



ARTICLE

Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features

Jonás Fouz-González

UCAM University, Spain (jfouz@ucam.edu)

Abstract

This study investigates the potential of a podcast-based approach aimed at helping foreign language learners improve their pronunciation of features that tend to be fossilised in their interlanguage. Training took place over a period of three weeks in which podcasts were used for perceptual and productive practice as well as for peer evaluation. Participants in this study ($N = 47$) were randomly assigned to two groups that acted as control and experimental at the same time. One group received training on the English /s – z/ contrast and the other one was trained on the pronunciation of English /b d g/ as stops in intervocalic position. Pre- and post-tests were used to measure the participants' perception and production of the target features. The results show that training had a positive impact on the participants' perception and production of the target sounds. Even though the differences between groups did not reach statistical significance for every sound in every task, the data reveal that the approach adopted could foster substantial improvements in the participants' pronunciation of features that tend to be fossilised, even after short periods of training.

Keywords: pronunciation; podcasts; second language acquisition; foreign language teaching; computer-assisted language learning; computer-assisted pronunciation training

1. Introduction

Over the last decades, numerous studies have explored different approaches to help learners improve their pronunciation of a second (L2) or foreign language (FL) (see Lee, Jang & Plonsky, 2015; Thomson & Derwing, 2015). Despite initial scepticism regarding the effectiveness of pronunciation instruction (see, e.g., Suter, 1976), research has demonstrated that pronunciation can be trained and that instruction can foster improvements in learners' intelligibility, comprehensibility, and accentedness (Derwing, Munro & Wiebe, 1997). However, given the need to address different language skills and because of limitations in classroom time, research should explore how different approaches can facilitate the integration of pronunciation most effectively in FL and SL courses. Researchers have long advocated the use of technology as a facilitating device, as it can enhance presentation styles and make materials more physically and psychologically accessible (Pennington, 1996), offer learners individualised practice with unlimited tries, and even provide them with instant, automatic feedback on their performance.

There are many ways in which technology can help FL learners work on their pronunciation (see Fouz-González, 2015). Nonetheless, despite the enormous potential technology holds for pronunciation instruction, and although some of the tools commonly explored have proven to be

Cite this article: Fouz-González, J. (2019). Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features. *ReCALL* 31(2): 150–169, doi:10.1017/S0958344018000174

very suitable for certain purposes and contexts (e.g. spectrograms, waveforms, or software using automatic error detection), many of them are not yet entirely appropriate for autonomous practice. Authors have often noted that the interpretation of some of these tools requires specialised training that may not be practical for every student or even for every teacher (Levis, 2007; O'Brien, 2006). An alternative that circumvents the problem of students' interpretation is to resort to tools that offer automatic feedback on the learners' pronunciation. Nevertheless, despite their promising potential for controlled practice (see, e.g., Liakin, Cardoso & Liakina, 2014; Neri, Cucchiariini & Strik, 2008), researchers have often criticised their limited reliability in pinpointing specific errors in spontaneous speech and, what is more important, the impossibility to inform learners of how to correct those errors (Levis, 2007; Neri, Cucchiariini, Strik & Boves, 2002; O'Brien, 2011). Hence, research needs to continue exploring tools and techniques that are easily interpretable and accessible for any learner, as this should allow teachers to incorporate pronunciation as an integral element of their syllabi more easily.

In light of the above, and because the perfect stand-alone tool does not exist, the approach adopted here consists in helping learners create accurate perceptual representations of the target features so that they can monitor their pronunciation and continue making progress autonomously. Adequate perceptual targets play a crucial role in the development of adequate language-specific articulatory patterns (Flege, 1995, 2009). In fact, studies have shown that perceptual training can help learners improve their perception of L2/FL sounds and that these improvements can be transferred to their production, even if production is not trained (Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Carlet, 2017; Thomson, 2011). Nevertheless, in FL settings, creating adequate pronunciation targets without instruction is not an easy task. On the one hand, learners' perception of the phonological system of an FL is strongly conditioned by their first language (L1) (Best & Tyler, 2007; Flege, 1995), which leads them to overlook phonetic information in the speech signal that is irrelevant in their L1 but not in the FL. On the other hand, FL learners normally have very little exposure to authentic input and few opportunities to produce output and consolidate the articulatory movements required for the FL.

Research suggests that directing the learners' attention to form facilitates the noticing of relevant phonetic information that helps learners improve their perception of FL sounds (Guion & Pederson, 2007). As Thomson (2011) notes, this should make it easier for learners to incorporate more of the input they receive into their emerging L2 categories. Moreover, research has shown that the adoption of form-focused training techniques combined with explicit instruction can maximise the benefits of training, fostering very positive gains in pronunciation even after short instructional periods (Couper, 2011; Saito, 2013). In this regard, one of the most common ways of directing the learners' attention to formal features of the language is through input enhancement (Sharwood-Smith, 1993). Enhancing the salience of the target features in the input is considered to facilitate the learners' noticing of those features and therefore acquisition (Schmidt, 1990). Regarding pronunciation, researchers have explored different ways of enhancing the learners' perception of the FL, such as acoustically modifying features of the input (Barreiro-Bilbao, 2013; Wang & Munro, 2004) or exposing learners to highly variable stimuli to help them notice differences between certain target contrasts (Logan, Lively & Pisoni, 1991; Thomson, 2011). However, input salience can also be enhanced by offering learners explicit information about the target features so that they can consciously direct their attention to them and notice aspects that they would otherwise ignore (see Fouz-González, 2017; Mompean & Fouz-González, 2016; Saito, 2013).

The approach adopted here follows Saito's (2013) recommendation to use Ranta and Lyster's (2007) pedagogical sequence (awareness > practice > feedback) for pronunciation instruction. Saito advocates using explicit instruction to help learners notice relevant information in the speech signal, which should help them restructure their representations of the FL sounds, and then offering students opportunities for production practice so that they proceduralise their representations of the FL segments and eventually automatise them.

2. Using podcasts for pronunciation training

Podcasts offer numerous possibilities for language learning (see Abdous, Camarena & Facer, 2009; Lomicka & Lord, 2011; O'Bryan & Hegelheimer, 2007; Rosell-Aguilar, 2009). In essence, they cater for two of the key ingredients necessary for language acquisition, namely input (see Krashen, 1982) and output (see Swain, 1985). Podcasts offer learners the possibility to access virtually unlimited input, something extremely beneficial in FL contexts, where exposure tends to be restricted to the classroom. This has important implications for pronunciation, as podcasts allow learners to listen to authentic recordings by multiple speakers, male and female, with different accents, and to multiple instantiations of the target features in different phonetic contexts. This variability plays a vital role in phonological acquisition, especially in FL settings where learners tend to be exposed to a wide range of non-native pronunciations (both by teachers and other students) that present a highly variable (and sometimes incorrect) model of L2 phonetic detail (see Best & Tyler, 2007; Flege, 2009). Moreover, because podcasting allows users to record their own podcasts and distribute them for free, they also provide learners with multiple opportunities for output production. This has obvious advantages for pronunciation too, as learners can record themselves, practise, and consolidate the articulatory movements required for the FL as well as critically analyse their own pronunciation and notice gaps in their performance, one of the first steps for self-monitoring (see Fraser, 2001). This is considered to be a way of empowering students to work on their pronunciation autonomously with any input they are exposed to, without being circumscribed to a specific set of materials.

A few researchers have already explored podcasts' potential to help FL learners improve their pronunciation. As a case in point, Lord (2008) implemented a podcasting project with 16 students in a Spanish phonetics course and found that learners improved their general pronunciation ability and their attitudes towards Spanish pronunciation. Moreover, students enjoyed the project, found it beneficial and felt that the experience had helped them become more aware of their own pronunciation. In a similar vein, Ducate and Lomicka (2009) investigated podcasts' potential to foster improvements in accentedness and comprehensibility with 22 students in German and Italian courses as well as the participants' attitudes towards pronunciation. Their data show that although training fostered improvements in some of the tasks, comprehensibility and accentedness did not improve significantly over the course of training, nor did their attitudes towards pronunciation.

Although Ducate and Lomicka (2009) also used podcasts as a model to imitate, the two studies mentioned previously focused mostly on the possibilities podcasts offer for output production and peer evaluation. Using recordings to critically analyse one's own pronunciation has long been recommended by researchers (Acton, 1984; Couper, 2003; Fraser, 2001; Smith & Beckmann, 2010; Walker, 2005), and studies have shown how this technique can help learners improve different aspects of their pronunciation (e.g. Lord, 2008; Luo, 2016). Even though this is undoubtedly useful for pronunciation practice, and despite the fact that the present study also required students to produce output, a key element in the approach adopted in this study is the combination of explicit instruction and the perceptual training provided with podcasts before asking learners to produce output and evaluate each other.

3. Method

3.1 Participants

Participants in this study were 47 native speakers of Spanish (35 female, 12 male; M age = 19.4, $SD = 0.66$).¹ They were recruited from a phonetics module in the second year of a four-year

¹The pre-test data from four participants in the perception tests were lost due to a problem with the computers (participants 6 and 48 from G1 and participants 40 and 45 from G2). In the production tests, the post-test recordings by participants 13 (G1) and 27 (G2) could not be evaluated due to a very poor audio quality. Therefore, these data are not considered in the analyses.

degree in English studies (formerly English philology). Students in this degree are considered to be very advanced EFL learners, as the degree includes numerous modules on English linguistics, translation, literature, history and culture, and, except for a few modules, is entirely taught in English. At the time of the study, the participants' level was B2 according to the Common European Framework of Reference for Languages (CEFR). It is important to point out that the participants were not the researcher's students, but participation in this project replaced one of the assignments in their phonetics course.

3.2 Target features

The target aspects addressed were the pronunciation of English /b d g/ as stops in intervocalic position and the English /s – z/ contrast. With regard to the first aspect, English and Spanish have the same three voiced stops in their phonemic repertoires. However, although English /b d g/ are realised as stops regardless of their phonetic environment, their Spanish counterparts are spirantised in intervocalic position, rendering three approximant² allophones [β ð ɣ] respectively (Hualde, 2014; Martínez-Celdrán, Fernández-Planas & Carrera-Sabaté, 2003). Hence, English /b d g/ are often realised as approximants in intervocalic position by Spanish EFL learners because they transfer their L1 spirantisation rule to English (Zampini, 1996). As for the second aspect, Spanish EFL learners normally fail to mark the distinction between English /s/ and /z/, as the latter does not exist in the phonemic inventory of Spanish³ and is often realised as the Spanish /s/ (Monroy-Casas, 2001).

The target aspects explained above were selected because they tend to be fossilised in the interlanguage of very advanced Spanish learners of English (see Monroy-Casas, 2001). Selinker's (1972) original definition of fossilisation states that a language feature is considered to be fossilised when learners reach a point beyond which no improvements are made, irrespective of the amount of instruction they receive or the amount of exposure to the target language they have. Because "fossilisation" can be interpreted to mean that features are impossible to change, researchers have proposed alternative terms such as "stabilisation" or "entrenchment" (see Pennington & Rogerson-Revell, 2019), which imply that even if features are extremely difficult to modify, they can be amenable to change through instruction. In this regard, fossilisation is used here to refer to aspects of FL learners' pronunciation that are expected to be very hard to change without training.

As Pennington (1998) points out, given the cognitive, perceptual, psychomotor, and affective factors that hinder pronunciation acquisition, it is very difficult to modify learners' pronunciation without explicit instruction once they have reached a point of fossilisation. This is especially so in FL contexts as, unlike L2 settings in which many pronunciation aspects can be expected to improve through exposure to the language and meaningful interactions, the opportunities for exposure and interaction are rather limited. In this respect, the target features described above were considered to be particularly convenient to test the potential of the approach adopted in this study, given that they are very difficult to modify for the target group and participants were not expected to make any progress without instruction.

As the time that can be devoted to pronunciation practice in language classes is limited, and because attainment of native-like mastery in pronunciation is not a realistic goal for the majority of learners, researchers have long advocated the prioritisation of pronunciation features that

²Although in some dialects of Spanish /b d g/ can be realised as stops in other positions (see Zampini, 1996), in Peninsular and Murcian Spanish, the varieties spoken by participants in this study, they are only pronounced as stops in word initial position after a pause, after a nasal consonant (as in *bombo*, *cuando*, or *mango*), or after /l/ in the case of /d/ (as in *molde* or *caldo*; see Hualde, 2014).

³Even though /s/ may sometimes be realised phonetically as [z] due to assimilation processes when followed by a voiced sound, Castilian Spanish only has one alveolar fricative in its phonemic repertoire: the voiceless /s/ (Hualde, 2014; Martínez-Celdrán *et al.*, 2003).

hinder the learners' intelligibility rather than those that could reduce or eradicate foreign accent (Levis, 2005; Munro & Derwing, 1995). Nevertheless, it is important to note that participants in this study were very advanced learners of English (the same student profile as that in Monroy-Casas's, 2001, study) and were therefore considered to be perfectly intelligible. B2-level students' phonological control is defined in the CEFR as "[h]as acquired a clear, natural, pronunciation and intonation" (p. 117).⁴ Students are supposed to be perfectly intelligible at B1, for which the CEFR states: "[p]ronunciation is clearly intelligible even if a foreign accent is sometimes evident and occasional mispronunciations occur" (p. 117). Additionally, participants in this study were the type of student who normally aspires to the highest proficiency possible when speaking in English.⁵ Thus, although intelligibility should indeed always be prioritised when there is limited time to address FL pronunciation, and even though the target aspects addressed in this study may not necessarily hamper intelligibility, they were considered to be suitable for this particular group.

3.3 Research questions

RQ1: Can this podcast-based approach foster improvements in the learners' perception of fossilised segmental features?

RQ2: Can the approach adopted foster improvements in the participants' production of fossilised segmental features?

3.4 Research design

Researchers have often pointed out that one of the biggest challenges in this type of study is finding participants for a control group, as students are usually volunteers and they are offered some kind of instruction in exchange for their participation (see, e.g., Lord, 2008; Thomson, 2011). Moreover, because the number of volunteers tends to be small, it does not always allow for a reduction of group size if the aim is to extrapolate conclusions from the sample. Additionally, if the researcher contemplates the treatment under examination as positive for students, it does not seem fair to deprive half the group of it. Hence, in this study, all participants acted as control and experimental at the same time. At the beginning of the study, participants were randomly assigned to one of two groups. Group 1 (G1) received training in the English /s – z/ contrast and Group 2 (G2) received training in the pronunciation of English /b d g/ (see Figure 1). There were 25 students in G1 and 22 in G2. This allowed the researcher to test the impact of instruction with a much bigger sample of participants ($N = 47$), and it also ensured that both groups were exposed to very similar training conditions (i.e. receiving the same amount of input and making the same effort during the study).

3.5 Instruments and procedure

3.5.1 Training stimuli

The training stimuli were obtained from the *6 Minute English* podcast series by the BBC. An effort was made to include podcasts that featured a considerable number of instantiations of each

⁴Retrieved from https://www.coe.int/t/dg4/linguistic/source/framework_en.pdf

⁵The questionnaires in this study (not reported here due to limitations of space) did not address the learners' intended careers. Nonetheless, this degree in English studies is the degree people normally study in order to become EFL teachers in Spain (who normally aspire to the highest proficiency possible). In fact, the students' responses to the final questionnaires reveal a mean score of 4.6 in a 5-point Likert scale question ($SD = 0.6$; $n = 41$) asking them about their willingness to acquire native-like proficiency.

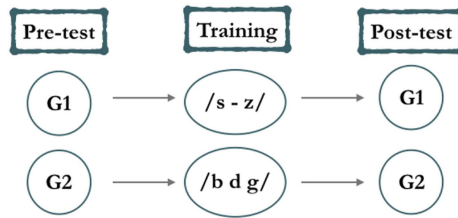


Figure 1. Study design

target sound. The target podcasts for /s – z/ were *Is silence golden?*, *Are you a winner?*, and *Modern offices*. The /b d g/ target podcasts were *Odd job interviews*, *Learn a thousand foreign words*, and *Young, British and sober*. Additionally, in order to offer learners productive practice with the target features, two short texts with multiple examples of the target sounds were created for learners to record at home (see online supplementary materials).

3.5.2 Training procedure

Participants met with the researcher for one hour a week over a period of three weeks. Training consisted of a four-stage procedure in which podcasts were used for input, output, and peer evaluation. Even though some of the activities were done in class, all the materials were shared with students through Edmodo (the podcasts that served as input, the students' recordings, and the links to the peer evaluations). More specifically, the approach was as follows:

1. Explicit instruction (in class): Participants received a brief explicit explanation about the target features. This covered the places and manners of articulation of the target sounds, the most common spellings for those sounds, the reasons why they are problematic for Spanish learners of English and tips on how to correct possible mispronunciations.
2. Input (in class): Participants listened to the weekly podcast and completed two activities online. These two activities required participants to listen closely to different instantiations of the target features. In the first activity, learners had to listen to a three-minute edited version of the podcast and find at least 10 words that contained each of the target sounds. In the second, learners were presented with short excerpts from the podcasts and they were asked to classify the target sounds as examples of English /s/ or /z/ in a selection of words (G1) or state whether they heard occlusion or not in different instances of English /b d g/ (G2). Activities 1 and 2 were administered through Google Forms (see Figure 2), which allowed the researcher to collect the data automatically, show students a summary of the choices they made, and offer feedback to the whole class.
3. Output (at home): After receiving the explicit instruction on the problematic features and having listened to different instantiations of the target sounds in the podcast, participants were asked to record a short text that contained a substantial number of instantiations of the target features and upload it onto Edmodo for peer evaluation. The texts were typographically enhanced, by underlining the orthographic representations of the target sounds and marking them in bold type.
4. Peer evaluation (in class): The peer evaluations followed the same format as Activity 2, and were also done using Google Forms, by asking learners to rate specific words from their colleagues' recordings. They were completed in class, after listening to each week's podcast and completing activities 1 and 2, which served as training for the subsequent peer ratings. However, this was only done in Weeks 2 and 3 of the study given that students had not recorded any podcasts during the first week.

Activity 1

Participant number *

Write down at least 10 examples of words containing /s/ in the podcast *

(Note that they don't need to be represented by <s> in spelling)

Write down at least 10 examples of words containing /z/ in the podcast *

(Note that they don't need to be represented by <z> in spelling)

Enviar

Activity 2

Participant number *

Which sound do you hear in "words"? *

- /s/
 /z/
 a very weak /z/

Which sound do you hear in "phraseS"? *

(the last <s>)

- /s/
 /z/
 a very weak /z/

Which sound do you hear in "associated"? *

(in the <c>)

- /s/
 /z/
 a very weak /z/

Which sound do you hear in "noise"? *

- /s/
 /z/
 a very weak /z/

Figure 2. Sample activities for the group receiving training in the /s – z/ contrast



Figure 3. Screenshot of the discrimination (left) and imitation (right) tasks

3.5.3 Testing procedure and instruments

Perceptual tests were administered with the open-source software application TP (Rato, Rauber, Kluge & dos Santos, 2015) in a quiet computer room at the university. Production tests were administered using a PowerPoint presentation that participants controlled on a MacBook Pro computer. They were recorded with Audacity, with a SAMSON C01U Microphone.

Perception tests

The learners' perception of the /s – z/ contrast was measured with an oddity discrimination task and an identification task (see Logan & Pruitt, 1995). In the discrimination task (see Figure 3, left), stimuli were presented in triads of minimally paired words in which one of the members was different (i.e. "change triads" – *sip-sip-zip*) or they all had the same phonological composition and should therefore be considered to be "the same" (i.e. "catch triads" – *zip-zip-zip*). The three stimuli in each triad were always pronounced by different speakers (male and female) in

order to measure the listeners' capacity to categorise sounds disregarding acoustic variations that are phonetically irrelevant to their identity. Interstimuli intervals were one second long.

In the identification task, participants were presented with one stimulus at a time and had to identify the sound they were hearing among four options: the two target sounds /s/ – /z/, the distractor /ʃ/, or the “I don't know” option (see Figure 3, right).

The above tasks were considered suitable to measure the participants' perception of the English /s – z/ contrast because the main problem Spaniards have with these sounds is that /z/ is absent in the phonemic inventory of Spanish. Hence, asking learners to either discriminate between /s/ and /z/ or to identify the sound they were hearing served as a measure of their ability to perceive those sounds. However, what needed to be tested for /b d g/ was not the learners' ability to perceive those sounds as different from each other (e.g. /b – d/ or /d – g/), but their ability to differentiate between English and Spanish realisations of those sounds. Therefore, two different tasks were used to measure the learners' awareness and perception of the stop realisation of English /b d g/. The first was a delayed accent-mimicry task in which learners had to imitate the way English speakers pronounce Spanish. This was meant to test the participants' implicit awareness of the realisation of English /b d g/ as stops in intervocalic position, not by immediate imitation of a model presented, but recalling their impressions from memory (see Flege & Hammond, 1982; Mora & Rochdi, 2016). In the second task, learners were given a list of Spanish words with the spellings for /b d g/ underlined and in bold. They were asked to read the words to themselves in Spanish while listening to their English counterparts over headphones and say whether they perceived the degree of occlusion of the underlined sounds to be the same or different. This task was intended to measure the participants' ability to perceive differences in occlusion in English and Spanish /b d g/. For example, learners heard the word *dagger* in English (not provided in its written form), and read the word *daga* in Spanish, having to decide whether /g/ had the same degree of occlusion in both languages.

Production tests

The participants' pronunciation of the target features was evaluated with three tasks that measured the participants' imitative, controlled, and spontaneous production, namely an imitation task, a sentence-reading task, and a timed picture-description task. Following Saito (2013), to ensure that participants pronounced a similar number of target items in the spontaneous task, each picture was accompanied by several word cues, including target words and distractors.

3.5.4 Testing stimuli

Perception

Testing stimuli for the perception tasks were obtained from several English dictionaries as well as specialised English pronunciation dictionaries. Stimuli in the identification task were divided into familiar ($n = 20$) and novel words ($n = 20$) to check whether potential improvements generalised to items that did not appear in training. Familiar words were selected from the most frequently occurring words in the podcasts used for training, although sometimes words were also included despite their low occurrence because of the spelling they exemplified (e.g. /z/ as represented by < s >).

Stimuli in the discrimination task consisted of 25 triads of minimally paired words (/s – z/), with 10 change triads, 10 catch triads, and five distractors (/s – ʃ/). Test items featured the target contrast in word initial, medial, and final position.

As for the delayed accent-mimicry task, each target sound was featured in five Spanish words embedded in carrier sentences, either in word-medial intervocalic position or in word-initial position flanked by vowels. Fifteen items featuring /p t k/ were used as distractors.

Finally, for the task measuring learners' perception of occlusion of English /b d g/, the researcher compiled a list of 48 English and Spanish cognates featuring voiced and voiceless stops in different

positions. Although the phonological structure of the cognates was not exactly identical (e.g. *labor* [la'βor] vs. *labour* ['leibə]), the criterion was to include pairs of words that were similar, featuring the target sounds in intervocalic position. Each target sound was featured in 10 words: five in word-initial position followed by a vowel and five in word-medial intervocalic position. However, in order to avoid possible biases towards the position of the sound in the word, nine more items (three per target sound) in which the target sounds were preceded by nasal consonants were included as “control items”. These control items should be perceived as having the same occlusion as their English counterparts, as /b d g/ are realised as stops in Spanish after nasal consonants even if they occur word medially. Finally, nine items featuring English /p t k/ were included as distractors. The testing stimuli for all tasks are available as online supplementary materials.

Production

The learners' imitative production of the target sounds was measured with five words per sound and three distractors featuring /ʃ/. /z/ was featured in initial and medial position and /b d g/ were always featured in word-medial intervocalic position.

For the sentence-reading task, stimuli were selected based on their frequency of occurrence in the training materials, although less commonly occurring words were also included in order to exemplify different orthographic representations. Stimuli were divided into familiar and novel words. The former were intended to test improvements in words learners had seen in training (i.e. words whose phonological make-up should be familiar to learners), and the latter were aimed at measuring whether improvements could generalise to words with which the participants had not practised. As Spanish students should not have problems with /s/, the testing items focused on /z/, with 10 familiar and 10 novel items. For /b d g/, there were 45 items in total. Each sound was featured in intervocalic position in 10 familiar (word-medial and word-final position) and five novel words (word medially).⁶

The stimuli for the timed picture-description task were chosen from the list of items previously selected for the sentence-reading task. Nevertheless, words were only included in the test if they featured the sound in word-medial intervocalic position, as the context in which learners produced the target words could not be controlled for. Given this, the stimuli for /b d g/ (15 words, five featuring each sound) had to combine familiar and novel items. The learners' spontaneous production of /z/ was measured with eight items. Five items were familiar stimuli featuring the sound with the spellings < s, se, x > and three were novel words with the spelling < z > .

3.5.5 Evaluation of stimuli

The participants' pronunciation was evaluated by three non-native judges expert in English pronunciation. A fourth expert was used to disambiguate disagreements. The rating sessions were conducted in a sound-attenuated professional studio at the university. The ratings were always dichotomous (1 if the target sound was pronounced adequately, 0 if it was mispronounced), and the raters could play every stimulus as many times as they needed. Interrater reliability was measured with Fleiss's kappa test, which yielded a reliability measure of 0.94 (0.81–1.00 range), interpreted as “almost perfect agreement”. Intrarater reliability was measured by comparing the raters' consistency in rating 20 extra items that had already been assessed, including five words from each target sound /z b d g/, with the same number of pre- and post-test productions, always by different speakers. There was only one item in which experts did not give the same rating, so no tests were conducted as intrarater reliability was considered to be almost perfect too.

⁶The only item in which the target sound was not entirely surrounded by vowels is *programme*, but it was included given that there were no more words in training (i.e. familiar) featuring the sound in medial position. However, this context (/gr/) also lends itself to the spirantisation of stop consonants. In fact, the word *programme* was one of the examples of consonant substitution (/g/ → [ɣ]) in the data by Monroy-Casas (2001).

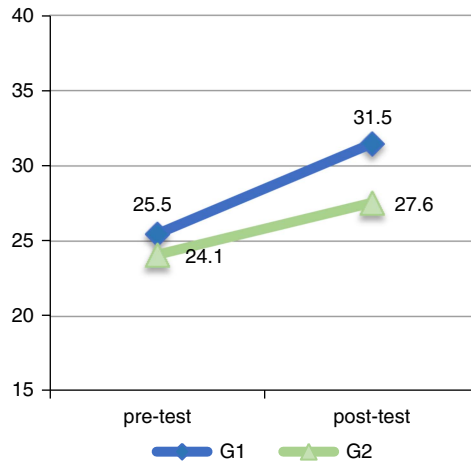


Figure 4. Mean scores for the /s - z/ contrast in the identification task

Table 1. Mean scores (SD) and improvement made (imp) in the identification and discrimination tasks

	Identification						Discrimination					
	pre	%	post	%	imp	%	pre	%	post	%	imp	%
G1	25.5 (6.2)	63.8	31.5 (4.3)	78.8	6	15	9.6 (1.9)	48	11.2 (2.4)	56	1.7	8.5
G2	24.1 (7.2)	60.3	27.5 (6.6)	68.8	3.4	8.5	9.8 (3.7)	49	11 (3.2)	55	1.2	6

4. Results

The data from pre- and post-tests were analysed with two-way mixed ANOVAs, with group as between-subjects factor and time as within-subjects factor.

4.1 Perception

The results of the identification task measuring the learners' perception of English /s - z/ revealed a significant interaction between the time and group variables, $F(1, 41) = 4.14$, $p = 0.048$, which indicates that the improvement made by one of the groups was significantly different from the one made by the other group (see Figure 4). G1 improved by 6 points (15%) as compared to G2, which improved by 3.4 points (8.5%) (see Table 1). However, the results from the discrimination task revealed no interaction effects between the time and group variables ($p > 0.05$), which indicates that the improvement made between groups was similar (see Table 1). G1's mean score improved by 1.7 points (8.5%) from pre- to post-test and G2's score improved by 1.2 points (6%).

An analysis of the scores obtained in the identification task for each sound separately reveals that the only significant Time x Group interaction was found for novel items featuring /s/, $F(1, 41) = 6.66$, $p = 0.014$. Although both groups made similar improvements in their ability to identify instances of /z/ correctly, only participants in G1 improved their ability to correctly identify instances of /s/ in items that had not appeared in training (see Table 2).

As for the delayed accent-mimicry task, considering the scores for /b d g/ as a whole, the pre-test data show that, on average, participants in both groups were somewhat aware that English-accented speakers would not spirantise /b d g/ in intervocalic position, although they did not consistently realise them as stops all the time (G1 66.11% and G2 54.29%; see Table 3). When comparing the improvement made between groups, no significant Time x Group interaction

Table 2. Mean scores (SD) and improvement (imp) in familiar (fam) and novel (nov) words for /s/ and /z/ in the identification task

		G1						G2					
		pre		%		post		pre		%		post	
		score	SD	score	SD	score	SD	score	SD	score	SD	score	SD
/s/	fam	7.6	(1.4)	75.7		8.3	(1.4)	82.6		0.7	7	6.9	(1.7)
	nov	6.9	(1.6)	68.7		8.3	(1.3)	83		1.4	14	6.8	(2)
/z/	fam	4.7	(2.8)	47		7.6	(2.1)	75.7		2.9	29	4.6	(2.5)
	nov	6.3	(2.3)	63.5		7.3	(2.1)	73.5		1	10	5.9	(2.1)

Table 3. Mean scores (SD) in pre- and post-tests and degree of improvement (imp) in the delayed accent-mimicry task

		/b/			/d/			/g/			/b d g/		
		pre	post	imp	pre	post	imp	pre	post	imp	pre	post	imp
G1		2.83	(1.6)	2.71	(1.7)	-0.13		3.88	(1.3)	3.63	(1.6)	-0.25	
%		56.67		54.17		-2.5		77.5		72.5		-5	
G2		2.14	(1.7)	2.76	(1.9)	0.62		3.38	(1.6)	3.52	(1.6)	0.14	
%		42.86		55.24		12.38		67.62		70.48		2.86	

effects were found. Participants in the group acting as control (G1) made no improvements from pre- to post-test, even showing decreases in their post-test scores. G1's mean score decreased by -0.4 points (-2.7%), whereas G2's improved by 0.81 points (5.4%). Nevertheless, G2's improvement was very modest and did not reach significance.

The results from the task evaluating the learners' perception of occlusion in English and Spanish /b d g/ show that although the group acting as experimental (G2) was the only group who made improvements from pre- to post-test (see Table 4), these were very small, with no significant interactions between time and group. Even though G2 generally outperformed G1, the differences were minimal.

4.2 Production

An analysis of the mean scores obtained by each group across tasks reveals that there are clearly observable differences between groups in the aspects in which they received training (see Figure 5). A significant Time x Group interaction was found in the learners' scores for /z/, $F(1, 43) = 7.43, p = 0.009$. G1 improved by 4 points (12%) and G2 improved by 1.3 points (3.9%). As regards /b d g/, considering the scores for the three sounds together, the Time x Group interaction was also found to be significant, $F(1, 43) = 5.33, p = 0.026$. G2 improved by 7 points (9.4%) and G1 improved by 2.6 points (3.5%). The total mean scores for each sound in the different production tasks can be found in the Appendix.

Focusing on the imitation task, the improvements made by both groups were very modest for both target aspects (see Figure 6). No interaction effects were found between the time and group variables for any of the sounds. G1's mean scores for /z/ improved by 0.5 points (9.2%) and G2's improved by 0.8 points (15.2%). As for /b d g/, the improvements made by G2 were very small, with 0.29 points for /b/ (5.7%) and 0.19 points for /d/ and /g/ (3.8%). G2 showed only slight improvements in their ability to imitate English /b d g/ as stops in intervocalic position, but it is

Table 4. Mean scores (SD) and improvement (imp) made for /b/, /d/, and /g/ in the task measuring learners' perception of occlusion

	/b/			/d/			/g/		
	pre	post	imp	pre	post	imp	pre	post	imp
G1	8.4 (1.7)	7.7 (2)	-0.7	7.8 (1.9)	7.7 (1.9)	-0.1	8.5 (2.3)	8.7 (1.9)	0.2
%	64.6	59.2	-5.4	60	59.2	-7.7	65.4	66.9	1.5
G2	8.5 (1.5)	8.8 (2.4)	0.3	8.1 (1.8)	8.3 (1.9)	0.3	8.3 (2.6)	8.45 (2.5)	0.15
%	65.4	67.7	2.3	62.3	63.8	2.3	63.8	65.4	0.8

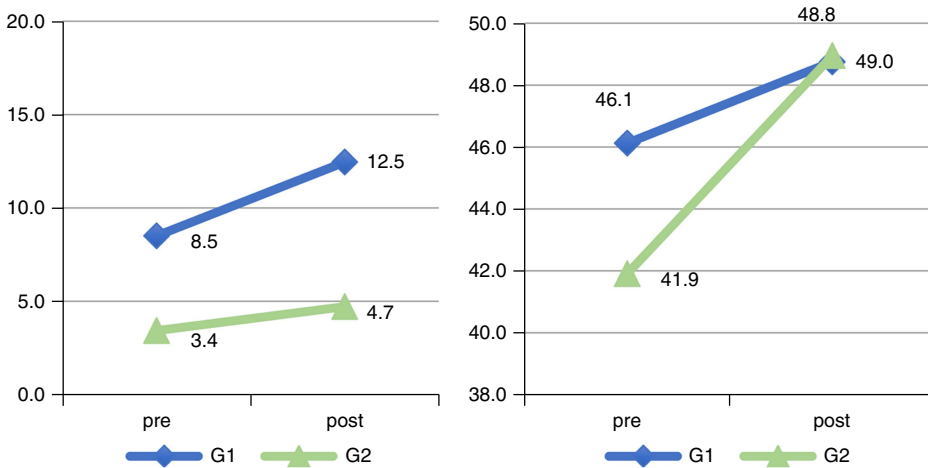


Figure 5. Pre- and post-test production scores for /z/ (left) and /b d g/ (right) across tasks

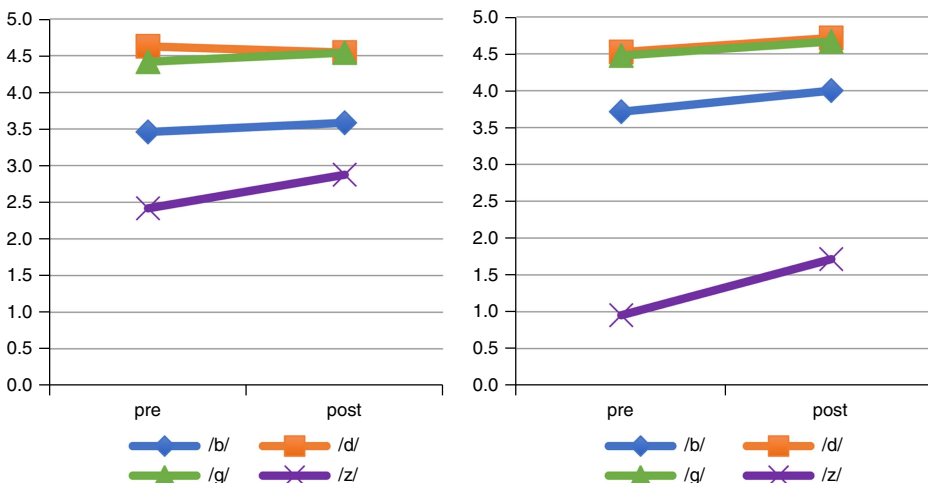


Figure 6. Mean scores for G1 (left) and G2 (right) in the imitation task

Table 5. Mean scores (SD) and improvement (imp) made in familiar (fam) and novel (nov) words in the sentence-reading task

		G1						G2					
		pre	%	post	%	imp	%	pre	%	post	%	imp	%
/b/	fam	5.96 (2.9)	59.58	6.63 (2.7)	66.25	0.67	6.67	4.95 (3)	49.52	7.00 (2.3)	70	2.05	20.48
	nov	2.25 (1.6)	45	2.71 (1.5)	54.17	0.46	9.17	2.24 (1.3)	44.76	2.62 (1.5)	52.38	0.38	7.62
/d/	fam	7.38 (1.8)	73.75	7.79 (2.1)	77.92	0.42	4.17	7.14 (1.9)	71.43	7.81 (1.5)	78.1	0.67	6.67
	nov	3.38 (1.4)	67.5	3.38 (1.6)	67.5	0	0	2.57 (1.7)	51.43	2.95 (1.4)	59.05	0.38	7.62
/g/	fam	6.21 (2.1)	62.08	6.21 (2.3)	62.08	0	0	5.38 (3)	53.81	6.67 (2.3)	66.67	1.29	12.86
	nov	2.21 (1.5)	44.17	2.33 (1.5)	46.67	0.13	2.5	1.86 (1.6)	37.14	2.43 (1.4)	48.57	0.57	11.43
/z/	fam	1.75 (2)	17.5	2.96 (2.3)	29.58	1.21	12.08	0.95 (2)	9.52	0.86 (1.4)	8.57	-0.1	-0.95
	nov	2.92 (3.3)	29.17	4.67 (3.2)	46.67	1.75	17.5	1.19 (2)	11.9	1.52 (2.2)	15.24	0.33	3.33

important to note that both groups' mean scores for /b d g/ were already very high in the pre-test, whereas for /z/, there was much room for improvement.

As for the sentence-reading task, the results obtained in familiar items show that the differences between groups (Time x Group interactions) were significant for the improvements made for /b/, $F(1, 43) = 4.33$, $p = 0.044$, /g/, $F(1, 43) = 11.33$, $p = 0.002$, and /z/, $F(1, 43) = 10.52$, $p = 0.002$. G1's mean improvement for /z/ (1.12 points, 12.08%) was significantly higher than the improvement by G2 (-0.1 points, -0.95%). Likewise, G2's mean improvement for /b/ and /g/ (2.05 points, 20.48% and 1.29 points, 12.86% respectively) was substantially higher than that of G1 (0.67 points, 6.67% and 0 points respectively).

The analysis of the scores for novel stimuli revealed that the differences in the improvement made between groups only reached significance for /z/, $F(1, 43) = 8.54$, $p = 0.006$, with G1 clearly outperforming G2 (see Table 5). Regarding /b d g/, no significant interactions were found between time and group.

Finally, the participants' scores in the timed picture-description task show that the improvements in the learners' spontaneous production of the target sounds were very limited. Considering the total scores in this task for each sound, G1's mean scores for /z/ improved by 0.54 points (6.8%), whereas G2's improved by 0.29 points (3.6%). Regarding /b d g/, /b/ was the only sound for which G2's improvement was higher than that of G1. The improvements made by participants in both groups were exactly the same for /d/, and almost the same for /g/ (see Appendix). No significant Time x Group interactions were found for any of the target sounds.

5. Discussion

This study explored the potential of a podcast-based approach combining input, output, and peer evaluation to help FL learners improve their pronunciation of segmental features that tend to be fossilised in their interlanguage. In line with the data reported by Monroy-Casas (2001) and Zampini (1996), the target aspects addressed also showed traits of fossilisation in the interlanguage of the Spanish participants in this study. Table 6 shows the percentage of items that were mispronounced on average in the pre-test. As a case in point, /z/ was mispronounced in 51.7% (G1) and 81% (G2) of the items in the imitation task, in 76.7% (G1) and 89.3% (G2) of the items in the sentence-reading task, and in 82.3% (G1) and 95.8% (G2) of the items in the timed picture-description task. As regards /b d g/, the participants' scores in the imitation task were very high from the beginning (see Appendix). This indicates that, overall, participants could

Table 6. Percentage of items that were mispronounced in the pre-test

	Imitative		Controlled		Spontaneous	
	G1	G2	G1	G2	G1	G2
/b/	30.8%	25.7%	45.3%	52.1%	65.8%	81.9%
/d/	7.5%	9.5%	28.3%	35.2%	40.8%	41.9%
/g/	11.7%	10.5%	43.9%	51.7%	68.3%	75.2%
/z/	51.7%	81%	76.7%	89.3%	82.3%	95.8%

attain the stop realisation of /b d g/ in production when imitating a model. This is not surprising given that [b d g] are allophonic variants of /b d g/ in Spanish, and learners should not find it difficult to articulate them as stops. As Table 6 shows, the number of mispronunciations is much higher in the controlled and spontaneous tasks. Participants spirantised /b g/ in intervocalic position in more than 40% of the items in the sentence-reading task and in more than 65% of the examples in the timed picture-description task. This indicates that although participants could realise these sounds as stops in most cases with relative ease when imitating a model, they failed to avoid spirantisation in more demanding tasks.

RQ1 addressed the potential of the approach to help learners improve their perception of the target features. The results show that the instruction had a positive impact on the participants' perception of the English /s – z/ contrast, but not of /b d g/. The data for /s – z/ show that the instruction fostered significant differences between the groups' ability to identify these sounds correctly, even in words that did not appear in training. The fact that the training and testing stimuli were pronounced by different speakers shows that the participants' improvements were not speaker dependent. However, the differences between groups did not reach statistical significance in the discrimination task. This indicates that although participants improved their ability to identify instantiations of /s/ and /z/ correctly in an identification task in which stimuli were presented individually (similar to the perceptual tasks they have been exposed to during training), they were not capable of improving their ability to perceive differences between /s – z/ in triads of minimally paired words in a discrimination task.

Regarding /b d g/, the data from the delayed accent-mimicry task and the task assessing the learners' perception of occlusion show that there were no significant differences in the improvement made between groups for these tasks. Both groups' scores remained relatively stable from pre- to post-test in both tasks, which suggests that learners did not develop their implicit awareness of the phonetic differences between Spanish- and English-voiced stops in intervocalic position, nor were they more capable of noticing differences in occlusion between English and Spanish /b d g/ after training. The absence of improvements in the task measuring the learners' perception of occlusion is rather surprising, given that the data from the imitation task (one of the production tasks) show that participants could indeed imitate (and therefore perceive) English /b d g/ as stops in intervocalic position. If learners are able to produce [b d g] correctly in an imitation task (which implies an adequate perception of the sounds imitated), they should be able to perceive differences in occlusion between the two languages. The instructions participants received specifically asked them to focus on the degree of occlusion, ignoring aspects such as aspiration, energy of articulation, etc. Nevertheless, it may be the case that learners did perceive occlusion adequately when imitating English speech (as evidenced in the imitation task) but failed to pay attention to the right cues when comparing English voiced stops and their Spanish counterparts. It is also possible that when participants read the Spanish words to themselves in an attempt to carefully analyse how they normally pronounce Spanish /b d g/, they artificially realised them as [b d g] in intervocalic position as a result of an excessively slow and careful pronunciation.

It is important to note that the fact that participants were taking a phonetics module at the time of the study may account for the fact that those acting as control made improvements in aspects in which they were not being trained in some tasks. However, this is considered to offer a very reliable measure of the effectiveness of the approach, as the only difference between groups was the focus of the instruction.

RQ2 investigated the effects of the instruction on the participants' production of the target features. Considering the total scores across production tasks, the data show that the instruction fostered substantial improvements in the aspects in which each group was trained. Nonetheless, the analysis of both groups' performance in the different tasks reveals that the differences between groups only reached statistical significance in the sentence-reading task (/b g/ and /z/ in familiar stimuli and /z/ in novel stimuli).

The difficulty these production tasks impose on the learners is not the same. Imitation should be easier than controlled production, and these two should in turn be easier than spontaneous production. An imitation task measures the listeners' perceptual and articulatory abilities, as participants listen to a model and have to repeat what they hear immediately afterwards. Nevertheless, in a sentence-reading task, participants need to be able to articulate the target sounds correctly by recalling them from memory, which also requires them to be aware of the phonological composition of words (in the case of English, often hindered by opaque sound-spelling correspondences). Finally, a timed picture-description task should be more difficult than controlled production in a sentence-reading task, as it requires learners to recall how sounds should be pronounced from memory, with the added difficulties of having to create sentences to describe the pictures and the timed nature of the task. The pre- and post-test data for each sound is in line with the hypothesised difficulty imposed by each task. The participants' imitation scores are the highest, followed by the sentence-reading task, and finally the timed picture-description task (see Appendix). The fact that significant differences between the two groups were only found in the controlled task suggests that for bigger improvements, especially in spontaneous production, longer training periods or more extensive exposure are needed, especially for sounds with various orthographic representations (e.g. noticing that the <s> in the word *reason* is pronounced with /z/).

It is important to note that training did not exert the same impact on every target sound, which could be explained by the different status of the target features in the participants' L1. The results suggest that, in general, it was much easier for learners to make improvements for /z/ than for /b d g/. As a case in point, the data from the identification task show that both groups were able to make similar improvements for /z/. Also, G2's improvements for /z/ in the imitation task were bigger than those of G1 (the group acting as experimental). G2's improvements in that task were even bigger for /z/ than for /b d g/, although it is important to note that both groups' scores for /b d g/ were very high from the pre-test. These similar rates of improvement can be observed in tasks that rely on the learners' perception of /z/; the improvements made in the tasks measuring the participants' controlled and spontaneous production were different. This can be explained by the fact that the participants' accurate production of /z/ in the latter tasks does not only depend on their ability to perceive and articulate the sound correctly, but also on their awareness of possible spellings for /z/, previous exposure to the target words (and noticing that they are pronounced with /z/ and not /s/), as well as automatising of the articulatory patterns required for /z/ in tasks that require learners to articulate this sound more independently (not after a previously heard model). On the contrary, the spellings for /b d g/ are much more transparent and should not pose this type of challenge for learners, as learners should soon become aware that every time they have to pronounce English /b d g/, they should pronounce them as stops. In spite of that, although the scores in the imitation task show that participants in both groups could perceive and articulate /b d g/ as stops in intervocalic position, participants in G2 still spirantised these sounds in many of the instances in the sentence-reading task and the timed picture-description task. The data are in line with Flege's (1995) speech learning model. As

/z/ is not present in the phonemic inventory of Spanish, it could be considered as a “new” sound and therefore it should be easier for students to make improvements. On the contrary, the differences between English and Spanish /b d g/ are phonetic rather than phonemic. Hence, they could be considered as “similar” sounds and therefore be more affected by equivalence classification (see Flege, 1987). Additionally, the results show that /d/ was generally less spirantised than /b g/ in all the production tasks (both by participants in the control and in the experimental group), which is in line with Zampini’s (1996) results. As Zampini (1996) points out, this may be due to the fact that EFL learners realise that [d] and [ð] have phonemic value in English and they soon reject [ð] as an acceptable pronunciation of English /d/.

Finally, it is important to point out that although the instruction fostered relevant improvements in the learners’ capacity to perceive and produce the target sounds, their production scores in the imitation task (for /z/) and in the sentence-reading and timed picture-description tasks (for /b d g/ and /z/) were still far from the maximum scores in each task. The results offer further support of Saito’s (2013) recommendation to adopt Ranta and Lyster’s (2007) pedagogical sequence for FL pronunciation instruction. FL learners do not only need to perceive and notice how FL sounds are pronounced, but also be exposed to different words featuring those sounds and notice their occurrence in those words (e.g. realise that *music* is pronounced with /z/ and not /s/) and, through production practice, gradually become capable of automatising the articulation of the target sounds in spontaneous production. The data for /b d g/ offer further support for this claim, as learners were perfectly able of perceiving and articulating these sounds in the imitation task, but were incapable of inhibiting the Spanish spirantisation rule when using them in English (i.e. they needed automatisation).

6. Conclusions

Research has shown that directing learners’ attention to phonetic information in the speech signal plays an important role in FL speech perception (Guion & Pederson, 2007) and that perceptual training can help learners improve their perception and production of features that are considered to be challenging, even when production is not trained (Bradlow *et al.*, 1997; Carlet, 2017; Lambacher *et al.*, 2005; Thomson, 2011). However, as the results reported above show, FL learners do not only need to be able to perceive the target sounds and attain them in production (e.g. in an imitation task), but also automatise the articulatory habits required by the FL and to be able to use the sounds adequately in controlled and spontaneous production – which is particularly challenging in English given the lack of transparency between sounds and spelling. The present study was set to explore the potential of a podcast-based approach combining explicit instruction and form-focused training to help FL learners improve their pronunciation of features that are considered to be very difficult to modify. Podcasts were used for perception, production, and peer-evaluation practice.

Numerous studies have investigated podcasts’ potential for language learning as well as learners’ perceptions towards them. However, research exploring their potential empirically is comparatively scarce (Lomicka & Lord, 2011). This study offers empirical evidence that the podcast-based approach adopted here can help adult FL learners improve their pronunciation of aspects that are fossilised in their interlanguage. It is important to note that the differences between groups did not reach statistical significance in every task and that there was still much room for improvement. Nevertheless, the findings are encouraging given that the training added up to a total of less than four hours and it fostered significant differences between groups in aspects that are considered to be very difficult to modify.

The study offers relevant implications for language teaching, as the approach adopted allows learners to work on their pronunciation with devices they already have and use. Podcasts allow FL learners to listen to the target features in real examples of connected speech, rather than in isolated words or sentences. If the approach adopted can enhance learners’ perception of the FL,

they should be able to practise autonomously anywhere, at any time, and with any podcast, and therefore also with the accent of their choosing. Besides, these analytic skills should be eventually extrapolated to any input learners are exposed to, including music, TV, or conversations with other speakers. Through focused listening, learners should be able to gradually improve their perceptual representations of the FL phonological system, become better able to monitor their pronunciation autonomously, and eventually incorporate the features they perceive in the input to their production. Moreover, the fact that podcasts offer learners the possibility to record themselves is a great asset to practise and automatise the articulatory movements required for the FL as well as to promote noticing through the learners' analysis of their recordings.

The fact that both groups acted as control and experimental proved to be very advantageous. First, it was a way of using a higher number of participants, as they all served to test the potential of the approach, therefore overcoming the common problem of using just half the sample. Second, as both groups received training, no one was deprived of instruction; even though they worked on different aspects, both groups were able to benefit from the approach. Finally, the two groups were required to make the same effort and were exposed to the same amount of extra input, the only difference being the focus of training.

In spite of the above, the study presents several limitations, which also offer directions for future research. First, the length of instruction was relatively short. Although the approach fostered significant differences between the groups in aspects that are considered to be particularly difficult to modify, longer training periods are needed for instruction to yield more positive results. Related to this is the fact that the amount of output practice was limited and that the learners' productions were scripted. This was done in an attempt to alleviate the students' workloads while also prompting numerous instantiations of the target sounds that every student should pronounce (which consequently facilitated the evaluation of those words in the peer-feedback activities). Nonetheless, to really exploit the potential of podcasting for pronunciation practice, students should be given the opportunity to design their own podcasts, building podcasting communities, and engaging in meaningful exchanges with other students. These two limitations are partly due to the fact that the participants in this study were not the researcher's students and their availability was limited. However, with longer instructional periods, podcasts could include a combination of scripted and extemporaneous tasks, like the ones used by Ducate and Lomicka (2009), or address different pronunciation aspects every week, as in Lord's (2008) study.

Finally, it is important to note that the target aspects addressed were not selected because of their impact on intelligibility, but because they represent features that tend to be fossilised in the interlanguage of advanced Spanish EFL learners (the target group addressed here, with a very specific profile and interested in attaining the highest mastery possible in terms of pronunciation). Nevertheless, the fact that the approach adopted could foster significant differences between the groups in features that tend to be fossilised after such a short training period suggests that it can also be potentially very advantageous for other pronunciation aspects.

Supplementary materials. For supplementary materials referred to in this article, please visit <https://doi.org/10.1017/S0958344018000174>

Acknowledgements. I would like to thank the journal's anonymous reviewers for their invaluable feedback on an earlier version of this paper. I am also extremely grateful to the students who volunteered to participate in this study and to Rafael Monroy-Casas and José Antonio Mompeán for helping me recruit the participants. Finally, my thanks also go to Pilar Mompeán and Ana Rosa Sánchez for their availability and patience in the evaluation of students' productions and to Aurora González Vidal and Antonio Maurandi López for their assistance with the statistical analyses.

Ethical statement. Participants in this study were volunteer students. The study was offered a free pronunciation course that would replace one of the compulsory assignments in the phonetics course in which students were enrolled. At the beginning of the study, each participant was assigned a participant number and all the data were treated anonymously.

References

- Abdous, M. Camarena, M. M. & Facer, B. R. (2009) MALL Technology: Use of academic podcasting in the foreign language classroom. *ReCALL* 21(1), 76–95. <https://doi.org/10.1017/S0958344009000020>
- Acton, W. (1984) Changing fossilized pronunciation. *TESOL Quarterly* 18(1), 71–85. <https://doi.org/10.2307/3586336>
- Barreiro-Bilbao, S. C. (2013) Perception of natural and enhanced non-native contrasts in clear speech. *Onomázein* 27: 207–219.
- Best, C. T. & Tyler, M. D. (2007) Nonnative and second-language speech perception: Commonalities and complementarities. In Bohn O & Munro M (eds.), *Language experience in second language speech learning: In honor of James Emil Flege*. Amsterdam: John Benjamins, 13–34. <https://doi.org/10.1075/llt.17.07bes>
- Bradlow, A. R. Pisoni, D. B. Akahana-Yamada, R. & Tohkura, Y. (1997) Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America* 101(4), 2299–2310. <https://doi.org/10.1121/1.418276>
- Carlet, A. (2017) *L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study*. Barcelona: Universitat Autònoma de Barcelona, unpublished doctoral dissertation.
- Couper, G. (2003) The value of an explicit pronunciation syllabus in ESOL teaching. *Prospect* 18(3), 53–70.
- Couper, G. (2011) What makes pronunciation teaching work? Testing for the effect of two variables: Socially constructed metalanguage and critical listening. *Language Awareness* 20(3), 159–182. <https://doi.org/10.1080/09658416.2011.570347>
- Derwing, T. M. Munro, M. J. & Wiebe, G. (1997) Pronunciation instruction for ‘fossilized’ learners: Can it help? *Applied Language Learning* 8(2), 217–235.
- Ducate, L. & Lomicka, L. (2009) Podcasting: An effective tool for honing language students’ pronunciation? *Language Learning & Technology* 13(3), 66–86. <http://doi.org/10125/44192>
- Flege, J. (1987) Effects of equivalence classification on the production of foreign language speech sounds. In James A & Leather J (eds.), *Sound patterns in second language acquisition*. Dordrecht: Foris, 9–39.
- Flege, J. (1995) Second-language speech learning: Theory, findings and problems. In Strange W (ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research*. Timonium: York Press Inc, 233–277.
- Flege, J. E. (2009) Give input a chance! In Piske T & Young-Scholten M (eds.), *Input matters in SLA*. Bristol: Multilingual Matters, 175–190.
- Flege, J. E. & Hammond, R. M. (1982) Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition* 5(1), 1–17. <https://doi.org/10.1017/S0272263100004563>
- Fouz-González, J. (2015) Trends and directions in computer-assisted pronunciation training. In Mompean JA & Fouz-González J (eds.), *Investigating English pronunciation: Trends and directions*. New York: Palgrave Macmillan, 314–342. https://doi.org/10.1057/9781137509437_14
- Fouz-González, J. (2017) Pronunciation instruction through Twitter: The case of commonly mispronounced words. *Computer Assisted Language Learning* 30(7), 631–663. <https://doi.org/10.1080/09588221.2017.1340309>
- Fraser, H. (2001) *Teaching pronunciation: A handbook for teachers and trainers: Three frameworks for an integrated approach*. Sydney: TAFE NSW Access Division.
- Guion, S. G. & Pederson, E. (2007) Investigating the role of attention in phonetic learning. In Bohn O-S & Munro MJ (eds.), *Second-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege*. Amsterdam: John Benjamins, 57–77. <https://doi.org/10.1075/llt.17.09gui>
- Hualde, J. I. (2014) *Los sonidos del español*. Cambridge: Cambridge University Press.
- Krashen, S. D. (1982) *Principles and practice in second language acquisition*. Oxford: Pergamon Institute.
- Lambacher, S. Martens, W. Kakehi, K. Marasinghe, C. & Mollholt, G. (2005) The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics* 26, 227–247. <https://doi.org/10.1017/S0142716405050150>
- Lee, J. Jang, J. & Plonsky, L. (2015) The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics* 36(3), 345–366. <https://doi.org/10.1093/applin/amu040>
- Levis, J. M. (2005) Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly* 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Levis, J. (2007) Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics* 27: 184–202. <https://doi.org/10.1017/S0267190508070098>
- Liakin, D. Cardoso, W. & Liakina, N. (2014) Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal* 32(1), 1–25. <https://doi.org/10.1558/cj.v32i1.25962>
- Logan, J. Lively, S. & Pisoni, D. (1991) Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America* 89(2), 874–886. <https://doi.org/10.1121/1.1894649>
- Logan, J. & Pruitt, J. (1995) Methodological issues in training listeners to perceive non-native phonemes. In Strange W (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Baltimore: York Press, 351–378.
- Lomicka, L. & Lord, G. (2011) Podcasting – past, present and future: Applications of academic podcasting in and out of the language classroom. In Facer BR & Abdous M (eds.), *Academic podcasting and mobile assisted language learning: Applications and outcomes*. Hershey: Information Science Reference, 1–20. <https://doi.org/10.4018/978-1-60960-141-6.ch001>

- Lord, G. (2008) Podcasting communities and second language pronunciation. *Foreign Language Annals* 41(2), 364–379. <https://doi.org/10.1111/j.1944-9720.2008.tb03297.x>
- Luo, B. (2016) Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction. *Computer Assisted Language Learning* 29(3), 451–476. <https://doi.org/10.1080/09588221.2014.963123>
- Martínez-Celdrán, E. Fernández-Planas, A. & Carrera-Sabaté, J. (2003) Castilian Spanish. *Journal of the International Phonetic Association* 33(2), 255–259. <https://doi.org/10.1017/S0025100303001373>
- Mompean, J. A. & Fouz-González, J. (2016) Twitter-based EFL pronunciation instruction. *Language Learning & Technology* 20(1), 166–190. <http://doi.org/10125/44451>
- Monroy-Casas, R. (2001) Profiling the phonological processes shaping the fossilised IL of adult Spanish learners of English as foreign language: Some theoretical implications. *International Journal of English Studies* 1(1), 157–217.
- Mora, J. & Rochdi, Y. (2016) Phonological awareness in Spanish-English interphonology: The case of spirantization. In Fernández Planas A (ed.), *53 Reflexiones sobre Aspectos de la Fonética y otros Temas de Lingüística*. Barcelona: Universitat de Barcelona, 395–403.
- Munro, M. J. & Derwing, T. M. (1995) Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Neri, A. Cucchiari, C. & Strik, H. (2008) The effectiveness of computer-based corrective feedback for improving segmental quality in L2 Dutch. *ReCALL* 20(2), 225–243. <https://doi.org/10.1017/S0958344008000724>
- Neri, A. Cucchiari, C. Strik, H. & Boves, L. (2002) The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning* 15(5), 441–467. <https://doi.org/10.1076/call.15.5.441.13473>
- O'Brien, M. G. (2006) Teaching pronunciation and intonation with computer technology. In Ducate L & Arnold N (eds.), *Calling on CALL: From theory and research to new directions in foreign language teaching*. San Marcos: CALICO Monograph Series, 127–148.
- O'Brien, M. G. (2011) Teaching and assessing pronunciation with computer technology. In Arnold N & Ducate L (eds.), *Present and future promises of CALL: From theory and research to new directions in language teaching*. San Marcos: CALICO Monograph Series, 375–406.
- O'Bryan, A. & Hegelheimer, V. (2007) Integrating CALL into the classroom: The role of podcasting in an ESL listening strategies course. *ReCALL* 19(2), 162–180. <https://doi.org/10.1017/S0958344007000523>
- Pennington, M. (1996) The power of the computer in language education. In Pennington MC (ed.), *The power of CALL*. Houston: Athelstan, 1–14.
- Pennington, M. C. (1998) The teachability of phonology in adulthood: A re-examination. *International Review of Applied Linguistics* 36(4), 323–341. <https://doi.org/10.1515/iral.1998.36.4.323>
- Pennington, M. C. & Rogerson-Revell, P. (2019) *English pronunciation teaching and research: Contemporary perspectives*. Basingstoke: Palgrave Macmillan
- Ranta, L. & Lyster, R. (2007) A cognitive approach to improving immersion students' oral language abilities: The awareness-practice-feedback sequence. In DeKeyser RM (ed.), *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York: Cambridge University Press, 141–160. <https://doi.org/10.1017/CBO9780511667275.009>
- Rato, A. Rauber, A. S. Kluge, D. C. & dos Santos, G. R. (2015) Designing speech perception tasks with TP. In Mompean JA & Fouz-González J (eds.), *Investigating English pronunciation: Trends and directions*. Basingstoke: Palgrave Macmillan, 295–313. https://doi.org/10.1057/9781137509437_13
- Rosell-Aguilar, F. (2009) Podcasting for language learning: Re-examining the potential. In Lomicka L & Lord G (eds.), *The next generation: Social networking and online collaboration in foreign language learning*. San Marcos: CALICO Monograph Series, 13–34.
- Saito, K. (2013) Reexamining effects of form-focused instruction on L2 pronunciation development: The role of explicit phonetic information. *Studies in Second Language Acquisition* 35(1), 1–29. <https://doi.org/10.1017/S0272263112000666>
- Schmidt, R. W. (1990) The role of consciousness in second language learning. *Applied Linguistics* 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
- Selinker, L. (1972) Interlanguage. *Internal Review of Applied Linguistics in Language Teaching* 10(1–4), 209–230.
- Sharwood-Smith, M. (1993) Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition* 15(2), 165–180. <https://doi.org/10.1017/S0272263100011943>
- Smith, J. & Beckmann, B. (2010) Noticing-reformulation tasks as a stimulus towards continued autonomous phonological development. *New Zealand Studies in Applied Linguistics* 16(1), 36–50.
- Suter, R. W. (1976) Predictors of pronunciation accuracy in second language learning. *Language Learning* 30(2), 271–287.
- Swain, M. (1985) Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In Gass SM & Madden CG (eds.), *Input in second language acquisition*. New York: Newbury House, 235–256.
- Thomson, R. I. (2011) Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal* 28(3), 744–765. <https://doi.org/10.11139/cj.28.3.744-765>
- Thomson, R. & Derwing, T. M. (2015) The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics* 36(3), 326–344. <https://doi.org/10.1093/applin/amu076>

- Walker, R. (2005) Using student-produced recordings with monolingual groups to provide effective, individualized pronunciation practice. *TESOL Quarterly* 39(3), 550–558. <https://doi.org/10.2307/3588495>
- Wang, X. & Munro, M. J. (2004) Computer-based training for learning English vowel contrasts. *System* 32(4), 539–552. <https://doi.org/10.1016/j.system.2004.09.011>
- Zampini, M. L. (1996) Voiced stop spirantization in the ESL speech of native speakers of Spanish. *Applied Psycholinguistics* 17(3), 335–354. <https://doi.org/10.1017/S0142716400007979>


Appendix

Mean scores (SD) and improvement (imp) made in the different production tasks

	G1						G2					
	Imitation task											
	pre	%	post	%	imp	%	pre	%	post	%	imp	%
/b/	3.46 (1.3)	69.2	3.58 (1.1)	71.7	0.13	2.5	3.71 (1.3)	74.3	4 (1.2)	80	0.29	5.7
/d/	4.63 (0.7)	92.5	4.54 (0.9)	90.8	-0.08	-1.67	4.52 (0.7)	90.5	4.71 (0.6)	94.3	0.19	3.8
/g/	4.42 (0.9)	88.3	4.54 (0.8)	90.8	0.13	2.5	4.48 (0.8)	89.5	4.67 (0.6)	93.3	0.19	3.8
/z/	2.42 (1.7)	48.3	2.88 (1.6)	57.5	0.46	9.17	0.95 (1.5)	19	1.71 (1.6)	34.3	0.76	15.2
	Sentence-reading task											
	pre	%	post	%	imp	%	pre	%	post	%	imp	%
/b/	8.21 (4.2)	54.7	9.33 (3.8)	62.2	1.13	7.5	7.19 (4)	47.9	9.62 (3.4)	64.1	2.43	16.2
/d/	10.75 (3.1)	71.7	11.17 (3.5)	74.4	0.42	2.8	9.71 (3.3)	64.8	10.76 (2.3)	71.7	1.05	7
/g/	8.42 (3.3)	56.1	8.54 (3.5)	56.9	0.13	0.8	7.24 (4.5)	48.3	9.1 (3.2)	60.6	1.86	12.4
/z/	4.67 (5.1)	23.3	7.63 (5.2)	38.1	2.96	14.8	2.14 (3.2)	10.7	2.38 (3.1)	11.9	0.24	1.2
	Timed picture-description task											
	pre	%	post	%	imp	%	pre	%	post	%	imp	%
/b/	1.71 (1.3)	34.2	2.08 (1.6)	41.7	0.38	7.5	0.9 (1.2)	18.1	1.52 (1.4)	30.5	0.62	12.4
/d/	2.96 (1.5)	59.2	3.25 (1.4)	65.0	0.29	5.8	2.9 (1.2)	58.1	3.19 (1.3)	63.8	0.29	5.7
/g/	1.58 (1.3)	31.7	1.71 (1.6)	34.2	0.13	2.5	1.24 (1)	24.8	1.38 (1.3)	27.6	0.14	2.9
/z/	1.42 (1.8)	17.7	1.96 (2.3)	24.5	0.54	6.8	0.33 (1)	4.2	0.62 (1.5)	7.7	0.29	3.6

About the author

Jonás Fouz-González holds a PhD in English Applied Linguistics from the University of Murcia (2015). His research interests are English phonetics and phonology, second language acquisition, computer-assisted pronunciation training and mobile-assisted language learning.

Author ORCID.  Jonás Fouz-González, <http://orcid.org/0000-0003-4952-772X>