Revista de Estudos da Linguagem, Belo Horizonte, v. 23, n.2, p. 311-334, 2015

The role of L1 knowledge on L2 speech perception: investigating how native speakers and Brazilian learners categorize different VOT patterns in English

O papel do conhecimento da L1 na percepção da fala em L2: investigando como falantes nativos e aprendizes brasileiros categorizam diferentes padrões de VOT em inglês

Bruno Schwartzhaupt
Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil.
schwartzhaupt.b@gmail.com

Ubiratã Kickhöfel Alves
Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil.
ukalves@gmail.com

Ana Beatriz Arêas da Luz Fontes
Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brasil.
ana.fontes@ufrgs.br

Abstract: The present study aimed to investigate how different *Voice Onset Time* (VOT) patterns are categorized by native speakers of American English and Brazilian Learners of English. American English and Brazilian Portuguese diverge as to the voicing patterns of plosive

eISSN: 2237-2083
DOI: 10.17851/2237.2083.23.2.311-334

consonants, for the VOT cue plays different roles in the distinction between voiced and voiceless consonant categories in each system. This study contrasted four VOT patterns (*Negative VOT*, *Zero VOT*, *Positive VOT* and a manipulated pattern, named *Artificial Zero VOT*) in two perceptual tasks (AxB discrimination and identification tests), and verified how the two groups of participants categorized these patterns. Results reinforce the idea that speech perception is multimodal and, therefore, the action of multiple cues must be taken into account when we consider phonetic-phonological processes.
Keywords: Voice Onset Time; Speech Perception; Discrimination; Identification.

Resumo: O presente estudo buscou investigar como diferentes padrões de *Voice Onset Time* (VOT) são categorizados por falantes nativos de Inglês Americano e aprendizes brasileiros de Inglês. O Inglês Americano e o Português Brasileiro divergem quanto ao padrão de vozeamento das consoantes plosivas, uma vez que a pista VOT desempenha diferentes papéis na formação de categorias de consoantes vozeadas e desvozeadas em cada sistema. Este estudo contrastou quatro padrões de VOT (*VOT Negativo, VOT Zero, VOT Positivo* e um padrão manipulado, nomeado *VOT Zero Artificial*) em duas tarefas perceptuais (testes de discriminação AxB e identificação), e verificou como os dois grupos de participantes categorizaram esses padrões. Os resultados corroboram a ideia de que a percepção da fala é multimodal e, portanto, a ação de múltiplas pistas acústicas deve ser levada em consideração quando abordamos processos fonético-fonológicos.
Palavras-chave: *Voice Onset Time*; Percepção da Fala; Discriminação; Identificação.

Recebido em 26 de agosto de 2014.
Aprovado em 27 de janeiro de 2015.

1 Introduction

The present study aimed to contribute to the understanding of how acoustic cues influence L2[1] speech perception in accordance with

[1]We consider it irrelevant to make a distinction between the terms Second Language

learners' L1 knowledge. In order to pursue this goal, we looked into the perception of different VOT patterns, in word-initial position in English, by both native speakers of American English and Brazilian L2 learners of English, as represented in the data from two different perceptual tasks.

Many studies have directed their attention to the acquisition of English aspirated consonants by Brazilians over the past few years (COHEN, 2004; ALVES, 2007; REIS E NOBRE-OLIVEIRA, 2008; FRANÇA, 2011; SCHWARTZHAUPT, 2012; PRESTES, 2013). The investigation of this phenomenon is justified by the fact that, in word-initial position, aspiration corresponds to a perceptually distinctive aspect in the production of stop consonants in English, accounting for the distinction between voiceless and voiced segments. In Brazilian Portuguese, however, stop consonants are not aspirated and aspiration does not play this distinctive role; thus, Brazilian learners face difficulties in producing this L2 aspect.

Since this phonetic-phonological aspect is perceptually distinctive in English but not in Brazilian Portuguese (BP), it can be hypothesized that there are different statuses given to aspiration as an acoustic cue in the two language systems. Even so, studies investigating perception of English stop consonants by Brazilians have suggested that discrimination between voiceless aspirated and voiced segments with *Zero VOT* or *Negative VOT* may be categorical (ALVES *et al.*, 2011). Nonetheless, as we intend to demonstrate from the tests conducted in this study, which include both natural and manipulated stimuli, speech perception is a process in which there is an interaction of multiple cues. Therefore, aspiration alone should not be regarded as the only cue in the distinction between voiced and voiceless plosives, and this phonetic-phonological aspect is expected to interact differently, and plays a different role with other cues across linguistic systems.

We begin this paper with a background on the theoretical assumptions underlying the present study, in which, among others, the concepts of Voice Onset Time and L1-L2 Transfer are presented. Next, we describe the methodology of this study, with information on

---

and Foreign Language, in order to pursue the aim of the present study. We also find it impossible to restrict the context in which this study was conducted to any of the terms alone. Therefore, in the reading of this paper, Second Language can be interpreted as a synonym of Foreign Language.

participants, target words selection, stimuli manipulation and recording, the two perceptual tasks used in the study, and the hypotheses established beforehand. The following section describes the results and the statistical analyses conducted with the data obtained from the perceptual tasks. Finally, in the last section, the results are discussed.

2 Background

2.1 Voicing Patterns in English and Brazilian Portuguese Plosive Consonants: The Voice Onset Time Distinction

The acoustic cue of *Voice Onset Time* (VOT) refers to the period of time between the stop consonant release and the vibration of the vocal folds of the vowel following this consonant. Three main VOT patterns can be found in the languages of the world (LISKER; ABRAMSON, 1964; COHEN, 2004; REIS; NOBRE-OLIVEIRA, 2008):

- ***Negative VOT*** (*pre-voicing*): in which vocal folds start vibrating before the stop consonant release, in an interval ranging from -125 ms to -75ms;

- ***Zero VOT***: in which the vibration of the vocal folds starts almost simultaneously to the plosive release, in an interval ranging from 0 ms to +35 ms;

- ***Positive VOT*** (*aspiration)*: in which a delay follows the plosive release, and vocal folds start vibrating after a 35 ms to 100 ms interval.

In accordance with the literature cited above, BP voiced stop consonants /b/, /d/ and /g/ are produced with *Negative VOT*, whereas voiceless plosives are produced with *Zero VOT*, with mean values of approximately 12 ms for /p/, 18 ms for /t/ and 38 ms for /k/. Nevertheless, recent studies investigating the production of stop segments in the Southern region of Brazil have shown higher VOT values, especially for the velar stop /k/ - with values ranging from 46.55 ms to 63.90 ms (REIS; NOBRE-OLIVEIRA, 2008; GEWEHR-BORELLA, 2010; FRANÇA, 2011; SCHWARTZHAUPT, 2012). As suggested by Schwartzhaupt (2012), such findings might indicate the existence of partial aspiration

of voiceless /k/ in Southern Brazilian Portuguese – in that case, native-like VOT production of /k/ would be facilitated for Southern Brazilian Portuguese speakers learning English as an L2.

In regard to the production of word-initial stop consonants in English, voiced plosives tend to be produced with *Zero VOT* (although productions with *Negative VOT* may also be found). Voiceless stops, on the other hand, are produced with *Positive VOT*: [p^h] with mean 55 ms, [t^h] with mean 70ms, and [k^h] with average 80 ms VOT. Considering the existing divergences between BP and English voicing patterns in word-initial plosive segments, the two languages belong to distinct groups concerning VOT patterns.

It is essential to notice, however, that VOT values are not absolute. VOT cannot be considered to be an isolated entity within a linguistic system. Several factors, which deserve consideration, might influence this phonetic-phonological aspect. Some studies show evidence of variation in VOT values as to the quality of the subsequent vowel (YAVAS, 2008; FRANÇA, 2011; SCHWARTZHAUPT, 2012; PRESTES, 2013) – essentially, it has been argued that a higher subsequent vowel causes VOT to be longer. The number of syllables of the target word has been said to affect VOT as well (YAVAS, 2008; FRANÇA, 2011). Other authors have argued that factors such as syllable stress, prosody, and speech rate should also be taken into account (COHEN, 2004; REIS; NOBRE-OLIVEIRA, 2008; ALVES, 2010).

2.2 L2 Phonetic-Phonological Acquisition as a Dynamic and Multimodal process

According to the emergentist view of language acquisition, both language and learner are regarded as *dynamic systems* (DE BOT *et al.,* 2007; ELLIS, 2011). Among other important characteristics, a dynamic system is composed of multiple agents – which interact and change one another –, it is also adaptive, and it is always evolving. In order to conceive this view, we first need to look at language as an ever-changing system, and bear in mind that such a constant change is a natural consequence of its use: individuals have their own language variety, and once they are inserted in a community, they interact, and thus change (and are changed by) the language of this community.

Under these circumstances, the system of an L2 learner is one that is bound to be changed with use – therefore, linguistic *input* is rich, and it plays a fundamental role in language acquisition. The input presents constraints and regularities; factors such as its frequency and saliency help shape the learner´s developing language system. By interacting with and using language, learners extract patterns of that system, these patterns *emerge* from communication, and so does the learner's awareness about them (ZIMMER; SILVEIRA; ALVES, 2009). However, it is important to notice that, in this perspective, cognitive functions are *domain-general* (BECKNER *et al.*, 2009): the same cognitive functions used to acquire any other type of knowledge (such as knowing how to drive or how to operate a computer) are also activated in first and second language acquisition.

More importantly than considering all these points, one should be aware that several different factors, linguistic and non-linguistic ones, have effects on the language acquisition process, and these factors cannot be considered in an isolated manner (DE BOT *et al*., 2007). It would be naïve, in this sense, for researchers to attribute problems in second language acquisition to factors such as learner's age, or L1 entrenched knowledge solely. Factors like these have an influence on language acquisition, but it is only through the interaction of these with a multitude of other factors that one may fully conceive the language acquisition process.

When we turn to one specific part of the second language acquisition process, L2 speech perception, we must consider that it occurs in a *multimodal* manner: multiple cues determine perception of segments, and these cues are not perceived by the learner in an isolated way (ZIMMER; SILVEIRA; ALVES, 2009; ZIMMER; ALVES, 2012, PEROZZO; ALVES, 2013). Moreover, certain cues – not only acoustic, but also visual, or of any other source in the environment – may not play the same relevant role in different L1 systems. In some cases, in order to acquire an L2 phonetic-phonological aspect, learners must perceive a cue which is not relevant in their L1 system, which makes this process even more difficult.

Furthermore, as explained by Zimmer and Alves (2008, 2010), oral L2 production also deals with the orchestration of multiple cues, which act together as a whole. The way cues interact in both production and perception of speech may, therefore, be distinct when we compare different linguistic systems. This process encompasses the physical and abstract levels, which go far beyond binary perspectives.

2.3 L1-L2 Phonetic-Phonological Transfer

The *Speech Learning Model* (FLEGE, 1995) and the *Perceptual Assimilation Model – L2* (BEST; TYLER, 2007) attempt to explain the segmental phonetic-phonological acquisition phenomenon of transfer between L1 and L2 knowledge. This investigation is fundamentally based on the model proposed by Best and Tyler (*Op. cit.*), for this is more compatible with the conception of phonetic-phonological acquisition underlying the present study (discussed in the previous subsection).

According to Best and Tyler (2007), the phonic elements of the learner's L1 and L2 systems interact in a common phonological space, and therefore the L2 learner tends not to perceive which articulatory features belong to their L1 and which belong to the L2 in question. This is to say that, once learners are faced with a "new" L2 sound, they might not extract information of their "new" articulatory gestures. The assumption is that, instead, learners assimilate the new sound to the L1 pattern, by following their L1 articulatory knowledge, thus considering it as an already existing sound from their L1 phonological space.

This premise allows us to explain the difficulties found in the acquisition of *Positive VOT* (aspiration) by Brazilian learners in the following manner: without formal instruction, these L2 learners tend not to perceive the differences between the BP and English voicing patterns in stop consonant production. Consequently, as *Positive VOT* (aspiration) is not a relevant acoustic cue in their L1 system, learners assimilate this pattern to the one from BP (with unaspirated plosive segments) and, therefore, do not produce the target aspiration. By conducting the present study, we expect to contribute with empirical evidence to support or refute this premise.

2.4 L1 – L2 Grapho-Phonic-Phonological Transfer

Another problem faced by L2 learners in the acquisition of the phonetic-phonological aspect in question is pointed out by Zimmer, Silveira and Alves (2009). This difficulty lies in the fact that BP and English, in spite of making use of the same alphabetical system, follow considerably different patterns concerning the relationship between orthography and sound. More specifically, the *grapho-phonic-*

phonological[2] relation in BP is rather transparent (orthography tends to represent pronunciation more straightforwardly), whereas this relationship in English is much more opaque. As a consequence of their entrenched L1 knowledge, learners tend to transfer the grapho-phonic-phonological patterns to their oral production in the L2 (ZIMMER; ALVES, 2006).

With regard to the acquisition of *positive VOT* by Brazilians, grapho-phonic-phonological transfer is a factor which reinforces the lack of assimilation of the target pattern. Considering that the graphemes 'p', 't' and 'k' correspond to *Zero VOT* stop consonants in the learner's L1 sound system, in his/her L2 oral production, this learner tends to associate the sounds represented by these graphemes in the target language (aspirated) to the ones they would represent in his/her mother tongue (unaspirated).

This is consistent with the multimodal conception of phonetic-phonological acquisition presented earlier in this paper: both the acoustic-articulatory and the orthographic stimuli (different sources of L2 input) can either work to oppose or to reinforce one another. Once learners assimilate L2 voicing patterns in accordance with their L1 knowledge, the orthographic stimulus may then be considered a source of reinforcement of the L1 pattern. If no assimilation occurred, it could be possible that both sources of input would be in competition, as the former would instantiate the L2 target forms, whereas the latter could be reinforcing the L1 pattern.

Therefore, when we consider the acquisition of English *Positive VOT* by Brazilian learners, we must observe that it might be impossible to consider the phonetic-phonological or the grapho-phonic-phonological transfer processes separately on theoretical grounds. Within a multimodal phonetic-phonological acquisition perspective, these factors (along with several others) make it more difficult for Brazilian learners to acquire the L2 voicing patterns.

---

[2]Zimmer and Alves (2006) describe this relation as *grapho-phonic-phonological* as an indication of the existence of a relationship between the orthographic form and the phones of the linguistic system in question. In this perspective, the traditional concepts of *phone* and *phoneme* correspond to a single reality. The authors (*Op. cit.*) believe that the use of this term is successful in expressing this relationship, for such a term, in this conception, does not refer to unities of a purely symbolic nature.

3 Method

3.1 Participants

Two groups of participants took part in this study. The first consisted of 20 adult native speakers of American English, all of whom were born in the state of Pennsylvania. The 20 subjects had acquired only English before reaching 6 years of age.

The second group was composed of 17 Brazilian speakers of English as an L2. All of them were born in the Brazilian state of Rio Grande do Sul, in the city of Porto Alegre and had only acquired Brazilian Portuguese before reaching 6 years of age. The learners were classified in the Oxford Online Placement Test[3] in the C1 and C2 levels of the Common European Framework of Reference for Languages (the two highest proficiency levels for this test), which are labeled "advanced" in the present study.

3.2 Selection of target words

Monosyllabic words initiated by the plosive consonants /p/, /b/, /t/, /d/, /k/ and /g/ were selected as targets. We also only included words whose initial plosive was followed by a high-front vowel /I/ - as pointed out by Yavas (2008), França (2011), Schwartzhaupt (2012) and Prestes (2013), aspiration is made clearer in this phonetic-phonological context, since high front vowels make VOT longer. Examples of those words included *peer, dip* and *kill.*

The number of words (*types*) was 12, which stands for 6 minimal pairs distinguished by the voicing of the initial plosive. Words were equally distributed in terms of place of articulation, as illustrated in *Box 1,* which follows:

[3]The Oxford Online Placement Test is a validated test taken online at www.oxfordenglishtesting.com. For more information, see Pollitt (2007) and Purpura (2007).

Box 1 – The 12 target words selected for this study

Place of Articulation	Voiceless	Voiced
bilabial	peer	beer
	pit	bit
alveolvar	tick	dick
	tip	dip
velar	kill	gill
	kit	git

3.3 Stimuli Recording, Analysis and Manipulation

The target words were presented to 6 native speakers of American English (3 adult men and 3 adult women), all of whom were living in Brazil at the time of the experiment[4], and had acquired only American English before reaching six years of age. The recordings were conducted in a professional studio with complete isolation from background noise. It is important to mention that the words were read in isolation (out of context) from a list, and that the speakers were instructed to maintain a regular pause time between words and to read them with the same intonation pattern.

The subsequent analysis of the stimuli recordings was conducted in software *Praat* (BOERSMA; WEENINK, 2013). Each word had the VOT of its initial plosive measured, and those productions which were considered to be the best instances of each plosive were selected for the perceptual tasks – by "best", we mean those whose VOT had the closest values to those predicted in the literature (*see subsection 2.1*).

[4]It is worth mentioning that the native speakers who had recorded the stimuli, therefore, are not the same American informants who took part in the perceptual task, since the former participants had been living in Brazil and the latter lived in the US at the time of data collection. The amount of time the participants of the former group had been living in Brazil varied widely, as well as the region of the country (United States) in which they were born. Although we acknowledge this fact as a limitation to the methodology employed in this study, since, according to a dynamic view of language acquisition, these American participants might have had their L1 system affected somehow by Brazilian Portuguese (L2), it is relevant to reinforce that all stimuli used in the perceptual task had their VOT measured, allowing us to select those tokens that best represented the VOT patterns of English (cf. LISKER; ABRAMSON, 1964; CHO; LADEFOGED, 1999).

The last stage consisted of the manipulation of some stimuli, which would belong to a fourth voicing pattern in this study – the *Artificial Zero VOT*. Productions of voiceless plosives – with *Positive VOT* – had their VOT cut out in software *Praat* (BOERSMA; WEENINK, 2013). Hypothetically, these stimuli should then sound like productions of voiced plosives, for they presented the same VOT pattern – *Zero VOT*[5] – which is typical of voiced segments in the target language. Nonetheless, these stimuli still maintained other acoustic cues from voiceless aspirated segments and, for that reason, a contrast with the other three "natural" voicing patterns (*Zero*, *Positive* and *Negative VOT*) was regarded as interesting for the observation of how multiple acoustic cues acted on the perception of these segments.

3.4 AxB Discrimination Task

The first of the two perceptual tasks was a discrimination test conducted on software *Praat* (BOERSMA; WEENINK, 2013). In this task, participants were exposed to a sequence of three productions, and were asked to determine whether the initial consonant was equal in the first two words of the sequence (AAB), in the last two words of the sequence (ABB), or if the initial consonant was equal in the three words of the sequence (AAA). Participants were first trained with a rehearsal task of identical procedures but different stimuli (contrasting other initial consonants than those investigated in the present study).

This test did not contain stimuli produced with *Zero VOT* due to a limitation in the number of stimuli produced with that pattern in the recordings[6] - the *Artificial Zero VOT* pattern was used instead, and therefore this test contrasted three VOT patterns. Specifically, the contrasts made in this test were *Negative VOT* versus *Artificial Zero*

---

[5]The pattern addressed as Zero VOT is not exactly 0 ms long, but a value below 35ms, as explained in subsection 2.1. More specifically, the manipulation aimed to obtain values of approximately 10ms for /p/, 15ms for /t/, and 25ms for /k/ productions.

[6]Most tokens of the target voiced segment were produced with *Negative VOT*. In spite of that, we do not consider this to be a methodological fault, since previous studies (such as Alves *et al.*, 2011) showed no discrimination between *Negative VOT* and *Zero VOT* as perceived by Brazilian learners.

VOT, *Negative VOT* versus *Positive VOT*, and *Artificial Zero VOT* versus *Positive VOT*. There were 36 trials in which there was a different initial consonant in the sequence, and 9 trials in which all the consonants were produced with the same VOT pattern[7]. The test had the same number of trials for each place of articulation (i.e., 15 trials per each of the three places of articulation). Each trial was heard only once, as participants were not allowed to repeat the trial. Data from 900 tokens (45 trials x 20 participants) were gathered from the test with native speakers of American English, whereas the test with Brazilian speakers provided 765 tokens (45 trials x 17 participants).

3.5 Identification Task

The second perceptual task, an identification test – also conducted on software *Praat* (BOERSMA; WEENINK, 2013)–, was composed of trials in which participants were exposed to only one production at a time. In this task, the participants' objective was to label the initial consonant of each production, within six possible answers: (/p/, /b/, /t/, /d/, /k/ or /g/). Participants were first trained with a rehearsal task of identical procedures but different stimuli (contrasting other initial consonants than those investigated in the present study).

This test had productions of all four VOT patterns – the three "natural" ones and the manipulated one. There were 24 tokens (6 per VOT pattern, equally distributed with the same number of trials for each place of articulation). Learners were not allowed to repeat any of the stimuli. Tests with native speakers provided a total of 480 tokens (24 trials x 20 participants), while those with Brazilian speakers provided 408 tokens (17 trials x 20 participants).

3.6 Hypotheses

As discussed previously in this paper,[8] since speech perception is a dynamic and multimodal process, the interaction of multiple cues

[7]In this paper, we do not report the results concerning the "catch trials", since our informants reached ceiling effects in their answers for these questions. This proves that participants really paid attention to the AXB task.

[8]See the introduction and *section 2*.

determines how segments are perceived. In the case of aspiration, we expected that it should be regarded as a primordial cue for native speakers of American English to categorize a stop consonant as voiceless; for Brazilian learners, however, other cues may account for this categorization. Thus, we established the following hypotheses:

> ***H1:*** In the AxB discrimination task, there will be significant differences between native speakers and learners in the accuracy levels contrasting '*Negative VOT* versus *Artificial Zero VOT*' and '*Artificial Zero VOT* versus *Positive VOT*' only. Native speakers will not discriminate between the patterns of the former contrast, but they will discriminate between those of the latter one successfully. The exact opposite is expected to happen in the Brazilian learners' performance.
>
> ***H2:*** In the identification task, there will be significant differences between native speakers and Brazilian learners only in the identification of the manipulated segments, presenting the *Artificial Zero VOT*. Considering the four VOT patterns altogether, native speakers will identify only segments with *Positive VOT* as voiceless, whereas Brazilian learners will identify plosives with both *Positive VOT* and *Artificial Zero VOT* as voiceless.

4 Results

4.1 Discrimination Results

Table 1 shows the descriptive analysis of the data obtained from the AxB discrimination task. The three possible answers to be assigned by the participants in this task are divided into three columns. The *accuracy* column provides the percentage of times in which the group of participants was able to successfully discriminate between the VOT patterns of the initial consonant in question. The *equality* column displays the percentage of times in which the given group of participants determined that the three productions in the AxB sequence were initiated by the same consonant – that is, there was no discrimination between VOT patterns in that amount of tokens. The *error* column provides the percentage of times in which subjects made a wrong discrimination of

stimuli, by giving an (ABB) response to an (AAB) sequence, for example (see *subsection 3.4* for a more comprehensive explanation).

Table 1 – AxB Discrimination Task Results

Contrast	Native Speakers			Brazilian Learners		
	accuracy	**equality**	**error**	**accuracy**	**equality**	**error**
Negative VOT vs Artificial Zero VOT	10.41% (25/240)	82.91% (199/240)	6.66% (16/240)	65.68% (134/204)	23.52% (48/204)	20.78% (22/204)
Negative VOT vs Positive VOT	92.08% (221/240)	2.05% (6/240)	5.41% (13/240)	91.66% (187/204)	1.96% (4/204)	6.37% (13/204)
Artificial Zero VOT vs Positive VOT	77.50% (186/240)	12.08% (29/240)	10.41% (25/240)	39.21% (80/204)	51.96% (106/204)	8.82% (18/204)

Aiming to test *Hypothesis 1* (H1 in *subsection 3.6*), a series of statistical tests were conducted, in which we tested whether Brazilians and Americans differed in their performance on the AXB discrimination task (see *Table 1*). We conducted a Mixed *Repeated Measures Analysis of Variance*[9] (hereafter *rANOVA*), with the three discrimination possibilities (accuracy, equality, error) as the within-participants variable and the two groups of participants (Brazilians and Americans) as the between-subjects variable. Follow-up *Paired Samples T-Tests* and *Independent Samples T-Tests* were conducted when necessary.

In regards to the *Negative VOT x Artificial Zero VOT* contrast, the rANOVA results indicated that there was a main effect of discrimination (*accuracy, equality*, and *error*), [$F(2,70) = 64.560$; $p < .01$]. Follow-up Paired T-Tests indicated that, when we consider the performance of all participants together, *equality* ratings ($M = 6.68$; $SD = 4.05$) were not

[9]The Repeated Measures Analysis of Variance is a parametric statistical test used for within-subjects designs with more than two independent variables, in which all participants are measured on every condition of the design.

significantly higher than *accuracy* ratings (M = 4.3; SD = 3.78), [t(36) = -1.869; p = .07]. A*ccuracy* ratings (M = 4.3; SD = 3.78) were, on the other hand, significantly higher than *error* responses (M = 1.03; SD = 1.28), [t(36) = 5.056; p < .01]; *equality* levels (M = 6.68; SD = 4.05) were also significantly higher than *error* (M = 1.03; SD = 1.28), [t(.36) = 7.351; p < .01]. The interaction between type of discrimination patterns and the two groups of speakers was also significant, [F(2,70) = 104.065; p < .01]. Follow-up Independent Samples T-Tests were conducted to verify the nature of the interaction. Results indicated that learners (M = 7.88; SD = 2.13) were significantly more accurate than native speakers (M = 1.25; SD = 1.16) in establishing the contrast [*accuracy*: t(35) = -11.264; p < .01]. Levels of *equality* attributed to the contrast by the participants were significantly higher for native speakers (M = 9,95; SD = 1.76), compared to learners (M = 2.82; SD = 2.03), [t(35) = -11.264; p < .01]. The two groups of participants did not differ as to the error rate in this contrast.

As to the *Negative VOT x Positive VOT* contrast, the rANOVA results showed that there was a main effect of discrimination (*accuracy, equality*, and *error*), [F(2,70) = 596.666; p < .01]. Follow-up Paired T-Tests indicated that, averaging across all participants, *accuracy* levels (M = 11.03; SD = 1.60) were significantly higher than *equality* levels (M = .27; SD = .65) [t(36) = 30.580; p < .01]. A*ccuracy* (M = 11.03; SD = 1.60) was significantly higher than *error* responses (M = .70; SD = 1.19) as well [t(36) = 22.759; p < .01]; *equality* (M = .27; SD = .65) levels were also higher than *error* responses (M = .70; SD = 1.19), [t(.36) = -2.462; p < .05]. The interaction between discrimination and the two groups of speakers was not significant [F(2,70) = .040; p = .961], showing that both groups of speakers had similar performance on the discrimination test. Because the interaction is not significant, follow-up Independent Samples T-Test were not conducted.

In regards to the *Artificial Zero VOT x Positive VOT* contrast, the rANOVA results indicated that there was a main effect of discrimination (*accuracy, equality*, and *error*), [F(2,70) = 50.932; p < .01]. Follow-up Paired T-Tests indicated that, taking all participants together, *accuracy* ratings (M = 7.19; SD = 3.29) were significantly higher than *equality* ratings (M = 3.65; SD = 3.34), [t(36) = 3.291; p < .01]. A*ccuracy* (M = 7.19; SD = 3.29) was significantly higher than *error* (M = 1.14; SD = 1.15) as well [t(36) = 10.215; p < .01]; the same being found for *equality* levels (M = 3.65; SD = 3.34) and *error* responses (M = 1.14;

SD = 1.15), [t(36) = 4.034; p < .01]. The rANOVA also showed that the interaction between discrimination and the two groups of speakers was significant [F(2,70) = 32.419; p < .01]. Follow-up Independent Samples T-Tests were conducted to verify the nature of the interaction. Results indicated that native speakers (M = 9.30; SD = 2.08) were significantly more accurate than learners (M = 4.71; SD = 2.68) in establishing the contrast [*accuracy*: t(35) = 5.859; p < .01]. Levels of *equality* attributed to the contrast by the participants were significantly higher for learners (M = 6.24; SD = 3.09), compared to native speakers (M = 1.45; SD = 1.43), [t(35) = -6.193; p < .01]. The two groups of participants did not differ as to the error rate in this contrast.

As a summary, the statistical analysis of the data obtained from the AxB discrimination task suggests that a) native speakers and Brazilian learners of English did not differ as to their capability of discriminating *Negative VOT* from *Positive VOT* – both groups were rather accurate in making the distinction; b) the two groups of participants were significantly different in their discrimination of *Negative VOT* from *Artificial Zero VOT* – learners were more accurate than native speakers; c) the groups differed significantly as to their capability of contrasting *Artificial Zero VOT* and *Positive VOT* – native speakers were more accurate than learners. This is what we had predicted in *Hypothesis 1* (see *subsection 3.6*), and therefore we state that the hypothesis was corroborated.

4.2 Identification Results

The descriptive analysis for the data extracted from the identification test is displayed in Table 2. The *voiceless* column provides the percentage of times in which the voicing pattern in question was labeled as a voiceless segment (/p/, /t/ or /k/); the *voiced* column, on the other hand, shows the percentage of times in which that voicing pattern was labeled as a voiced segment (/b/, /d/ or /g/). The *error* column displays information on the percentage of times in which subjects could not identify the correct place of articulation of the stimulus, regardless of its voiceless or voiced feature – an instance of that case would be the one in which a participant assigned a /p/ response to an aspirated [t] production.

Table 2 – Identification Task Results

Voicing Pattern	Native Speakers			Brazilian Learners		
	voiceless	**voiced**	**error**	**voiceless**	**voiced**	**error**
Negative VOT	0.83% (1/120)	99.16% (119/120)	0% (0/120)	0% (0/102)	100% (102/102)	0% (0/102)
Zero VOT	0.83% (1/120)	95.83% (115/120)	3.33% (4/120)	22.54% (23/102)	75.49% (77/102)	1.96% (2/102)
Artificial Zero VOT	19.16% (23/120)	76.66% (92/120)	4.16% (5/120)	76.47% (78/102)	17.64% (18/102)	5.88% (6/102)
Positive VOT	99.16% (119/120)	0% (0/120)	0.83% (1/120)	99.01% (101/102)	0% (0/102)	0.98% (1/102)

Aiming to test *Hypothesis 2* (H2 in *subsection 3.6*), the same statistical tests from the analysis with the discrimination task were conducted in order to determine whether Brazilians and Americans differed in their performance on the identification task (see *Table 2*). Specifically, we conducted a Mixed *Repeated Measures Analysis of Variance* (hereafter *rANOVA*), with the three identification outcomes as the within-participants variable (voiceless, voiced and error) and the two groups of participants (Brazilians and Americans) as the between-subjects variable. Follow-up *Paired Samples T-Test* and *Independent Samples T-Test* were conducted when necessary.

With respect to the identification of segments produced with *Negative VOT*, the rANOVA results indicated that there was a main effect of identification (*voiceless, voiced*, and *error*), [$F(2,70) = 16048.073$; $p < .01$]. Follow-up Paired T-Tests indicated that, averaging across all participants, identification as *voiced* ($M = 5.97$; $SD = .16$) was significantly higher than identification as *voiceless* ($M = .03$; $SD = .16$) [$t(36) = -110.000$; $p < .01$]; identification as *voiced* ($M = 5.97$; $SD = .16$) was significantly higher than *error* responses ($M = .00$; $SD = .00$), [$t(36) = -221.000$; $p < .01$]. Identification as *voiceless* ($M = .03$; $SD = .16$) and *error* responses ($M = .00$; $SD = .00$) were not, on the other hand, significantly different [$t(36) = -1.000$; $p = .324$]. The interaction between identification outcomes and the two groups of speakers was not significant, [$F(2,70) = .846$; $p = .433$], indicating that groups did not differ in their responses to this VOT pattern and that there was no need for the conduction of follow-up Independent Samples T-Tests.

As to the identification of segments produced with *Positive VOT*, the rANOVA results indicated that there was a main effect of identification (*voiceless, voiced*, and *error*), [$F(2,70) = 7943.017$; $p < .01$]. Follow-up Paired T-Tests indicated that, when we consider the performance of all participants together, identification as *voiceless* ($M = 5.95$; $SD = .22$) was significantly higher than identification as *voiced* ($M = .00$; $SD = .00$) [$t(36) = 157.770$; $p < .01$]; identification as *voiceless* ($M = 5.95$; $SD = .22$) was significantly higher than *error* responses ($M = .05$; $SD = .22$) as well [$t(36) = -78.168$; $p < .01$]. Identification as *voiced* ($M = .00$; $SD = .00$) and *error* responses ($M = .05$; $SD = .22$), on the other hand, were not significantly different [$t(36) = 1.434$; $p = .160$]. The interaction between identification outcomes and the two groups of speakers was not significant here either, [$F(2,70) = .013$; $p = .987$], indicating that groups did not differ in their responses to this VOT pattern and that there was no need for the conduction of follow-up Independent Samples T-Tests.

As to the identification of segments produced with *Zero VOT*, the rANOVA results indicated that there was a main effect of identification outcomes (*voiceless, voiced*, and *error*), [$F(2,70) = 283.440$; $p < .01$]. Follow-up Paired T-Tests indicated that, considering performance of all participants, identification as *voiced* ($M = 5.19$; $SD = 1.19$) was significantly higher than identification as *voiceless* ($M = .65$; $SD = 1.03$), [$t(36) = -12.592$; $p < .01$]; identification as *voiceless* ($M = .65$; $SD = 1.03$) was significantly higher than *error* responses ($M = .16$; $SD = .44$), [$t(36) = -2.834$; $p < .01$]. Identification as *voiced* ($M = 5.19$; $SD = 1.19$) was significantly higher than *error* ($M = .16$; $SD = .44$) as well [$t(36) = -20.645$; $p < .01$]. There was a significant interaction between the identification outcomes and the two groups of speakers [$F(2,70) = 15.157$; $p < .01$]; therefore, follow-up Independent Samples T-Tests were conducted to verify the source of this interaction. Results indicated that learners ($M = 1.35$; $SD = 1.16$) identified *Zero VOT* as *voiceless* significantly more times than native speakers ($M = .05$; $SD = .22$), [$t(35) = -4.890$; $p < .01$]. The level of identification of the VOT pattern as *voiced* was significantly higher for native speakers ($M = 5.75$; $SD = .55$), compared to learners ($M = 4.53$; $SD = 1.41$), [$t(35) = 3.552$; $p < .01$]. The two groups of participants did not differ as to the error rate in the identification of this VOT pattern.

Finally, concerning the identification of segments produced with the *Artificial Zero VOT*, the rANOVA results indicated that there was

a main effect of identification outcomes (*voiceless, voiced*, and *error*), [$F(2,70) = 44.178$; $p < .01$]. Follow-up Paired T-Tests indicated that, when we consider the performance of all participants together, identification as *voiceless* ($M = 2.73$; $SD = 2.09$) and *voiced* ($M = 2.97$; $SD = 2.21$) were not significantly different from one another [$t(36) = -.347$; $p = .73$]. Identification as *voiceless* ($M = 2.73$; $SD = 2.09$) was, however, significantly higher than *error* ($M = .30$; $SD = .66$), [$t(36) = -6.827$; $p < .01$]; identification as *voiced* ($M = 2.97$; $SD = 2.21$) was also significantly higher than *error* ($M = .30$; $SD = .66$), [$t(36) = -6.466$; $p < .01$]. There was a significant interaction between the identification outcomes and the two groups of speakers [$F(2,70) = 62.190$; $p < .01$]; therefore, follow-up Independent Samples T-Tests were conducted to verify the source of this interaction. Results indicated that learners ($M = 4.59$; $SD = 1.17$) identified *Artificial Zero VOT* as *voiceless* significantly more times than native speakers ($M = 1.15$; $SD = 1.18$) [$t(35) = -8.839$; $p < .01$]. The level of identification of the VOT pattern as *voiced* was significantly higher for native speakers ($M = 4.60$; $SD = 1.46$), compared to learners ($M = 1.06$; $SD = 1.14$), [$t(35) = 8.082$; $p < .01$]. The two groups of participants did not differ as to the error rate in the identification of this VOT pattern.

Summarizing the analysis of our second perceptual task, we may suggest that participants do not differ as to their categorical identification of *Negative VOT* and *Positive VOT* as voiced and voiceless plosives, respectively. With respect to *Zero VOT*, groups differ significantly: although both groups tend to associate the VOT pattern with the production of a *voiced* plosive, the native speakers' association is more categorical, since a significant amount of learners identify the stimuli as a *voiceless* segment. In regards to the manipulated *Artificial Zero VOT*, we can state that groups diverge considerably, since learners tend to identify the stimuli with that VOT pattern as *voiceless*, but native speakers identify it as *voiced* – a difference which was statistically significant. *Hypothesis 2* (see *subsection 3.6*) was, therefore, partially corroborated: although there were significant differences between the groups in the identification of segments produced with the *Artificial Zero VOT* as predicted, there were also significant differences in the identification of plosives produced with *Zero VOT*. Additionally, we can confirm a tendency which was hypothesized, that native speakers would identify only segments with *Positive VOT* as voiceless, whereas the Brazilian learners would also do so as to segments with *Artificial Zero VOT*.

5 Discussion

The results presented in the previous section seem compatible with the dynamic and multimodal phonetic-phonological acquisition perspective underlying the present study. Firstly, native speakers of American English and Brazilian learners did not differ as to their perception of *Negative VOT* and *Positive VOT* as standard cues of voiced and voiceless plosives, respectively.

With respect to the perception of *Zero VOT,* the two groups also behave similarly in their perception of *Zero VOT* as associated with voiced plosives, although we see that the tendency for native speakers to make this association is stronger. If we were to suggest an explanation as to why native speakers' identification is more categorical, we might think of the fact that *Zero VOT* is actually the standard pattern for voiceless plosives in BP. It is perfectly possible that some instances of voiced segments produced with this VOT pattern may be perceptually associated with voiceless consonants by Brazilians, since at least one relevant cue – VOT – is typical of their L1 voiceless stops in those productions. That being the case, it is also possible that this association takes place in productions of one place of articulation more than the others – something which further verification of our data may reveal.

Above all, the difference which allows us to have interesting insights into the action of multiple acoustic cues lies in how the two groups of participants perceive the manipulated *Artificial Zero VOT*: for speakers whose L1 system is American English, the absence of the long lag of *Positive VOT* as an acoustic cue affects the discrimination between voiced and voiceless plosives (*voiceless* becomes *voiced*). On the other hand, this shift from *voiceless* to *voiced* does not happen for speakers whose L1 system is Brazilian Portuguese.

The answer for this equality between *Positive VOT* and the *Artificial Zero VOT* as patterns that stand for voiceless plosives in Brazilian learners' perception may lie in the action of other cues (such as burst intensity or the verified F0 value in the following vowel)[10]. It is quite reasonable to assume that these other cues – which presumably were not altered with the stimuli manipulation (see *subsection 3.3*) – are more relevant than *Positive VOT* to the Brazilian learners' perception of

[10]Such factors have been suggested by Sundara (2005), Oh (2011) e Kong *et al.* (2012).

these segments, in the sense that they are the ones that determine if the segment is to be perceived as voiceless. Furthermore, the fact that those manipulated segments seem to "confuse" learners' perception, leading to a higher error rate in their performance in these tasks, may deserve attention. Thus, we are addressing different statuses of acoustic cues across linguistic systems.

In addition, it is interesting to notice that recent studies (REIS; NOBRE-OLIVEIRA, 2008; ALVES *et al*., 2011; FRANÇA, 2011; SCHWARTZHAUPT, 2012; PRESTES, 2013) suggest that Brazilian learners in high proficiency levels produce what may be called "partial" aspiration of voiceless plosive consonants. However, as suggested in this paper, they still do not attribute a significant status to *Positive VOT* as determinant for the *voiceless* versus *voiced* distinction. Therefore, it may be necessary for learners to receive formal instruction, in order to draw their attention to this cue (ALVES, 2010; ALVES; MAGRO, 2011).

The results discussed above may serve as evidence for us to reinforce the idea that speech perception is guided by the action of multiple cues, and that these cues interact differently in separate linguistic systems, assuming a different status in each system. Therefore, this should be regarded as a fundamental assumption that should underlie any and all investigations in L2 phonetic-phonological acquisition we conduct, as well as the teaching of a foreign language.

References

ALVES, Ubiratã Kickhöfel. Uma discussão conexionista sobre a explicitação de aspectos fonético-fonológicos da L2: dados de percepção e produção da plosiva labial aspirada do inglês. In: POERSCH, José Marcelino; ROSSA, Adriana Angelim (Org.). *Processamento da Linguagem e Conexionismo.* Santa Cruz do Sul: Editora da UNISC, 2007, p. 155-185.

_____. *Efeitos da Instrução Formal na Aquisição de Aspectos Fonético Fonológicos do Inglês (L2) por Brasileiros*. 2010. 37f. Projeto de pesquisa. Porto Alegre: Universidade Federal do Rio Grande do Sul (UFRGS), 2010.

_____; 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* – Vol. 4. Pelotas, RS: Editora e Gráfica Universitária da UFPel, 2011. p. 3-4.

_____; MAGRO, Vivian. Raising awareness of L2 phonology: explicit instruction and the acquisition of aspirated /p/ by Brazilian Portuguese speakers. *Letras de Hoje*, Porto Alegre, v. 43, n. 6, 2011.

BECKNER, C; ELLIS, N; BLYTHE, R; HOLLAND, J; BYBEE, J; KE, J; CHRISTIANSEN, M; LARSSEN-FREEMAN, D; CROFT, W; SCHOENEMANN, T. Language is a Complex Adaptive System: Positional Paper. *Language Learning*, v. 59, s. 1, p. 1-26, 2009.

BEST, C. T.; TYLER, M. D. Nonnative and second-language speech perception: Commonalities and complementarities. In: BOHN, Ocke-Schwen; MUNRO, Murray J. *Language Experience in Second Language Speech Learning*: Studies in honor of James Emil Flege. Amsterdam: John Benjamins, 2007, p. 13-34.

BOERSMA, P.; WEENINK, D. *Praat*: Doing Phonetics by Computer. Version 5.3.48.

2013. Disponível em: <www.praat.org>. Acesso em: 20 jan. 2015.

CHO, T.; LADEFOGED, P. Variation and universals in VOT: evidence from 18 languages, *Journal of Phonetics*, v. 27, p. 207-229, 1999.

COHEN, G. V. *The VOT Dimension: a bi-directional experiment with English Brazilian Portuguese stops.* 2004. 70f. Dissertação (Mestrado em Língua Inglesa) - Programa de Pós-Graduação em Língua Inglesa, Universidade Federal de Santa Catarina, 2004.

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

ELLIS, N. The emergence of language as a complex adaptive system. In: SIMPSON, J. (ed.). *Handbook of Applied Linguistics*. London: Routledge, 2011, p. 666-679.

FLEGE, J. E; MUNRO, M. J.; MacKAY, I. R. A. Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, v. 97, n.5, p. 3125-3134, 1995.

FRANÇA, K. *A aquisição da aspiração das plosivas surdas do inglês por falantes do português brasileiro: implicações teóricas decorrentes de duas diferentes formas de descrição dos dados*. 2011. 112f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras, Universidade Católica de Pelotas, 2011.

GEWEHR-BORELLA, S. *A influência da fala bilíngue Hunsrückisch-Português Brasileiro na escrita de crianças brasileiras em séries iniciais*. 2010. 205f. Dissertação (Mestrado em Letras) – Programa de Pós-Graduação em Letras, Universidade Católica de Pelotas, 2010.

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*, v. 40, p. 725-744, 2012.

LISKER, L; ABRAMSON, A. A Cross-language study of voicing in initial stops: Acoustical measurements. *Word*, New York, United States, v.3, n.20, p. 384-422, 1964.

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

PEROZZO, R. V.; ALVES, U. K. Implicações Dinâmicas para a Formação da Fonologia em L2. *Revista Signo*, vol. 38, n. 65, p. 247-260. UNISC, 2013.

POLLITT, A. *The meaning of OOPT Scores*. 2007. Retrieved: August 5th, 2014, from: <www.oxfordenglishtesting.com>.

PRESTES, S. P. C. Produção de consoantes oclusivas iniciais do Inglês por falantes nativos de PB. 2013. 139f. Dissertação (Mestrado em Letras) - Programa de Pós-Graduação em Letras, Universidade Federal do Paraná, 2013.

PURPURA, J. *The Oxford Online Placement Test: What does it measure and how*? 2007. Retrieved: August 5th, 2014, from: <www.oxfordenglishtesting.com>.

REIS, M.; NOBRE-OLIVEIRA, D**.** Effects of perceptual training on the identification and production of English voiceless plosives aspiration by Brazilian EFL learners. INTERNATIONAL SYMPOSIUM ON THE ACQUISITION OF SECOND LANGUAGE SPEECH, 5, Florianopolis, SC, 2008. *Anais*... Florianopolis, UFSC, 2008. p. 372-381.

SCHWARTZHAUPT, B. M. *Factors influencing voice onset time: analyzing Brazilian Portuguese, English and Interlanguage data*. 2012. 65f. Trabalho de Conclusão de Curso (Graduação em Letras) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, 2012.

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

YAVAS, M**.** Factors influencing the VOT of English long lag stops and interlanguage phonology. In: RAUBER, Andrea S.; WATKINS, Michael A.; BAPTISTA, Barbara O. (Ed.). *New Sounds* 2007. INTERNATIONAL SYMPOSIUM ON THE ACQUISITION OF SECOND LANGUAGE SPEECH, 5, Florianópolis, SC. *Anais*... Florianópolis, SC, UFSC, 2008, p. 492-498.

ZIMMER, M. C.; SILVEIRA, R.; ALVES, U. K. *Pronunciation instruction for Brazilians*: bringing theory and practice together. Newcastle upon Tyne: Cambridge Scholars Publishing, 2009.

ZIMMER, M. C.; ALVES, U. K. A produção de aspectos fonético-fonológicos da segunda língua: instrução explícita e conexionismo. *Revista Linguagem & Ensino*, Pelotas, v. 9, n.2, p. 101-143, jul. / dez. 2006.

_____; _____. *On the Status of Terminal Devoicing as an Interlanguage Process among Brazilian learners of English.* Ilha do Desterro*,* v. 55, p. 41-62, 2008.

_____; _____. *Learning to orchestrate time: Voicing patterns and gestural drift in L2 speech production. Resumos*… São Paulo School of Advanced Studies in Speech Dynamics. São Paulo, 2010, p. 47-48.

_____; _____. Uma visão dinâmica da produção da fala em L2: o caso da dessonorização terminal. *Revista da Abralin*, número especial 2, 2012.