

PERCEPTION AND PRODUCTION OF ENGLISH VOT PATTERNS BY BRAZILIAN LEARNERS: THE ROLE OF MULTIPLE ACOUSTIC CUES IN A DST PERSPECTIVE

Ubiratã Kickhöfel ALVES*
Márcia Cristina ZIMMER**

- **ABSTRACT:** In this study, departing from a dynamic conception of L2 phonetic-phonological acquisition, we investigate 34 Southern Brazilian learners' perception (identification and discrimination) and production of VOT patterns of initial stops in English. We initially hypothesized that, especially among learners with a basic level of L2 proficiency, VOT was not the main acoustic cue employed in the perception of voicing distinctions. Our results show that, regardless of the learners' proficiency level (basic or advanced), VOT is not a sufficient cue for the distinction between /p/, /t/, /k/ and /b/, /d/, /g/. These results, which have an influence on the lower VOT values found in our production data, conform with a dynamic view of L2 acquisition, according to which multiple acoustic cues play a role in language acquisition, forcing learners to tune in to the most important cue(s) in the target language.
- **KEYWORDS:** VOT. Second Language Acquisition. Acoustic Cues.

Introduction

The process of learning phonetic-phonological aspects of a second language (L2)¹ is complex and dynamic, for many variables, acting conjointly, are fundamental to understand this process. Regarding the perception and production of the target language sounds, multiple acoustic cues are at play in establishing the functional differences among the sounds to be acquired. In that respect, learning an L2 implies the learner's skill not only to perceive the acoustic cues which are productive in the target system, but also to use them in order to establish phonological differences in the foreign language system.

As an example to the challenges to be faced by learners, we may consider the task Brazilians undertake when learning English *Voice Onset Time* (VOT)

* UFRGS – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Letras. Porto Alegre – RS – Brasil. 91501-970 – ukalves@gmail.com.

** UNIRITTER – Centro Universitário Ritter dos Reis. Programa de Pós-Graduação em Letras. Porto Alegre – RS – Brasil. 90840-440 – marcia.zimmer@gmail.com.

¹ In this study, the terms 'Second Language' and 'Foreign Language' are treated as synonyms, as well as the terms 'acquisition' and 'learning'.

patterns. In English, voiceless plosives /p/, /t/, /k/ are produced with a long VOT interval, which is also called Positive VOT (aspiration). This is the main phonetic cue employed in the distinction between voiceless and voiced stops (SCHWARTZHAUPT; ALVES; FONTES, 2013). However, in a previous pilot study, Alves and Zimmer (2012) suggested that among Brazilian learners VOT duration did not seem to be a fundamental cue for the distinction between voiced and voiceless stops in English, contrary to what is observed among native speakers of that language. Brazilian Portuguese speakers seem to pay more attention to other acoustic cues, such as burst intensity and the F₀ value of the vowel following the stop, when establishing functional differences between voiceless and voiced plosives in English. This might also account for the fact that Brazilian learners, even in advanced levels of proficiency, are not able to produce VOT patterns similar to the ones found among natives (ALVES; SCHWARTZHAUPT; BARATZ, 2011).

In other words, following the hypotheses raised in the pilot study carried out by Alves and Zimmer (2012), it is possible that Brazilian learners do not use VOT as their main cue in distinguishing voiceless from voiced stops in the target language. This considered, it might be the case that other acoustic cues are being primarily employed in the perception and production of voice distinctions. Similar cases have been discussed in recent studies (SUNDARA, 2005; OH, 2011; KONG; BECKMAN; EDWARDS, 2012), which investigate Canadian French, Korean and Japanese, respectively. In these languages, additional cues, such as burst intensity and fundamental frequency (F₀) in the following vowel, take the lead as the main acoustic correlates employed to distinguish voiceless from voiced plosive segments in perception and production.

Such findings have a direct impact on the understanding of the process of phonetic-phonological learning of a foreign language (FL). The acquisition of the two-way voicing system of English (L2) will imply that learners focus their attention on VOT, so as to learn the new pattern (aspiration) which occurs in English. The acquisition of English aspiration by learners of these L1 systems, therefore, would require a double task: before learning the L2 VOT pattern itself, learners have to “listen to” this cue, which does not play such an important role in their first language.

In terms of L2 acquisition perceptual models, the L2 re-structuring of acoustic cues can be explained by Best and Tyler's *Perceptual Assimilation Model-L2* (BEST; TYLER, 2007). According to Antoniou et al. (2011), this model is based on the theory of Gestural Phonology (BROWMAN; GOLDSTEIN, 1992, 1993, 2000). Indeed, following Goldstein and Fowler (2003), we can postulate the notion of phonological gesture as “common currency” of analysis between phonological knowledge, perception and production. In that sense, “[...] by acquiring an L2, learners are exposed to a new set of articulatory gestures, including new phasing

relations and patterns of coordination between these gestures.” (ANTONIOU et al., 2011, p.560).

Departing from the statement that “[...] if phonological atoms are public actions, then they directly cause the structure in the speech signals, which, then, provides information directly about the phonological atoms.”, (GOLDSTEIN; FOWLER, 2003, p.179), our aim in this study is to explain how the exposure to an acoustic cue which is conveyed by a distinctive gesture in the target language can cause changes in the perception and production of Brazilian learners’ interlanguage system.

Therefore, based on perception and production tests, we discuss the possible redundant feature of VOT for the distinction between voiceless and voiced stops in English. Our main goals are: (i) to assess whether Brazilian learners of English in two different levels of proficiency present distinct response rates in the perception of the VOT patterns produced by native speakers of English; (ii) to investigate whether learners in the two proficiency groups produce VOT patterns similar to those found in the target language; (iii) to discuss the role of VOT as the main acoustic cue used by Brazilian learners in the functional distinction between voiceless and voiced initial stops.

Method

Participants

34 participants², learners of English residing in the Southern Brazilian city of Porto Alegre, took part in the study. After having taken the Oxford Online Placement Test³, learners were placed as elementary (24 learners whose levels ranged from A1 to A2 in the Common European Framework) and advanced (10 learners whose levels ranged from C1 to C2). All learners took perception (Identification and Discrimination) and production tests.

Perception tests

The stimuli were recorded in a professional studio by six native speakers of North American English (three male and three female) who had been living in Southern Brazil for less than 6 months. These six speakers read a set of three

² All the participants filled in a Free and Informed Consent Form in which they received information on the procedures involved in the data collection, as well as the risks and benefits of the study. Participants were also informed they could quit at any phase of the experiments.

³ The Oxford Online Placement Test is a validated test, taken online from the website <www.oxfordenglishtesting.com>. For further details on the test, see Purpura (2007) and Pollitt (2007).

minimal pairs (bit – pit; dick – tick; gill – kill), each pair starting with a different place of articulation, followed by a high vowel (YAVAS; WILDERMUTH, 2006). In order to guarantee the quality of the audio stimuli, each speaker was asked to record their word list three times, so that the best tokens could be chosen for the perception tests.

The stops produced by the six native speakers of English presented three different VOT patterns. Voiceless stops (pit, tick, kill) were always produced with Positive VOT, whereas /b, d, g/ (bit, dick, gill) were produced either with pre-voicing (Negative VOT) or with Zero VOT, as VOT patterns in initial /b, d, g/ are variable in English (LISKER; ABRAMSON, 1964; ABRAMSON; LISKER, 1973; DOCHERTY, 1992; SIMON, 2010).

Besides these three VOT patterns, productions of voiceless plosives were also manipulated in Praat – Version 5.2.9 (BOERSMA; WEENINK, 2013), so that we could obtain the Manipulated Zero VOT pattern: as the VOT of the plosives was reduced, the resulting manipulated consonant would have the same VOT duration as that of a voiced segment, but could maintain the other acoustic cues found in a voiceless stop in English. From the contrast between the manipulated Zero and the Natural Zero VOT patterns, we can assess whether VOT plays the role of key acoustic cue in the distinction between English voiceless and voiced stops by Brazilian learners.

Therefore, both Identification and Discrimination tests were built with these four VOT patterns: Negative VOT, Positive VOT, Non-Manipulated Zero VOT and Manipulated Zero VOT. Both tests were also built and administered in Praat. The sections that follow provide more detail on each one of the tests.

Identification Test

In the identification test, the participants were presented with individual word stimuli (a member of one of the three minimal pairs described above) and were invited to click on a button indicating the initial consonant of the word they heard (/p/, /b/, /t/, /d/, /k/ or /g/). Stimuli with the four VOT patterns (Negative VOT, Natural Zero, Artificial Zero and Positive VOT) were presented in a random order. The task had a total of 48 stimuli words to be identified, and each one of the four VOT patterns was presented in 12 tokens (4 for each place of articulation).

Discrimination Test

The discrimination test consisted of an AxB task. In this task, the stimuli presented to learners were made up of word triads. The participants were provided

with multiple choice questions and were asked to indicate if the initial consonant of the second word was similar to that of the first word (e.g. bit – bit – pit), to that of the third word (e.g. bit – pit – pit), or if the three words began with the same consonant (e.g. pit – pit – pit).

Three kinds of contrasts were tested in the AxB task: Negative VOT vs. Manipulated Zero VOT (12 questions – four for each place of articulation); Negative VOT vs. Positive VOT (12 questions), Manipulated Zero VOT vs. Positive VOT (12 questions). Other possible contrasts, such as Non-Manipulated Zero VOT vs. Negative VOT, as well as Non-Manipulated Zero VOT vs. Artificial Zero VOT, were not included in this experiment for delimitation purposes, since it had already been reported that learners tend to discriminate the latter, but not the former, of these two contrasting pairs (ALVES; SCHWARTZHAUPT; BARATZ, 2011). Besides the three kinds of contrasts used in the present experiment, the test also presented nine (three for each place of articulation) “catch trial” questions – that is, triads that presented the same initial consonants (e.g. pit – pit – pit), so that we could test the participants’ attention to the task.⁴

Production tests

The same learners who participated in the two perception tests also took two production tests, one in English and another one in Brazilian Portuguese (BP).

Word production test in Brazilian Portuguese

The participants were asked to read words starting with the segments /p/, /k/, /b/, /g/ and followed by a high front vowel, corresponding to the same phonetic-phonologic context used in the perception tests. Words starting with /t/ and /d/ were not included in the instrument because the alveolar stop is palatalized before [i] in the dialect spoken by the participants (KAMIANECKY, 2002). In this study, we will present the results concerning the voiceless stops /p/ and /k/.

For each one of the target consonants, there were two types (apart from eight **types** distractor words, starting with non-plosive segments) in the test. Each **type** was produced twice, which adds up to 4 **tokens** for each consonant, produced by each one of the participants. The words were presented in a Microsoft PowerPoint file (.ppt), each one in a different **slide**, on a Sony Vaio laptop, model PCG-31311X. The participants’ production was recorded with the aid of a Philips SHM 3550 headset, via *Software Audacity* (2015) – version 2.0.5.

⁴ Since the answers provided in the catch trials presented high accuracy levels, indicating, therefore, that participants were in fact paying attention to the task, these results are not going to be presented in this article.

Word production test in English

In this test, which also consisted of reading words presented individually on slides of a .ppt file, the target words started with the segments /p/, /t/, /k/, /b/, /d/, /g/ and were followed by a high front vowel (ex. **pit, tip, kit**). Bearing in mind the goals of this study, we will report only the VOT values of voiceless plosives /p/, /t/, /k/.

Apart from distractors, the test comprised three *types* per consonant. Each *type* was produced twice, which adds up to six tokens per consonant for each participant. Similarly to the Brazilian Portuguese test, the participants' production was recorded with the aid of a Philips SHM 3550 headset, on a Sony Vaio laptop.

Hypotheses

In this section, we present the hypotheses outlined for each of the tests (Identification, Discrimination and Production). All the hypotheses follow the assumption that, in a more elementary proficiency level, learners do not follow VOT as their main cue to distinguish between voiceless and voiced stops, whereas advanced learners will give that cue a priority status.

Identification test hypotheses

H1: "With regard to the identification of Negative VOT (productions of English /b/, /d/, /g/) and Positive VOT (aspiration of English /p/, /t/, /k/) patterns, there will be no significant differences between the two proficiency groups."

Motivation for the hypothesis: Even if not guided by VOT, elementary level learners of English will identify Negative VOT as voiced and Positive VOT as voiceless, as they rely on additional acoustic cues, such as burst intensity, which leads them to a correct characterization of voicing in these consonants.

H2: "As for the identification of Non-manipulated Zero VOT, there will not be a significant difference between students in the two proficiency levels."

Motivation: Learners in a more elementary level will tend to identify these consonants as voiced stops, as they might be guided by burst intensity when identifying consonants. More advanced learners, in turn, will be guided by the L2 VOT pattern, so the short lag VOT will also lead learners to identify these consonants as voiced.

H3: "Concerning the identification of Manipulated Zero VOT, there will be a significant difference between students at the two proficiency levels."

Elementary learners of English will base their judgments on L1 cues which seem to be more important than VOT, since all acoustic features of this manipulated sound – apart from the reduced interval of aspiration – lead to its identification as /p/, /t/, /k/; therefore, elementary learners will identify this pattern as voiceless. Advanced learners, in turn, will identify these consonants as voiced, as they already have VOT as their main acoustic cue and will tend to follow the pattern found in English in order to distinguish voiceless from voiced consonants.

Discrimination test hypotheses

H4: “As for the discrimination between Negative VOT vs. Positive VOT, we hypothesize that there will not be a significant difference between elementary and advanced learners of English.”

Motivation: Both groups of learners are going to discriminate between English tokens of /b/, /d/, /g/ and /p/, /t/, /k/, even if through different main acoustic cues, as only more advanced learners are expected to use the L2 VOT pattern as their main acoustic cue in voicing distinctions.

H5: “Regarding the discrimination between Negative vs. Manipulated Zero VOT, a significant difference between the two proficiency groups will be found.”

Motivation: Elementary level learners are expected to present high discrimination rates, as they might be guided by their L1 acoustic cues, leading them to consider Negative VOT (weak burst intensity) as voiced and Manipulated Zero VOT (strong burst intensity) as voiceless. More advanced learners, however, will be guided by the VOT cue: as both patterns are characterized by a short lag VOT, we do not expect them to be discriminated by these learners who follow the L2 pattern.

H6: “Regarding the discrimination between Manipulated Zero and Positive VOT, we also hypothesize that there will be a significant difference between the results found in the two groups.”

Motivation: Elementary level learners are expected to present lower discrimination levels, as they are guided by their L1 cues and disregard VOT as the main cue in their responses. More advanced learners, on the other hand, guided by the English VOT pattern as their main cue, will discriminate these two patterns, as the former presents a short lag VOT, while the latter is characterized by aspiration.

Production test hypotheses

H7: “In each of the groups (considered separately), there will be a significant difference between the VOT values of /p/ and /k/ found in Brazilian Portuguese and in English.”

Motivation: although the learners are not likely to have achieved VOT patterns similar to the ones produced in English, mainly due to the fact that VOT tends not to be the main acoustic cue followed in their L1, they already make partial use of the VOT cue to signal, with a longer duration, the English plosives.

H8: “There will be no significant differences between the two groups in their production of long-lagged VOT of each of the English stops.”

Motivation: this hypothesis follows previous trends in the literature (ALVES; SCHWARTZHAUPT; BARATZ, 2011) which suggest that, regardless of participants' proficiency level, their VOT patterns do not reach the native ones. Although both basic and advanced learners are able to identify aspirated plosives as voiceless (as laid out in our first hypothesis), and even when they use VOT as their main cue (advanced learners), we claim that such facts do not necessarily account for significant differences in the production of VOT patterns between the two groups.

Results and Discussion

This section is divided in three parts, which deal with the description and discussion of the identification, discrimination and production data, respectively.

Identification

The data concerning the Identification test are presented in what follows. As we can see in Table 01, regardless of the participants' level of proficiency, the Negative VOT pattern (pre-voicing) is nearly categorically identified as voiced (99,31%, $m=12$ – elementary level, and 96,67%, $m=12$ – advanced level). Mann-Whitney tests showed no significant differences between the two proficiency levels in the identification of the segments as voiceless ($U=108,00$; $p=,121$) or voiced ($U=112,5$; $p=,487$). This is not surprising, as in the participants' mother language, prevoicing is already a cue for the presence of a voiced segment. Data recently collected by our research group suggest that, at least in the dialect of Brazilian Portuguese (BP) spoken in Rio Grande do Sul, prevoicing in /b/, /d/, /g/ is not categorical, and instances of production of these showing VOT Zero patterns have been found. This fact gives support to our argument that prevoicing may

not be robust enough to enable the distinction between voiced and voiceless segments in that dialect. Investigations on the identification of BP /b/, /d/, /g/ without prevoicing are fundamental for a more consistent discussion concerning the effective role played by this acoustic cue in the functional sonority distinctions in BP (ALVEZ; ZIMMER, 2012). Thus, as VOT is not the sole acoustic cue employed in the identification of these segments as voiced, additional cues, such as burst intensity, could account for our findings, as voiced segments, both in English and in Brazilian Portuguese, are produced with a weaker burst intensity (LISKER; ABRAMSON, 1964).

Table 1 – Results of the Identification⁵ test (12 questions for each VOT pattern).

<i>VOT patterns</i>	<i>Elementary Group</i>		<i>Advanced Group</i>	
	<i>Voiceless</i>	<i>Voiced</i>	<i>Voiceless</i>	<i>Voiced</i>
<i>Negative</i>	0% 0/288	99,31% 286/288	3,33% 4/120	96,67% 116/120
<i>Natural Zero</i>	27,43% 79/288	69,1% 199/288	25% 30/120	71,67% 86/120
<i>Manipulated Zero</i>	57,29% 165/288	39,93% 115/288	76,67% 92/120	16,67% 20/120
<i>Positive</i>	91,67% 264/288	2,78% 8/288	98,33% 118/120	0% 0/120

Source: Made by the author.

As to the identification of Positive VOT patterns (aspiration), the results also confirm what was expected: in both proficiency levels, the identification of such segments as voiceless seems to be categorical (elementary – 91,67%, m=12; advanced – 98,33%, m=12). Mann-Whitney tests did not show any significant differences between the groups in the identification of the segments as voiced (U=76,500; p=,051) or voiceless (U=100,00; p=,175). Such findings are in line with what has been observed in previous studies (ALVES; SCHWARTZHAUPT; BARATZ, 2011; ALVES; ZIMMER, 2012), which suggest that aspirated segments in English are easily identified as voiceless by learners. Based on these findings, Hypothesis 1 is confirmed: regardless of which cue is taken as the main cue, there were no significant differences between the

⁵ In this table, we only present the rates of those answers that consisted of accurate choices concerning place of articulation (i.e., learners may hear a bilabial stop and identify it as [b] or [p], but not as [t], [d], [k] or [g]). This explains why the sum of voiceless and voiced responses does not correspond to 100% of the answers.

two proficiency levels in what regards the identification of voicing in Negative and in Positive⁶ VOT patterns.

As to the Natural Zero VOT pattern, we found that the participants in both groups of proficiency prefer to identify it as voiced (elementary: 69,71%, $m=8,00$; advanced: 71,67, $m=10,00$). Mann-Whitney tests did not show significant differences between the groups in the identification of the segments as voiced ($U=111,00$; $p=,727$) or voiceless ($U=103,00$; $p=,510$). The results corroborate our hypothesis, and additional remarks must be made concerning the motivation for our initial hypothesis. As we had predicted, both groups would identify the Natural Zero VOT pattern as voiced by following distinct paths: the elementary learners would answer based on cues other than VOT (such as burst intensity), whereas advanced participants – who already follow VOT as their main cue – would use the L2 patterns to accomplish this task. In order to assess whether the responses provided by the participants were actually based on the possibilities raised above, the data regarding the Manipulated Zero pattern will be really clarifying.

The data on Table 1 show that, when faced with this artificial pattern, the learners show a higher degree of difficulty to identify the segments as voiced or voiceless. This difficulty is more obvious in the elementary level, in which 57,29% of the tokens ($m=6,00$) are identified as voiceless, while 39,95% of the **tokens** ($m=5,00$) are identified as voiced. The preference becomes a little clearer among advanced participants, who identify 76,67% of the tokens as voiceless ($m=9,00$). According to Hypothesis 3, advanced learners would follow the VOT cue and would therefore identify such segments as voiced, unlike elementary learners, who would identify segments containing Zero Manipulated VOT as voiceless due to the fact that they were still relying on cues such as burst intensity. That did not occur; on the contrary, advanced learners present slightly higher rates of preference to identify such segments as /p/, /t/, /k/. Thus, our third hypothesis was not corroborated. In addition, our data suggest that other cues other than VOT may be playing a role in the identification of voiceless and voiced stops. It is worth mentioning that the same experiment, when carried out with native speakers of American English, showed high levels of identification of the Zero Manipulated patterns as voiceless, which confirms the tendency the native speakers show to use the absence/presence of aspiration to identify voicing, even when they listen to segments of a hybrid nature (SCHWARTZHAUPT; ALVES; FONTES, 2013).

Now we turn to the results regarding the Natural Zero pattern, for which we had not predicted a significant difference between the groups. In fact, if we had

⁶ In the discussion about the identification results of the next patterns, we will resume this issue to suggest that VOT is not effectively followed by either of the groups.

taken into consideration the fact that this pattern is present in the consonants /p/, /t/ and /k/ of the learners' L1 system⁷, we could have predicted a preference for the identification of the consonants showing the Natural Zero pattern as voiceless, but that was not observed in the data, either. The identification of that pattern as voiced serves as an argument for the possibility that the learners are being guided by cues other than VOT. Furthermore, the data concerning Manipulated Zero VOT are highly suggestive that learners do not discriminate voicing based on the presence of aspiration, which once again suggests that VOT is not the main cue employed by learners, and that neither the L1, nor the L2 VOT pattern, works as the main criterion for discrimination between English /p/, /t/, /k/ and /b/, /d/, /g/. We can also conclude that neither the consonants with Negative VOT patterns nor the ones presenting Positive VOT patterns are identified as voiced and voiceless via the VOT cue, regardless of the learners' level of proficiency.

Further evidence for the conclusion that VOT is not the main cue the two proficiency groups pay attention to can be found in the next section.

Discrimination

In this section, we present the results regarding the Discrimination test, laid out on Table 02. As to the contrast Negative VOT *versus* Positive VOT, the data on Table 02 report high levels of discrimination in the elementary group (Accuracy=76,74%, m=9,00; No Discrimination=9,72%, m=.50) and in the advanced group (Accuracy=93,33%, m=11,50 ; No Discrimination=2,5%; m=0). Wilcoxon tests showed no significant differences between the groups (Accuracy: U=76,500, p=.087; No Discrimination: U=79,000, p=0,082). In fact, as we had predicted, the VOT negative and positive patterns were highly discriminated.

⁷ As we have already mentioned, at least in a smaller degree, this pattern can also be found in some productions of /b/, /d/ and /g/, in the dialect spoken in Rio Grande do Sul (gaucho dialect), which is considered as additional evidence for the fact that not even in the learners' L1 is VOT the main cue.

Table 2 – Results of the Discrimination Test (12 questions for each VOT pattern)⁸

VOT patterns	Elementary		Advanced	
	<i>Accurate</i>	<i>No Discrimination</i>	<i>Accurate</i>	<i>No Discrimination</i>
Negative x Artificial Zero	45,49% 131/288	29,51% 85/288	64,17% 77/120	25% 30/120
Negative x Positive	76,74% 221/288	9,72% 28/288	93,33% 112/120	2,5% 3/120
Artificial Zero x Positive	34,03% 98/288	45,49% 131/288	38,33% 46/120	50,83% 61/120

Source: Made by the author.

In what regards the contrast between Negative VOT vs. Manipulated Zero VOT, we expected to find a significant difference between the groups, for we hypothesized that, as the advanced learners followed VOT as the main cue, they would not perceive the difference between those two patterns, whereas elementary learners, when guided by cues such as burst intensity, would be able to discriminate between Negative VOT (with a weak burst) and Manipulated Zero (which requires a strong burst). However, results in Table 02 seem to contradict the fifth hypothesis, although the percentages of “No discrimination” answers are lower for advanced learners than those of elementary participants.

Both groups (with an advantage to the advanced group – 64,17%, m=8,00) tend to judge the two patterns as distinct from each other. This can be taken as an additional argument for the proposal that VOT is not the main cue for the distinction between voiced and voiceless plosives.

Finally, the data concerning the contrast between Manipulated Zero VOT and Positive VOT suggest that both elementary learners (Accuracy= 34,03%, m= 4,00 ; No Discrimination= 45,49%, m= 5,00) and the advanced participants (Accuracy= 38,33%, m = 5,50; No Discrimination= 50,83%, m= 6,50) felt highly insecure in their answers. In addition, both groups showed a tendency to consider such patterns as the same. Mann-Whitney tests showed no significant statistical differences between the groups (Accuracy: U=104,500, p=,555; No

⁸ In this table, ‘accuracy’ corresponds to the rates of correct answers provided by the participants in the AxB task (e.g. in [p]at, [p]at and [b]at, learners should say that X is the same as A, not B); ‘No Discrimination’ refers to the occurrences in which learners did not discriminate X from A or B (all equal). In this table, we do not present the rates of incorrect discrimination answers, which correspond to those instances in which participants did not consider the three members of the triad to be equal, but chose A instead of the correct option B (or vice-versa) in the AxB task.

Discrimination: $U=105,000$, $p=,569$). Hypothesis 6 was thus refuted, for learners of both proficiency levels tended to consider both patterns as the same.

In sum, the results in the Discrimination test are in accordance with the ones found in the Identification test, and they provide further evidence that VOT is not the main cue followed by elementary and advanced learners of English, for whom we hypothesized that, as is true of native speakers (SCHWARTZHAUPT; ALVES; FONTES, 2013), the presence/absence of aspiration would correspond to the main element to rely on when answering the tests. In the next section, we detail the implications of such results in the production data of plosive segments.

Production tests

The results concerning the word production test in Brazilian Portuguese (BP) are presented on Table 03, where the reported values for VOT are surprisingly high for BP. Even though previous studies (GEWEHR-BORELLA; ZIMMER; ALVES, 2011; VEIGA-FRANÇA, 2011; ALVES; SCHWARTZHAUPT; BARATZ, 2011; SCHWARTZHAUPT, 2012) have already pointed out a possible ‘semi-aspiration’ in the velar plosive, the mean values of 24,33 ms ($m = 24,00$) found in the elementary group, and of 33,00 ms ($m = 33,5$) in the advanced group for the production of the bilabial /p/ are really surprising.

Table 3 – Results of the word production test in BP

Consonant	Elementary (24)		Advanced (10)	
	<i>Tokens</i>	<i>Mean (SD)</i>	<i>Tokens</i>	<i>Mean (SD)</i>
/ p /	133	24,13 (5,44)	51	33,00 (7,18)
/ k /	139	55,71 (20,09)	57	51,7 (23,09)

Source: Made by the author.

When faced with such results, we cannot deny the possibility that – mainly among advanced learners – the participants’ L1 speech is being affected by the transfer of L2 VOT patterns (SANCIER; FOWLER, 1997; COHEN, 2004), so that the VOT intervals in BP displayed on Table 3 do not reflect the durations effectively produced by monolingual speakers of the gaucho dialect spoken in the South of Brazil. This possibility does not seem hard to conceive of in a dynamic perspective of language acquisition, according to which any change in one of the linguistic systems of the speaker can cause substantial changes in all the other language systems. Therefore, this can be seen as reciprocal influences not only from the L1 into the L2, but also from the L2 into the L1, or any other additional languages

spoken by the speaker, such as the L3, and so on (DE BOT; LOWIE; VERSPOOR, 2007; BECKNER et al., 2009; BLANK, 2013).

Table 4 displays the results of the word production test in English, in which higher values are found in the L2 than in the L1 for /p/ (elementary: 45,04, m=45,50; advanced: 34,4, m=31,5) and /k/ (elementary: 68,87, m=67,50; advanced: 79,8, m=82,5). Therefore, the seventh hypothesis is only partially corroborated. According to that hypothesis, there would be a significant difference between the VOT values found in the L2 – English – and in the L1 – BP – within each group. In fact, Wilcoxon tests found a significant difference in the productions in the two languages both in the elementary group ($Z=-2,702$, $p=,007$) and in the advanced group ($Z=-2,193$, $p=,028$) in the productions of /k/. This result was not fully replicated for /p/: a significant difference between the production of that consonant in the two languages was found only within the elementary group (/p/: $Z=-4,03$, $p=,000$), probably because the VOT values for this consonant were already high in the L1 of the advanced group.

Table 4 – Results of the word production test in English

Consonant	Elementary (24)		Advanced (10)	
	<i>Tokens</i>	<i>Mean (SD)</i>	<i>Tokens</i>	<i>Mean (SD)</i>
/ p /	136	45,04 (16,71)	53	34,40 (15,25)
/ t /	131	59,04 (13,56)	57	58,40 (17,52)
/ k /	131	68,87 (21,42)	60	79,80 (14,95)

Source: Made by the author.

Based on the data shown above, we could enquire about the significant differences found. As VOT is not the main cue used by learners in perception, how could we explain the fact that the L2 productions present longer aspiration intervals than the ones found in the L1? In order to answer this question, we should first consider the fact that the participants do not follow VOT as their main acoustic cue, as we have discussed previously. That does not imply that the learners cannot perceive and recognize VOT as an aspect of English phonology. In other words, it is plausible that aspiration is already perceived as an allophonic detail which is necessary for “unaccented” speech production, but which is not the primary cue for the phonological distinction between voiced and voiceless segments. Following this thread of thought, we might consider that learners could, up to a certain extent, produce aspiration as a detail to make them reduce their degree of accent, but not as a necessary cue to establish phonological distinctions. Such a functional distinction would be instantiated

from the conjoint action of multiple acoustic cues, supporting a dynamic view of language acquisition.

Besides, it is also necessary to take into consideration the fact that even though the durations of L2 aspiration seem to be longer than the time intervals found in the L1 data, the L2 productions have not yet reached the native VOT values. According to Cho and Ladefoged (1999), standard VOT patterns in English are 55 ms for /p/, 70 ms for /t/ and 80 ms for /k/. Therefore, although the participants no longer produce the VOT values found in their L1, the duration of aspiration intervals produced in their L2 correspond to an intermediate value between the ones they present in their L1 and their L2. This serves as an additional argument for the fact that VOT seems to have an allophonic character, so that it does not take priority over other acoustic cues in voice distinctions in L2 English by Brazilians.

This intermediate value for VOT, already described in the Brazilian Portuguese-English literature (ALVES; SCHWARTZHAUPT; BARATZ, 2011), provided motivation for the outline of our eighth and last hypothesis, according to which there would be no significant differences between the VOT values produced by elementary and advanced students of English. When we formulated that hypothesis, we had departed not only from the aforementioned studies, but also from the dynamic approach to language acquisition that guides this study. Hence, although learners can identify and discriminate the target language sounds, the production of segments in the target language requires that learners abandon the “timing” of the L1 articulations, so that they can orchestrate their articulators according to the L2 tempo and rhythm (ZIMMER; ALVES, 2012).

Therefore, it seemed plausible that, despite the fact that proficient learners already followed VOT as a main cue, they still did not seem to have acquired the temporality of VOT in the L2, as the perception of VOT patterns would be necessary, but not sufficient, for the production of aspiration.

Mann-Whitney tests were run and did not show any significant difference between the two levels of proficiency for /p/ ($U=74,000$, $p=.082$), /t/ ($U=114,500$, $p=.835$) or /k/ ($U=84,000$, $p=.173$). The discussion on the perception data provided in the last section, however, enables us to conceive another explanation for the data: as VOT is not being used as main cue for the perception and discrimination of voicing, perhaps it is reflected in the production data as well. In a nutshell, the little difference found between the VOT patterns produced by participants in the two groups may not be related exclusively to the difficulty to acquire timing distinctions in the L2, but are probably due to the fact that, both in terms of perception and production, the voice distinction is being instantiated by acoustic cues other than VOT, so that perception and production are highly intertwined for the learners in the two groups.

Learners may be assuming that the production, or the partial production, of long-lagged VOT in English does not necessarily play a detrimental role in intelligibility. This can be reinforced in a context of communication among Brazilians, who share the same main acoustic cues in Portuguese, and thus could go without VOT in the distinction between /p/, /t/, /k/ e /b/, /d/, /g/ in English. A real need to employ cues such as VOT would only be felt necessary when Brazilians had to communicate with native speakers of English, who really make use of VOT to functionally distinguish voicing among such segments. Explicit instruction on how to use VOT can also contribute to elucidate the phenomenon (MOTTA; ALVES, 2013), whose effects need to be more widely investigated.

In sum, the results of the production test indicate that, in the target language, the length of VOT produced by learners from both proficiency groups are higher than those produced in their L1. However, this length is not yet as high as the ones found in the native speech. That led us to suggest that – although learners already recognize the need to produce a longer VOT to reduce accent, so that such acoustic cue acquires an “allophonic” character – Voice Onset Time is still not the main cue to voicing distinction among plosives.

Final considerations

In this study, we departed from the assumption that elementary learners of English would not follow VOT as their main acoustic cue, as additional cues seem to play a more decisive role in distinguishing voicing in Brazilian Portuguese. All the working hypotheses relied on the assumption that there would be a difference between elementary and advanced learners. Hence, VOT would not be followed by elementary learners of English (L2), whereas advanced learners, on the other hand, would follow the presence/absence of aspiration to distinguish voiced from voiceless English stops both in perception and in production.

However, the findings of this study reveal that not even participants with an advanced level of English proficiency used VOT as their main cue to distinguish between /p, t, k/ and /b, d, g/. Regardless of the learners’ degree of proficiency, it seems that without formal instruction the learners will continue to use the acoustic cues which are relevant in the distinction of voicing in their L1, which was found both in the perception and in production data.

It is necessary to make clear that this study aimed to find whether VOT could be characterized as the cue which would suffice to distinguish voicing among Brazilian learners of English. As we found out that other acoustic cues were acting as relevant cues, it remains to be investigated which acoustic(s) cue(s)

is/are being primarily followed by Brazilian learners in these voice distinctions. As already mentioned, studies by Sundara (2005), Oh (2011) and Kong, Beckman e Edwards (2012) have raised the possibility that both burst intensity and F₀ seem to play a decisive role in this sense. What remains to be determined, in further studies, is which of these two acoustic cues plays a more decisive role in Brazilian Portuguese and, as a likely consequence, in Brazilian Portuguese-English interlanguage. Even though we do not focus on this issue in the present article, we can say that the employment of burst intensity as the main acoustic cue by BP learners seems to be a very likely possibility. For now, it suffices to say that, similarly to what has already been found in other language systems, VOT by itself does not seem to be a sufficient cue for voicing distinctions by Brazilian learners of English.

The findings presented here can be accounted for by the gestural approach to phonology, according to which a common currency for perception and production could be the phonological gesture (GOLDSTEIN; FOWLER, 2003). A possible explanation for the fact that the simple exposure to VOT did not support the acquisition of the L2 gestures involved in aspiration is the difficulty, reported in various studies on L1 acquisition, to acquire contrasts generated from the action of less visible organs (GOLDSTEIN; FOWLER, 2003). In the case reported in this study, the distinctive role played by the larynx has not possibly been learned because the learners have proceduralized their phonological distinctions from their L1 gestural constellations and timing (duration or gestural phasing). The gestural orchestration resulting in aspiration, for Brazilian learners, may be playing a merely allophonic role, such that the action of the larynx is not playing a distinctive role in the learners' gestural score. It may be possible that – in the case of the L2 – the abstraction of the movements required to reach an L2 articulatory target is suffering the influence of the L1 gestural abstraction, which is already automated in the procedural memory. Thus, in the case of VOT, the learners may well interpret the longer duration generated by the opened constriction of the larynx as a non-distinctive feature.

The results of this study are in consonance with several others previously carried out in our research group. First, our production data suggest that the effects of the L2 over the production of the native language need to be more widely investigated. Moreover, as VOT does not take priority over other cues in the voicing distinctions among English stops, we find it necessary to highlight the beneficial roles of explicit instruction (MOTTA; ALVES, 2013) and perceptual training (ALVES, 2012) in the perception and production of English /p/, /t/, /k/ by Brazilian learners. In that direction, we can mention another contribution from the area of Interlanguage Phonetics-Phonology: to provide insight into Applied Linguistics for the teaching of Foreign Languages (ALVES, 2012). The findings we garnered in this study will certainly pave the way for a broad agenda

of investigation on the role of acoustic cues that are effectively employed by Brazilian learners in establishing functional distinctions both in the L1 and in the interlanguage systems.

ALVES, U. K.; ZIMMER, M. C. Percepção e produção dos padrões de VOT do inglês por aprendizes brasileiros: o papel de múltiplas pistas acústicas sob uma perspectiva dinâmica. **Alfa**, São Paulo, v.59, n.1, p.155-175, 2015.

- *RESUMO: Neste trabalho, a partir de uma concepção dinâmica de aquisição fonético-fonológica de L2, investigamos a percepção (identificação e discriminação) e a produção dos padrões de Voice Onset Time (VOT) das plosivas iniciais do inglês por 32 aprendizes do Sul do Brasil. Partimos da premissa de que, sobretudo entre aprendizes com nível básico de proficiência, o VOT não se mostra como pista acústica prioritária para as distinções funcionais de sonoridade. Os resultados dos testes de percepção mostram que, independentemente do nível de proficiência dos aprendizes (básico ou avançado), o VOT tomado unicamente não se faz suficiente para a distinção entre /p/, /t/, /k/ e /b/, /d/, /g/. Tais resultados, que exercem influência sobre os dados de produção, corroboram uma visão dinâmica de aquisição de L2, a partir da qual múltiplas pistas acústicas agem em conjunto nas distinções entre sons, cabendo ao aprendiz saber selecionar aquelas pistas com caráter mais primordial no sistema a ser adquirido.*
- *PALAVRAS-CHAVE: Percepção de L2. Produção de L2. Língua inglesa. VOT. Pistas Acústicas.*

REFERENCES

ABRAMSON, A.; LISKER, L. Voice-Timing Perception in Spanish Word-Initial Stops. **Journal of Phonetics**, London, n.1, p.1-8, 1973.

ALVES, U. K. Pesquisa em aquisição de L2 e ensino: um relacionamento possível (mas não necessariamente garantido). In: LEFFA, V.; ERNST, A. (Org.). **Linguagens: Metodologias de Ensino e Pesquisa**. Pelotas: EDUCAT, 2012. P.233-252.

ALVES, U. K.; SCHWARTZHAUPT, B. M.; BARATZ, A. H. Percepção e produção dos padrões de VOT do inglês (L2) por aprendizes brasileiros. In: FERREIRA-GONÇALVES, G.; BRUM-DE-PAULA, M. R.; KESKE-SOARES, M. **Estudos em Aquisição Fonológica**. Pelotas: Ed. da UFPel, 2011. p.3-4.

ALVES, U. K.; ZIMMER, U. K. The Dynamics of Perception and Production of VOT Patterns in English by Brazilian Learners. In: MELLO, E.; PETTORINO, M.; RASO, T. (Ed.). **Proceedings of the VIIth GSCP International Conference: Speech and Corpora**. Firenze: Firenze University Press, 2012, p.223-227.

ANTONIOU, M. et al. Inter-Language Interference in VOT Production by L2-Dominant Bilinguals: Asymmetries in Phonetic Code-Switching. **Journal of Phonetics**, London, v.39, p.558-570, 2011.

AUDACITY. Software livre. Disponível em: <www.audacity.sourceforge.net>. Acesso em: 20 jan. 2015.

BECKNER, C. et al. Language is a Complex Adaptive System: Position Paper. **Language Learning**, Ann Arbor, v.59, suppl.1, p.1-26, 2009.

BEST, C. T.; TYLER, M. D. Nonnative and Second-Language Speech Perception: Commonalities and Complementarities. In: BOHN, O.-S.; MUNRO, M. J. **Language Experience in Second Language Speech Learning**: Studies in Honor of James Emil Flege. Amsterdam: John Benjamins, 2007. p.13-34.

BLANK, C. A. **A influência grafo-fônico-fonológica na produção oral e no processamento de priming em multilíngues**: uma perspectiva dinâmica. 2013. 226f. Tese (Doutorado em Letras) – Programa de Pós-Graduação em Letras, Universidade Católica de Pelotas, Pelotas, 2013.

BOERSMA, P.; WEENINK, D. **Praat**: Doing Phonetics by Computer. Version 5.3.48. 2013. Disponível em <www.praat.org>. Acesso em: 20 jan. 2015.

BROWMAN, C. P.; GOLDSTEIN, L. Competing Constraints on Intergestural Coordination and Self-Organization of Phonological Structures. **Bulletin de la Communication Parlee**, Cedex, n.5, p. 25-34, 2000.

_____. Dynamics and Articulatory Phonology. In: VAN GELDER, T.; PORT, R. F. (Ed.). **Mind as motion**. Cambridge: MIT Press, 1993, p.51-62.

_____. Articulatory Phonology: An overview. **Phonetica**, Basel, n.49, p.155-180, 1992.

CHO, T.; LADEFOGED, P. Variation and Universals in VOT: Evidence from 18 Languages. **Journal of Phonetics**, London, n.27, p.207-229, 1999.

COHEN, G. **The VOT Dimension**: a Bidirectional Experiment with English and Brazilian-Portuguese Stops. 2004. 96f. Dissertação (Mestrado em Língua Inglesa) – Universidade Federal de Santa Catarina, Florianópolis, 2004.

DE BOT, K.; LOWIE, W.; VERSPOOR, M. A Dynamic Systems Theory approach to second language acquisition. **Bilingualism: Language & Cognition**, Cambridge, v.10, n.1, p.7-21, 2007.

DOCHERTY, G. J. **The Timing of Voicing in British English Obstruents**. Berlin; New York: Foris Publications, 1992.

GEWEHR-BORELLA, S.; ZIMMER, M. C.; ALVES, U. K. Transferências grafo-fônico-fonológicas: uma análise de dados de crianças monolíngues (Português) e bilíngues (Hunrückisch-Português). **Gragoatá**, Niterói, v.30, p.201-219, 2011.

GOLDSTEIN, L.; FOWLER, C. A. Articulatory Phonology: a Phonology for Public Language Use. In: MEYER, A. S., SCHILLER, N. O. (Ed.). **Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities**. Berlin: Mouton de Gruyter, 2003. p.159-207.

KAMIANECKY, F. **A palatalização das oclusivas dentais /t/ e /d/ nas comunidades de Porto Alegre e Florianópolis: uma análise quantitativa**. 2002. 114f. Dissertação (Mestrado em Letras) – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2002.

KONG, E. J.; BECKMAN, M. E.; EDWARDS, J. Voice Onset Time is Necessary but not Always Sufficient to Describe Acquisition of Voiced Stops: The Cases of Greek and Japanese. **Journal of Phonetics**, London, v.40, p.725-744, 2012.

LISKER, L.; ABRAMSON, A. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. **Word**, New York, n.20, p.384-422, 1964.

MOTTA, C.; ALVES, U. K. Percepção de padrões de Voice Onset Time por aprendizes brasileiros de inglês: dados de discriminação e identificação. In: JORNADA DE JÓVENES LINGUISTAS, 2., 2013, Buenos Aires. **Resumos...** Buenos Aires, 2013. v.1. p.140.

OH, E. Effects of speaker gender on voice onset time in Korean stops. **Journal of Phonetics**, London, n.39, p.59-67, 2011.

POLLITT, A. **The meaning of OOPT Scores**. 2007. Disponível em: <https://www.oxfordenglishtesting.com/uploadedFiles/Buy_tests/oopt_meaning.pdf>. Acesso em: 20 jan. 2015.

PURPURA, J. **The Oxford Online Placement Test: What does it Measure and How?**. 2007. Disponível em: <http://www.oxfordenglishtesting.com/uploadedfiles/6_New_Look_and_Feel/Content/oopt_measure.pdf>. Acesso em: 26 ago. 2013.

SANCIER, M. L.; FOWLER, C. A. Gestural drift in a bilingual speaker of Brazilian Portuguese and English. **Journal of Phonetics**, London, n.25, p.421-436, 1997.

SCHWARTZHAUPT, B. **Factors influencing Voice Onset Time: analyzing Brazilian Portuguese, English and Interlanguage data**. 2012. 65f. Trabalho de Conclusão de Curso (Graduação em Letras) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

SCHWARTZHAUPT, B.; ALVES, U. K.; FONTES, A. B. L.; O VOT como pista suficiente para a distinção surdo/sonoro: dados de falantes do inglês americano. In: BRUM DE PAULA, M. (Org.). **4º Seminário de Aquisição Fonológica: Resumos e Programação**. Pelotas: Ed. da UFPel, 2013. p.26.

SIMON, E. **Voicing in Contrast**: Acquiring a Second Language Laryngeal System. Ghent, Belgium: Academia Press, 2010.

SUNDARA, M. Acoustic phonetics of coronal stops: a cross-language study of Canadian English and Canadian French. **Journal of the Acoustical Society of America**, New York, n.118, p.1026-1037, 2005.

VEIGA-FRANÇA, K. V. **A aquisição da aspiração das plosivas surdas do inglês por falantes do Português Brasileiro**: Implicações teóricas decorrentes de duas formas de descrição dos dados. 2011. 100f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras, Universidade Católica de Pelotas, Pelotas, 2011.

YAVAS, M.; WILDERMUTH, R. The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers. **IRAL**, Heidelberg, n.44, p.251-263, 2006.

ZIMMER, M. C.; ALVES, U. K. Uma visão dinâmica da produção da fala em L2: o caso da Dessonorização Terminal. **Revista da Abralín**, Brasília, v.11, n.1, p.221-272, 2012.

Received January 2014

Accepted April 2014

