Accent Imitation on the L1 as a task to improve L2 pronunciation

Charlotte Everitt

MA Thesis 2015

Supervised by

Dr. Joan Carles Mora

Applied Linguistics and Language Acquisition in Multilingual Contexts

Department of English and German Philology

Universitat de Barcelona

Table of Contents

Abstract ... 3
INTRODUCTION ... 3
Research questions ... 8
Hypotheses ... 8
METHOD ... 9
Participants ... 9
Research Design ... 9
Testing ... 10
Identification ... 10
Rated Dissimilarity ... 12
Picture Naming ... 13
Delayed Sentence Repetition ... 14
Training ... 15
Word Imitation ... 15
Sentence Imitation ... 16
Dialogue ... 16
Spontaneous Conversation ... 17
RESULTS ... 17
Comparing groups ... 18
Research Question 1 ... 18
Research Question 2 ... 22
Research Question 3 ... 25
Research Question 4 ... 30
DISCUSSION ... 31
Research Question 1 ... 31
Research Question 2 ... 32
Research Question 3 ... 33
Research Question 4 ... 34
Student Evaluation ... 35
Limitations ... 35
CONCLUSION ... 37
Acknowledgements ... 39
REFERENCES ... 40
APPENDICES ... 43
Appendix 1 Training Stimuli (Group 1) ... 43
Appendix 2 Training Stimuli (Group 2) ... 45
Appendix 3 Testing ... 47
Appendix 4 Questionnaire ... 49

Abstract

Three groups of Spanish/Catalan L1 adolescent learners of English as a foreign language took part in the present study. The aim of the study was to assess the effects of accent imitation training in the L1 on L2 Voice Onset Time (VOT) perception and production. One experimental group underwent accent imitation training on their L1 and another experimental group received L2 training. A control group did not receive any training. Results showed that both training groups improved in their English VOT perception and production, but the L1 accent imitation group saw wider benefits than the L2 training group. These results imply that L2 VOT perception and production can be modified after relatively little training, and that accent imitation on the L1 is an effective training methodology.

INTRODUCTION

Anyone learning a second language (L2) will encounter multiple obstacles that may prevent them from achieving native-like attainment, particularly after the so-called 'Critical Period' (e.g. Johnson & Newport 1989). Accent is a notoriously difficult obstacle to overcome; although studies have shown that large amounts of native input and/or phonetic training can help significantly (Flege, Bohn & Jang 1997; Aliaga-García 2007, among others). However, in the majority of foreign language learning, access to native-input and/or phonetic training is extremely scarce. Often the teachers are not even native-speakers of the language, and large class sizes prevent adequate output. This is worsened further if the learners are of a low proficiency, as all of their attention is on meaning and accent is discarded and seen as unnecessary effort. However, when accurate perceptual representations for L2 are not developed, L1 categories may be used (Best & Tyler, 2007; Flege, 1995), meaning speech production will remain heavily accented and potentially unintelligible, and should therefore not be overlooked.

For native-speakers of Spanish learning English as an L2, the acquisition of plosives can be particularly difficult. This is because the voice onset times (VOT) of the two languages differ significantly in usage, but this difference often goes unnoticed as the consonants are often used in the same way phonologically. VOT is defined as "the

timing relation between the first distinct pulse in the amplitude (plosive release) and the zero crossing of the first periodic pulse (onset of voicing)" (Neuhauser 2011: 1463).

The plosives /b d g/ in Spanish are prevoiced, meaning the VOT is actually negative, up to -40ms, whereas the same consonants in English would have a VOT of approximately 0ms (Benki 2005, Docherty 1992). With voiceless plosives /ptk/, a native Spanish speaker would normally produce a VOT of between 0-10ms (Lisker and Abramson 1964), whereas English native speakers use aspiration to form voiceless stops and so VOT would be around 30ms or more (Benki 2005). Although this distinction is made at the phonetic level, phonologically these voiced and voiceless oral stops are used in English and Spanish in similar ways to convey differences in meaning (Mora, Rochdi & Kivistö 2014). Consequently the functional load of this voicing contrast is high, as the L2 phonetic categories may be mapped onto pre-existing L1 categories in the long-term memory (Best & Tyler 2007). Without accurate perception, accurate L2 pronunciation may be much harder, and Spanish/Catalan bilinguals have been found to produce these English stops inaccurately (Flege *et al*. 1997). The ability to perceive cross language differences in VOT is likely to help learners develop L2-specific phonetic categories distinct from those of the L1, therefore raising awareness of cross-language VOT differences may eventually help learners produce L2 stops more accurately.

Flege et al. (1998) studied VOT in the production of /t/ of native-English speakers and Spanish speaking learners of English, and found that longer VOT values were produced preceding high vowels than low vowels. Similarly, VOT was longer in one-syllable words than in two syllables words. Yavaş and Wildermuth (2006) found similar results for vowel height, and also found differences in VOT according to place of articulation, with VOT increasing from bilabial (/p/) to alveolar (/t/) to velar (/k/).

As /b p/ and /g k/ are articulated in the same place for both English and Spanish (/b p/: bilabial, /g k/: velar), VOT is the only characteristic differentiating these plosives in the two languages. However, articulation of /d t/ are alveolar in English but dental in Spanish, making this contrast easier for learners to hear and to start to modify. Once place of articulation is changed, VOT is likely to become more target-like too.

In view of the above findings, this study included a variety of vowel heights and syllable lengths (mono- and di-syllabic words), as well as an equal number of bilabial, alveolar and velar stops, to ensure that participants were exposed to these variations both during training and testing.

In order to help these learners acquire this non-distinctive phonetic difference, it is necessary to provide enough opportunities to receive native-input and produce ample output, as currently both of these components are often missing from the English as a Foreign Language (EFL) classroom. Accent imitation has only been used thus far for the purpose of testing rather than training (although see Hilton 2005 for an anecdotal account of 'reverse accent mimicry'). Flege & Hammond (1982) found that native English speakers were able to modify their VOT to become more Spanish-like when reading English sentences with a Spanish accent. Similarly, Neuhauser (2011) analysed the VOT of German speakers when imitating a French accent and found that participants reduced their VOT of voiceless plosives during imitation. The same was found for voiced stops, with some participants even using voicing (e.g. 0ms VOT or less). Likewise, Mora *et al*. (2014) compared VOT of word-initial pre-vocalic /p t k/ of Spanish and English words, as well as Spanish words pronounced with an English accent. Spanish-speakers were found to produce longer VOTs in English *and* English-accented Spanish than Spanish words, showing phonological awareness of cross-language VOT differences. These studies seem to provide evidence that this non-distinctive phonetic difference between languages may remain malleable and receptive to learning in second language acquisition. It therefore seems pertinent to study the extent to which this cross-linguistic category can be trained. One way of doing this is to expose learners to high-variability input in which native-English speakers are speaking Spanish with a strong English accent. High-variability of input in terms of talkers during training has been shown to improve the generalizability of learned stimuli into novel contexts (Lively et al. 1993, Bradlow et al. 1997).

This study uses an accent imitation training paradigm to promote awareness of cross-language VOT-based differences between English and Spanish stops. This involves the imitation of Spanish words and sentences produced by 6 native speakers from a variety of native-English backgrounds, using a strong English accent. Although

this may seem counterintuitive as the learners will be speaking in their first language (L1), it enables low proficiency learners to produce much more output, which enables both intensive and extensive practice of the L2-specific articulatory features of English stops. In addition, the training allows learners to concentrate solely on the differences in sound, raising awareness of cross-language VOT differences and aiding the formation of new L2 sound categories. It also provides articulatory practice that should in turn facilitate automatization of the articulatory gestures involved in L2 sound production. It was hoped that the above advantages would allow the accent imitation group to improve their perception and production more than the L2 training group.

Much variation exists between studies regarding the length of cross language training. Long-term studies often range from 6 to 45 sessions, and appear to be "necessary for learners to perceive many non-native phonetic categories" (Logan & Pruitt 1995: 365). However, though many studies have used a considerable number of training sessions, the largest gains are likely to be found early in training (Lively, Logan & Pisoni 1993). Carlet & Cebrian (2014) found significant improvements in perception of vowel and consonant contrasts after a three-week training period. Even short-term studies (those lasting only one session) have shown participants to significantly improve (e.g. Carney *et al*. 1977, Pisoni *et al*. 1982). It is likely that the perception of VOT may be "easier to modify" (Strange & Dittman 1984: 142) than other distinctions, as it varies along temporal rather than spectral dimensions (Logan & Pruitt 1995: 365). Given the time constraints imposed on this study and the significant results found from previous short-term research, it was decided that four training sessions would be adequate for the purpose of this study.

Previous research on the subject of phonetic training appears almost solely in a laboratory setting. Although the results from these studies are extremely important and give an insight into the possibilities phonetic training has to offer, research in this area desperately needs to move into the EFL classroom, so that results can be generalized to a setting where a much larger proportion of language learning actually occurs. For this reason the present study chose an EFL classroom as the context for research.

Previous studies have mainly used identification and discrimination tasks both in the training and the testing of participants when studying cross-language VOT

perception. Identification tasks have been used in cross-language research to show 'how listeners develop phonetic categories and how they treat within-category variability' (Logan & Pruitt 1995, 358). Pisoni *et al*. (1982) used identification tasks to train American English listeners in the perception of prevoicing using a continuum, and saw results after only 10 minutes of training. Carney *et al*. (1997) used identification and discrimination tasks to test participants' perception of bilablial stop consonants differing in VOT before and after training. The training used a very similar procedure to that of the testing tasks. Participants were found to improve their within-category discrimination, even close to phonetic boundaries. More recently, Collet *et al*. (2015) studied "changes in voicing identification, discrimination, and categorical perception induced by identification training" (463). Participants were shown to modify their phonological perception after short-term training, even for VOT values not used in training. Identification tasks have been shown to be more effective than discrimination tasks at generalizing the stimuli that were not presented during training (Logan & Pruitt 1995, but see Flege 1995). For this reason, the present study used an Identification task alongside a Rated Dissimilarity task (see Cebrian *et al*. 2011), which required participants to rate two stimuli for *amount* of perceived difference on a 1-9 scale. This allowed for a much more detailed insight into participants' perception of phonetic categories between languages.

To assess the production gains of training, it is beneficial to use tasks eliciting productions in different contexts. Picture Naming Tasks have been used in previous studies to measure VOT productions (e.g. Olson 2013), particularly for low proficiency learners, as the learners do not have to produce spontaneous speech (e.g. Simon 2009). As the training focused mostly on imitation, it was important to use a test that showed if the effects of training transferred to a more spontaneous context, showing their *generalizability* (Logan & Pruitt 1995: 353).

Similarly, delayed mimicry paradigms have also been used previously to avoid direct imitation (see Flege & Hammond 1982, Mora *et al*, 2014). Instead, participants must use long-term phonetic representations of differences between the L1 and L2. Unlike in a spontaneous conversation setting, all participants produce the same sentences, allowing for an easy comparison between subjects. The ability of participants

to generalize their gains to items not present in training is extremely important. Logan & Pruitt (1995: 371) state that "[w]hen subjects show generalization to novel stimuli or tasks, we can be more confident that robust learning has occurred".

The training and testing tasks in this study have been used to allow for more generalizable and reliable perception and production data to be collected and analysed.

In light of the above, the following *research questions* are proposed:

1. Do learners' perception of English voiced and voiceless stops /b/ and /p/ become more target-like after training?

2. Do learners perceive greater differences in VOT of the voiceless stop /p/ between languages after training?

3. Do learners' productions of English voiceless stops /ptk/ become more target-like after training?

4. Are there differential gains in VOT perception and production as a function of training group?

Hypotheses:

1) Both groups receiving training will:

   a) Become more target-like in their perception of the English word-initial voiced and voiceless oral stops /b/ and /p/.

   b) Perceive greater differences in VOT of the voiceless stop /p/ between languages.

   c) Produce VOTs of English word-initial voiced and voiceless oral stops /ptk/ that are more target-like.

2) The group receiving L1 accent imitation training will outperform the group doing L2 training in terms of perception and production gains.

METHOD

Participants

Participants consisted of three groups of adolescent (aged 14-15) L1 Spanish-Catalan EFL learners.

Group 1: Accent imitation training on the L1 (n=17)

Group 2: L2 training (n = 16)

Group 3: Control group (n=16)

The Control Group only took part in the testing and did not do any of the training, and instead continued with their regular EFL classes. Due to class constraints, participants could not be randomly assigned to experimental and control groups. However, the teacher provided a proficiency score (out of 10) for each student to control for differences in L2 proficiency. All participants attended the same public school taught by the same teacher to reduce teaching effects.

To back up the teacher's report, a vocabulary size test (X_lex) was also administered, along with a linguistic background questionnaire to control for any large differences in background and exposure to English.

Research Design:

<table>
<tr><th></th><th>Pre-test</th><th>Session 1</th><th>Session 2</th><th>Session 3</th><th>Session 4</th><th>Post-test</th></tr>
<tr><td>Group 1</td><td rowspan="3">1) Identification
2) Rated Dissimilarity
3) Picture Naming
4) Delayed Sentence Repetition</td><td colspan="4" rowspan="2">- Word Imitation
- Sentence Imitation
- Dialogue
- Spontaneous conversation</td><td rowspan="3">1) Identification
2) Rated Dissimilarity
3) Picture Naming
4) Delayed Sentence Repetition</td></tr>
<tr><td>Group 2</td></tr>
<tr><td>Group 3</td><td colspan="4">Normal EFL classes continue</td></tr>
</table>

Testing *(see Appendix 3)*

Perception was tested using both an Identification and a Rated Dissimilarity Task while production was tested using both a Picture Naming Task and a Delayed Sentence Repetition Task. Testing of all 3 groups was carried out before and after training over a 4 week period in April and May 2015. Instructions for all tasks were given in English as students should be in English 'mode' but translations into Spanish or Catalan were provided by the teacher to ensure all students understood before beginning the task.

Identification Task:

Materials

A continuum ranging from /ba/ (VOT -120ms) to /pa/ (+120ms) was created using stimuli recorded by a native speaker of (American) English. These syllables were taken from the English words 'pack' (with aspiration) and 'back', and the Spanish word 'bajo' (with prevoicing). These monosyllabic words were digitally recorded in a soundproof booth at the Phonetics Laboratory of the Universitat de Barcelona in mono using a Marantz Solid State Recorder (PMD660) at a 44.1kHz sampling rate, with a Shure SM58 unidirectional microphone. The words were inserted into carrier phrases, to ensure that falling intonation was elicited, but multiple repetitions were also used to guarantee that all the items were recorded in the necessary manner. Once the stimuli were recorded, the WAV files were then opened in Praat (Boersma 2001) and the best token of each syllable was segmented at zero crossings (to prevent clicking sounds) and extracted.

A 16-step continuum was created, ranging from -120ms to +120ms of voicing. This range and interval size was chosen as it is large enough so that each extreme will only receive unambiguous answers, with intervals small enough to show subtle changes in the phonetic category boundaries. The aspirated part of the spectrum was made by taking the /pa/ syllable and lengthening/shortening the amount of aspiration in intervals of 16ms, ranging from 0ms to +120ms. It was decided that the original step at -8ms VOT should be changed to 0ms VOT, as the researcher was interested to know what differences would exist in perception of this part of the continuum, where cross-language

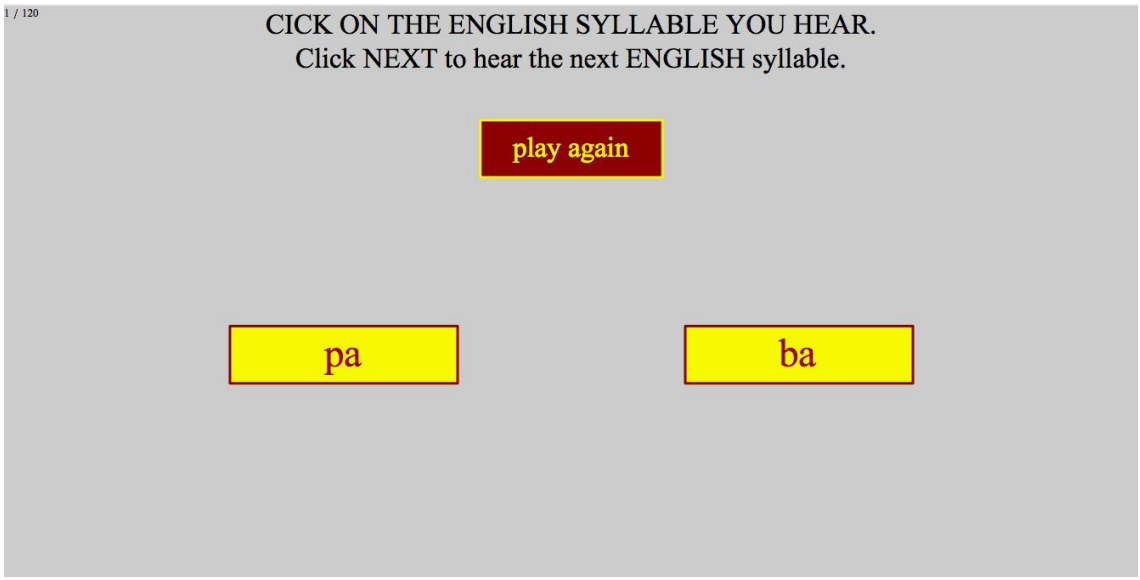
differences in perception are most noticeable. Prevoicing was taken from 'bajo' and added to the /ba/ syllable made using the word 'back'. The amount of prevoicing was then lengthened/shortened to the required length, ranging from 0ms to 120ms (a VOT of -120ms).The end of each audio file was then ramped for 20ms at zero crossings so that the stimuli would sound more authentic and would not end so abruptly. In order to ensure all stimuli were of the same volume, the sounds were then preprocessed by removing 50-Hz energy and normalizing amplitude. Both of these processes were carried out using GSU Tools (Owren 2008). The continuum was then listened to by a native speaker of English and of Spanish/Catalan to ensure that they heard both a /b/ and /p/ at some point along the continuum.

It was decided that the final two steps at each end of the continuum would not be included in this task, as they were too exaggerated in comparison with what native-speakers of each language would produce. This left 12 steps in total.

Procedure

Before the identification task took place, participants were told they would hear syllables and they had to decide if they heard /pa/ or /ba/. It was stressed that these were English syllables, as it was important to ensure participants were focusing on their L2 representations. The practice round presented each of the 12 stimuli once, so that the listeners could become accustomed to the task. They could then take a short break if necessary, which also provided an opportunity to ask any questions they may have had. The real test presented each of the 12 stimuli ten times (120 presentations), although participants were allowed to take a short break after half of the stimuli had been presented. This was done to decrease the chances of any fatigue effect. 10 repetitions of each stimuli were presented to ensure that the data collected are an accurate representation of the phonetic categories that the learners have, as too few repetitions makes it difficult for a reliable mean to be calculated. The participants were asked to select a response after listening to the stimuli either once or twice depending on their preference, and then click 'NEXT' to proceed. In both the practice and the real test, all stimuli were randomised, with doublets avoided so that any answers given would not be affected by the preceding stimuli.

Figure 1: Screenshot of Identification Task



Analysis

Participants' responses for each stimulus presentation were categorized: ba = 0, pa = 1. This enables a mean score to be calculated for each of the 12 stimuli, ranging from 0 to 1. These results were then used to calculate phonetic category boundary location and steepness (categoricality of /b-p/ separation), using logistic regression.

Rated Dissimilarity Task

Materials

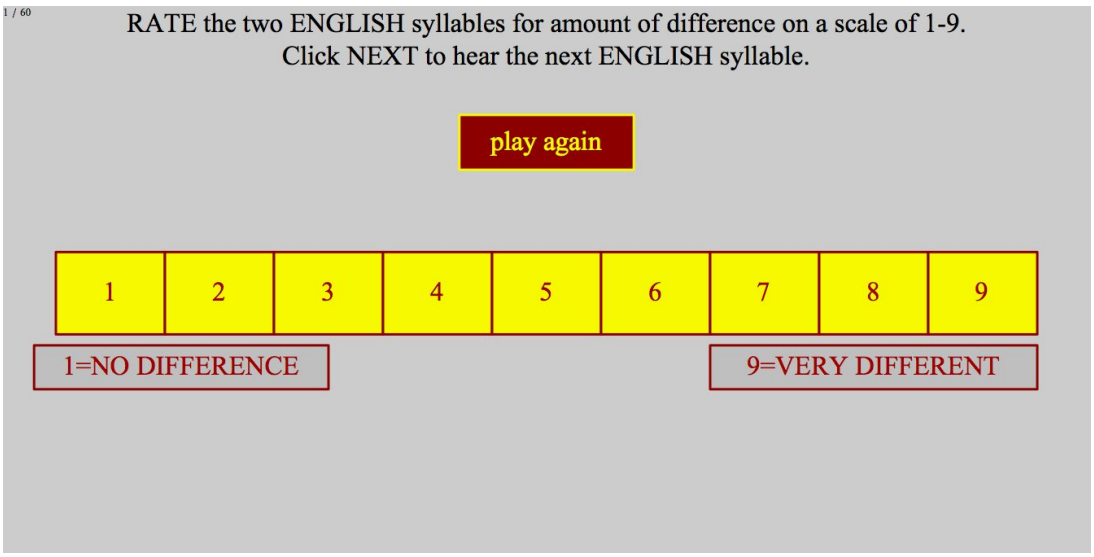
The same stimuli were used as for the Identification task (the /b-p/ continuum), but presented in pairs. However for this task all 16 stimuli were used, as a large range was necessary to encourage the use of the entire rating scale.

Procedure

Participants were asked to rate how different the English syllables sounded to them on a scale of 1-9 (1=No difference, 9= Very different). A practice round presented 10 pairs that represented a variety of differences in VOT, so that participants would become aware of the range of the scale. The test round presented 30 pairs of stimuli with a range of VOT distances, each presented twice. In both the practice and test, stimuli

were randomised without doublets and each stimuli pair could be repeated once before a selection had to be made and 'NEXT' clicked, as for the identification task.

Figure 2: Screenshot of the Rated Dissimilarity Task



Analysis

The responses from this task (ratings on a scale of 1 to 9) were categorized into 'Cross-language' and 'Within-language' pairs. Differences in the responses for cross-language pairs were analyzed at pre- and post-test to assess improvements in cross-language perception.

The aforementioned perception tests were accompanied by two production tests. Productions of isolated words have been found to differ in VOT from productions within sentences (Lisker & Abramson 1967). Therefore two tasks eliciting both production types were used, a picture naming task and a delayed sentence repetition task.

Picture Naming Task

Materials

Twelve target pictures and four distractor pictures were chosen. All pictures were taken from the Snodgrass & Vanderwart (1980) database of standardised pictures. For the target pictures, two pictures were chosen for each of the consonants /bdg/ and /ptk/, with a variation of mono- and disyllabic words as well as high, mid and low vowels (as this has been shown to influence VOT (Flege & Munro 1994).

Procedure

The task started with a familiarisation round, where each stimulus was presented once randomly with its name on the screen. Participants were told to name the picture as well as they could and to remember it. The results from the familiarization round were not included in the analysis. In the test round, each stimulus was presented twice, and participants were told to name the picture as *well* and not as fast as they could, to avoid hurried pronunciation which may affect VOT. In both cases the stimuli were randomised. Elicitations were recorded automatically through DmDx (Forster & Forster 2003).

Analysis

Productions from this task were used to calculate group gains in VOT of the consonants /ptk/ as a whole and individually in the context of 'isolated words'.

Delayed Sentence Repetition Task

Materials

This task also required stimuli to be recorded by native speakers. In this case ten sets of questions and answers were used which contained voiced and voiceless word-initial stops in English, with syllable and vowel variation, as above. Target words were at the end of a sentence and were stressed, as this has been found to increase VOT within sentences (Lisker & Abramson 1967).

These were also recorded in the same way as the perception stimuli, but by six native speakers (three male, three female) with differing regional accents as high variability in input has been shown as beneficial (Lively et al. 1993, Bradlow et al. 1997). These recordings were opened in Praat, then segmented and extracted using a TextGrid. Each audio file contained a 1000ms period of silence at beginning and end to allow participants enough time to process and understand the contents. All sound files were then preprocessed to standardise amplitude.

Procedure

This task presented recordings of a question followed immediately by the answer. The same question was then presented alone, and the participants had to repeat the answer they had previously heard. Whenever a sentence was played, the written text would accompany it, to ensure that participants were able to memorise and repeat the sentence despite low levels of proficiency. A familiarisation round was not included in this task but each question and answer pair was randomly presented three times. DmDx was used to record productions, as with the Picture Naming task.

Analysis

Productions were analyzed in the same way as for the previous task, but this time in the context of 'words in sentences'.

Training *(see Appendices 1 & 2)*

Both experimental groups completed four training sessions at weekly intervals, each lasting approximately 40 minutes. Group 1 received training in the L2 (English) and Group 2 received accent imitation training in the L1 (Spanish). Group 3 (Control) continued with their normal EFL classes.

Word Imitation

Materials

96 words (12 x 4 sessions x 2 languages) were recorded by the same six native speakers as for the Delayed Sentence Repetition task. These words targeted English word-initial voiced and voiceless stops with a variety of vowel height and syllable length (mono- and disyllabic). It was decided that distractors were unnecessary as the participants were naïve listeners, and the second half of the sessions provided plenty of variety. Praat was used to decrease differences in speed between talkers and languages using the 'Convert' > 'Lengthen (overlap-add)' function. A rate of 0.75 was used to slow

words down or 1.25 to speed words up. A 1000ms period of silence was added to give participants time to process the information and preprocessed to balance amplitude.

Although the native speakers recorded all stimuli, only two words from each speaker were presented to participants per session, with the consonant changing each week, to ensure high variability of input and an equal distribution of all talkers.

Procedure

Three randomized repetitions were presented to participants individually through DmDx. Participants' imitations during training were recorded via DmDx. Although their imitations were not examined in this study, recording their elicitations motivated the students to do their best (as they thought the recordings were important) and will allow for future analysis to be carried out at a later stage. So as not to distract the learners, feedback was not giving during imitation.

Sentence Imitation

Materials

48 sentences (6 x 4 sessions x 2 languages) were recorded by the same native speakers as above. These stimuli had the same characteristics as for word imitation, and were modified in the same way. Similarly, each native speaker produced one sentence per session, alternating consonants.

Procedure

The procedure was identical to that used for word imitation.

Dialogue

Materials

Each mini-dialogue targeted English voiced and voiceless word-initial stops. This activity was done in pairs and was included to make the training more interesting

for adolescent learners, as well as giving them an opportunity to put the knowledge gained from imitation into practice.

Procedure

Students were first put into pairs, which changed every session. An example reading was given by the researcher and pairs were then given five minutes to practice both speaking parts with their partner before recording. Learners were told to read the dialogues in their best English-accent, regardless of whether the dialogue was in Spanish or English. Recordings were made using Praat. The researcher walked around the classroom listening to participants' productions and giving constructive feedback to motivate students to continue doing their best.

Spontaneous Conversation

Materials

No specific materials were provided for this task, only a conversation topic. As for the dialogue, this task was provided to make the training more interesting and provide an extra opportunity for practice and output.

Procedure

The same pairs were used as for the dialogue, changing weekly. No practice time was given for the spontaneous conversation; learners were asked to immediately record a short conversation with their partner in Praat using a very strong English accent. As above, the researcher provided constructive feedback.

RESULTS

Unfortunately, participants had to be excluded due to missing a training or testing session (n=8), doing a task incorrectly (n=3), not saving the task properly (n=2) or problems with the recording (n=2). It was decided to only include participants who had data for all of the tasks, to enable a fair comparison. The only exception to this was the

X_lex, which had two missing data points but the Teacher's Score did not suggest that these two participants were outliers. The subsequent groups were the following:

Group 1: L2 training (*n*=11)

Group 2: Accent imitation training on the L1 (*n* = 13)

Group 3: Control group (*n*=10)

Comparing groups

For effects of training to be comparable across groups, participants needed to have a similar proficiency level. The teacher provided a subjective score for each participant out of 10 for overall proficiency, and the X_lex vocabulary size test was administered at pre-test. The scores from these tests were then submitted (separately) to a one-way ANOVA, with Group type (L2 training, accent imitation on L1 and Control Group) as the independent variable and Teacher Score/X_lex score as the dependent variable. Results showed no significant main effect of Group for Teacher Score ($F(2,31)=.22$, $p=.803$), or for X_lex score ($F(2,29)=.79$, $p=.461$). The linguistic background questionnaire did not suggest large differences in exposure to English.

As no significant differences were found across groups for either of the proficiency measures or the questionnaire responses, it was concluded that all groups were comparable.

The majority of the data was not normally distributed, requiring the use of non-parametric tests. Where these were used, a *mean* score was calculated for individual responses/productions, but the *median* was used to compare groups.

Research Question 1

Does learners' perception of English voiced and voiceless stops /b/ and /p/ become more target-like after training?

Participant response in the Identification task was used as the Dependent variable, with Stimulus as the Independent variable to perform logistic regression (see Collet *et al*. 2015) for each group (*Graphs 1-3)*. The logistic regression curve provided values corresponding to phonetic boundary location and categoricality. A more

target-like perception would be illustrated by the curve moving to the right and becoming steeper from pre- to post-test.

Related-samples Wilcoxon Signed Rank tests showed no significant difference between pre- and post-test for boundary location or for boundary categoricality in perception for any of the three groups (see *Tables 1 & 2*).

Table 1: Boundary categoricality

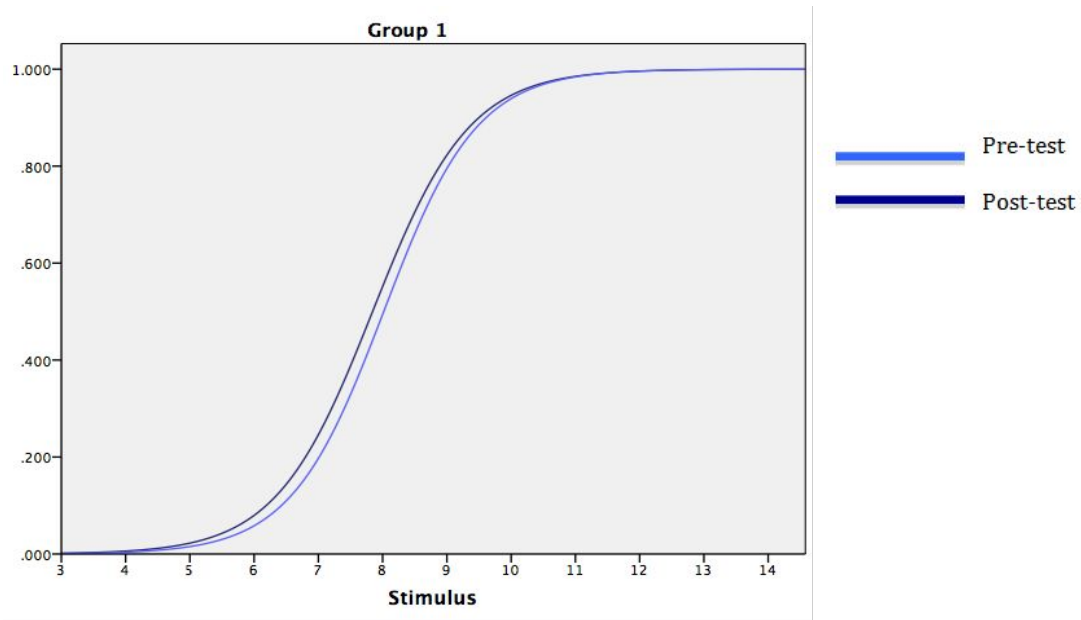
	Median (pre)	Median (post)	*T*	*p*
Group1	0.2	0.19	23	0.646
Group2	0.2	0.19	29	0.249
Group3	0.2	0.21	23	0.646

Table 2: Boundary Location

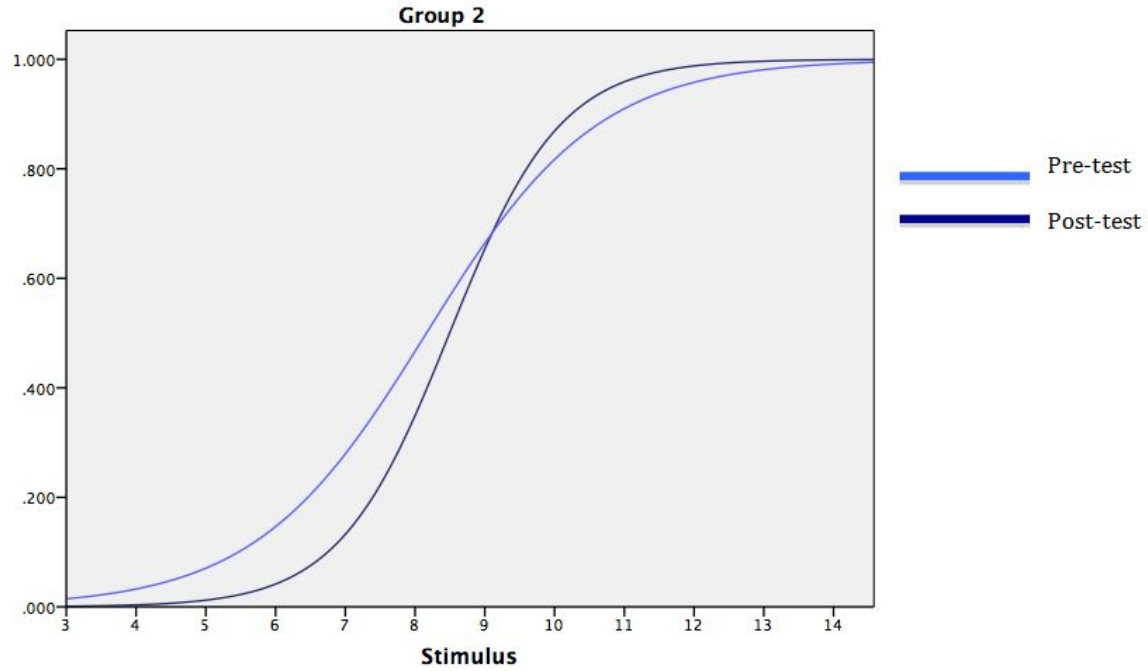
	Median (pre)	Median (post)	*T*	*p*
Group1	6.04	5.82	23	0.959
Group2	6.05	6.2	55	0.209
Group3	6.05	6.02	23	0.646

Logistic curves (showing mean response per stimulus (/ba/=.000, /pa/=1.000)):

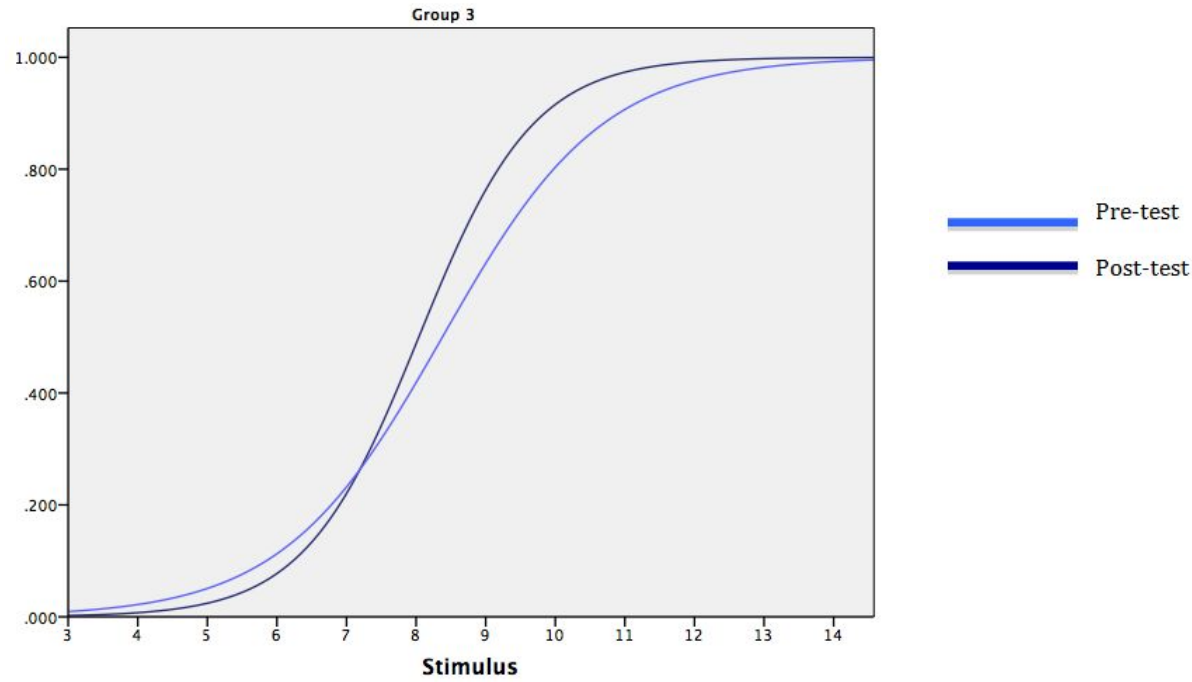
Graph 1: Logistic Curve at pre- and post-test, Group 1



Graph 2: Logistic Curve at pre- and post-test, Group 2



Graph 3: Logistic Curve at pre- and post- test, Group 3



As no significant changes were found in the boundary location or categoricality, it was decided that the most 'critical values' should be analyzed. This involved the steps of the continuum ranging from -24ms VOT to +24ms VOT (Steps 7,8,9 and 10). These steps were chosen as they generally represent the cross-over point between Spanish-like VOT and English-like VOT. Mean responses to the ten presentations for each step were calculated per participant, then group medians were compared at pre-test and post-test using a series of paired-samples Wilcoxon Signed Rank tests, to ascertain if any of the groups became more target-like. Neither Group 1 nor Group 3 showed significant differences for any of the 'critical' steps from pre-test to post-test (see *Tables 3 & 5)*. Group 2 did show significant differences between pre- and post- test for Step 9 (*Mdn*=0.9 vs. *Mdn*=0.7 respectively; *T*=6.0, *p*=.023 (see *Table 4)*.

Graph 4: ID responses Step 9

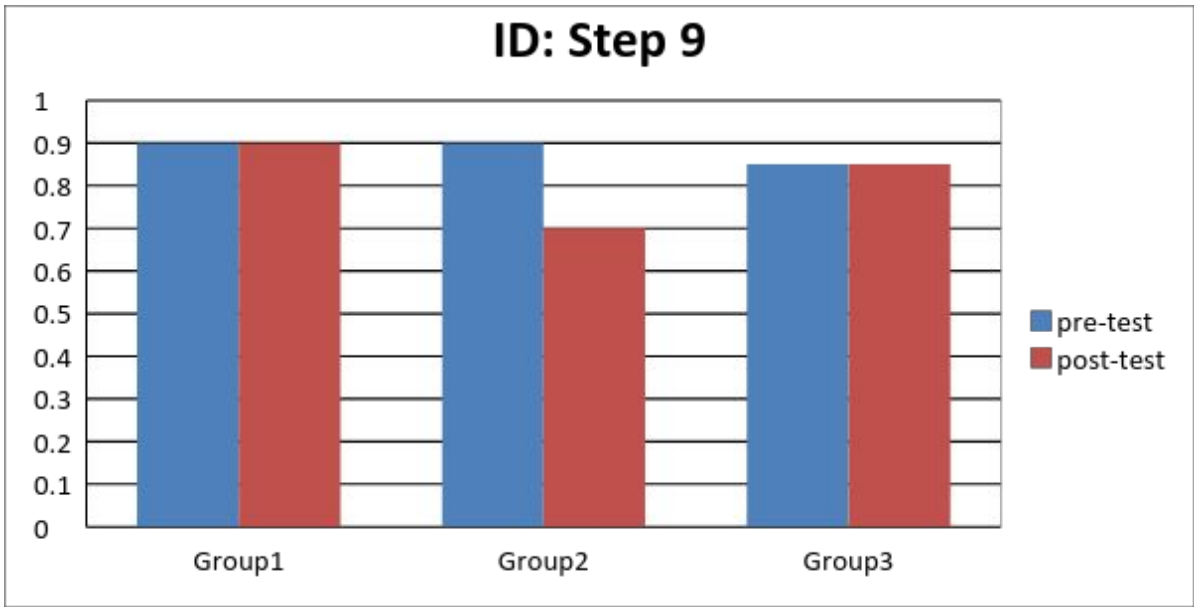


Table 3: ID Critical Values, Group 1 (L2 accent training)

VOT	Median (pre)	Median (post)	*T*	*p*
-24ms	0.1	0	20.5	0.72
0ms	0.8	0.80	21.0	0.86
8ms	0.9	0.90	16.0	0.78
24ms	1	1	6.0	0.10

Table 4: ID Critical Values, Group 2 (Accent imitation on the L1)

VOT	Median (pre)	Median (post)	*T*	*p*
-24ms	0	0	5.0	0.125
0ms	0.8	0.80	21.0	0.857
8ms	0.9	0.70	6.0	0.023
24ms	1	1	8.5	0.197

Table 5: ID Critical Values, Group 3 (Control Group)

VOT	Median (pre)	Median (post)	*T*	*p*
-24ms	0.1	0.28	14.5	0.608
0ms	0.9	0.90	14.5	0.399
8ms	0.85	0.99	14.5	0.618
24ms	1	1	7.0	0.461

Research Question 2

Do learners perceive greater differences in VOT of the voiceless stop /p/ between languages after training?

Data from the Rated Dissimilarity task were first analyzed by calculating a mean response (from the 1 to 9 scale) for each participant for all cross-language /p/ pairs. These mean ratings were submitted to a mixed-design ANOVA to assess changes in cross-language perceptual ratings over time. Group type (L2 imitation, L1 accent imitation and Control Group) was the between-groups factor and Time (pre-test,

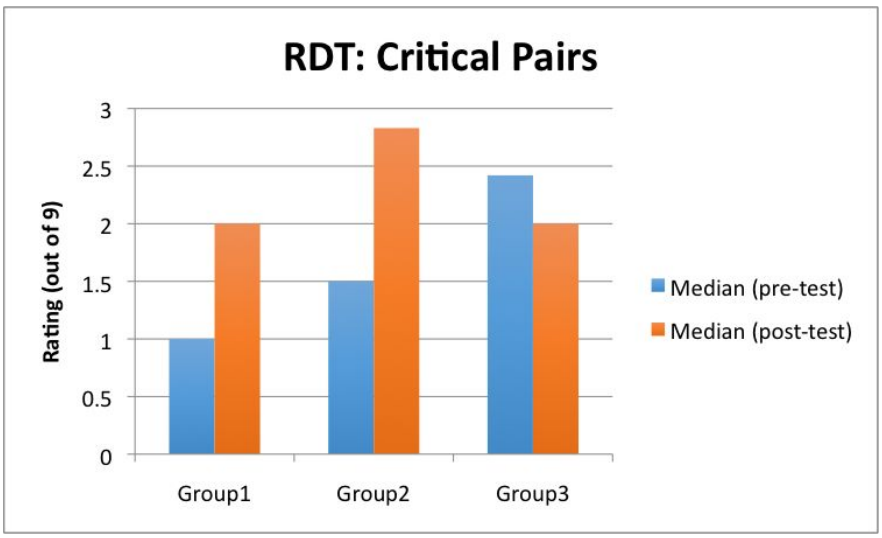
post-test) was the within-groups factor. No significant main effect was found of Time (pre- to post-test) on mean cross-language /p/ ratings ($F(1,31)=3.28$, $p=.08$). The interaction Time x Group type was not found to be significant ($F(2, 31)=.20$, $p=.82$). This shows that overall mean ratings for cross-language /p/ pairs did not significantly change over time.

However, as these measures looked at all cross-language pairs, some of the VOT differences to be rated were very large whereas others were very small. Ratings at pre- and post-test may therefore have cancelled each other out. For this reason, it was decided that only the 'critical' cross-language pairs should be analyzed (i.e. VOT 0ms (Spanish /p/ vs. English /p/ at 24, 40 and 56ms). The results show that both training groups significantly increase their dissimilarity ratings at post-test (see *Table 6 & Graph 5)*.

Table 6: RDT Critical Pairs

	Median (pre)	Median (post)	*T*	*p*
Group1	1	2	434.0	.001
Group2	1.5	2.83	49.5	.025
Group3	2.42	2	29.5	.838

Graph 5: RDT Critical Pairs



As significant differences were found from pre- to post-test for both Group 1 (*Mdn*=1 vs. *Mdn*=2 respectively; *T*=434, *p*=.001) and for Group 2 (*Mdn*=1.5 vs. *Mdn*=2.83 respectively; *T*=49.5, *p*=.025), it was decided that each of the critical cross-language /p/ pairs should be analyzed separately. Paired-samples Wilcoxon Signed Rank tests showed significant differences for Group 1 in the ratings for the critical cross-language pair with a VOT difference of 24ms from pre-test to post-test (*Mdn*= 1 vs. *Mdn*=1 respectively; *T*=1.5, *p*=.039), but not for the other two pairs (see *Table 7a*).

Group 2 (*Table 7b)* showed significant differences from pre-test to post-test for the ratings of the VOT pairing with a difference of 40ms (Mdn=1 vs. *Mdn*=2 respectively; *T*=55, *p*=.005), and with a difference of 56ms (Mdn=2 vs. *Mdn*=3.5, respectively; *T*=61, *p*=.012). Although this group did not significantly improve for other critical pair (24ms), it is worth noting that the difference almost reaches significance (Mdn=1 vs. *Mdn*=3.5 respectively; *T*=53.5, *p*=.066). It is possible that this difference may have reached significance if the training had been longer. No significant differences were found for the control group for any of the critical cross-language /p/ pairs (see *Table 8*).

Table 7a: RTD Critical pairs, Group 1

VOT difference	Median (pre)	Median (post)	*T*	*p*
24ms	1	1	15	.039
40ms	1	1	14	.080
56ms	1	2	35.5	.122

Table 7b: RDT Critical pairs, Group 2

VOT difference	Median (pre)	Median (post)	*T*	*p*
24ms	1	3.5	53.5	.066
40ms	1	2	55	.005
56ms	2	3.5	61	.012

Table 8: RDT Critical Pairs, Group 3

VOT difference	Median (pre)	Median (post)	*T*	*p*
24ms	1.5	1	15	.865
40ms	1.75	1	15	.672
56ms	1.5	2.25	21.5	.203

Research Question 3

Do learners' productions of English voiceless stops /ptk/ become significantly more target-like after training?

Originally this study was intended to assess the effect of training on both voiced and voiceless word-initial stops. However, due to the nature of the testing, the recordings elicited from the students contained a considerable amount of background noise. Unfortunately this meant that it was impossible to measure prevoicing accurately and reliably. Therefore, only voiceless stops were analyzed.

Before further analyses were carried out, mean VOT values for /p t k/ at pre-test for isolated words (Picture Naming task) and for words within sentences (Delayed Sentence Repetition) were submitted to a Kruskal-Wallis test to check for significant differences across groups. Group type was used as the between subjects factor and VOT as the dependent variable. The results revealed no significant main effect of Group Type on VOT of /p t k/ at pre-test for isolated words (H (2) = 2.137, p=.343), or for words in sentences (H (2) = 4.316, p=.116). As participants did not differ significantly in their pre-test VOT times across groups, it was assumed that all groups were starting from a very similar standard and so could be easily compared at post-test.

Group 1 (L2 training) were found to improve most in their productions of /p/ across groups, in both contexts. Group 2 (L1 accent imitation) made the most gains for /k/ in both contexts and for /t/ in Isolated Words. However, for Words in Sentences /t/ gains were almost equal across training groups. Group 2 was also found to gain most for the average of all consonants /ptk/. Group 3 (Control) did not change their VOT by more than 4ms for any of the consonants.

Graph 6: Gains, Isolated Words

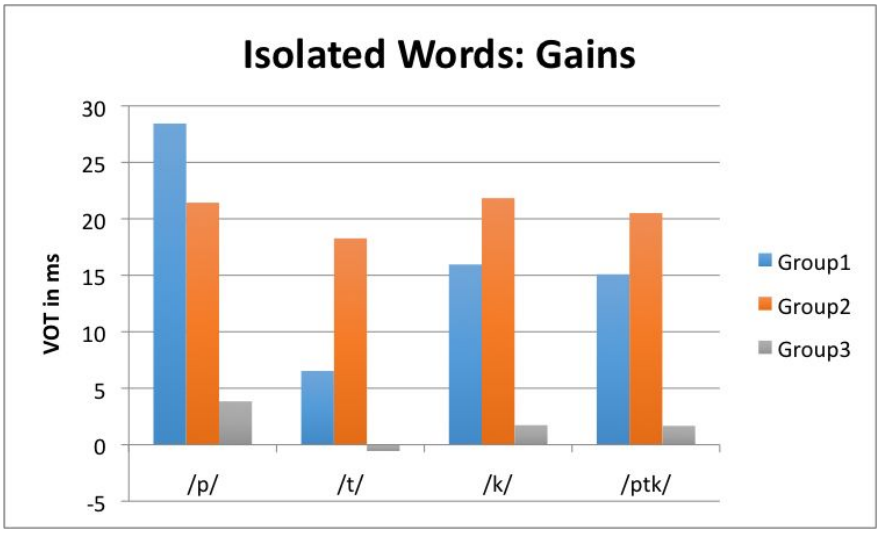


Table 9: Gains, Isolated Words

Mean gains (ms)	/p/	/t/	/k/	/ptk/
Group1	28,44	6,53	15,96	15,09
Group2	21,43	18,27	21,83	20,51
Group3	3,84	-0,56	1,73	1,67

Graph 7: Gains, Words in Sentences

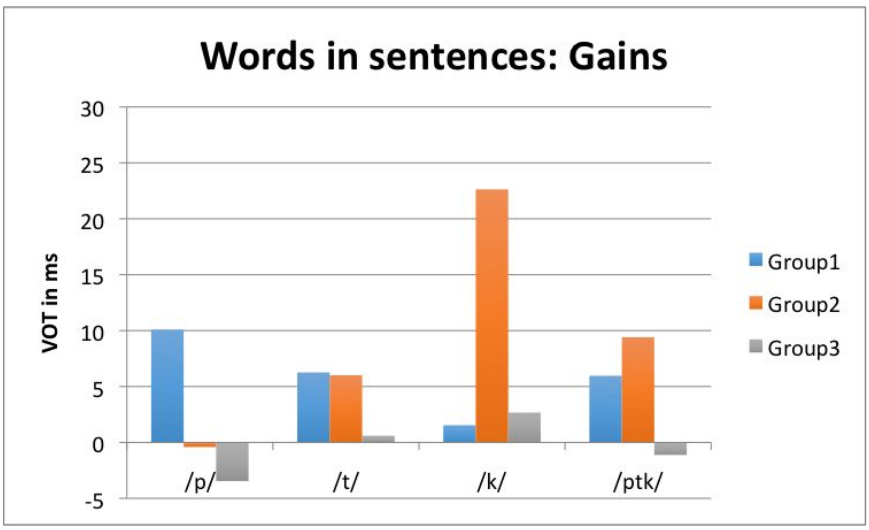


Table 10: Gains, Words in sentences

Mean gains (ms)	/p/	/t/	/k/	/ptk/
Group1	10,09	6,26	1,54	5,96
Group2	-0,41	6,01	22,63	9,41
Group3	-3,46	0,6	2,67	-1,12

Paired-samples Wilcoxon Signed Rank tests were carried out on the mean VOT values for all consonants /p t k/ of isolated words and of words in sentences for each group to compare pre- and post-test, assessing the effectiveness of the training in terms of significant increases in VOT.

Isolated words

The mean VOT values for each student of words beginning with /p t k/ elicited from the Picture Naming task were compared at pre- and post- for each group. Significant differences in VOT between pre- and post- test were found for both the L2 training group (*Mdn*=35.74 vs. *Mdn*=48.85 respectively; *T*=60, *p*=.016) and for the group doing accent imitation training on their L1 (*Mdn*=32.29 vs. *Mdn*=44.74 respectively; *T*=89, *p*=.002). The control group's median VOT of /p t k/ for isolated words actually decreased from pre- to post-test, but this change was not significant (*Mdn*=30.93 vs. *Mdn*=29.25, respectively; *T*=32, *p*=.646). This suggests that both training types were effective in increasing VOT towards a more target-like value.

As significant gains were found for both of training groups for all consonants, further analysis was carried out to assess gains of each consonant individually. Results can be seen in *Tables 11, 12 & 13*.

Table 11: VOT of /p/, Isolated Words

Isolated words: /p/	Median (pre)	Median (post)	*T*	*p*
Group1	23,32	40,5	49	0,28
Group2	18,74	33,57	87	0,004

Table 12: VOT of /t/, Isolated Words

Isolated words: /t/	Median (pre)	Median (post)	*T*	*p*
Group1	36,36	45,12	40	0,534
Group2	27,81	39,23	83	0,009

Table 13: VOT of /k/, Isolated Words

Isolated words: /k/	Median (pre)	Median (post)	*T*	*p*
Group1	42,94	58,56	57	0,033
Group2	53,86	67,94	87	0,004

As the above tables show, both training groups became significantly more target-like in terms of VOT values for the consonant /k/, whereas only the group who received accent imitation training on their L1 significantly improved their VOT productions of the consonants /p/ and /t/, although Group 1 are approaching significance for /t/.

Words in sentences

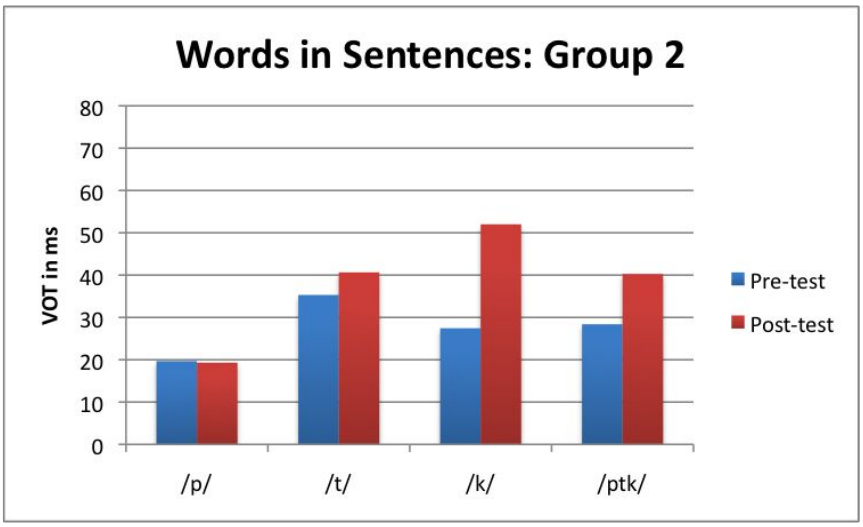
The mean VOTs for each participant of the consonants /p t k/ of the target words with the sentences elicited in the Delayed Sentence Repetition task were compared at pre- and post-test. No significant differences were found for the group who undertook L2 training (*Mdn*=36.20 vs. *Mdn*=42.13 respectively; *T*= 54, *p*=.062), but significant differences were found for the group who did accent imitation training on their L1 (*Mdn*=28.39 vs. *Mdn*=40.26 respectively; *T*= 88, *p*=.003). As above, the control group

did not change significantly between pre- and post- test (*Mdn*=32.17 vs. *Mdn*=33.8 respectively; *T*= 22, *p*=.575). As significant increases in VOT for words in sentences were only found for the accent imitation group, analysis of individual consonants was only carried out for this group, as shown below in *Table 14*, *Graph 8*:

Table 14: VOT /p/ /t/ /k/, Group 2

Words in sentences	Median (pre)	Median (post)	*T*	*p*
Group2 /p/	19,68	19,31	38	0,6
Group2 /t/	35,3	40,65	65	0,173
Group2 /k/	27,45	52,01	88	0,003

Graph 8: VOT /p/ /t/ /k/, Group 2



The above table and graph show that the previous significant increases in VOT found for all consonants for this group were in fact mainly due to increases in VOT for /k/, although /t/ did improve, but not significantly.

Research Question 4

Are there differential gains in VOT perception and production as a function of training group?

Gains were calculated for each group in terms of perception and production by subtracting pre-test from post-test scores. For perception, these gains were in terms of boundary location and categoricality, Identification critical values, and Rated Dissimilarity critical pair values. *Table 15* shows the number of participants who made gains in all of these areas.

Table 15: Perception Gains

	Raw numbers (improved/total)	Percentage
Group 1	2/11	18.2%
Group 2	4/13	30.8%
Group 3	0/10	0%

For production gains, pre-test VOT values were subtracted from post-test VOT values for /ptk/ in isolated words and words in sentences. *Table 16* shows the number of participants in each group who made gains in both contexts.

Table 16: Production Gains

	Raw numbers (improved/total)	Percentage
Group 1	7/11	63.6%
Group 2	11/13	84.6%
Group 3	2/10	20.0%

The above tables show that more participants improved from the training in the L1 accent imitation group (Group 2) than in the L2 training group (Group 1).

DISCUSSION

The main aim of this study was to assess the effectiveness of L1 accent imitation training on the L2 in the perception and production of VOT. The training aimed to make participants' perception and production of English word-initial voiced and voiceless plosives more target-like. Another training group was included which received L2 training to allow for comparisons between training types, and a learner group not receiving any accent training was included for control purposes.

As the control group did not significantly improve in any of the perception or production tests at post-test, it was assumed that no significant testing effect existed (from taking the same tests twice) and that any results found for the two training groups were due to the training itself.

Unfortunately only productions of voiceless word-initial stops could be analyzed due to high levels of background noise, as previously mentioned.

Research Question 1

Does learners' perception of English voiced and voiceless stops /b/ and /p/ become more target-like after training?

The results of the identification task did not show a significant improvement from pre- to post-test for any of the groups in terms of phonetic category boundary placement or categoricality calculated using logistic regression. This would suggest that the training did not have a significant effect on participants' perception of the English voiced and voiceless stops /b/ and /p/. It is likely the proximity of English /b/ and /p/ to their Spanish counterparts made differences difficult to perceive and therefore difficult to learn, as proposed by the Speech Learning Model (Flege 1995). However, other phonetic categories of word-initial plosives would have to be analyzed before suggesting that this type of training is not capable of making learners' L2 phonetic categories more target-like. It is also possible that the orthographic similarity between English and Spanish caused complications for participants to complete this task reliably. Although

explicit instructions were given that the syllables to be heard were in English, the options 'ba' and 'pa' on the screen could easily be read as Spanish syllables.

When the 'critical value' stimuli were analyzed, only Group 2 showed a significantly more target-like response at post-test at Step 9 (8ms VOT). This would be perceived as a /p/ in Spanish but a /b/ in English (Benki 2005), so this significant move from pre- to post-test (0.9 to 0.7), although far from being target-like, does suggest that the accent imitation training on the L1 can be effective. It is possible that these participants improved their ability to use 'proximal stimulus details' (Best & Tyler 2007, 26) to discern differences between the consonants of the two languages. However, more training sessions would be necessary to see if these perceptions could eventually lead to the creation of a new phonetic category for the L2 sound.

Research Question 2

Do learners perceive greater differences in VOT of the voiceless stop /p/ between languages after training?

The results from the Rated Dissimilarity Task showed no significant effect of time or overall interaction of time and group on the mean rating of cross-language pairs. This implies that training did not lead to learners perceiving greater differences in VOT of the voiceless stop /p/ between languages.

When the 'critical' cross-language pairs were analyzed, significant differences were found between pre- and post-test for both training groups (and no significant differences for the Control Group). When these pairs were analyzed individually, Group 1 (L2 training) were found only to significantly improve for pair 'Step8, Step10', which had a VOT difference of 24ms. However for Group 2 (L1 accent imitation) the significant differences were found for 'Step8, Step11' (40ms difference) and 'Step8, Step12' (56ms). It is interesting that the two groups appear to have received different benefits from their trainings.

These results show that both types of training were effective in improving participants' perception of cross-language differences in VOT for the voiceless stop /p/, even if the improvements were small.

It is important to remember when interpreting the perception results that the training undertaken by both groups was a production-based training. It is therefore not very surprising that not all perception tests showed significant improvements. However, the significant results that were found are important as they show that participants were able to generalize some of the knowledge gained from the training to a novel context, which suggests 'robust learning' has taken place (Logan & Pruitt 1995: 371).

Research Question 3

Do learners' productions of English voiceless stops /ptk/ become more target-like after training?

For isolated words, VOT of all consonants /ptk/ was found to become significantly more target-like at post-test for the two groups who received training, suggesting that both training types were effective. However, the benefits of training found for isolated words only generalized to words in sentences for the group who received accent imitation training on their L1. In both cases the control group did not significantly improve. The lack of gains found for words in sentences may not be surprising. VOT has been found to be shorter for words in sentences than for isolated words (Lisker & Abramson 1967), which may have hindered significant increases on the part of the participants. Although VOT is usually greater for voiceless word-initial stops of stressed words in sentences (Lisker & Abramson 1967), it is possible that the low proficiency of these learners prevented target-like intonation, rendering this effect null.

When consonants were analyzed individually for the groups who showed significant improvements as a whole, the results showed that Group 1's (L2 training) productions of isolated words in fact only significantly improved for /k/. However, Group 2 (L1 accent imitation) improved significantly for each consonant in this context. For words in sentences, the overall significant results previously found for Group 2 were also only significant for /k/. It is interesting that /k/ appears to have been affected more

than the other consonants. This could be due the fact that /k/ has a longer VOT than /p/ or /t/ (Yavaş & Wildermuth 2006), and may therefore be more noticeable to learners, or its aspiration may be easier to exaggerate.

As group 2 showed significant production gains in both production contexts, it is possible that accent imitation on the L1 is a more effective method than L2 training at improving VOT production of L2 word-initial voiceless stops. However, it is also possible that a greater length of training may have led to significant gains for group 1 too.

Research Question 4

Are there differential gains in VOT perception and production as a function of training group?

The results show that a greater number improved among those who did Accent Imitation training on their L1 (Group 2) compared with the L2 training group (Group 1). A possible explanation for this may be proficiency. Low levels of English proficiency may have prevented learners in Group 1 from benefitting significantly from the training. For those students who partook in Accent Imitation training in their L1, L2 proficiency did not present an obstacle, and therefore phonetic training may have been of more help. Larger group sizes in future research would allow for further investigation of correlations between gains and proficiency.

The results found in this study both for perception and production imply that L1 accent imitation is more effective than accent training in the L2. However, one may have expected the L2 training to do better than their counterparts, as they are able to lexically encode the phonetic information they are receiving. As this was not the case then either the benefits of lexical encoding were not significant, or the benefits L1 accent imitation training overcame this. Future research would have to include tests of lexical encoding to investigate this relationship in greater depth.

Student Evaluation

At the end of the post-test, the students in both training groups were given an evaluation questionnaire. Of 33 students, 31 said they would like this training to be incorporated into their normal classes, and only 1 said he did not find the course useful (as he does not speak English outside of school). 28 students said they thought their English accent had improved.

Many of the students who were imitating their L1 accent whilst speaking English said they found the training funny and interesting. Among both training groups, word imitation was often found very repetitive and boring whereas the dialogues appeared to be much more popular. These responses suggest that further consideration needs to be given to designing tasks which engage the students.

Limitations

As with any research, this study had various limitations. For the identification task many repetitions of the stimuli were needed in order to calculate a reliable mean result. However, this task can become very monotonous, particularly for adolescent learners, causing participants to lose concentration and potentially to pick a 'random' answer rather than focusing on their decision. This may have decreased the reliability of results. Contrary to the Identification task, the Rated Dissimilarity task may have used too few repetitions: 30 stimulus pairs were presented only twice each. Although the participants seemed to find this test much less boring and seemed better able to concentrate, calculating a mean from two data points may not be very reliable. Clearly a balance needs to be struck between holding young participants' attention and ensuring reliability of data.

Although the perception tasks give an insight into category boundary placement, categoricality and cross-language differences, results only apply to the /b/-/p/ continuum and may not be transferable to other phonetic category boundaries. Future research in this area would have to include continua representing a wider range of phonetic categories to make more generalizable conclusions.

Regarding the production task stimuli, increasing variability further by using a greater number of stimuli along with more repetitions may have lead to more reliable

results. The DSR task did not include two instances of the consonants /d/ or /k/, which may have had an effect on the means and overall results. Although all target-words for testing were selected to be relatively simple, using a frequency list compatible with X_lex scores would have increased the probability of all students' familiarity, which may have provided more reliable data.

Furthermore, baseline data could have been provided by native speakers of English, to provide a more accurate definition of 'target-like' that was comparable to that of learners in this testing context.

For the training stimuli, unfortunately, two /b/ words for the fourth session in Spanish had to be replaced with words from the first session, due to data loss. This meant that the group did not receive the amount of variation they should have done, which may also have had an impact on the results, albeit small.

As well as limitations pertaining to the methodology itself, many limitations also arose due to the context of research. The classroom is not a controlled environment for research: many participants were doing the tests simultaneously and a considerable amount of background noise was present. As previously mentioned, this prevented the analysis of voiced plosive data, which would have been an extremely interesting dimension to assess. Future research in this area should aim to create the conditions necessary for this type of data to be collected and analyzed in a classroom context.

The age of the participants also caused some complications. Adolescent learners often have trouble concentrating for prolonged periods of time, especially if the nature of the task is fairly monotonous. This caused some participants to become distracted, during both testing and training. Also, some participants missed a session due to illness or extra-curricular activities, which was beyond the control of the researcher. Unfortunately this meant some participants had to be removed from the analysis, leading to smaller group sizes than originally intended. Due to prior or unforeseen commitments on the part of the school, the original schedule was not entirely possible, meaning not all sessions were exactly one week after the other, and less time was available for data analysis. Had more time been available, more training sessions could have been conducted, which may have led to different results. Furthermore, a delayed post-test

would have provided an interesting insight into how many benefits were retained by the participants long-term (Thomson & Derwing 2014).

CONCLUSION

This study investigated the effectiveness of accent imitation training in the L1 on the perception and production of L2 English voiced (/b/) and voiceless (/ptk/) word-initial stops. In English, voiced stops preceding a vowel are usually produced without voicing (e.g. 0ms VOT), and aspiration is used to create voiceless stops. However in Spanish, this differentiation is inversed: voiceless stops are produced without voicing (0ms VOT) and voiced stops use prevoicing (negative VOT). The aim of this research was to examine the effects of accent imitation on the L1 on the phonetic categories of participants, measured in terms of perception and production. Results were compared with a group who undertook accent training in their L2 as well as a control group, to ensure that gains found from the training could be contributed to this specific training methodology.

Learners' phonetic categories of English voiced and voiceless stops /b/ and /p/ were not found to become significantly more target-like in terms of boundary placement or steepness for any of the groups. This may have been too ambitious considering the number of training sessions and the production-focused method. However, the group whose training consisted of accent imitation on the L1 were found to become more target-like in their perception of English /ba/ with a VOT of 8ms, suggesting that perception may have started to become more target-like, although more sessions would have been needed to test this further.

Much more promising results were found for production. Both training groups became significantly more target-like in their production of English voiceless stops in isolated words. For the group who received accent training in the L2 this was only significant for /k/, but the L1 accent imitation group became significantly more target-like in their production of all voiceless consonants assessed. This latter group also improved significantly in their production of English voiceless stops of words in

sentences for /k/, suggesting robust learning that can be applied to more fluent contexts, where VOT is usually shorter.

In both perception and production, participants who undertook accent imitation training in their L1 benefitted more than those who received L2 training. The control group did not significantly improve on any of the tests.

As both training methodologies were effective in some aspects, their pedagogical implications should be seriously considered. It is possible that a combination of the two methods might be easiest to implement into a classroom, as teachers are likely to resist allowing students to speak in their L1 for any considerable length of time, although students would be training their L2 production through L2 articulatory targets. The positive feedback given by the students both in terms of utility and enjoyment reasserts the above statements.

Although conducting this research in a classroom context presented a large number of limitations, significant results were found, even in the short time available. Accent imitation training on the L1 is a novel concept but one that merits further study, in both classroom and laboratory settings. It is possible that even more significant gains may be found amongst adults in a more controlled laboratory setting. Future research in this area would benefit from an increased number of training sessions, whilst looking at a greater variety of L2 speech aspects on a wider population. It would also be interesting for a perceptual aspect to be added to the training, to see what effect this has on both perception and production.

Acknowledgements

I would first like to thank my supervisor Dr. Joan Carles Mora for his invaluable support and expertise. This thesis would not have been possible without him.

I would also like to express my enormous gratitude towards the school, teachers and students who participated in this study. A training study greatly disrupts the academic schedule, and I am extremely grateful for the cooperation and effort made on the school's behalf.

Finally, I would like to thank my friends and family for the support they have given me during this process. Particularly my mother, whose continuous encouragement and guidance made a seemingly impossible task, possible.

REFERENCES

Aliaga-García, C. (2007). The role of phonetic training in L2 speech learning. In *Phonetics Teaching & Learning Conference (PTLC)*.

Benkí, J.R. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In J. Cohen, K.T. McAlister, K. Rolstad, J. MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism*, (pp. 240-248) Cascadilla Press, Somerville.

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.

Boersma, Paul (2001). Praat, a system for doing phonetics by computer. *Glot International* 5:9/10, 341-345.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (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.

Carlet, A. & Cebrian, J. (2013). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *Proceedings of the 7th International Symposium on the Acquisition of Second Language Speech. Concordia University Working Papers in Applied Linguistics*, 5, 85-98.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America*, *62*(4), 961-970.

Cebrian, J. Mora, J. C., & Aliaga-Garcia, C. (2011) Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In Dziubalska-Kołaczyk, K., Wrembel, M., Kul, M. (eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010, Volume I.* Frankfurt am Main: Peter Lang. 41-52.

Collet, G., Colin, C., Serniclaes, W., Hoonhorst, I., Markessis, E., Deltenre, P., & Leybaert, J. (2015). Changes in voicing perception by adult French speakers after identification training. *Applied psycholinguistics*, *36*(02), 463-483.

Docherty, G. J. (1992). *The timing of voicing in English obstruents*. Berlin: Foris.

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. 229–273). 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*(4), 437-470.

Flege, J.E., & Hammond, R. (1982). Non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition, 5*, 1–17.

Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition*, *20*(02), 155-187.

Flege, J. E., & Munro, M. J. (1994). The word unit in second language speech production and perception. *Studies in Second Language Acquisition*, *16*(04), 381-411.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, & Computers, 35,* 116-124.

Hilton, L. M. (2005). Reverse Accent Mimicry: An Accent Reduction Technique for Second Language Learners.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive psychology*, *21*(1), 60-99.

Lisker, L. and Abramson, A.S. (1964) A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422.

Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and speech*, *10*(1), 1-28.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*(3), 1242-1255.

Logan, J.S. & Pruitt, J.S. (1995). Methodological issues in training learners to perceive non-native phonemes. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 351-378). Timonium: MD: York Press.

Mora, J. C., Rochdi, Y., & Kivistö-de Souza, H. (2014). Mimicking accented speech as L2 phonological awareness. *Language Awareness*, 23(1-2), 57-75.

Neuhauser, S. (2011). Foreign accent imitation and variation of VOT and voicing in plosives. In *International Congress of Phonetic Sciences, XVII* (pp. 1462-1465).

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, *41*(6), 407-420.

Owren, M. J. (2008). GSU Praat Tools: Scripts for modifying and analyzing sounds using Praat acoustics software. *Behavior Research Methods*, *40*(3), 822-829.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, *8*(2), 297.

Simon, E. (2009). Child L2 development: A longitudinal case study on Voice Onset Times in word-initial stops. *Journal of child language*, *37*(01), 159-173.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /rl/ by Japanese adults learning English. *Perception & Psychophysics*, *36*(2), 131-145.

Yavaş, M. & Wildermuth, R. (2006). The effects of place of articulation and the vowel height in the acquisition of English long lag stops by Spanish speakers. International Review of Applied Linguistics, 44(3), 251-263.

APPENDICES

Appendix 1: Training, Group 1

Session1 (English)

Words		**Sentences**
panel	paw	Tonight, we will have pizza
tear	tartan	She likes to talk.
keel	cabin	It is cold in winter
beast	barge	I am scared of bees.
debt	dab	I go to dance class on Saturdays.
guilt	gallop	Shall we play a game?

DIALOGUE

It's boiling today, what shall we do?
Let's go to the beach!
Great idea, what time?
Let's say two?
Ok, I'll bring a ball to play with!
Good idea, so we don't get bored.
Perfect. Come to mine first?
Sure, I'll pack a bag and head over.
Cool!

CONVERSATION TOPIC

Talk about your daily routine

Session 2 (English)

Words		**Sentences**
Pain	peach	Some parrots can talk
teach	Tarmac	I am feeling very tired.
Cop	Candle	That song is very catchy.
bold	banker	Do you like the bag I bought?
dean	dashing	It is very dark in here.
geese	garment	He looks very guilty.

DIALOGUE

We should get a cab to the station
Yes, you're right. Can you call for one?
Ok, do you have cash?
Yes, there's some in the kitchen.
I'll get it.
Please grab my tie too!
Got it.
Let's go! I don't want to be late!
Calm down, we have lots of time.

CONVERSATION TOPIC

What did you do this weekend?

Session 3 (English)

Words		**Sentences**
Peeling	painter	He is allergic to peanuts.
teepee	tender	I have a very nice teacher.
keen	cuddle	Keep going, you can do it!
beaming	bap	She is behaving badly.
defect	daft	I dare you to do it.
geezer	gawk	Be nice, we have a guest!

DIALOGUE

Would you like to come to my party?
Yes please, when is it?
It's the day after tomorrow
At what time?
At ten o clock.
What will you wear?
I will wear a pink dress/shirt.
Should I bake a cake?
No, don't worry, I have one!
Ok, well I'll bring a gift.
That's very kind.
See you then!
Bye!

CONVERSATION TOPIC

What do you do in your free time?

Session 4 (English)

Words		**Sentences**
pickle	pardon	Be careful, the pan is hot!
tar	teapot	I am a teenager.
keeper	cap	I saw them kissing.
busted	beacon	It's the second building on the left
demon	dart	Spiders can be deadly
gang	geeky	She is a pretty girl

DIALOGUE

Do you have any pets?
Yes I have a dog!
When did you get him?
When he was just a puppy.
How cute!
Yes he was very cute. Now he barks all the time.
oh dear.
Do you have any animals?
No, but I'm begging my dad to get a kitten.
Oh I'm very jealous! But we can't have one because of my dog.
That makes sense.
I guess.

CONVERSATION TOPIC

What will you do this summer?

Appendix 2: Training, Group 2

Session1 (Spanish)

Words		**Sentences**
pico	palo	Tiene un piso muy grande
tasa	tira	Ahora te toca a ti
cabo	kilo	Quedé el quinto en el concurso
barra	beso	Voy al colegio en bici
digno	dama	Grabamos un disco
gamba	guiño	Me gustan tus gafas

DIALOGUE

¡Pon la mesa por favor!
Ahora la pongo
Utilizamos las servilletas verdes
Me faltan dos tazas
Estoy lavándolas
Vale… ¿qué comemos?
Tacos de pollo y queso
¡Qué rico!
Llama a tu padre, está lista la cena

CONVERSATION TOPIC

Habla de tu rutina diaria

Session 2 (Spanish)

Words		**Sentences**
pozo	pavo	Estoy a punto de llegar
tio	tuyo	Espera tu turno, por favor
cobra	codo	La cama está muy cómoda
bajo	baza	Después de cenar, fuimos a un bar
dé	dato	¿Te ha dado los billetes ya?
guita	gozo	Tenemos una amplia gama de productos

DIALOGUE

¿Tienes animales en casa?
Si, tengo un perro. ¿Y tú?
Tengo dos peces, pero me gustaría tener un gato
Yo también, me encantan
¿Cómo se llama tu perro?
Se llama Carlos. ¿Y tus peces?
Se llaman Quique y Paolo
¿Cuales tipos de peces son?
Son peces dorados las dos

CONVERSATION TOPIC

¿Qué hiciste este fin de semana?

Session3 (Spanish)

Words		**Sentences**
pista	paz	Los cactus tienen muchos pinchos.
todo	torta	Se oyó un tiro y después alguien gritó.
coma	calle	Coge el tren que sale a las once
bici	base	Tiene una barba muy grande
dial	danza	Las llaves estan dentro del cajón
gallo	goma	Gasto demasiado en ropa.

DIALOGUE

¿Qué haces hoy?
No mucho, tengo deberes
¿Quieres ir al parque?
No sé, ya es tarde
Pero hace calor todavia
Es verdad
¿Traigo vino y pica pica?
Tengo que cenar con mi padre.
Vale, quedamos otro día entonces.
Por supuesto.

CONVERSATION TOPIC

¿Qué haces en tu tiempo libre?

Session 4 (Spanish)

Words		**Sentences**
pata	pala	Voy al cine un par de veces a la semana
toro	tabla	Prefiero el té negro al té verde.
kiwi	caña	Toma la segunda calle a la derecha
beso	barra	Estoy de baja por enfermedad
diva	duro	La ducha está rota
gano	gafas	No comía con hambre, comía con gula

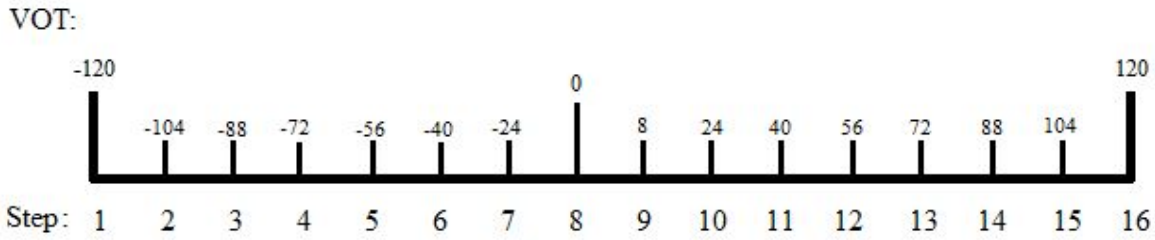
DIALOGUE

¿Has estado en Vigo?
Si una vez, pero hace mucho tiempo
¿Y cómo fue?
Sólo tuve dos días, pero me encantó
¿Qué hiciste?
Tomé un guía privada de la cuidad, en coche y en barco
¡Qué chulo! ¿Por qué fuiste?
Tengo un amigo que vive allí
¿Tienes ganas de volver?
Si claro, es una ciudad bella

CONVERSATION TOPIC

¿Qué harás este verano?

Appendix 3: Testing



Identification: Steps 3 through 14

Rated Dissimilarity

Type	Pairing	VOT difference (ms)
Cross-language	Step08,Step10	32
	Step08,Step10	32
	Step08,Step11	48
	Step08,Step12	64
	Step08,Step13	80
	Step08,Step13	80
	Step08,Step14	96
	Step08,Step15	112
	Step08,Step16	128
	Step08,Step16	128
	Step03,Step13	160
	Step04,Step14	160
	Step01,Step15	224
	Step02,Step16	224
Within-language	Step01,Step02	16
	Step07,Step08	16
	Step07,Step08	16
	Step08,Step09	16
	Step08,Step09	16
	Step15,Step16	16
	Step06,Step08	32
	Step06,Step08	32
	Step08,Step06	32
	Step04,Step08	64
	Step03,Step08	80
	Step03,Step08	80
	Step05,Step08	80
	Step02,Step08	96
	Step01,Step08	112
	Step01,Step08	112

Picture-Naming

/b/	baby	boy
/d/	doctor	dog
/g/	girl	goat
/p/	pen	peanut
/t/	table	teeth
/k/	cat	key
distractors	frog	glass
	hand	moon

Delayed Sentence Repetition

Target consonant	**Question**	**Answer**
/b/	Where did you go this weekend? What is that?	I went to the beach It is a beehive
/d/	Who is that man?	That is my dad
/g/	Do you see the ducks? What did you say?	No, those are geese Mind the gap
/p/	What is that fruit? Can I take your order?	It is a peach I'll have the pancakes
/t/	Would you like a coffee? Did you enjoy your class?	No, I'll have tea Yes, I liked the teacher
/k/	What shall we do?	Let's go camping!

Appendix 4: Questionnaire

1 **What is your name?**

2 **When is your birthday? e.g. 5 February 1991**

3 **Which group are you in?** *Mark only one.*

Tuesday 9am-10am

Tuesday 10am-11am

Wednesday 1.30pm-2.30pm

4 **In what language did you learn to speak?** *Mark only one.*

Spanish

Catalan

Other:

5 **What languages do you speak and how well?**

1= I do not speak this language, 10 = I speak like a native

	1	2	3	4	5	6	7	8	9	10
Spanish										
Catalan										
French										
English										
Other_____										
Other_____										
Other_____										

6 **How old were you when you started to learn English?**

7 **Have you ever been to an English speaking country?**

Yes. If yes, when and how long for? _____

No

8 **How often do you watch television or films in English?**

More than once a week

More than once a month

Every couple of months

Once a year

Never

9 **Do you spend any time outside of school speaking English?**

Yes. Where and how often? _____

No

10 **Do you take English classes outside of school?**

Yes. How often? _____

No

11 **Did you enjoy this accent-training course?** Please say why.

Yes because

_____

No because

_____

12 **What did you like most about this course?**

_____

13 **What did you like least about this course?**

_____

14 **Do you think the course was useful?** Please say why.

Yes because

_____

No because

_____

15 **Do you think your English accent improved?** Please say why.

Yes because

_____

No because

_____

16 **Would you like this kind of training to be part of your normal classes?**

Yes

No

Thank you!