

EFFECTS OF PERCEPTUAL TRAINING ON THE IDENTIFICATION AND PRODUCTION OF WORD-INITIAL VOICELESS STOPS BY ARGENTINEAN LEARNERS OF ENGLISH

Ubiratã Kickhöfel Alves*

Universidade Federal do Rio Grande do Sul
Rio Grande do Sul, BR

Pedro Luis Luchini**

Universidad Nacional de Mar del Plata
Buenos Aires, ARG

Abstract

In this study, we investigate the effectiveness of perceptual training, administered to Argentinean learners, in the perception and production of word-initial voiceless stops in English. 24 participants were divided into 3 groups: (i) Group 1, which participated in 3 training sessions; (ii) Group 2, which, besides performing the same training tasks, was explicitly informed about the target item; (iii) Group 3 (control). All participants took part in a pre-test, a post-test and a delayed post-test. In all these tests, they participated in a consonant identification task and took part in a read-aloud task. Our results show a significant increase of both experimental groups in identification. As for production, Group 2 exhibited a significant increase in /p/ and /t/ after training. These results are indicative of the effectiveness of perceptual training tasks in helping learners focus on Voice Onset Time.

Keywords: Perceptual training; Awareness raising; Acoustic cues; Voice Onset Time; English as a Foreign Language.

Introduction

Many studies focusing on the role of L2 pronunciation teaching have been proposed in the last few years. In these studies, teaching practices such as explicit instruction (Alves, 2004; Silveira, 2004; Lima Júnior, 2010; Alves & Magro, 2011; Kissling, 2013; Echelberger, 2013; Perozzo, 2013; Sangüesa, 2016) and perceptual training (Nobre-Oliveira, 2007; Bettoni-Techio, 2008; Reis & Nobre-Oliveira, 2008; Aliaga-Garcia, 2010; Brawerman-Albini, 2012; Wong, 2012; Rato, 2013; Carlet, 2017) have been investigated, in order to verify the effectiveness of these practices in the acquisition of a second language sound system.

In order to evaluate the role of these practices, factors such as the first language, the target language, the learners' proficiency level and the phonetic aspect under investigation, among many others, should be considered.

Bearing this in mind, in this study we investigate the role of perceptual training in the acquisition of aspirated initial stops by Argentinean learners of English. English has a two-way voice distinction for stops in word-initial position. Voice Onset Time (VOT) is the main acoustic cue employed by speakers of English when distinguishing /p, t, k/ from /b, d, g/. This distinction is clear as, in word-initial position in English, voiced plosives are generally produced with

* Full time Professor at the Graduate Program in Linguistics at Universidade Federal do Rio Grande do Sul (Brazil). Researcher at Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-Brazil). His e-mail address is ukalves@gmail.com.

** Full time Adjunct Professor, English Teacher Training Program, Humanities Division, Universidad Nacional de Mar del Plata (Argentina). Head of Language Studies at Universidad CAECE Mar del Plata (English/Spanish Translation Program). His e-mail address is luchinipedroluis@gmail.com.



short (or zero) VOT, whereas voiceless /p, t, k/ exhibit voicing lag or positive VOT (aspiration). These patterns, however, are not the same ones found in Argentinean Spanish. Even though Spanish also exhibits a two-way distinction for voicing, the VOT patterns through which this distinction is instantiated are different from those found in English, as aspirated plosives are not found in this language. Indeed, according to the literature on Argentinean Spanish (Lisker & Abramson, 1964; Abramson & Lisker, 1973; RAE, 2011), voiced plosives exhibit pre-voicing (or negative VOT), whereas voiceless plosives would be characterized by a short lag or zero VOT.

Given the characterization above, as we consider Argentinean learners of English, the acquisition of the two-voice distinction in English would imply a modification in the VOT patterns found in these learners' L1 (Yavas & Wildermuth, 2006; Alves & Luchini, 2016; Tobin *et al.*, 2017), leading these learners to produce aspirated voiceless initial stops.¹ However, recent studies carried out by our research group, with both Brazilian (Alves & Motta, 2014; Alves & Zimmer, 2015; Schwartzhaupt *et al.*, 2015) and Argentinean learners (Alves & Luchini, 2016) of English, have suggested that acquiring word-initial voiceless stops is an even more complex process. We have shown that, unlike native speakers of English, who follow VOT as their main cue in the distinction between voiceless and voiced stops in word-initial position, VOT does not seem to be the sole cue Argentinean and Brazilian learners attend to in voicing distinctions.

Therefore, it might be the case that, despite its recognized importance, the acoustic cue of negative VOT might not be the only phonetic aspect which accounts for voice distinctions in Argentinean Spanish, as it is possible that other acoustic cues are being primarily employed in the perception and production of voice distinctions. Similar cases have been found in Canadian French (Sundara, 2005), Korean (Oh, 2011) and Japanese (Kong *et al.*, 2012). In these languages, additional cues, such as burst intensity and F0 in the following vowel, take the lead as the main acoustic correlates employed by speakers in order to distinguish plosive segments in perception

and production. VOT, in these language systems, plays the role of an additional cue, which cannot function by itself in distinguishing the voicing of consonants, unlike what occurs in English.

The data presented in Alves & Luchini (2016) confirm the claim above. In this study, the perception of three different VOT patterns was investigated, among intermediate and advanced Argentinean learners of English: negative VOT (found variably in English /b/, /d/, /g/, cf. Lisker & Abramson, 1964; Simon, 2010), positive VOT (found in English /p/, /t/, /k/, cf. Lisker & Abramson, 1964; Cho & Ladefoged, 1999; Simon, 2010) and zero VOT, which may be found variably in English /b, d, g/ (cf. Lisker & Abramson, 1964; Simon, 2010) and categorically in Spanish /p, t, k/ (cf. Lisker & Abramson 1964; Abramson & Lisker 1973; RAE 2011). We also included a manipulated pattern, which was built as we took tokens of aspirated /p, t, k/ and removed their long-lag VOT completely, so that these new stimuli presented the VOT pattern of a voiced consonant in English, but at the same time preserved all of the acoustic cues (such as burst intensity and F0 frequency) that are found in voiceless stops in this language. Results from Alves & Luchini (2016) demonstrated that learners showed ceiling effects in the identification of negative and positive VOT patterns. However, even though natural zero VOT was already identified as voiced, consonants with artificial zero VOT were still identified as voiceless, suggesting that learners attended to something else besides VOT, in the identification of the L2 voicing patterns. It is also relevant to mention that, in a previous study (Schwartzhaupt *et al.*, 2015), the same identification test had been applied to monolingual speakers of English, who showed high rates in the identification of both zero VOT patterns (natural or manipulated) as voiceless.

The results above might have direct implications in the fields of second language acquisition and teaching. With regard to L1 systems in which positive VOT might not be taken as the main cue in voicing distinctions, such as Argentinean Spanish (Alves & Luchini, 2016) and also Brazilian Portuguese (Alves & Motta, 2014; Alves & Zimmer, 2015; Schwartzhaupt *et al.*, 2015), the acquisition of the two-way voicing system of L2 English

will imply that, firstly, learners focus their attention on positive VOT, so as to learn the new pattern which occurs in English (aspiration). The acquisition of English aspiration by learners of these L1 systems, therefore, would imply a double task: before learning how to produce the L2 VOT pattern itself, students have to learn how to “listen to” this cue, which does not play such an important role in their first language.

The importance of this new “tuning in” is quite clear when we consider the consequences of this lack of focus on positive VOT not only in perception, but also in production, especially if we assume a perceptual model such as the Speech Learning Model (Flege, 1995), which connects the processes of sound perception and production. If L2 learners of English do not focus on positive VOT, but rather attend to those other sources of information that are present in the acoustic signal, they are very likely not to have perception problems regarding the identification and discrimination of English initial /p/, /t/, /k/ and /b/, /d/, /g/; indeed, these other acoustic cues which are being primarily considered may lead them to a correct identification either way (voiceless consonants /p/, /t/, /k/, for example, present higher burst intensity and F0 values than /b/, /d/ /g/ in English, as well as in those languages in which VOT is not the main cue). The fact that the two-voicing distinction in English may be perceived appropriately, regardless of the acoustic cue which is being focused on, might at first allow us to conclude that it would not be necessary for learners to focus on positive VOT. However, should we consider the possibility that positive VOT is not considered in perception, there is a strong possibility that learners are not going to make use of this cue in production and, consequently, will not find it necessary to aspirate voiceless plosives in English, as the voicing distinction might be maintained through other cues. This non-aspiration in learners’ production might have consequences in intelligibility (cf. Schwartzhaupt, 2015), given the fact that speakers of English follow positive VOT (aspiration) to distinguish voiceless from voiced plosives, as our studies have suggested (Schwartzhaupt *et al.*, 2015). It is therefore necessary to lead learners to focus on positive VOT, as the

intelligibility of their oral productions might be affected if they do not.

Perceptual training tasks have been an important aid in the teaching of second language sounds, and current research has shown its positive effects in both perception and production (Nobre-Oliveira, 2007; Reis & Nobre-Oliveira, 2008; Aliaga-Garcia, 2010; Rato, 2013; Carlet, 2017). When planning training sessions, both researchers and teachers must consider not only the target language, but also the learners’ first language system. We therefore enquire if, in the case of learners whose L1 systems tend not to attend to VOT as their main acoustic cue, perceptual training and feedback on aspiration might be effective. Since, in this study, perceptual training has the role of exposing learners to a cue that tends to be unattended, it is also important to investigate the effect of associating awareness raising through explicit instruction (cf. N. Ellis, 2005; Andringa & Rebuschat, 2015) to perceptual training. Therefore, in the present study, we investigate whether informing students about the target item they should focus on might make training more effective. Following Guion & Pederson (2007) and Pederson & Guion-Anderson (2010), we also investigate whether learners who are explicitly told to direct their attention to VOT present better results in their perception and production.

Starting from these assumptions, in this study we focus on the role of high variability perceptual training² (with or without explicit awareness raising) on the perception and production of aspiration by learners from the city of Mar del Plata (state of Buenos Aires), Argentina.³ Twenty-four participants were divided into three groups: (i) an experimental group, which took part in 3 training sessions (40 min. each); (ii) another experimental group, which, besides participating in the three training sessions, was informed about the L2 aspect to be focused on; (iii) a control group. The stimuli in the training sessions consisted of data produced by six different speakers of American English, and included two of the four VOT patterns whose identification had been previously studied in Alves & Luchini (2016): positive VOT (voiceless stops in English) and artificial/manipulated zero VOT (aspirated plosives whose

VOT had been cut off). With this hybrid pattern, we aimed to train learners on identifying these consonants as voiced, by concentrating on VOT as their main acoustic cue. All participants sat for (i) a pre-test; (ii) a post-test (three days after the last training session); and (iii) a delayed post-test (one month later), in which identification and production tasks were administered. With this methodology, we were able to investigate the generalization effects of perceptual training to production, as well as the possible long-term effects of this laboratorial practice.

Following the perceptual training studies carried out by Nobre-Oliveira (2007), Reis and Nobre-Oliveira (2008) and Rato (2013), we hypothesize that (i) High Variability Perceptual Training (with or without explicit awareness raising) promotes higher levels of Identification of natural zero VOT and artificial zero VOT after training, helping learners tune in to positive VOT as the main acoustic cue in voicing distinctions in English⁴; (ii) High Variability Perceptual Training promotes generalization to production (especially in Group 2, whose members had their attention directed to positive VOT), leading to higher VOT values in the production of /p/, /t/ and /k/ after training; (iii) The positive effects of perceptual training in perception and production remain one month after the end of the training sessions, indicating its long-term effects.

Method

Participants

Twenty-four students took part in the study, 17 women and 7 men. Participants were randomly divided into three groups of 8 students. Group 1 participated in the training sessions but was not told about the phonetic aspect to focus on. Group 2 participants, besides taking part in the training sessions, were asked to focus on aspiration and were taught that initial voiceless stops in English are aspirated (these instructions were repeated in the beginning of each one of the three training sessions). Group 3 served as control.

Participants were all taking their last high school year, and at the time of the investigation were attending

5 hours of English classes per week. They were taking a preparation course for the TOEFL exam. Before taking part in the experiment, all participants took the Oxford Online Placement Test,⁵ which indicated that all of them presented a C1 or a C2⁶ level of proficiency in English, according to the Common European Framework.

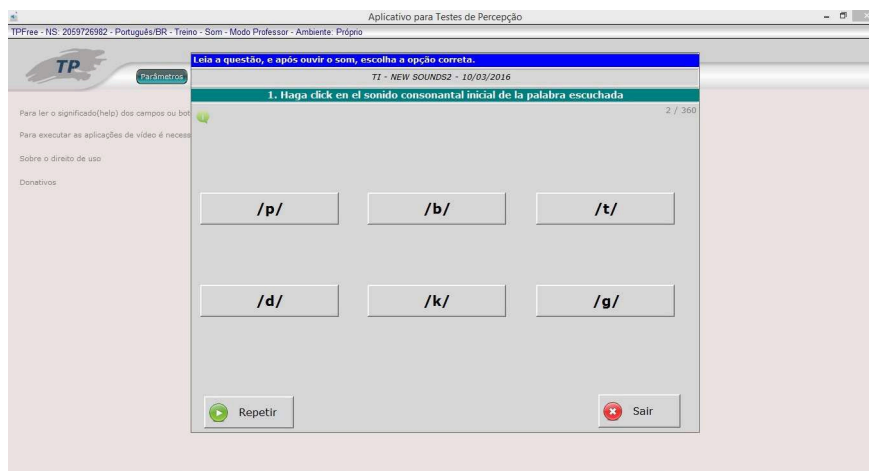
Perceptual training sessions

The training sessions consisted of the administration of an identification task with immediate feedback, built and administered on TP Software (Rauber *et al.*, 2013), and repeated in each session. The stimuli had been produced by six different native speakers of American English (3 men and 3 women).⁷

The task presented 18 audio files. The lexical items used in the training sessions were 'pee', 'tip' and 'kit'.⁸ Following Yavas and Wildermuth (2006) and Schwartzhaupt (2012), we used stimuli followed by a high vowel, since this environment fosters higher levels of aspiration and its perception. There were six different audio files for each one these lexical items, one of which produced by a different speaker. From these 6 stimuli, 3 of them had their aspiration cut off, so that we could build the artificial zero VOT pattern (a hybrid consonant, as already described). Each one of these 18 stimuli (9 with zero VOT and 9 with positive VOT) was repeated 20 times in a random order, which led to 360 tokens heard in each session. Pauses were allowed after 90 tokens each.

In the training sessions, which consisted of an Identification task, learners had to choose the initial consonant of the word they had just heard, as seen in Figure 1.

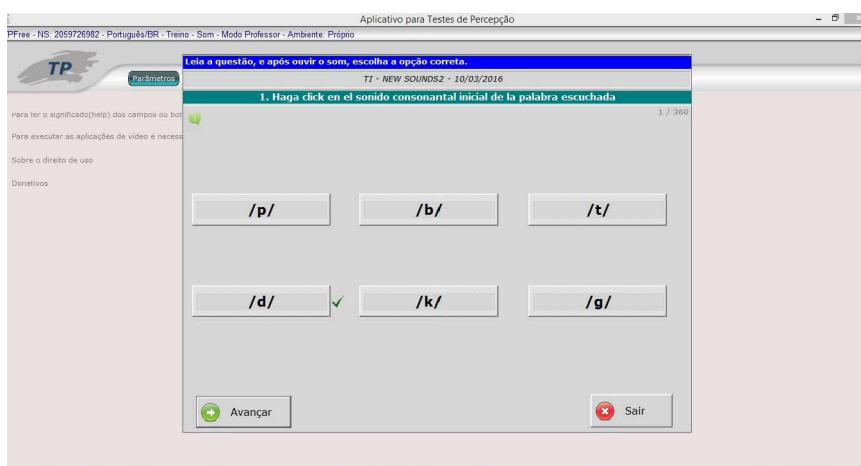
Figure 1. Training sessions: identification test choices on TP



Immediate feedback was offered after each one of the answers provided by the learners. Stimuli with artificial zero VOT were considered to be correct if learners identified the consonant they had just heard as voiced

(and if its place of articulation was correct). By doing so, we expected to train learners to pay attention to positive VOT, as the presence/absence of aspiration was decisive to their answers.

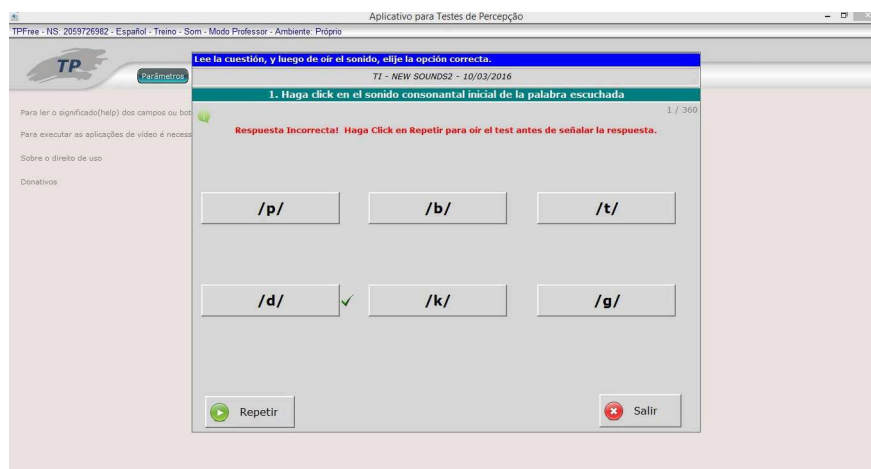
Figure 2. Training sessions: identification test on TP – positive feedback



When answers were not correct, learners were informed of the correct answer immediately, and were forced to

listen to the stimulus again before pressing the correct button, as shown in Figure 3.

Figure 3. Training sessions: identification test on TP – negative feedback



Each training session lasted around 30 minutes. The training tasks were administered at the school lab, and students heard the stimuli with earphones. As already mentioned, in the beginning of each session, participants who belonged to Group 2 were asked to base their identification on the presence/absence of aspiration, and were taught that initial /p/, /t/ and /k/ are aspirated in English.

Data collection instruments – Pre and Post-Tests

As mentioned, participants sat for a pre-test (which took place two days before the beginning of the training sessions), a post-test (which took place three days after the last training session) and a delayed post-test (which took place one month after the first post-test). In all these three data collection sessions, learners performed an identification and a production task.

Identification Task

The identification task follows a similar design to the tests employed in Alves & Motta (2014), Alves & Zimmer (2015), Schwartzhaupt *et al.* (2015) (with Brazilian learners and native speakers of English), and Alves & Luchini (2016) (with Argentinean learners of English). For this study, the identification tasks were built on TP (Rauber *et al.*, 2013).

In the Identification Test administered in the pre-test and in the two post-tests, learners were presented with individual word stimuli and were invited to click on a button indicating the initial consonant of the word they heard (/p/, /b/, /t/, /d/, /k/ or /g/). No immediate feedback was provided. In the beginning of the test, three trial runs were provided. After the trial runs, stimuli with the four VOT patterns (negative VOT, natural zero, artificial zero and positive VOT) were included and presented in a random order. In the task, which comprised 48 stimuli words to be identified, each one of the four VOT patterns was presented in 12 tokens (4 for each place of articulation, the same word produced by a different speaker,⁹ as in [b]it, [d]ick, and [g]ill, for the negative VOT pattern, for example).¹⁰ Tests were taken at the language lab.

Production Task

The production task was also the same one employed in Alves & Zimmer (2015) (with Brazilian learners of English). This test consisted of reading isolated words presented on individual slides of a .ppt file. The target words employed were 'peer', 'pit', 'pee', 'team', 'tick', 'tip', 'kit', 'keel', and 'kill',¹¹ that is, three different lexical items for each place of articulation. Each target word was produced twice, which adds up to six tokens per consonant for each participant.

Participants took the test individually, in a silent room. The participants' production was recorded with a Philips SHM 3550 headset, on a DELL Inspiron laptop computer. Productions were recorded on Audacity 2.0.¹² After collected, the data were analyzed acoustically on Praat *version 5421* (Boersma & Weenink, 2015). The statistics were carried out in SPSS-18.

Results and discussion

Identification Task

The results of the identification task are presented in Table 1. In this table, we present the percentage of

correct answers for each one of the patterns investigated,¹³ as well as the results of the intragroup analysis that we carried out.

Table 1. Accuracy rates (percentage of accuracy in first line, average and standard deviation in second line and median in third line of each column) in the Identification tasks (Pretest, Post-test and Delayed Post-test) and Friedman test results for the three groups.¹⁴

	Pattern	Pre	Post-test 1	Post-test 2	X2 (df)
Group 1 (Training)	Negative¹⁵ VOT	94.79%	98.96%	97.92%	1.50 (2)
		11.38 (1.06)	11.88 (.35)	11.75 (.463)	
		Md:12.00 91/96	Md: 12.00 95/96	Md:12.00 94/96	
	Positive VOT	91.67%	100%	98.96%	8.40 (2) *
		11.00 (.93)	12.00 (.00)	11.88 (.35)	
		Md:11.00 88/96	Md:12.00 96/96	Md:12.00 95/96	
	Zero VOT	75%	89.58%	89.58%	14.25 (2) ***
		9.00 (1.31)	10.75 (1.58)	10.75 (1.76)	
		Md:9.00 72/96	Md:11.00 86/96	Md:12.00 86/96	
	Artificial Zero	40.63%	52.08%	59.38%	7.55 (2) *
		4.88 (1.73)	6.25 (1.04)	7.12 (2.48)	
		Md:4.50 39/96	Md:6.00 50/96	Md:8.00 57/96	
Group 2 (Training + Instruction)	Negative VOT	96.88%	97.92%	100%	3.5 (2)
		11.62 (.52)	11.75 (.46)	12.00 (.00)	
		Md:12.00 93/96	Md:12.00 94/96	Md:12.00 96/96	
	Positive VOT	90.63%	97.92%	96.88%	4.63 (2) *?
		10.88 (1.36)	11.75 (.46)	11.62 (.74)	
		Md:11.50 87/96	Md:12.00 94/96	Md:12.00 93/96	
	Zero VOT	79.17%	94.79%	87.5%	6.08 (2) *
		9,50 (1.77)	11.38 (1.06)	10.50 (1.77)	
		Md:9.00 76/96	Md:12.00 91/96	Md:11.00 84/96	
	Artificial Zero	39.58%	62.5%	54.17%	5.87 (2) *?
		4.75 (1.83)	7.50 (2.33)	6.50 (2.88)	
		Md:5.00 38/96	Md:7.50 60/96	Md:7.00 52/96	

Control Group	Negative VOT	94.79% 11.38 (1.06) Md:12.00 91/96	96.88% 11.62 (.52) Md:12.00 93/96	97.92% 11.75 (.71) Md:12.00 94/96	2.92 (2)
	Positive VOT	98.96% 11.88(.35) Md:12.00 95/96	95.83% 11.50 (1.07) Md:12.00 92/96	93.75% 11.25 (2.12) Md:12.00 90/96	.29 (2)
	Zero VOT	68.75% 8.25 (1.83) Md:9.00 66/96	71.88% 8.62 (2.07) Md:8.50 69/96	69.79% 8.38 (2.20) Md:8.50 67/96	.54 (2)
	Artificial Zero	33.33% 3.88 (1.73) Md:3.50 32/96	34.38% 4.12 (2.90) Md:4.00 33/96	26.04% 3.12 (1.96) Md:2.50 25/96	3.47 (2)

Note. *? p < .10 (marginally significant), *p < .05, ** p < .01, ***p < .001

The descriptive results in Table 1 serve as evidence to our claim (Alves & Luchini, 2016) that additional cues besides VOT are important in the voicing distinctions of English by Argentinean learners. If voicing status was based solely on VOT, both zero VOT and artificial zero VOT would have been identified as voiceless in the pre-test already. However, learners seem to prefer to identify the natural zero VOT pattern as voiced, but the manipulated pattern exhibiting a hybrid consonant as voiceless. This suggests that other cues might be at play in this decision.

We ran Friedman tests¹⁶ (intra-group analyses) in order to verify if there were significant differences among the correct responses in the pre-test, the post-test and the delayed post-test, considering each one of the groups of participants. As expected, no significant differences concerning negative VOT responses in any of the groups were found; this had already been predicted, since voiced stops in Argentinean Spanish are pre-voiced. We had also predicted that a significant difference would not be found for positive VOT, as previous studies (Alves & Luchini, 2016) had shown almost-near ceiling effects in the identification of this pattern as voiceless. Surprisingly, the significant difference found in Group 1 and the marginally significant difference (p=.053) shown in Group 2 indicated that there was still room for improvement, and training helped learners increase their accuracy rates.

Following our first hypothesis, we had predicted that training would prove effective in the identification of (natural) zero VOT and artificial zero VOT. In other words, training would help learners attend to the fact that, unlike what happens in their L1, zero VOT characterizes voiced, not voiceless stops, in the target language. In the same fashion, a significant difference was also hypothesized for artificial zero VOT, as we expected training to help learners focus on VOT as the main acoustic cue responsible for voicing distinctions in the target language. The results of the Friedman tests with Groups 1 and Group 2 confirm this hypothesis: in Group 1, the increase in the accuracy rates of zero VOT was highly significant, and a significant difference was also found in the perception of artificial zero VOT. The effects of training could also be noticed in Group 2, which exhibited a significant increase for zero VOT and a marginally significant difference (p=.053) for artificial zero VOT. Moreover, another source of evidence for the role of perceptual training can be found in the results of the Control Group – no significant differences were found in any of the VOT patterns tested.

In Table 2, we present the significance values of the post-hoc Wilcoxon Tests (employing Bonferroni correction), which compares the pre-test and the immediate post-test, the post-test and the delayed post-test, as well as the pre-test and the delayed post-test.

Table 2. Post-hoc Wilcoxon (Bonferroni) test results – Identification Task.

	Pattern	Pre-test vs. Post-test 1	Post-test 1 vs. Post-test 2	Pre-test vs. Post-test 2
Group 1 (Training)	Negative VOT	-----	-----	-----
	Positive VOT	*	n.s.	n.s.
	Zero VOT	*	n.s.	*
	Artificial Zero	n.s.	n.s.	*
Group 2 (Training + Instruction)	Negative VOT	-----	-----	-----
	Positive VOT	n.s.	n.s.	n.s.
	Zero VOT	*	n.s.	n.s.
	Artificial Zero	*	n.s.	n.s.
Control Group	Negative VOT	-----	-----	-----
	Positive VOT	-----	-----	-----
	Zero VOT	-----	-----	-----
	Artificial Zero	-----	-----	-----

Note: ----- not applicable (Friedman test results were not significant), n.s. not significant, *significant (p<.017)

For Group 1, results of the post-hoc test revealed significant differences between the pre and the post-test in the identification of positive VOT. As already mentioned, this had not been predicted, since learners were expected to present very high accuracy levels in the identification of this pattern right in the pre-test. Still regarding Group 1, significant differences were also found in the identification of zero VOT as voiced, as can be easily seen in the descriptive data shown in Table 1. These significant differences were found between the pre-test and each one of the two post-tests, but not between the two post-tests themselves. These results might be suggestive that, at least for the zero VOT pattern, the results found in the immediate pre-test were maintained in the post-test. Finally, as for the perception of the manipulated VOT pattern by Group 1, significant differences were found between the pre-test and the delayed post-test only. As for this VOT pattern, the descriptive accuracy rates tend to increase (but not significantly) from the pre-test to the post-test, and increase even more in the delayed post-test, indicating that the effects of training may even increase with time.

In Group 2, significant increases for zero VOT and artificial zero VOT were found between the pre and the first post-test. It is interesting to consider that significant results were not found between the pre and the delayed post-test in this group, which prevents us from fully confirming our third hypothesis on the long-term effects of training, as will be discussed later; despite this fact, the descriptive results in Table 1 show that the delayed post-test rates are still higher than those found in the pre-test, but not as high as those found in the immediate post-test. The finding of significant differences only between the pre and the first post-test seems to characterize an opposite pattern to that found in Group 1, in which we found a significant difference between the pre and the delayed post-test, but not between the pre and the first post-test. We may speculate that this difference might be the result of the type of training (with or without explicit instruction) received by each one of the groups. In the group that received instruction (Group 2), the difference in accuracy rates between the pre and the post-test seems to have been more abrupt right in the first post-test, indicating

that the provision of instruction might contribute to immediate effects. In turn, Group 2, which was not instructed on what to pay attention to, needed some more time (and, maybe, a larger amount of input) to “discover” what aspect should be focused on. Although additional studies are undoubtedly necessary for this puzzle to be solved, the possibility that the addition of instruction to training sessions might contribute to more significant differences in a shorter period of time must not be disregarded.

We also ran inter-group tests, in order to verify significant differences between the three groups of participants in each one of the tests. In Table 3, we report the results of the three Kruskal-Wallis tests.

Table 3. Kruskal-Wallis Test Results – Identification Test

	Pre-test X2 (df)	Post-test 1 X2 (df)	Post-test 2 X2 (df)
Negative VOT	.05 (2)	1.28 (2)	2.02 (2)
Positive VOT	4.56 (2)	2.3 (2)	.52 (2)
Zero VOT	.90 (2)	8.65 (2)*	6.11 (2)*
Artificial Zero	1.64 (2)	6.26 (2)*	8.81 (2)*

Note. * < .05

The results show that there were no significant differences among the three groups in the pre-test, indicating that their rate of correct responses tended to be statistically equivalent before the training sessions. As expected, in both post-tests, significant differences were found for zero VOT and artificial zero VOT only. The results of the post-hoc Mann-Whitney tests (with Bonferroni correction) are shown in the following table.

Table 4. Post-hoc Mann-Whitney (Bonferroni) test results – Identification Task

	Group 1 vs. Group 2	Group 2 vs. Group 3	Group 1 vs. Group 3
Negative VOT	Pre-test	-----	-----
	Post-test 1	-----	-----
	Post-Test 2	-----	-----

Positive VOT	Pre-test	-----	-----	-----
	Post-test 1	-----	-----	-----
	Post-Test 2	-----	-----	-----
Zero VOT	Pre-test	-----	-----	-----
	Post-test 1	n.s.	**	*
	Post-Test 2	n.s.	*	*
Artificial Zero	Pre-test	-----	-----	-----
	Post-test 1	n.s.	*	n.s.
	Post-Test 2	n.s.	*	**

Note. ----- not applicable (Kruskal-Wallis test results were not significant), n.s. not significant, *p<.017, ** p<.01

As for zero VOT, both experimental groups (1 and 2) outperformed the Control Group in both post-tests. As for the identification of artificial zero VOT, only Group 2 outperformed the Control Group statistically in the first post-test, but both Groups 1 and 2 outperformed the Control Group in the delayed post-test. This may be understood if we consider the descriptive data shown in Table 1, which indicates that, although there was an improvement in the descriptive accuracy rates of artificial zero VOT in Group 1 between the pre and the post-test, accuracy values are even higher for Group 1 in the delayed post-test. Once again, we should speculate that, with no explicit instruction, it might take longer to “discover” the acoustic cue learners should focus on in the input they received.

Finally, it is also important to highlight that Table 4 shows no significant differences between the results of Group 1 and Group 2, in any of the data collection sessions. Besides reinforcing the effects of perceptual training, these results seem to suggest that both forms of training (with or without instruction provided) might be effective in developing perception.

Summing up, the results of the statistical tests tend to confirm our first hypothesis, which predicted positive effects of training for both experimental groups in the perception of zero VOT and artificial zero VOT. Indeed, training also helped learners perfect their perception of positive VOT. The results seem to suggest that perceptual training (whether accompanied by instruction on aspiration or not) helps learners focus on VOT as a decisive cue, leading them to listen

to the presence/absence of aspiration as a key factor to determine voicing status.

Production Results

In our second hypothesis, we had predicted that the effects of perceptual training could be generalized to production. In Table 5, we present the mean VOT

values of the three groups, as well as their standart deviation and median values. The results of the Friedman tests for each of the groups are also shown.

Table 5. Production test results (average (in ms) in first line, standard deviation in second line and median in third line of each column) and Friedman test results¹⁷

		Pre-test	Post-test 1	Post-test 2	X2 (df)
Group 1 (Training)	/p/	27.77 (15.08) Md:23.86	27.16 (14.77) Md: 23.02	36.17 (20.51) Md: 31.35	4.75 (2) *?
	/t/	50.72 (15.69) Md: 49.09	55.29 (15.37) Md:60.77	59.73 (17.69) Md: 62.28	3.00 (2)
	/k/	67.24 (20.28) Md: 60.22	72.68 (13.99) Md: 77.19	81.98 (20.08) Md: 74.91	5.25 (2) *?
Group 2 (Training + Instruction)	/p/	41.33 (9.25) Md: 42.83	58.08 (15.62) Md: 59.62	58.62 (12.92) Md: 57.17	12.25 (2) **
	/t/	56.14 (14.61) Md: 54.89	64.71 (19.19) Md: 58.46	63.10 (16.13) Md: 62.66	12.25 (2) **
	/k/	77.32 (14.38) 74.94	86.03 (22.15) 81.39	86.66 (20.01) 79.91	4.75 (2) *?
Control Group	/p/	38.45 (21.99) 38.86	40.49 (24.28) 45.04	45.11 (27.63) 35.89	1.75 (2)
	/t/	50.75 (17,41) 46.07	49.69 (16.33) 49.37	56.70 (19.48) 59.29	4.75 (2) *?
	/k/	65.23 (17.8) 64.98	65.83 (19.27) 70.45	70.78 (19.57) 70.61	1.75 (2)

Note. Md = median; Standard deviations are presented in brackets; *? p < .10 (marginally significant), * p<.05, ** p<.01

As we had previously done in the perceptual test results, we ran intra-group analysis to verify if there were going to be significant differences between the three tests, considering each group separately. Although the descriptive data reveal some improvement after training in the production values presented by Group 1, only marginally significant

differences were found in the production of /p/ (p=.093) and /k/ (p=.072). Significant differences (p<.001) were found for /p/ and /t/ in Group 2. As for this group, a marginally significant difference was found for /k/ (p=.093). Surprisingly, the Control Group also showed a marginally significant difference for /t/, with p=.093 (almost reaching 1.0).

In Table 6, we present the results of the post-hoc Wilcoxon tests (Bonferroni correction):

Table 6. Post-hoc Wilcoxon Test (Bonferroni) results – Production Test

		Pre-test vs. Post-test 1	Post-test 1 vs. Post-test 2	Pre-test vs. Post-test 2
Group 1 (Training)	/p/	n.s.	n.s.	n.s.
	/t/	-----	-----	-----
	/k/	n.s.	n.s.	n.s.
Group 2 (Training + Instruction)	/p/	*	n.s.	*
	/t/	*	n.s.	*
	/k/	n.s.	n.s.	n.s.
Control Group	/p/	-----	-----	-----
	/t/	n.s.	n.s.	n.s.
	/k/	-----	-----	-----

Note. ----- not applicable (Friedman test results were not significant), n.s. not significant, * p<.017, ** p<.01

This table indicates a significant difference between the pre-test and the two post-tests in the productions of /p/ and /t/ by Group 2. Even though the results of the production test are not as clear as those found in the perception test, as the production data do not fully confirm our second hypothesis, the results presented in Table 6 detail some important aspects that must be taken into consideration. Firstly, as for the production of /p/ and /t/ by Group 2, significant differences were found between not only the pre-test and the first post-test, but also the pre-test and the delayed post-test. Secondly, as we concentrate on the results for the production of /p/ and /k/ by Group 1, or /t/ by the Control Group (whose significant differences had been set marginally), we find no significant differences in the post-hocs. In other words, the only significant differences which showed post-hoc effects were the ones related to Group 2.

It is also worth mentioning that, even though few significant differences were shown in Table 5, the descriptive data presented in that very same table indicate some increase in VOT values between the pre-test and post-test results, especially for Group 1 (see,

for example, the results for /k/ produced by this group). Despite this descriptive difference, statistical differences were not found. One possible explanation for this fact might be in the low number of participants for each group, which can be considered to be a limitation of the present study. Future replications of this study, with a larger number of participants in each group, might yield significant differences.

Still concerning the intra-group analysis, it has to be considered that no significant differences between the two post-tests were found in any of the groups or consonants. The lack of significant differences between the results of the two post-tests was also noticeable in Table 2, which described the results obtained in the perception test. This might also be regarded as an indicator of the long-term effects of the training sessions.

In what follows, we present the inter-group analysis. Table 7 presents the results of the Kruskal-Wallis tests, which correspond to each one of the three data collections. In Table 8, we present the results of the post-hoc Mann-Whitney tests.

Table 7. Kruskal-Wallis test results – Production

	Pre-test X2 (df)	Post-test X2 (df)	Post-test 2 X2 (df)
/p/	4.22 (2)	8.3 (2) *	4.16 (2)
/t/	1.22 (2)	1.75 (2)	.18 (2)
/k/	2.38 (2)	2.8 (2)	2.14 (2)

Note. * p<.05

Table 8. Post-hoc Mann-Whitney (Bonferroni) test results - Production

		Group 1 vs. 2	Group 2 vs. 3	Group 1 vs. 3
/p/	Pre-test	-----	-----	-----
	Post-test 1	**	n.s.	n.s.
	Post-test 2	-----	-----	-----
/t/	Pre-test	-----	-----	-----
	Post-test 1	-----	-----	-----
	Post-test 2	-----	-----	-----

/k/	Pre-test	-----	-----	-----
	Post-test 1	-----	-----	-----
	Post-test 2	-----	-----	-----

Note. ----- not applicable (Friedman test results were not significant), n.s. not significant, ** $p < .01$

The Kruskal-Wallis tests showed significant differences for Group 2 only, in the production of the bilabial stop /p/. The post-hoc tests show a significant difference between the two experimental groups in the first post-test, which can be confirmed by a visual inspection of the descriptive data presented in Table 5. Whereas Group 2 presented a significant increase between the pre and the first post-test, the first group did not seem to show an increase in the VOT values for this consonant. The results outlined in Tables 7 and 8 conform the intra-group analysis, and do not allow us to confirm our second hypothesis fully. Indeed, significant differences were noticeable in Group 2 only.

While we must consider the possibility that the small number of participants might have played a role in these non-significant differences, it is also important to find some speculative reasons why a significant increase was found only in Group 2, but not in Group 1. In fact, although both groups showed significant intragroup differences with regard to perception, the production results show a significant improvement in only one of the groups, whose participants had been instructed on what to focus on in the training sessions. Given these results, we cannot disregard the possibility that explicit instruction might have had a role in this significant difference. As the production test allows for a high level of monitoring, the provision of explicit knowledge on the phenomenon to be focused on might be used in monitored production. In other words, it might be the case that this significant difference is not the direct result of perceptual improvement, but the use of explicit knowledge in monitored production. Additional studies, with a larger number of participants and some production test designs that allow for less monitored production, might be relevant in providing a more definite answer to the possibility raised here.

Final considerations

As we analyze the perception and production results by the groups in the three tests (pre-test, immediate post-test and delayed post-test), the hypotheses proposed in the Introduction of this paper must be revisited. Hypothesis 1 predicted that perceptual training, with or without explicit instruction, would lead to an improvement in the identification of zero VOT and artificial zero VOT as voiced. This hypothesis was confirmed, as both experimental groups showed significant differences in these two patterns. Perceptual training was also relevant in the identification of positive VOT as voiceless, helping learners reach ceiling effects in the correct identification of this VOT pattern.

As for the second hypothesis, which predicted that learners would be able to generalize this growth to production, this could not be fully corroborated. Indeed, only marginally significant differences (with no post-hoc significant differences) were found in Group 1. In the intra-group analysis, Group 2 presented a significant increase concerning the production of /p/ and /t/, so we cannot disregard the possibility that instruction played a more decisive role in these results. In this sense, instruction might have proved useful in allowing learners to monitor themselves and achieve higher VOT results, even when they are not developmentally ready to do so. Further studies investigating the role of instruction isolated from perception training might also be useful, as they might show that students receiving instruction might present better production levels even before an increase in perception, challenging the canonical perception-production developmental order (a possibility raised in Flege, 1995). It might be the case, therefore, that this increase in production might be the reflection of conscious monitoring, and might not be reflected in more natural speech settings.

Finally, our third hypothesis predicted that the improvements found in both perception and production would be maintained one month after the last training session. Once again, this hypothesis was only partially corroborated. As for the perception of both zero VOT and artificial zero VOT, our intra-group analysis showed no significant differences between the

pre-test and the delayed post-test in Group 2 (which received instruction), despite the significant difference found between the pre-test and the immediate post-test. Despite this fact, it is well true that the descriptive rates found in their delayed post-test are still much higher than those found in their pre-test. As for the accuracy rates for Natural VOT by Group 1, significant differences are found between the pre-test and each one of the two post-tests, which would allow us to corroborate this hypothesis; however, with regard to the artificial zero pattern, a significant difference is found between the pre-test and the delayed post-test only. All of these perceptual results lead us to speculate that the combination of explicit instruction and perceptual training might lead to immediate changes in the learners' perceptual rates; these changes might be so abrupt that such high rates are not maintained one month later. In turn, it might be the case that learners that receive no instruction need a longer period of time in order to 'tune in' to the right cue. As for the production results, the intra-group analysis indicated that the significant increase in the production of /p/ and /t/ by Group 2 also presents a long-term status. All these factors considered, it is undeniable that, even in those cases in which no significant differences between the pre-test and the delayed post-test had been found, the descriptive values found in the delayed post-test were still closer to those found in the immediate post-test than to those found in the pre-test, which allows us to suggest some positive (descriptive) effects of the training in the post-test. As a result of this fact, significant differences between the immediate and the delayed post-test were never found in perception or production, suggesting that the effects of training might still be felt one month later.

It is undeniable that the present study shows a considerable number of limitations, most of which have already been pointed out throughout this article. Firstly, the number of participants might have contributed to the absence of significant differences in the production test. Secondly, the number of training sessions (only three) might not have been enough to foster generalization to production. Indeed, this small number of sessions is a result of time constraints faced with the group of

learners investigated, and are a consequence of problems that are frequently faced by experimental studies which deal with classroom realities. In this study, we aimed at minimizing such a limitation with the provision of awareness raising to Group 2, which would accelerate the processing of the target item being trained. Finally, it might be the case that our delayed post-test should have taken place at some time later than one month. This would have allowed us to say whether the supposed perceptual improvement found in the delayed post-test in Group 1 (training only) would be maintained after a longer period of time. A more delayed post-test would have also helped us say whether the improvements in production found in Group 2, which were considered to be the result of a more monitored production, would be maintained at some time longer. We have to reinforce, once again, that this short period of time between the two post-tests was a result of the time constraints imposed by the classroom environment in which our research study took place.

These limitations open new avenues for further investigations and research questions. With regard to perception, further studies on the effects of place of articulation in the perception of zero VOT and artificial zero VOT might be of great importance. As for production, further analyses of the generalization to novel items also prove relevant.¹⁸ Finally, the effects of explicit instruction combined with perceptual training need additional research studies. It is also important to investigate the role of these two classroom interventions individually; this will allow us to verify if the effects of training are fostered by instruction, or if instruction by itself might be relevant, regardless of any perceptual practice. In this sense, variables such as the number of training sessions in perceptual studies, as well as the kind of awareness raising task provided (with a more or less metalinguistic/communicative tone) are also important aspects to be considered and investigated.

In conclusion, the results presented in this paper indicate beneficial effects of perceptual training in foreign language classrooms, even in situations in which time constraints might represent an impediment for a higher number of training sessions. The provisions of instruction added to perception might not only

contribute to an increase in perception, but also foster production. Considering the results of the study, we may say that perceptual training not only helped improve the perception of a given acoustic cue that proved difficult to learners; indeed, it guided learners to focus on a new cue which, in their first language, does not play a decisive role.

Notes

1. Voiced stops in word-initial position in English may be produced variably with a zero VOT pattern or pre-voicing. Therefore, Argentinean learners, who produce a negative VOT pattern in word-initial voiced stops, do not need to change their VOT patterns in word-initial /b, d, g/, as far as their production is concerned. Previous studies (e.g. Simon & Leuschner, 2010) have shown that learners whose L1 systems exhibit pre-voiced stops do not tend to change this pattern in the development of L2 English. For this reason, in this study, we concentrate on the training and testing of voiceless stops only.
2. Rato (2014, p. 531) defines High Variability Perceptual Training (HVPT) as that “with multiple talkers and stimuli”.
3. As we acknowledge the fact that spectral and timing cues interact perceptually as they are integrated in the perception of stops (Dmitrieva *et al.*, 2015; Francis *et al.*, 2008; Kingston *et al.*, 2008), one might ask why we have isolated the VOT cue in our training and testing experiments. As explained above, given the fact that learners attend to other cues besides positive VOT in perception, they find no difficulties in discriminating and identifying voiced and voiceless initial stops in English (Alves & Motta, 2013; Alves & Zimmer, 2015; Alves & Luchini, 2016). Although no perceptual problems are found, when it comes to production, learners also use these other cues and do not attend to positive VOT. This lack of word-initial aspiration leads to identification and intelligibility problems among native speakers of English (Schwartzhaupt, 2015; Schwartzhaupt *et al.*, 2015). Therefore, in line with Abramson & Whalen (2017), by focusing on VOT alone and by providing a manipulated pattern which “forces” learners to focus on the presence of positive VOT, we expect learners to focus on positive VOT in perception; as a consequence, this should lead to higher VOT values in the production of word-initial voiceless stops.
4. In the identification pre and post-tests, we also investigated the perception of negative VOT and positive VOT in English. However, given the ceiling effects found in Alves & Luchini (2016), we did not include these two patterns in this hypothesis, as we expected high accuracy levels in perception in the pre-test already.
5. For further information on the Oxford Online Placement Test Online, see Purpura (2007) and Pollitt (2007).
6. According to the *Common European Framework of Reference for Languages* (CEFR), proficiency is characterized in six levels: A1, A2, B1, B2, C1, C2, being these latter two the most advanced ones. Participants in levels C1 and C2 are considered proficient users. For more information, see <http://www.examenglish.com/CEFR/cefr.php>.
7. These speakers were the same whose stimuli were used in previous studies, such as Alves & Motta (2014), Alves & Zimmer (2015) and Schwartzhaupt *et al.* (2015). They are the same speakers whose stimuli were used in the identification pre and posttests (even though the identification task in the pre and post-tests was carried out with other target words).
8. We can justify the low number of lexical items due to the fact that, in the stimuli obtained by the six speakers, tokens of word-initial /b/, /d/, /g/ with zero VOT were not frequently produced, as negative and zero VOT may occur variably in word-initial voiced stops in English. These were the lexical items whose productions were more frequently produced with zero VOT.
9. The same speakers whose stimuli were presented in the training task.
10. The lexical items in the identification task in the pre and post-tests are different from those stimuli used in the training sessions. Therefore, should there be an improvement in the accuracy rates in the identification test, this indicates the learners’ ability to generalize their perceptual ability to different lexical items.
11. From the three lexical items that represent each one of the places of articulation, one of them had been used in the training task (*pee, tip, kit*), another one had been employed in the perceptual pre and post-tests (*pit, tip, kill*) and one was a novel lexical item (*peer, team, keel*). With this design, we aim at investigating whether there are higher VOT values in those lexical items with which learners have already been trained. For delimitation purposes, we leave this verification for a future study.
12. Free software, obtained on <<http://www.audacity.sourceforge.net>>.
13. As already mentioned, for stimuli starting with positive VOT, answers identifying the consonants as voiceless (/p/, /t/, /k/) were considered to be correct. For stimuli

starting with the other three patterns (negative VOT, zero VOT and artificial zero VOT), answers identifying the consonants as voiced (/b/, /d/, /g/) were considered to be correct. Mistakes concerning place of articulation (for example, when aspirated /p/ was perceived as /t/, although the voicing of the initial consonant was identified correctly) were not computed as correct answers.

14. In this table, perception results for all places of articulation are averaged together, since we found no place of articulation effects on perception.
15. As already shown in Alves & Luchini (2016), the perception of negative VOT and positive VOT by Argentinean learners tend to exhibit ceiling effects. This is justified as negative VOT occurs in word-initial voiced stops in Spanish, and learners tend to focus on other acoustic cues (such as F0 and burst intensity), instead of aspiration, to identify aspirated stops as voiceless. As stated in our fourth footnote, this is the reason why no hypotheses were proposed for these two patterns. These results reinforce the need of a perceptual training approach focusing solely on the presence/absence of aspiration.
16. In this study, we ran non-parametric tests, as the Normality Tests of Kolmogorov-Smirnov and Shapiro-Wilk indicated that the dependent variables tested did not show a normal distribution.
17. Unlike the data shown in Table 1 (perception), in this table each place of articulation is presented separately, since differences regarding place of articulation can be clearly shown in production. Although data on word-initial voiced stops were also collected, these data are not presented in this paper, as all of the students' productions tended to produce pre-voiced consonants (cf. Simon & Leuschner, 2010). As pre-voiced stops occur variably in word-initial position in English, we interpret that the production of negative VOT by learners does not affect intelligibility and, therefore, they need not acquire the zero VOT pattern in word-initial /b, d, g/. This also justifies why our training sessions focused on the presence or absence of Positive VOT only.
18. As mentioned in the Method, our production test allowed for the investigation of the effect of both trained and novel words. This investigation corresponds to the next step in our analysis.

References

- Abramson, A., & Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1, 1-8.
- Abramson, A., & Whalen, D. H. (2017). Voice Onset Time (VOT) at 50: Theoretical and practical issues in measuring voicing distinctions. *Journal of Phonetics*, 63, 75-86.
- Aliaga-Garcia, C. (2010). Measuring perceptual cue weighting after training: A comparison of auditory vs. articulatory training methods. In K. Dziubalska-Kolaczyk, M. Wrembel & M. Kul (Eds.), *Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 12-18). Poznan, Poland.
- Alves, U. K. (2004). *O papel da instrução explícita na aquisição fonológica do inglês (L2) – evidências fornecidas pela Teoria da Otimidade*. Unpublished Masters' Thesis. Pelotas: Universidade Católica de Pelotas.
- Alves, U. K. & Luchini, P. L. (2016). Percepción de la distinción entre oclusivas sordas y sonoras iniciales del inglés (LE) por estudiantes argentinos: datos de identificación y discriminación. *Revista Lingüística (ALFAL)*, 32, 25-39.
- Alves, U.K. & Magro, V. (2011). Raising awareness of L2 phonology: explicit instruction and the acquisition of aspirated /p/ by Brazilian Portuguese speakers. *Letras de Hoje* 46 (3), 71-80.
- Alves, U. K. & Motta, C. S. (2014). Focusing on the right cue: Perception of voiceless and voiced stops in English by Brazilian learners. *Phrasis – Studies in Language and Literature* (5), 31-50.
- Alves, U. K. & Zimmer, M. C. (2015). Perception and production of English VOT patterns by Brazilian learners: the role of multiple acoustic cues in a DST perspective. *Alfa Revista de Linguística*, 59 (1), 155-175.
- Andringa, S., & Rebuschat, P. (2015). New directions in the study of implicit and explicit learning – an introduction. *Studies in Second Language Acquisition*, 37, 185-196.
- Bettoni-Techio, M. (2008). *Perceptual Training and word-initial /s/-clusters in Brazilian Portuguese/English Interphonology*. Unpublished Doctoral Dissertation. Florianópolis: Universidade Federal de Santa Catarina.
- Boersma, P. & Weenink, D. (2015) *Praat – Doing Phonetics by Computer – Version 5.3.48*. 2013. <http://www.praat.org>.
- Brawerman-Albini, A. (2012). *Os efeitos de um treinamento de percepção na aquisição do padrão acentual pré-paroxítono da língua inglesa por aprendizes brasileiros*. Unpublished Doctoral Thesis. Curitiba: Universidade Federal do Paraná.
- 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. Unpublished Doctoral Dissertation. Barcelona: Universitat Autònoma de Barcelona.
- Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27, 207-229.
- Dmitrieva, O., Llanos, F., Shultz, A. A. & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset F0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49, 77-95.
- Echelberger, A. D. (2013). *Explicit pronunciation instruction and its impact on the intelligibility of literacy level adult EL learners*. Unpublished Master's Thesis. Saint Paul, MN: Hamline University.
- Ellis, Nick C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition*, 27, 305-352.
- Flege, J. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.). *Speech perception and language experience: Issues in cross language research* (pp. 233-278). Maryland: York Press.
- Francis, A. L., Kaganovich, N., & Driscoll-Huber, C. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *The Journal of the Acoustical Society of America*, 124(2), 1234-1251.
- Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O-S. Bohn & M. J. Munro (Eds.). *Language Experience in Second Language Learning* (pp. 57-77). Amsterdam: John Benjamins.
- Kingston, J., Diehl, R. L., Kirk, C. J., & Castleman, W. A. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *Journal of Phonetics*, 36(1), 28-54.
- Kissling, E. (2013). Teaching Pronunciation: is explicit phonetics instruction beneficial for FL learners? *Modern Language Journal*, 97(3), 720-744.
- Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice Onset Time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of Phonetics*, 40, 725-744.
- Lima Junior, R. (2010). Uma investigação dos efeitos do ensino explícito da pronúncia na aula de inglês como língua estrangeira. *Revista Brasileira de Linguística Aplicada*, 10(3), 747-771.
- Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384-422.
- Nobre-Oliveira, D. (2007). *The effect of perceptual training on the learning of English vowels by Brazilian Portuguese speakers*. Unpublished Doctoral Dissertation. Florianópolis, SC: Universidade Federal de Santa Catarina.
- Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39, 59-67.
- Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *Journal of the Acoustical Society of America*, 127, EL54-EL59.
- Perozzo, R. V. (2013). *Percepção de oclusivas não vozeadas sem soltura audível em codas finais do inglês (L2) por brasileiros: o papel da instrução explícita e do contexto fonético-fonológico*. Unpublished Master's Dissertation. Porto Alegre: Universidade Federal do Rio Grande do Sul.
- Pollitt, A. (2007). *The meaning of OOPT Scores*. Available on <http://www.oxfordenglishtesting.com>. Access on August 26, 2013.
- Purpura, J. (2007). The Oxford Online Placement Test: What does it measure and how?. *Disponível em* <http://www.oxfordenglishtesting.com>. Access on August 26, 2013.
- RAE – Real Academia Española. (2011). *Nueva Gramática de la lengua española – Fonética y Fonología*. Barcelona: Espasa Libros.
- Rato, A. (2013). *Cross-language perception and production of English vowels by Portuguese learners: the effects of perceptual training*. Unpublished Doctoral Dissertation. Braga: Universidade do Minho.
- _____. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Concordia Working Papers in Applied Linguistics*, 5, 529-546.
- Rauber, A. S., Rato, A., Santos, G. R., Kluge, D. C., & Figueiredo, M. (2013). *TP: Testes de Percepção e Treinamento Perceptual com Feedback Imediato – Versões 3.1*. Available in http://www.worken.com.br/tp_regfree.php. 2013.
- Reis, M. & Nobre Oliveira, D. (2008). Effects of perceptual training on the identification and production of the English voiceless plosives aspiration by Brazilian EFL learners. In A. S. Rauber; M. A. Watkins; B. O. Baptista. *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianópolis: Universidade Federal de Santa Catarina, 398-407.
- Sangüesa, V. M. (2016). *The second time around: the effect of FI upon return from SA*. Unpublished Master's Thesis. Barcelona: Universitat Pompeu Fabra.

- Schwartzhaupt, B. M. (2012). *Factors influencing Voice Onset Time: analyzing Brazilian Portuguese, English and Interlanguage data*. Unpublished Essay. Porto Alegre: Universidade Federal do Rio Grande do Sul.
- Schwartzhaupt, B. M. (2015). *Testing intelligibility in English: the effects of Positive VOT and contextual information in a sentence transcription task*. Unpublished Master's Thesis. Porto Alegre: Universidade Federal do Rio Grande do Sul.
- Schwartzhaupt, B. M., Alves, U. K. & Fontes, A. B. A. L. (2015). The role of L1 knowledge on L2 speech perception: investigating how native speakers and Brazilian learners categorize different VOT patterns in English. *Revista de Estudos da Linguagem* (23), 311-334.
- Silveira, R. (2004). *The influence of pronunciation instruction in the perception and production of English word-final consonants*. Unpublished Doctoral Thesis. Florianópolis: Universidade Federal de Santa Catarina.
- Simon, E. (2010). *Voicing in Contrast: Acquiring a Second Language Laryngeal System*. Ghent: Academia Press.
- Simon, E., & Leuschner, T. (2010). Laryngeal systems in Dutch, English and German: a contrastive-phonological study on second and third language acquisition. *Journal of Germanic Linguistics*, 22(4), 403-424.
- Sundara, M. (2005). Acoustic phonetics of coronal stops: A cross-language study of Canadian English and Canadian French. *Journal of the Acoustical Society of America*, 118, 1026-1037.
- Tobin, S. J., Nam, H., & Fowler, C. A. (2017). Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics*, 65, 45-59.
- Wong, J. (2012). Training the Perception and Production of English /e/ and /æ/ of Cantonese ESL Learners: a comparison of low vs. high variability phonetic training. *Proceedings of the 14th Australasian International Conference on Speech Science and Techniology*. Sydney, Australia, p. 3-6.
- Yavas, M., & Wildermuth, R. (2006). The effects of place of articulation and vowel height in the acquisition of English aspiration stops by Spanish speakers. *IRAL*, 44, 251-263.

Recebido em: 11/02/2017

Aceito em: 09/05/2017