.
So, how do you pay your utilities in your new place?	I only pay bills online.
How old are your sons?	My kids are two and four.

Appendix B. Questionnaire for Native English Raters

Rater ID: _____

Taking into account the sentences that you perceived as less comprehensible, what specific criteria did you use to rate those sentences as less comprehensible? Please be as specific as possible.

_____

_____

_____

_____

_____

Taking into account the sentences that you perceived as more comprehensible, what specific criteria did you use to rate those sentences as more comprehensible than others? Please be as specific as possible.

Are there any other issues you would like to comment on the sentences you heard and rated? Any extra information you would like to give us will be greatly appreciated.

Thank you for your participation!

Authors' addresses

Joshua Gordon
Universidad Nacional, Sede Región Brunca,
Campus Pérez Zeledón
Oficina de Direccción de Areas
San Isidro de Perez Zeledon
San José, Costa Rica

joshgordonz@gmail.com

Isabelle Darcy
Department of Second Language Studies
1021 E. Third St.
Memorial Hall 315
Indiana University
Bloomington, IN 47405
USA

idarcy@indiana.edu