UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
MESTRADO EM LETRAS
LINGUÍSTICA APLICADA

BRUNO MORAES SCHWARTZHAUPT

TESTING INTELLIGIBILITY IN ENGLISH: THE EFFECTS OF POSITIVE VOT AND CONTEXTUAL INFORMATION IN A SENTENCE TRANSCRIPTION TASK

Porto Alegre
2015

BRUNO MORAES SCHWARTZHAUPT

Testing Intelligibility in English: the Effects of Positive VOT and Contextual Information in a Sentence Transcription Task

Dissertação de Mestrado em Letras, vinculada à área de Estudos da Linguagem, na especialidade Linguística Aplicada e linha de pesquisa Aquisição da Linguagem, apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Ubiratã Kickhöfel Alves

Porto Alegre
2015

CIP - Catalogação na Publicação

Schwartzhaupt, Bruno Moraes
Testing Intelligibility in English: the effects of Positive VOT and contextual information in a sentence transcription task / Bruno Moraes Schwartzhaupt. -- 2015.
86 f.

Orientador: Ubiratã Kickhöfel Alves.

Dissertação (Mestrado) -- Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras, Porto Alegre, BR-RS, 2015.

1. Aprendizes brasileiros de Inglês. 2. Inteligibilidade. 3. Percepção de padrões de VOT. 4. Consoantes Plosivas. 5. Informação Contextual. I. Alves, Ubiratã Kickhöfel, orient. II. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os dados fornecidos pelo(a) autor(a).

A presente dissertação de Mestrado em Letras, de **Bruno Moraes Schwartzhaupt**, sob o título **Testing Intelligibility in English: the effects of Positive VOT and contextual information in a sentence transcription task** foi apresentada como requisito parcial para a obtenção do título de Mestre pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul. O trabalho encontra-se vinculado à área Estudos da Linguagem, na especialidade Linguística Aplicada e linha de pesquisa Aquisição da Linguagem, tendo sido devidamente defendido e aprovado no dia 30 de Março de 2015, pela seguinte banca examinadora:

Prof. Dr. Ubiratã Kickhöfel Alves (orientador)
Universidade Federal do Rio Grande do Sul

Profa. Dra. Adelaide Hercília Pescatori Silva
Universidade Federal do Paraná

Profa. Dra. Ana Beatriz Arêas da Luz Fontes
Universidade Federal do Rio Grande do Sul

Profa. Dra. June Campos
Universidade Federal do Rio Grande do Sul

Dedico esta dissertação aos meus pais, Rosita e Ercilio, pelo apoio incondicional em todas as etapas da minha vida.

ACKNOWLEDGEMENTS

I would first like to thank present-day Bruno for his dedication, so that when I look back on this I will know that I did my best. After all, that is all that matters.

I must thank my parents and sisters for all the love and support they have always given me. It is all too cliché, I know, but I would never have been able to accomplish any of this without them. Rosita and Ercilio, thank you very much; you have done for me more than I could ever ask for.

I would also like to thank my friends, who will be there for me, because I am there for them, too. This goes to Bento, Felipe, G. da Silva, Geisson, Rodrigo, and, of course, my BFF Davi. I hope to still have beers with you after a FKN12 F.C. match when I am 50.

To my advisor and friend for almost 5 years now, Ubiratã Kickhöfel Alves, thank you very much, for everything you taught me, and for all your dedication. Not only are you exceptionally good at what you do, you are also a great person whom I truly admire.

I must thank Ana Fontes, too. Ana, it is admirable to see how much you care for what you do, and I hope you know how much of an example of advisor and researcher you are. I owe much of this thesis to your unique methodological advice.

Special thanks to Ingrid Finger, June Campos and Ian Alexander, who have played fundamental roles in my journey as an English learner, teacher and researcher. I will never forget how much you cared for me as your student, and you should know that you are role models to me.

I would also like to thank professor Adelaide Silva, for giving me the honor of accepting the invitation to be part of this committee.

Some friends made the conduction of this study a lot easier. I have to thank my friend Gibran, and Ian again, for having the ideas that made the whole data collection process about 90% easier and faster.

I also want to thank Reiner, Letícia, Juliana, Sandro and Vinícius for being the nicest classmates in the Master's courses.

I must thank each and every one of the participants who made this study possible by lending me a little of their precious times.

Finally, I have to thank CAPES for the financial support in the last two years.

“[...] Então o homem, flagelado e rebelde, corria diante da fatalidade das coisas, atrás de uma figura nebulosa e esquiva, feita de retalhos, um retalho de impalpável, outro de improvável, outro de invisível, cosidos todos a ponto precário, com a agulha da imaginação; e essa figura, — nada menos que a quimera da felicidade, — ou lhe fugia perpetuamente, ou deixava-se apanhar pela fralda, e o homem a cingia ao peito, e então ela ria, como um escárnio, e sumia-se, como uma ilusão.”

(Machado de Assis)

ABSTRACT

The present study aimed to investigate the effects of *Positive VOT* (with or without contextual information) on the intelligibility of short English sentences produced by Brazilians to native speakers of American English (6) and proficient Brazilian learners (12). These 18 participants transcribed sentences produced with voiceless unaspirated and voiceless aspirated stop segments, with and without contextual information. Each participant transcribed 112 sentences through software *E-Prime 2.0* (SCHNEIDER, ESCHMAN & ZUCCOLOTTO, 2012) with a response-time limit of 20 seconds. The data obtained from the task was analyzed objectively, assigning transcriptions to binary accurate (high intelligibility) or inaccurate (low-to-no intelligibility) categories. Overall, accuracy levels did not vary considerably amongst the 18 participants, although it can be said that native speakers of American English performed slightly better. The analysis of the transcriptions suggests that, while factors external to the variables controlled by this study might have played fundamental roles in the overall performance of the task, contextual information may possibly have remedied the absence of *Positive VOT* as a cue for achieving sentence intelligibility.

Keywords: Proficient Brazilian learners of English; Native Speakers of American English; Voice Onset Time; Intelligibility.

RESUMO

O presente estudo buscou investigar os efeitos do *VOT Positivo* (com ou sem informação contextual) na inteligibilidade de sentenças curtas em Inglês, produzidas por brasileiros, por parte de falantes nativos de Inglês Americano (6) e aprendizes proficientes de Inglês (12). Esses 18 participantes transcreveram sentenças com e sem informação contextual, produzidas com plosivas-alvo sonoras, surdas sem aspiração e surdas com aspiração. Cada participante transcreveu 112 sentenças através do software *E-Prime 2.0* (SCHNEIDER, ESCHMAN & ZUCCOLOTTO, 2012) com um tempo-limite para respostas de 20 segundos. Os dados obtidos através da tarefa foram analisados objetivamente, dividindo as transcrições em categorias binárias "correta" (alto grau de inteligibilidade/inteligibilidade plena) e "incorreta" (baixo grau de inteligibilidade/inteligibilidade praticamente inexistente). Em geral, os níveis de acuidade não variaram consideravelmente entre os 18 participantes, embora possamos dizer que falantes nativos de Inglês Americano tiveram desempenho relativamente melhor. A análise das transcrições sugere que, enquanto é possível que fatores externos às variáveis controladas por este estudo tenham tido papel fundamental no desempenho dos participantes, a informação contextual talvez tenha remediado a falta do *VOT Positivo* como uma pista para atingir inteligibilidade.

Palavras-chave: Aprendizes brasileiros de inglês em nível proficiente; falantes nativos de Inglês Americano; Voice Onset Time; Inteligibilidade.

LIST OF BOXES

Box 1 - Different levels of *understanding* in communication as conceived in this study 36

Box 2 - The 24 target words for the transcription task 43

LIST OF FIGURES

Figure 1 - Illustration of the production of a stop consonant 21

Figure 2 - E-Prime instructions screen 1 47

Figure 3 - E-Prime instructions screen 2 48

Figure 4 - E-Prime typing field screen 49

LIST OF TABLES

Table 1 - Group 1 participants' (n=6) background information. (SD: standard deviation) 39

Table 2 - Group 2 Brazilian participants' (n=12) background information. 40

Table 3 - Mean VOT levels and standard deviation (SD) of the stimuli for the transcription task. N = 4. 45

Table 4 - Overall accuracy in the transcription task considering contextual information 56

Table 5 - Overall accuracy in the transcription task considering VOT..... 58

Table 6 – Mean accuracy in the transcription task according to six types of stimulus. SD = standard deviation..... 59

CONTENTS

1 INTRODUCTION 15

2 THEORETICAL BACKGROUND 20

2.1 Voice Onset Time 20

2.2 Language as a CAS and L2 phonetic-phonological acquisition under this perspective 23

2.3 The role of L1 knowledge on L2 speech perception 25

2.3.1 L2 Speech perception: the matter of L1 transfer 25

2.3.2 The acquisition of *Positive VOT* by Brazilians 26

2.4 On Intelligibility 30

2.4.1 ELF and the need for a clear definition of "understanding" 30

2.4.2 Intelligibility defined and assessed: the value of contextual information 33

3 METHOD 37

3.1 Goals of this study 37

3.2 Participants 38

3.2.1 Group 1: stimuli recording 38

3.2.2 Group 2: sentence transcription task 40

3.3 Selection of stimuli for the transcription task 41

3.3.1 Target words 41

3.3.2 Target sentences 43

3.3.3 Stimuli recording and selection 44

3.4 The transcription task 46

3.5 Criteria for data analysis 49

3.6 Hypotheses 52

4 RESULTS AND DISCUSSION 56

4.1 Overall descriptive analysis 56

4.2 Hypotheses tested 59

4.2.1 Testing Hypothesis 1 60
4.2.2 Testing Hypothesis 2 61
4.2.3 Testing Hypothesis 3 62
4.2.4 Testing Hypothesis 4 63
4.2.5 Testing Hypothesis 5 65
4.2.6 Testing Hypothesis 6 66
4.2.7 Testing Hypothesis 7 67
4.3 Summary of findings 68
5 CONCLUSION 71
REFERENCES 74
APPENDICES 78
Appendix 1 – Termo de Consentimento Livre e Esclarecido (Group 1) 78
Appendix 2 – Ficha de Informações do Participante 80
Appendix 3 – Termo de Consentimento Livre e Esclarecido (Group 2) 81
Appendix 4 – Consent Form 83
Appendix 5 – Participant Information Form 85
Appendix 6 – List of Target and Distractor Sentences 86

1 INTRODUCTION

Pronunciation has received different attention in the history of Second Language (SL)[1] teaching. In its changing roles, pronunciation has gone from being taught with a strong focus on the repetition of a native accent to being completely neglected from L2 classrooms, reaching a status in which it receives some attention, while it still lacks a clear goal and method (WALKER, 2010; DERWIN & MUNRO, 2005). Today, in academia, there seems to be an ongoing effort to advocate the importance of focusing on some (not all) pronunciation aspects of the target language in SL teaching in order for the learner to achieve successful communication[2] (JENKINS, 2000; DERWIN & MUNRO, 2005, 2009; DAVIES, 2009; WALKER, 2010; CELCE-MURCIA *et al.*, 2010; NELSON, 2011; to name a few). Such an agreement is perhaps the result of an extensive body of research, mostly dedicated to the learning of L2 English, which has aimed to discuss both the need for pronunciation teaching and the syllabus it should adopt.

As regards the teaching of L2 English, the contemporary research work cited above stresses that it is essential to consider the *lingua franca*[3] status of the target language. The general claim is that, inasmuch as today the majority of speakers of English are nonnative speakers, learners face a wide range of different accents in English oral communication (DERWIN & MUNRO, 2009; DAVIES, 2009; WALKER, 2010). That being the case, it seems only reasonable that a standard pronunciation target be questioned: billions use English as their L1 or L2 in their daily oral communication, and these speakers have various L1 phonetic-phonological systems – which means they will have several different *accents*. Common sense tells us that such a variety of accents does not harm communication, once the context (in a broad sense) more than often suffices to eradicate possible interpretation problems derived from the "misproduction" of certain segments. Within this context,

[1] I consider it irrelevant to make a distinction between the terms Second Language, Foreign Language or Additional Language Learning in order to pursue the aim of the present dissertation, neither do I find it possible to restrict the context in which this study should be conducted to any of them alone. Therefore, in the reading of this dissertation, Second Language (SL) learning can be interpreted as a synonym for both Foreign and Additional language learning.

[2] For a discussion on this topic, see *subsection 2.4.1.*

[3] I further discuss this concept in *subsection 2.4.*

demanding that L2 learners achieve native-like English pronunciation is unnecessary. In addition, it has been argued that native-like pronunciation attainment in SL late acquisition is an unrealistic goal (JENKINS, 2000).

For that reason, current research discusses the necessity for pronunciation teaching to focus on *intelligibility*[4]. While the concept is rather far from a categorical, objective definition, it is commonly associated with the successful "understanding"[5] of an utterance. However, there is no consensus as to what linguistic domains are involved – meaning or sound, for example. As far as pronunciation is concerned, the claim is that learners should not aim (or be taught) to sound like native speakers, but simply to be understood in their L2 (JENKINS, 2000; DERWIN & MUNRO, 2005, 2009; WALKER, 2010). Nonetheless, it is important to observe that productions with certain deviant phonetic aspects in relation to L1 English native speech may, depending on the context, cause intelligibility problems. Aiming to promote pronunciation teaching of English as a Lingua Franca (hereafter ELF) with the intelligibility goal, Jenkins (2000) proposed a set of aspects of the language which should be addressed in the L2 classroom, namely the *Lingua Franca Core* (LFC). The LFC lists some pronunciation aspects of the English language which, according to the author, should be produced by the L2 learner in order for him/her not to face intelligibility problems.

One of the phonetic-phonological segmental aspects of the English language listed in the Lingua Franca Core by Jenkins (2000) is aspiration of voiceless stops /p/, /t/ and /k/. Walker (2010) also gives this aspect an important status for intelligibility. As regards the voicing patterns in plosive consonants, it can be stated that English and Brazilian Portuguese (hereafter BP) belong to two distinct classes (COHEN, 2004). When comparing the mean *Voice Onset Time* (VOT)[6] values of word-initial plosive segments in both languages, we verify that, as a general rule, voiced /b/, /d/ and /g/ are produced with *Negative VOT* in BP, but with *Zero VOT* in English. In relation to voiceless /p/, /t/ and /k/, they are produced with *Zero VOT* in BP, whereas the pattern for those plosives in English is *Positive VOT* (aspiration).

[4] This concept is broadly discussed in *subsection 2.4*.

[5] From the beginning of this dissertation, it should be clear that the term *understanding*, as discussed by Nelson (2011), has been used vaguely and there is no clear definition of what it is to be understood (it is usually unclear, for example, what linguistic domains are involved in the process). This is further discussed in *subsection 2.4*.

[6] VOT definition and patterns are presented in *subsection 2.1.*

Several factors pose challenges for Brazilian learners in the process of their acquisition of the English voicing pattern of plosive consonants. Models of speech perception, such as the *Speech Learning Model* by Flege (1995) and the *Perceptual Assimilation Model – L2* by Best & Tyler (2007), suggest that the perception of these L2 categories is influenced by L1 transfer. Moreover, Zimmer, Silveira & Alves (2009) have argued that orthography also plays a role in this process, for learners face L1-L2 *Grapho-Phonic-Phonological Transfer*[7]. The phenomenon has often figured in Brazilian research on phonetic-phonological SL acquisition (e.g. COHEN, 2004, REIS & NOBRE-OLIVEIRA, 2007; MAGRO, 2010; ALVES & MAGRO, 2011; FRANÇA, 2011; SCHWARTZHAUPT, 2012; PRESTES, 2013; ALVES & MOTTA, 2014; ALVES & ZIMMER, 2015). What the literature in the field suggests, therefore, is that Brazilian learners tend not to produce the necessary aspiration in English – and that has been said to pose a problem for intelligibility.

Within this context, the present study aims to bring together studies of speech perception with special attention to the acquisition of *Positive VOT* by Brazilians and the discussion on intelligibility and the factors that are important for this property to be achieved. It seems clear that connecting these frameworks is the most appropriate way of objectively addressing the matter of the actual relevance of *Positive VOT* for communication in English involving Brazilian learners. In this sense, we are thinking pronunciation teaching, and the role *Positive VOT* may or may not play in it. Moreover, as I intend to discuss throughout the writing of this dissertation, such investigations that so objectively scrutinize the phonetic detail have a lot to contribute to the discussion on how to assess and define intelligibility.

In order to build these links, this study investigates the intelligibility of sentences with or without contextual information, produced with or without *Positive VOT* by Brazilians as transcribed by both native speakers of American English and Proficient Brazilian learners. The contrasts established between contextual information and *Positive VOT* and their statuses as cues responsible for intelligibility in the sentences, taking different listener's L1 backgrounds into consideration, are expected to shed light on the issues related to the relevance of *Positive VOT* for intelligibility in English.

The goal of this study is, therefore, to investigate the effects of *Positive* VOT (with or without contextual information) on the intelligibility of short English sentences produced by Brazilians to native speakers of American English and proficient Brazilian learners. These

[7] This is discussed in *subsection 2.3*.

effects are determined from the levels of accuracy achieved by native speakers and Brazilian learners in a sentence transcription task. In this sense, the present study has the following specific objectives:

- To investigate the accuracy of both native speakers of American English and proficient Brazilian learners in the transcription of sentences with or without contextual information, produced with or without Positive VOT by Brazilians.

- To verify, considering the data provided by native speakers and proficient Brazilian learners, whether there are significant differences in accuracy in the transcription of sentences considering…

  (a) only the participant's L1 as a variable;

  (b) only the VOT of the target word as a variable;

  (c) only the presence or absence of contextual information as a variable*;*

  (d) the participant's L1 and the VOT of the target word as variables;

  (e) the participant's L1 and the presence or absence of contextual information as variables; and

  (f) the interaction of all the three variables tested in this study as factors: the participant's L1, the VOT of the target word and the presence or absence of contextual information as variables;

- To discuss, based on the results of the transcription task, the relationship between contextual information and *Positive VOT* as cues promoting intelligibility;

- To debate the question of how intelligibility may be defined and assessed, based on the results of the transcription task;

- To discuss, based on the results of the transcription task, the need for pronunciation instruction on English *Positive VOT* for Brazilian learners;

The present research study is composed of five main sections. I begin by presenting the theoretical background, in which I will discuss questions related to *Voice Onset Time*, the conception of language as a *Complex Adaptive System*, the role of the L1 knowledge on L2 speech perception and intelligibility (*section 2*). Next, I address the methodology with which the present study was carried out, detailing information about the goals of this study, its

participants, the selection of target words and sentences, the transcription task and the criteria for analyzing its data, as well as the hypotheses tested in this study (*section 3*). In the following section, I present and discuss the data obtained from the transcription task, testing hypotheses established for the present study (*section 4*). Finally, in the last section, I conclude discussing the main findings, implications, limitations and future directions of this study (*section 5*).

2 THEORETICAL BACKGROUND

In this section, I explain key concepts to support the development of the present study. I begin by addressing the concept of *Voice Onset Time* (*subsection 2.1*). The following subsection explains the dynamic perspective of language acquisition underlying this study (*subsection 2.2*). Next, I discuss the question of the transfer of L1 knowledge to the L2 and how it affects speech perception (*subsection 2.3*). In the last subsection, discussions on the understanding of nonnative speech and the definition of intelligibility are addressed (*subsection 2.4*).

2.1 Voice Onset Time

This subsection addresses the phonetic-phonological aspect whose production is a crucial variable for the phenomenon investigated in this study.

Voice Onset Time (VOT) refers to a time interval in the production of a stop consonant (/p/, /b/, /t/, /d/, /k/, /g/). The production of a stop consonant takes place in three stages. In the first stage, the articulators produce a total obstruction of the air in the oral cavity (stage of *closure*). Following the closure, the consonant is released, which corresponds to the articulatory phase in which the obstruction is undone, and air flows through the oral cavity. Finally, in the last stage, the vocal folds of the segment which follows the consonant start vibrating (KENT & READ, 2002; ASHBY & MAIDMENT, 2005).

VOT names the period of time between the second and third stages – the stop consonant release and the vibration of the vocal cords of the segment following this consonant. The literature in the field customarily divides VOT into three categories (LISKER & ABRAMSON, 1964; COHEN, 2004; REIS & NOBRE-OLIVEIRA, 2007; SCHWARTZHAUPT, 2012; PRESTES, 2013):

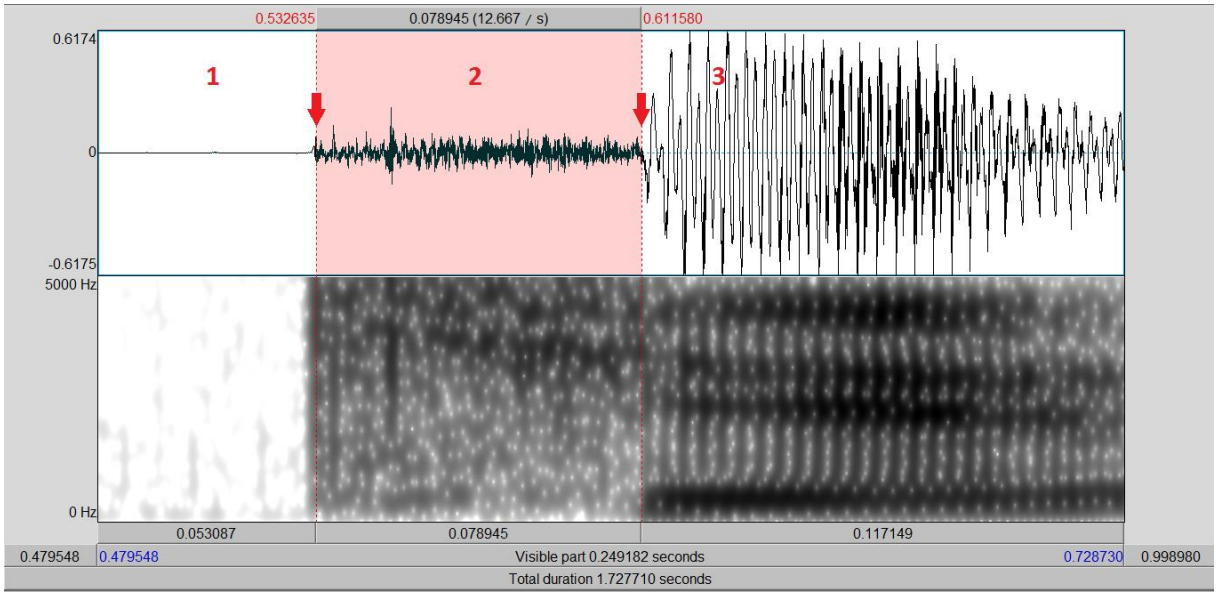
- ***Negative VOT*** (also *pre-voicing*): the vibration of the vocal folds starts before the stop consonant release, in an interval ranging from -125 ms to -75ms;

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

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

The following *Figure 1*, a printed screen from software *Praat* (BOERSMA & WEENINK, 2014), illustrates the three stages of the production of the aspirated initial stop consonant in the word "*kit*" (preceded by a pause) by a native speaker of American English. The stages are labeled from numbers 1 to 3: 1 being the closure period (silent); 2 standing for a VOT of 78,94 ms; and 3 corresponding to the following [ɪ] vowel.

Figure 1 - Illustration of the production of a stop consonant



Source: SCHWARTZHAUPT (2012, p.19)

As regards the production of stop consonants in English, the aforementioned literature states that voiced plosives (/b/, /d/, /g/) tend to be produced with *Zero VOT* – i.e. no pre-voicing –, but productions of those segments with *Negative VOT* may also occur as variants (SIMON, 2006; SCHWARTZHAUPT, ALVES & FONTES, 2013; *in press*). Voiceless stops (/p/, /t/, /k/), on the other hand, are produced with *Positive VOT*. Mean VOT values for those segments tend to be 55ms for [p^h] productions, 70ms for [t^h], and an average 80ms for [k^h] production (CHO & LADEFOGED, 1999; COHEN, 2004; ALVES, 2010).

In contrast to those of English, BP voiced stop consonants tend to be produced with *Negative VOT*, whereas voiceless plosives tend to be classified as being produced with *Zero VOT* – i.e. no aspiration –, with mean VOT values of 12ms for [p], 18ms for [t] and 38ms for [k] (COHEN, 2004; REIS & NOBRE-OLIVEIRA, 2007; ALVES, 2010). Nonetheless, most

recent studies have shown rather higher VOT values for stop consonants in the Southern region of Brazil (REIS & NOBRE OLIVEIRA, 2007; GEWER-BORELLA, 2010; FRANÇA, 2011; SCHWARTZHAUPT, 2012). While bilabial /b/ and alveolar /t/ showed higher VOT values, but not to a large degree, the values found for the velar /k/ segment were considerably higher: mean 46.55ms were found by Reis & Nobre-Oliveira (*op. cit*); Gewehr-Borella (*op. cit*) measured mean VOT values for /k/ in two cities, finding 52.43ms and 63.90ms[8]; França (*op. cit*) verified a mean 47.20 ms VOT; mean values found by Schwartzhaupt (*op. cit*) were 58,05. On account of that, one can suggest that some varieties of BP from the South present *partial aspiration* (SCHWARTZHAUPT, 2012).

It should be clear, as addressed by several authors (e.g. COHEN, 2004; REIS & NOBRE-OLIVEIRA, 2007; ALVES, 2010), that VOT values are not absolute, and they cannot be considered to be an isolated entity within any linguistic system. Rather, VOT values are influenced by several factors which deserve attention, such as syllable stress, prosody, speech rate, the quality of the vowel following the plosive segment. For that reason, some studies (PRESTES, 2013; SCHWARTZHAUPT, 2012) have employed alternative measurements for VOT, such as the one of relative values: target words with initial stops are inserted in sentences which then have their entire length measured. Both VOT length and sentence length are calculated with the objective to determine what percentage of the sentence is occupied by VOT – that is, the relative value of VOT in the sentence.

Also, as perhaps demonstrated by the divergent mean VOT values for BP stops discussed earlier, it is important to clarify that no concrete assertions can be made as to how far variations in VOT values within both languages – if in any languages – go. One could argue, for example, that the partial aspiration we see in southern varieties of BP is not to be found in northern varieties – empirical evidence to support this idea is, however, scarce.

Nevertheless, the difference in the voicing patterns of stop segments in BP and English is widely acknowledged by the literature: BP has unaspirated voiceless stops, whereas English has aspirated voiceless stops. For that reason, it can be stated that the languages belong to distinct groups in what comes to VOT patterns (COHEN, 2004; ALVES, 2010).

[8] The former mean VOT value for /k/ was verified in *Rio Grande*, and the latter in *Picada Café*, both cities from the Brazilian Southernmost state, Rio Grande do Sul. The author analyzed production of monolingual and bilingual (BP + Hunsrückisch) participants, the values presented here were found in segments produced by monolingual participants.

In the next subsection, I turn to the view of language and SLA which is the basis for the present study, discussing the conception of language as a *complex adaptive system*.

2.2 Language as a CAS and L2 phonetic-phonological acquisition under this perspective

In this subsection I present the conception of language and SLA underlying this research study. It is important to begin with a definition of language which encompasses its complex and unpredictable behavior: language as a *Complex Adaptive System* (CAS). Several authors have established analogies between the behavior of languages and complex nonlinear systems[9] in the natural world (e.g. LARSEN-FREEMAN, 1997; DE BOT *et al.*, 2007; BECKNER *et al.*, 2009; ELLIS, 2011; ALBANO, 2012; ZIMMER & ALVES, 2012; SCHWARTZHAUPT, 2013).

Such systems are fundamentally defined as composed of *multiple agents*, whose interaction and adaptation to one another determine the behavior of the system as a whole. Whatever characteristics we observe in the system are thus a product of the interaction between all of its agents, which, alone, do not stand for the system. Another important characteristic of a CAS is that it is perpetually *dynamic* (ELLIS, 2011): the system is the product and cause of constant change, a requirement to account for the diversity of its multiple agents.

A language system is also composed of multiple agents who interact to modify the system – we may think of speakers inserted in a language community, for example. In this sense, speakers constantly and perpetually modify the language system of the community in which they are inserted, and have their own system modified by the community. Supporting this analogy, authors have argued that the Dynamic Systems Theory (DST) is the one which most adequately accounts for language variation (DE BOT *et al.*, 2007; BECKNER *et al.*, 2009).

Another important characteristic of dynamic systems is also regarded as meaningful for explaining language variation: *nonlinearity* (LARSEN-FREEMAN, 1997; DE BOT *et al.*, 2007). Changes in the system tend to result in effects which are unpredictable and not directly

[9] Adaptive, complex, dynamic and nonlinear are all characteristics of the same type of system – which I first presented as CAS – and, therefore, all of them can be used interchangeably to name such a system throughout this dissertation.

proportional to the dimensions of the alteration. All agents can be affected in different ways, so the outcome of change is unpredictable. Language variation is, however, not random: it reflects the system's need for alteration, mostly in search of facilitated interaction between its agents. In this sense, when we discuss variation, the implications are that every instance of language use provokes unpredictable changes in some level of the system, whether we observe them or not.

Therefore, it should be clear that what drives language variation is its use. This perspective explicitly denies the predictability of language variation and acquisition through universal rules; instead, it recognizes language as a *domain-general* capacity (BECKNER et *al.*, 2009), product of human cognition and experience with the world.

Turning, then, to acquisition, it should be clear that the perspective adopted in the present study considers linguistic input to be fundamental to language acquisition. It is from input that learners extract patterns of the language system, its constraints and regularities – factors such as input frequency and saliency help shape the learner's developing system. This process takes place through interaction, and it is when patterns *emerge* from communication (ZIMMER, SILVEIRA & ALVES, 2009).

It is also important to mention that, under the emergentist view, different factors, linguistic and non-linguistic, have effects on the language acquisition process, and these factors cannot be considered in an isolated manner (DE BOT *et al.*, 2007). This is consonant with the complex, dynamic and nonlinear characteristics which have been previously mentioned in this writing. In this sense, research cannot define difficulties in L2 acquisition as matters related solely to learner's age or L1 transfer, for example. It should, however, search for tendencies within a multitude of factors that may affect the product of acquisition.

When we consider L2 phonetic-phonological acquisition, a small part of the complex, dynamic process which is SLA, we must first consider its reliance on speech perception of a *multimodal* character: multiple cues (of different nature) determine perception of segments, and these cues are not perceived by the learner in an isolated way (ZIMMER, SILVEIRA & ALVES, 2009; ZIMMER & ALVES, 2012). Moreover, certain cues 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 an acoustic cue which is not relevant in their L1 system, which makes this process even more difficult.

Zimmer & Alves (2008, 2010) stress that oral L2 production also deals with the orchestration of multiple aspects, which act together as a whole. This process comprehends

the physical and abstract levels, which go beyond binary perspectives. In other words, in the phonetic-phonological acquisition of an L2, learners have to perceive and produce cues which are relevant for the target system, facing interference from divergent roles these clues might play in the learner's L1.

In this subsection I briefly presented the main tenets of the conception of language and SLA permeating the present study. Next, I will focus on two topics whose thorough discussion is fundamental to the development of this investigation: speech perception and intelligibility. I begin by discussing speech perception and the acquisition of *Positive VOT* under this account of perception.

2.3 The role of L1 knowledge on L2 speech perception

This subsection is dedicated to the matter of L2 speech perception and, more specifically, to how it affects the acquisition of the phonetic-phonological aspect under investigation in this study: *Positive VOT*. I first discuss L2 speech perception and how it is affected by L1 knowledge. Next, the acquisition of positive VOT is addressed.

2.3.1 L2 Speech perception: the matter of L1 transfer

Two models are fundamental to explain the phenomenon of L2 segmental phonetic-phonological acquisition: the *Speech Learning Model* (FLEGE, 1995) and the *Perceptual Assimilation Model* (PAM) – *L2* (BEST & TYLER, 2007). In this study, I will regard speech perception as proposed by Best & Tyler (*op. cit.*), for this is more compatible with the dynamic conception of phonetic-phonological acquisition underlying the present study (discussed in the previous subsection).

A fundamental assumption is posited by Best & Tyler (2007) when they state that a learner's L1 and L2 phonological systems are not completely separate. The phonic elements of the learner's L1 and L2 systems interact in a common phonological space. The implication of this assumption is that, in the process of SLA, the learner tends not to perceive which acoustic categories belong to their L1 and which belong to the target language in question. Underlying this view is the conception of the complementary actions of two levels: a lower level (phonetic detail) and a higher level (phonological categories).

In the premises of PAM-L2, perceptual learning of L2 sounds is closely dependent on "the contrastive phonological and gradient phonetic properties of the L1s" (BEST & TYLER, 2007, p. 19). Once learners are faced with an L2 sound which does not belong to their L1 system, they might not *tune* their focus of attention to the appropriate "transformational invariants in the locations, amplitudes and phasings of speaker's vocal tract gestures" in the L2 (BEST & TYLER, 2007, p. 25). In other words, learners tend to follow mother-tongue cues and the role they play in their L1 systems while categorizing target-language sounds. It is important to clarify, in this sense, that within the PAM-L2 framework, the nature of L1 influence on L2 perception is based on phonetic-articulatory patterning, both in the abstract and in the phonetic-detail levels.

It should be clear, however, that, in this study, the phenomenon of L2 acquisition of sounds is regarded as *phonetic-phonological.* This is to say that this study conceives the distinction between the phonetic and phonological levels differently than proposed by PAM-L2. I will therefore situate the study in this context.

What I propose is the conciliation between the predictions of PAM-L2 on speech perception and what is put forth by Albano (2001). The author conceives the existence of a gradient between the physical phone and the phoneme. Distal articulatory gestures are the basic units of analysis, but they exist as an association between the phonetic and phonological levels – consequently, they are both physical and functional categories. This puts the phonetic and phonological levels far from a dichotomical perspective; rather, they interact with one another, in a continuum. Throughout this dissertation, I may refer to the phonetic or phonological level solely when explaining or quoting what is set by PAM-L2, but the idea of a continuum between these levels should be borne in mind.

Having situated this study in the grounds for the L2 speech perception phenomenon, I will next address the matter of the acquisition of *Positive VOT* by Brazilians under the perspective presented here.

2.3.2 The acquisition of *Positive VOT* by Brazilians

The premise discussed in the previous subsection 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 initial 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, thus, may not perceive or produce the target aspiration.

Another problem faced by L2 learners in the acquisition of the phonetic-phonological aspect in question is pointed out by Zimmer, Silveira & Alves (2009). Such 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 established between orthography and sound. More specifically, the *grapho-phonic-phonological*[10] relation in BP is transparent (orthography tends to represent pronunciation more clearly), whereas this relationship in English is much more opaque. As a consequence of the entrenched knowledge from their L1, learners tend to transfer the grapho-phonic-phonological patterns to their oral production in the L2 (ZIMMER & ALVES, 2006).

As regards the acquisition of *positive VOT* by Brazilians, grapho-phonic-phonological transfer is a factor which reinforces the lack of assimilation of acoustic cues from the target language: 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 phones represented by these graphemes in the target language (aspirated) to the ones they would represent in his/her mother tongue (unaspirated).

This is consonant with the dynamic and multimodal conception of phonetic-phonological acquisition presented in the previous subsection: both the acoustic 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 stated, both sources of input could be in competition, as the former would instantiate the L2 target forms, whereas the latter could be reinforcing the L1 pattern.

[10] Zimmer & 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.

Therefore, when considering 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 separately on theoretical grounds. Within a multimodal phonetic-phonological acquisition perspective, these factors (along with others) may contribute in equal and simultaneous ways for Brazilian learners not to acquire the L2 voicing pattern.

Another factor which reinforces the expectation that learners will not find relevance on *positive VOT* as an acoustic cue is related to the lack of negative feedback when learners interact with speakers with whom they share the same L1. It is reasonable to assume that most Brazilian learners of English who live in Brazil have most of their interaction in English with other Brazilian speakers. Since, in this context of communication, unaspirated voiceless plosives are not expected to be interpreted as voiced segments (as they might do if the interlocutor was a native speaker of English), there is no need for learners to draw their attention to this aspect and change their production towards the L2 pattern.

One more source of the aforementioned lack of negative feedback is the context in speech. Were plosive segments or words initiated by them presented in isolation, unaspirated segments would be more likely to have negative effects on communication. That is not the case, however: in most situations, the context of communication already suggests what segment is expected. If, for instance, a speaker of L2 English is driving his car, and he asks someone on the street "*Do you know a place where I can park?",* but he does not aspirate the initial /p/ in the word 'park' as a native speaker would do, because of context, it is still very unlikely that his interlocutor will understand that the L2 speaker wants to go somewhere to "bark". In this case we might be discussing *intelligibility* rather than speech perception – this is one of the main questions discussed in this study, which I will discuss in detail in the next subsection.

Overall, learners' interactions in both of these situations support the entrenchment of the L2 deviant form: firstly, there is no perception of the L2 acoustic cue due to L1 influence; in addition, there is no production of the aspect in question and, with the lack of negative feedback, there is no need for learners to change their production. Therefore, without formal instruction on this phonetic-phonological aspect of the target language, Brazilian learners of English are unlikely to produce *positive VOT*.

Recent studies investigating the acquisition of English *Positive VOT* by Brazilian learners have shown, however, that proficient learners' productions may actually be close to

the target language pattern, especially in relation to the velar consonant /k/ (REIS & NOBRE-OLIVEIRA, 2007; ALVES, SCHWARTZHAUPT & BARATZ, 2011; FRANÇA, 2011; M. ALVES, 2011; SCHWARTZHAUPT, 2012). In this case, considering the velar stop consonant, a reasonable explanation lies in the fact that, as I addressed in *subsection 2.1*, VOT values found for that place of articulation in southern BP are already distant from what can be regarded as a *Zero VOT* pattern. This would be to say that these learners might have their acquisition of English *Positive VOT* facilitated in the case of velar plosive segments. As for the production of bilabial and alveolar segments, we might think of learners as being in an *intermediate* stage of acquisition of *Positive VOT* (ALVES, SCHWARTZHAUPT, & BARATZ, 2011), which tackles the whole set of adversities discussed in this subsection.

In addition, important findings obtained from identification and discrimination tasks have suggested the action of multiple acoustic cues in the perception of VOT patterns by Brazilians (ALVES, SCHWARTZHAUPT & BARATZ, 2011; SCHWARTZHAUPT, ALVES & BARATZ, 2012; ALVES & MOTTA, 2014; ALVES & ZIMMER, 2015; SCHWARTZHAUPT, ALVES & FONTES, 2013; *in press*). Authors have argued that, while for native speakers of American English *Positive VOT* seems to be the primordial acoustic cue to account for the distinction between voiced and voiceless plosives in English, for Brazilian learners, that might not be the case: acoustic cues other than *Positive VOT* (such as burst intensity) may act together to account for this distinction. To put it simply, native speakers seem to categorize unaspirated /p/, /t/, /k/ as voiced segments, while Brazilian learners still categorize those segments as voiceless.

In this subsection, I approached the matter of speech perception at the segmental level. As it will be made clearer in the next subsection, in the present study, this is conceived as a more basic level to intelligibility; however, the boundaries between these two concepts might be somewhat blurred in some cases, as we shall see in what follows.

2.4 On Intelligibility

The main issue investigated by the present study is whether and how *Positive VOT* affects *intelligibility*. We are dealing, however, with a concept which is not at all straightforward: intelligibility has been defined in different ways by different authors. As I will discuss in the next subsections, the definition should be determined by how we assess intelligibility. I will begin by addressing the need for a clear definition of intelligibility, and then discuss the most appropriate definition for this study.

2.4.1 ELF and the need for a clear definition of "understanding"

The theoretical background I presented so far supports the claim that Brazilian learners tend not to perceive or produce the typical *Positive VOT* in their communication in L2 English. Nonetheless, as I have addressed in the introduction, it is important that we bear in mind that English is a *lingua franca*. English plays a fundamental role in global communication. It has more nonnative speakers than native ones, and it is intrinsically destined to be produced with a wide variety of different accents. To minimize the communicative value of a given accent as opposed to another is senseless; however, there must be boundaries to variation (especially when we think of pronunciation teaching), for it becomes a problem at the point in which it harms intelligibility in communication.

We therefore need empirical studies to show us what aspects of the language are fundamental for pronunciation teaching. This is important for us not to aim to simply eradicate accentedness, promoting native-like pronunciation as the only goal. Such a goal would tackle issues like the practical uneasiness of achieving it, or questions related to learner's identity (DERWIN & MUNRO, 2009).

For that reason, Jenkins (2000) proposed the LFC, which is based on empirical studies analyzing communication between nonnative speakers. Not only Jenkin's (2000) LFC, but other authors (CELCE-MURCIA *et al.*, 2010; WALKER, 2010) have given *Positive VOT* the status of a phonetic-phonological aspect which, if absent from oral production in English, may cause learners to face intelligibility problems. Several minimal pairs (*e.g.* "bet" and "pet") are distinguished through the presence or absence of aspiration in the production of

plosive consonants. In specific contexts, learners who do not produce this L2 feature may not be understood.

If we assume that L2 language teaching aims to make learners achieve intelligibility, it is crucial that we have a clear definition of what intelligibility is. Nevertheless, "to be understood" is not a straightforward concept. Many scholars have addressed the *understanding* of an utterance in different levels and contexts as *intelligibility*, *comprehensibility* or *interpretability*. I will now address each of the terms separately, comparing the main conceptions, before I can assert the definition of intelligibility adopted in the present study.

Derwin & Munro (2005; 2009) state that **intelligibility** should be defined in terms of the extent to which a listener actually *understands* an utterance. Simply put, a learner should face intelligibility problems when the listener does not understand his/her utterance. The authors recognize that there is no fully adequate way for assessing intelligibility, but propose one that should be straightforward: a sentence transcription task, in which a high percentage of sentences accurately transcribed is associated with high intelligibility. The Smith framework (1992), and also Kachru & Smith (2008), define intelligibility similarly to Derwin & Munro (*op. cit.*), although there appears to be a distinction between understanding and word recognition. The authors refer to the term as "the recognition of a word or another sentence-level element of an utterance" (KACHRU & SMITH, 2008, p. 61) – in this sense, knowledge of word usage or interpretation of speakers' intentions does not seem to be required, as highlighted by Nelson (2011). Rather, under this perspective, intelligibility seems to refer to a simple identification of the lexical items present in the utterance.

On a more complex level of communication, involving more variables, we find **comprehensibility**. Derwin & Munro (2005; 2009) define comprehensibility in terms of the effort made by a listener to understand an utterance – something to be measured on a scale from "easy" to "difficult", meaning that comprehensibility is high when the utterance is easy to understand. As suggested by Nelson (2011), this definition does not seem to go beyond a simple measurement of how difficult it is to achieve intelligibility, and is perhaps less useful as a separate term. In the Smith framework (1992), comprehensibility is said to comprise a level beyond sound identification, involving knowledge of word usage and context. This is made clearer in Kachru & Smith (2008), who state that, in this level, contextual meaning is fundamental, and there is the listener's crucial role of recognizing the speaker's intent. The authors propose that comprehensibility be measured by asking the listener to paraphrase the

utterance, or by questioning the listener with what is known as a listening comprehension question.

Finally, at the most complex level of understanding in communication, the Smith Framework (1992) and Kachru & Smith (2008) put **interpretability**[11]. This relies on the pragmatic level, and requires more than intelligibility or comprehension for one to be *understood*. Interpretability may depend on several factors, such as culture-specific ones, so that the listener can really apprehend what is meant by his/her interlocutor. In an example provided by Kachru & Smith, as a response for a caller who asks "Is Sean there?", the listener may reply "Yes, he is." – which "would be evidence of high intelligibility and comprehensibility but low interpretability" (KACHRU & SMITH, 2008, p. 64).

For all of these, it must be clear that, as observed by Nelson (2011), there is a *shared responsibility* between speakers and listeners. No speaker is perpetually intelligible to all listeners under all circumstances, especially because a great deal of *understanding*[12] relies on the listener's experiences and his/her expectations built over communication. Among other factors, a more experienced listener has had more contact with different vocabulary, various accents and situations of communication. Expectations, on the other hand, may be related, for instance, to the degree of effort that a listener is willing to invest on the interaction, especially because of a perceived accent in his/her interlocutor. This means that a speaker tends to be more intelligible to a more experienced listener, and even more so to one whose expectations are met.

As for intelligibility between speakers who share the same L1, although a shared L1 is expected to facilitate intelligibility (mostly because interlocutors share common underlying phonological systems), Kachru & Smith (2008) make an interesting point highlighting the fact that different speech communities use language differently, especially in the case of speakers with different ethnicities. Different experiences with language may lead listeners into having different expectations as to their interlocutors' language, and, therefore, intelligibility problems may, therefore, occur.

Next, I will discuss these conceptions under the perspective adopted in the present study, defining intelligibility and discussing how it can be assessed.

[11] Derwin & Munro (2005; 2009) do not discuss interpretability.

[12] *Understanding* (italicized) will hereafter be used to refer to intelligibility (which will be defined in the terms of the present study in the next subsection), comprehensibility and interpretability together; that is, to refer to all levels of understanding in communication.

2.4.2 Intelligibility defined and assessed: the value of contextual information

The previous subsection presented the background on intelligibility which is the basis for the analysis to be conducted in this study. In the first place, it should be clear that this study does not intend to test intelligibility, comprehensibility or interpretability according to the views presented. These terms are important for me to establish the definition of intelligibility adapted for the current investigation.

As I stated earlier, literature has defined intelligibility differently (DERWIN & MUNRO, 2005, 2009; SMITH, 1992; KACHRU & SMITH, 2008; NELSON, 2011). The growing consensus as to the necessity for pronunciation instruction to focus on *being understood* (as discussed in the previous subsection), led to the conduction of important empirical studies investigating intelligibility in Brazil (CRUZ, 2008; BECKER, 2013; GONÇALVES, 2014), which motivate further discussions as to how intelligibility should be defined and, especially, as to how it should be assessed.

Trying to define intelligibility, as we analyze these more traditional definitions, and refer to what was discussed in *subsection 2.3*, it might be possible to order different levels of *understanding*: *speech perception* would precede *intelligibility* by Kachru & Smith (2008), for the processing of distal articulatory gestures should be more basic than the recognition of sentence-level elements of an utterance; *intelligibility* by Kachru & Smith (*op. cit.*) would precede *intelligibility* by Derwin & Munro (2005, 2009), for the former is more clearly concerned with "recognition" of elements, while the latter relies on "actual understanding of an utterance", and that, we may assume, includes more variables. These would all obviously precede two less basic levels, as explained in the previous subsection, *comprehensibility* and *interpretability* (KACHRU & SMITH, *op. cit.*).

If we discuss these definitions individually, one might argue that the concept of intelligibility by Kachru & Smith (2008) is not very different from what other authors refer to as speech perception: Cohen (2004), Reis & Nobre-Oliveira (2007), Alves & Magro (2011), Alves *et al.*(2011) investigate VOT acquisition by assessing something addressed as *perception* through *identification* and *discrimination* tasks. Kachru & Smith (*op. cit.*), in turn, define intelligibility in terms of the recognition of "a word or another sentence-level of an utterance." It might be possible to separate speech perception as a level which is inferior to the word (the most basic domain of intelligibility) and refers to its constituents (segments). One might also question what is meant by Derwin & Munro (2005, 2009) in "actual

understanding of an utterance" – a clearer explanation of the linguistic levels involved may be required. Since we are dealing with a rather complex phenomenon, it is natural to find difficulty in assigning categorical definitions to *understanding* in communication. As I pointed out earlier, one fundamental need is that we define intelligibility in terms of how we assess it.

In the present investigation, I will adopt two concepts of intelligibility, which should be appropriate, considering the intelligibility task proposed here: **word intelligibility** and **sentence intelligibility**[13]. Word intelligibility, in this study, may be compared to the definition put forth by Kachru & Smith (2008): it relies simply on the capability of the listener to recognize the lexical items in an utterance, and it is thus assessed by means of a sentence transcription task – in that case, if the listener transcribes the utterance correctly, we assume there is high intelligibility. The distinction between segment and sentence intelligibility lies in the contextual information of the sentences in which target words are inserted. Sentence intelligibility still relies a great deal on word recognition and is still assessed in the same way, but the recognition of a target word in specific is facilitated by the contextual information in the sentence.

For example, contrasting the sentences – chosen for the conduction of the transcription task in this study (see *subsection 3.3*) – (i) "The kid drew a large *cod*" and (ii) "The fisherman caught a huge *cod*", a few observations can be made.

In sentence (i), the target *cod* could be replaced by virtually any concrete noun and the sentence would still be equally meaningful – therefore, if sentence (i) were heard in isolation, nothing in it would help the listener identify the target except for the recognition of words, which, in its turn, may rely on segmental recognition solely and is therefore analogous to speech perception (as conceived in this study) – practically, an identification task. In this sense, if the target word were pronounced with an initial plosive deviant in form from its "standard" English pronunciation – [kɑd], with no aspiration – intelligibility problems might occur, since the word in question may be confused with *God* [gɑd]. Moreover, if the target word is not recognized by an L2 learner of English, the absence of contextual information does not provide him/her with any clue as to the meaning of this word. Therefore, in

[13] Gonçalves (2014) proposed a similar approach, with a two-level analysis (one concentrated on the segmental level), in order to verify what the author addressed as intelligibility of English vowels.

sentences like (i), intelligibility may be considered to be in the word level (word intelligibility).

In sentence (ii), the target *cod* might still be replaced for a great number of concrete nouns. However, the contextual information in the sentence leads the listener into a more restricted semantic field of vocabulary, suggesting that the target word is related to *fishing* – even more so, some listeners may easily assign it to a species of fish. Considering the problematic cases presented for sentence (i), it is reasonable to assume that the contextual information facilitates the recognition of the target word in spite of the lack of aspiration in the initial plosive or even in the case in which the word is not recognized as a whole by an L2 learner of English – while the learner does not know the meaning of the word, he/she will probably consider it to be related to fishing (provided that the other words are recognized). For that reason, we may assume that intelligibility can be, in cases like this, in the sentence level (sentence intelligibility), due to a semantic interdependence between the words in the utterance, which facilitate the *understanding* of one another.

Summarizing the order of levels of *understanding* in communication I tried to establish earlier in this subsection, we might think of the following summary *Box 1*:

Box 1 - Different levels of *understanding* in communication as conceived in this study

	Level	**Author(s)**	**Definition**
Less reliance on the sentence contextual information	*Speech Perception*	Best & Tyler (2007)	Identification and Discrimination of distal articulatory gestures
	(Word) Intelligibility	Kachru & Smith (2008)	Recognition of words in an utterance; not dependent on the sentence contextual information
	Sentence Intelligibility	proposed and tested in the present study	Recognition of words[14] in an utterance; dependent on the sentence contextual information
	Comprehensibility	Kachru & Smith (2008)	Recognition of speaker's intent; dependent on contextual meaning
More reliance on the sentence contextual information	*Interpretability*	Kachru & Smith (2008)	Recognition of speaker's intent and response at the pragmatic level; highly dependent on context

In conclusion, it is important to notice that this conception of intelligibility in complementary levels should be consonant with the conception of language as a CAS. These theoretical levels are an attempt to capture nuances of intelligibility in a continuum. Multiple factors (of linguistic and extralinguistic nature) act together as a whole to determine how communication takes place and interlocutors *understand* one another.

[14] For further information as to the boundaries of what is considered as recognition of words, see *subsection 3.5.*

3 METHOD

This section presents the methodology which guided this investigation[15]. I begin by stating the goals of the study (3.1). Next, I present information about the participants who performed stimuli recording and the subsequent transcription task (3.2). The process of selection of target words and sentences for the transcription task is detailed in the next subsection (3.3). I then describe the sentence transcription task (3.4). Next, the criteria for analyzing the data obtained in the sentence transcription task are elucidated (3.5). Finally, I present the hypotheses to be tested in the present study (3.6).

3.1 Goals of this study

As already pointed out in the introductory chapter, the present study aims to investigate the effects of *Positive* VOT (with or without contextual information) on the intelligibility of short English sentences produced by Brazilians to native speakers of American English and proficient Brazilian learners. By achieving this goal, the study is expected to contribute to the discussions on the definition of intelligibility and on the necessity of instruction on *Positive VOT* for Brazilians. For that to be accomplished, this study has the following specific objectives:

- To investigate the accuracy of both native speakers of American English and proficient Brazilian learners in the transcription of sentences with or without contextual information, produced with or without Positive VOT by Brazilians.

- To verify, considering the data provided by native speakers and proficient Brazilian learners, whether there are significant differences in accuracy in the transcription of sentences considering…

    (a) only the participant's L1 as a variable;

    (b) only the VOT of the target word as a variable;

[15] The research project referring to the present study has been approved by the university's and the government's Ethics in Research Committee (*Plataforma Brasil*) under the title "*A relevância do VOT Positivo para a comunicação oral em Inglês (L2): testando o efeito da informação contextual sobre a inteligibilidade"* (code 40755015.4.0000.5347).

(c) only the presence or absence of contextual information as a variable*;*

(d) the participant's L1 and the VOT of the target word as variables;

(e) the participant's L1 and the presence or absence of contextual information as variables; and

(f) the interaction of all the three variables tested in this study as factors: the participant's L1, the VOT of the target word and the presence or absence of contextual information as variables;

- To discuss, based on the results of the transcription task, the relationship between contextual information and *Positive VOT* as cues promoting intelligibility;
- To debate the question of how intelligibility may be defined and assessed, based on the results of the transcription task;
- To discuss, based on the results of the transcription task, the need for pronunciation instruction on English *Positive VOT* for Brazilian learners.

3.2 Participants

Participants were divided into two groups: one group was responsible for recording the stimuli for the sentence transcription task, which was, in turn, performed by a second group of participants. The sentence transcription task and the procedures for data collection are explained in the next subsections.

3.2.1 Group 1: stimuli recording

The first group of participants was composed of four advanced Brazilian learners of English. Two of these participants (a male and a female) were undergraduate students of the teaching course on English Language and Literature at the *Federal University of Rio Grande do Sul* (UFRGS-Brazil) from level 5/6 required for obtaining the university degree, but no previous instruction on English phonetics. As I will explain later, these participants were expected not to produce *Positive VOT* in their reading of the target sentences for stimuli recording – therefore, they were only asked to read the sentences, with no further instruction.

The other two participants in the first group (a male and a female) were graduate students in Applied Linguistics at UFRGS. These students had had instruction on English phonetics and, at the time of stimuli recording, were explicitly instructed as to the need of producing *Positive VOT* (a concept with which they were already familiar) in the appropriate contexts.

All of the four participants were born in the Brazilian southernmost state, Rio Grande do Sul – three of them in the capital city of Porto Alegre, and one in the town of Vacaria. None of them reported to have acquired a language other than Brazilian Portuguese before reaching 6 years of age. The two undergraduate students who had had no instruction on English phonetics stated that they did not speak a Foreign Language other than English. As for the two graduate students with instruction on English phonetics, both of them stated that they were basic learners of French, while one of them also spoke Spanish (basic level) and Italian (intermediate level). The following *Table 1* summarizes other background information about the four participants, with average age, age at which they started learning English, length of time studying English, and length of residence (LOR) in an English speaking country:

Table 1 - Group 1 participants' (n=6) background information. (SD: standard deviation)

Information	Mean	SD	Minimum	Maximum
Age	24,75	2,87	21	28
Age at which started learning English	12,5	1,73	11	14
Length of time studying English (years)	12,25	2,06	10	14
LOR in an English-speaking country (months)	3	6	0	12

The four participants signed a Consent Form (*Termo de Consentimento Livre e Esclarecido – Appendix 1*) specific for their group in the study. They also filled out a Participant Information Form (*Ficha de Informações do Participante – Appendix 2*), which provided the information presented in this subsection.

3.2.2 Group 2: sentence transcription task

The second group of participants was further divided into two other groups: proficient Brazilian learners of English (hereafter PBLE) and native speakers of American English (hereafter NSAE).

The group of PBLE was composed of 12 learners (8 male and 4 female), all of whom were born in the Brazilian southernmost state of Rio Grande do Sul. None of them reported to have acquired a language other than Brazilian Portuguese before reaching 6 years of age. All of the participants were classified by the *Oxford Placement Test 1*[16] (ALLAN, 2004) as either *Proficient* or *Highly Proficient* learners of English. Most of them reported to speak a foreign language other than English, but none in an advanced level.

The following *Table 2* summarizes other background information about the twelve participants, with average age, age at which they started learning English, length of time studying English, and length of residence (LOR) in an English speaking country:

Table 2 - Group 2 Brazilian participants' (n=12) background information.

Information	Mean	SD	Minimum	Maximum
Age	26,58	6,82	21	47
Age at which started learning English	9,33	3,49	5	15
Length of time studying English (years)	17,25	6,49	10	35
LOR in an English-speaking country (months)	3,66	6,87	0	22

[16] The *Oxford Placement Test 1* consists of two main tasks: the *Grammar Test* and the *Listening Test*. In the *Grammar Test*, learners are evaluated as to their knowledge of grammatical structures of the English language through the choice of a correct answer between three which may fill in a blank in a text. In the *Listening Test,* learners are evaluated as to their capability of understanding isolated utterances, assigning one of two possible answers, which correspond to a keyword pronounced in the target sentence. Learners are given a 100 point score for each task (200 total), according to their accuracy level. Finally, learners are classified in 1 out of 10 levels, the lowest being *Beginner* (less than 75 points) and the highest being *Functionally Bilingual* (more than 198 points). Participants in the study were all classified as *Proficient* (150 to 169 points) or *Highly Proficient* (170 to 189 points). The test has been validated in over 150 countries.

The twelve Brazilian participants signed a Consent Form (*Termo de Consentimento Livre e Esclarecido – Appendix 3*) specific for their group in the study. They also filled out a Participant Information Form (*Ficha de Informações do Participante – Appendix 2*), which provided the information presented in this subsection.

As regards the NSAE group, it was composed of 6 participants (4 male and 2 female), all of whom were born in the United States of America[17] and had only acquired American English before reaching six years of age. Two of the participants stated they were proficient speakers of Brazilian Portuguese; whereas the other four were basic or intermediate learners. Furthermore, three of them reported to be speakers of Spanish in an advanced level, while another one stated he was an intermediate learner; the other two were basic learners of Spanish (1) and French (1). Participants' mean age was 37,33 (standard deviation: 15,27), ranging from 22 to 56. They were living in Brazil for a mean 1,75 years (standard deviation: 2,82). While two of the participants had been in the country for less than a month, others had been living in Brazil for years (a maximum 7).

The six North-American participants signed a Consent Form (*Appendix 4*) specific for their group in the study. They also filled out a Participant Information Form (*Appendix 5*), which provided the information presented in this subsection.

In the next subsection, I will detail the process through which the target sentences were chosen to compose the present study. I will also discuss the process of stimuli recording.

3.3 Selection of stimuli for the transcription task

3.3.1 Target words

The transcription task which provided data for the analysis conducted in this study was composed of 12 minimal pairs of monosyllabic English words. The selection was conducted with help from the *sound search* function of the *MacMillan English Dictionary for Advanced*

[17] Since all of the participants were contacted in Porto Alegre (Brazil), there was a limited number of participants and it was not possible to select participant from a single dialect of American English. The participants were from Minnesota (2), California (1), Illinois (1), Ohio (1), and Washington (1).

Learners (2007). All of these words are nouns in singular form, and none of them was presented in the dictionary as exclusive of a dialect other than American English.

Several choices were made in the selection process of these words in order not to let the words be in themselves a factor of influence in the results of the transcription task. The selected words have a C+V+C syllable structure, with an initial plosive consonant. Also, the same number of target words for each place of articulation – bilabial, alveolar and velar was selected. Furthermore, inasmuch as these are minimal pairs distinguished only by the voicing of the initial plosive, the task is going to have the same number of voiced and voiceless segments in that position.

The same equal distribution was expected as to the quality of the vowel following the plosive segment. . In accordance with what was observed by Yavas & Wildermuth (2006), França (2010), Schwartzhaupt (2012) and Prestes (2013), both the height and the frontness of the subsequent vowel have an influence on the VOT values of the stop consonant: the tendency is for subsequent high-front vowels to make VOT longer. However, a problem found with this process was that the selection of the target words does not seem to be able to account for front and back, high, mid and low vowels for all plosives. Taking that into consideration, I selected the same number of minimal pairs for each of the subsequent vowels available in English – 3 pairs for each of the 4 vowels: /ɪ/, /æ/, /ʌ/ and /ɑ/.

It is also important to explain that I chose words in which the spelling would not favor a different syllabification from its target one when read by the Group 1 of participants – the word “tomb”, for example, might be pronounced [tumb] by Brazilian learners instead of the target [tum]. Moreover, taking the same example into account, this would help prevent misspelling in the transcription made by Brazilians: when [tum] is heard, a spelling other than “tomb” might be given in response, such as “toom”.

The following *Box 2* presents the 24 target words selected for the transcription task conducted in the present study:

Box 2 - The 24 target words for the transcription task

Place of articulation	Subsequent Vowel	Target Word
bilabial	/ɪ/	pill
		bill
	/æ/	pan
		ban
	/ʌ/	pug
		bug
	/ɑ/	pox
		box
alveolar	/ɪ/	tip
		dip
	/æ/	tab
		dab
	/ʌ/	tuck
		duck
	/ɑ/	tot
		dot
velar	/ɪ/	kill
		gill
	/æ/	cap
		gap
	/ʌ/	cut
		gut
	/ɑ/	cod
		god

3.3.2 Target sentences

As I will explain in this subsection, the target words were presented to the participants in carrier sentences. After the target words were selected, each of them was inserted in a grammatical sentence for the stimuli recordings to take place. Sentences were of two types: *without contextual information* and *with contextual information* (see *subsection 2.4.2* for an explanation on how they are related to *word intelligibility* and *sentence intelligibility*, respectively).

Each of the 24 target words (e.g. *cod*) was therefore inserted in two carrier sentences: one without contextual information (e.g. “The girl drew a large cod”) and another one with contextual information (e.g. “The fisherman caught a huge cod”). Target sentences were thus

48. In order for differences between the formal linguistic aspects of the sentences not to affect participant's performance, the sentences had the same number of words[18] (six), and the number of syllables varied slightly (from 6 to 9). The sentences also had the same stress pattern, with primary stress falling on the second and last words of the sentences (subject and direct object, respectively) – "The boy found a large pill.", for example. It is also essential to clarify that the target word was always the last word of the sentence.

Distractor sentences, in which the last word was not initiated by a plosive consonant, were also created, in order to attempt to divert participants' attention from the fact that target words ended in minimal pairs with plosive segments. The number of distractor sentences was 16. Examples of final words in distractor sentences include *fan, ring, train, house, whip* and others. A complete list of target and distractor sentences is found in *Appendix 6.*

3.3.3 Stimuli recording and selection

Once the process of selection of target words and sentences was finalized, the oral recording of those sentences was performed. The recordings took place in a professional audio studio, with total acoustic isolation, in order to assure that stimuli sound quality did not interfere on listener's judgment during the transcription task.

Two groups of participants recorded the stimuli (see *subsection 3.2.1* for a thorough explanation): the first was composed of two Brazilian speakers (a male and a female) who were instructed to produce VOT levels similar to those which, according to the literature, represented the L2 standard (see *subsection 2.1)* for such words. The second group of speakers was composed of two other Brazilian speakers (two male and two female) who were not expected to produce the standard L2 VOT levels – that is, speakers who produced *Zero VOT*. It is also fundamental to make it clear that the four speakers who recorded the stimuli were not part of the group of 12 participants who performed the transcription task.

Once the recordings were conducted, the acoustic analysis of all the data was performed in order to determine the VOT values in the plosive segments produced by the participants in the target sentences. The analysis was conducted through software *Praat*

[18] There is the exception of two sentences: "the woman got another tuck", in which there are 5 words, and "the maid gave it a quick dab", in which there are 7. Such exceptions were necessary due to difficulties faced while trying to assign these target words to sentences following the overall criteria used for the other sentences.

(BOERSMA & WEENINK, 2014), and it aimed to obtain stimuli with VOT levels closest to those discussed earlier. This verification was conducted manually, and its measurement criterion was as follows: VOT was measured from the point at which the consonant burst starts to the point at which the first glottal pulse of vocal fold vibration of the following vowel was observed in the spectrogram.

The total number of stimulus-sentences used in the study, including target and distractor ones, was 112. For each minimal pair, I selected two stimuli of sentences with target words initiated by voiced segments and other two of sentences with target words initiated by voiceless segment (one with *Zero VOT* and one with *Positive VOT*). As verified in other studies (see *subsection 2.1*), however, the VOT levels produced by Brazilian learners without instruction on English phonetics tend to be higher than those typical of a *Zero VOT* pattern (although they were not high enough to reach the level of aspiration typical of the L2 for each place of articulation). In this sense, it is important to make it clear that such high VOT values were expected. For that reason, the stimuli with the lowest VOT values were selected, even though some of them were not considered as *Zero VOT* productions. As for the plosive segments produced by the group of speakers who were instructed to produce *Positive VOT*, it can be said that many productions were exaggerated (that is, were produced with higher levels of VOT than required), and some of these had to be selected for the experiment. It is also important to clarify that there was approximately the same number of stimuli per speaker (see *subsection 3.1.1*). The following *Table 3* presents the mean VOT levels of the voiceless segments used as stimuli in this investigation according to their place of articulation:

Table 3 - Mean VOT levels and standard deviation (SD) of the stimuli for the transcription task. N = 4.

Plosive segment	Sentence context	Mean VOT (ms) unaspirated	SD unaspirated	Mean VOT (ms) aspirated	SD aspirated
/p/	*without info*	13,5	4,62	72,93	18,27
	with info	13,36	3,90	76,49	18,94
/t/	*without info*	45,85	12,82	95,45	9,56
	with info	43,32	3,61	80,44	9,20
/k/	*without info*	58,99	6,16	89,02	8,87
	with info	54,48	9,13	94,42	3,06

It is important to clarify that, although half of the target stimuli are sentences initiated by voiced plosive segments, the VOT of these productions was not measured. This was the case because, in the first place, these segments were not expected to have any effect on perception between groups: as verified by Schwartzhaupt *et al.* (2013; *in press*) NSAE and PBLE do not differ as to their perception of segments with either *Negative VOT* or *Zero VOT* for voiced segments. Moreover, measurement of such VOT patterns is made almost impossible in the context of the sentences selected for this study, in which the initial plosive of the target word is preceded by a vowel.

In the next subsection, I will detail the sentence transcription task which provided data for the analysis conducted in this study.

3.4 The transcription task

The analyses conducted in this study were based on the results obtained from a sentence transcription task. The task was elaborated in software *E-Prime* 2.0 (SCHNEIDER, ESCHMAN & ZUCCOLOTTO, 2012)[19]. It is important to mention that the use of this software has an innovative character in relation to other studies conducted for the investigation of factors affecting intelligibility (e.g. CRUZ, 2008; BECKER, 2013; GONÇALVES, 2014). Among others, research can benefit from the use of this software for a) typed transcriptions may be more objectively verified than handwritten ones; and b) the software allows the researcher to determine a limited response time, after which the participant can no longer transcribe the sentence.

The participants' task was to read the instructions presented in the computer screen, hear the 112 stimulus-sentences (48 sentences with contextual information, 48 sentences without contextual information and 16 distractor sentences) through headphones[20] and type them in the computer. Before the 112 experimental trials – which were presented in a different random order for each participant –, participants had 3 practice trials to help them

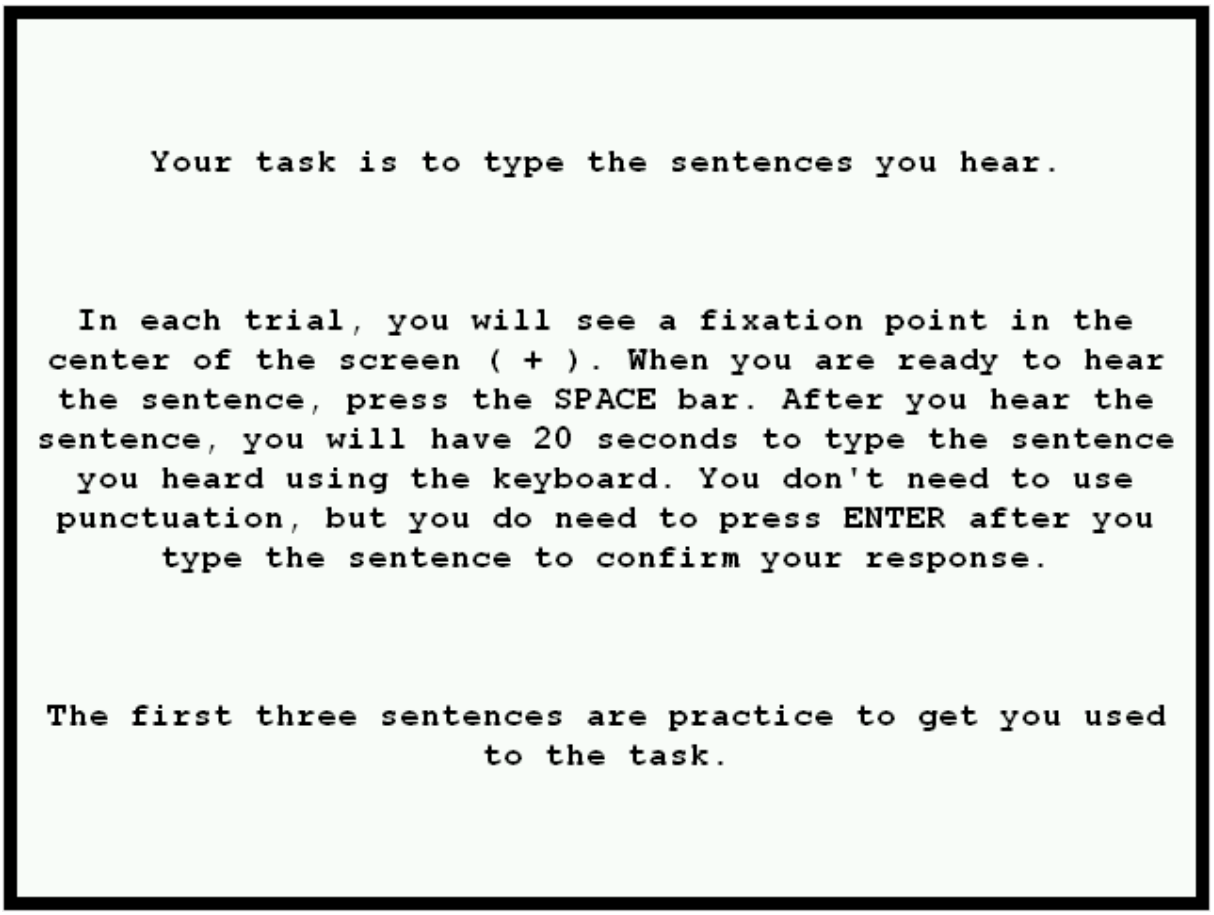
[19] *E-Prime* is a set of applications developed by *Psychology Software Tools Inc.* used to design, generate and run behavioral experiments in computer environment. The program allows researchers to collect, edit and analyze data in its interface, comprising audio, image and video files.

[20] All participants used a *Goldship* headphone model 1742. Its specifications are as follows: frequency response: 20 – 20 KHz; impedance: 32 Ohm; maximum power input: 100 mW; sensitivity: 91 dB.

get used to the task. Practice trials were composed of three distractor sentences which would later appear on the experimental trials. The procedure lasted for about 30 minutes, depending on each participant's performance. The next paragraphs detail the procedure step-by-step.

On the first screen, participants had general instructions as to how to proceed during the transcription task. At this point, they were already wearing headphones. The following *Figure 2* displays the task's initial screen:

Figure 2 - E-Prime instructions screen 1

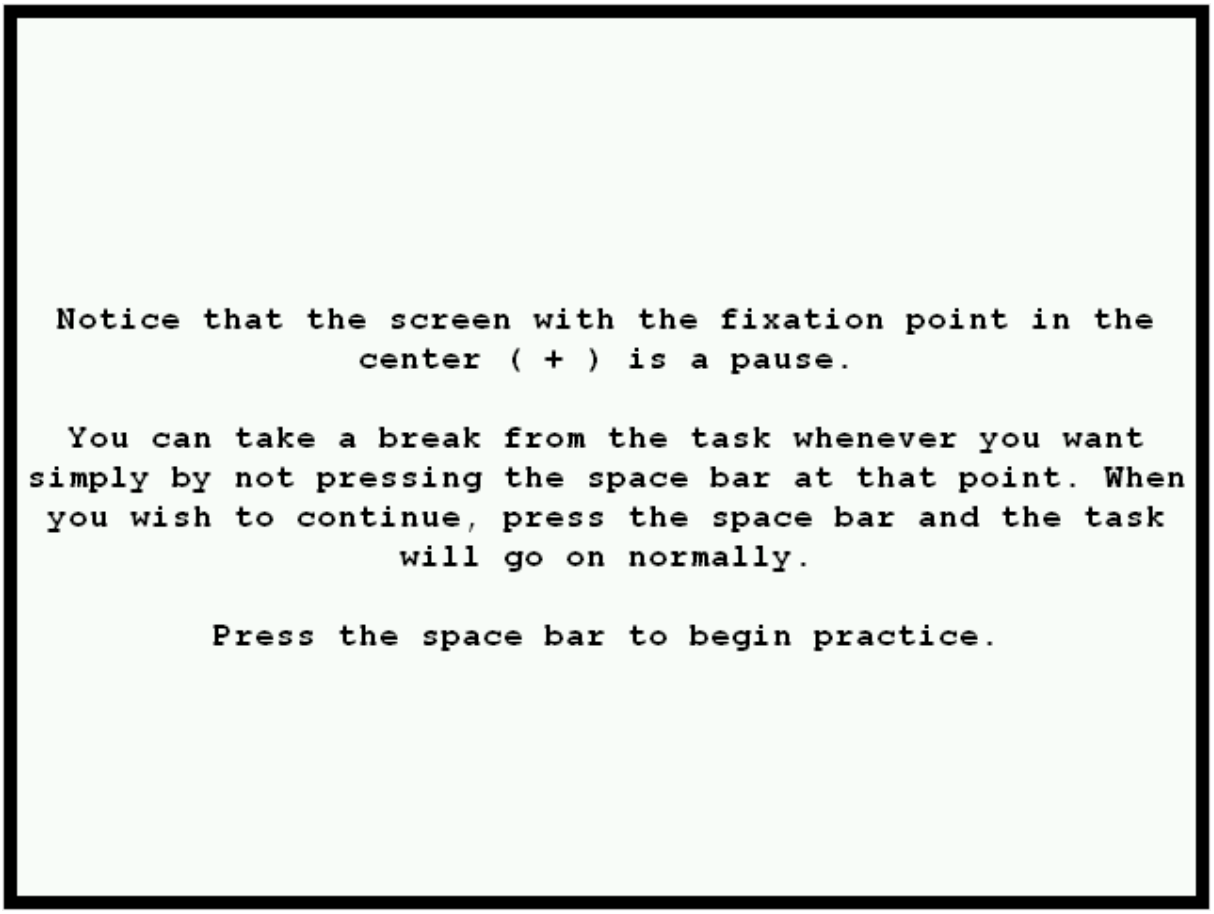


As it can be observed on *Figure 2*, participants had 20 seconds to transcribe the sentences – in case ENTER was not pressed after 20 seconds, whatever had been typed would be saved as a response. This choice was made with the intent to obtain more automatic responses, whilst giving the participants enough time to transcribe each sentence[21]. The use of capitals and punctuation was not considered important for this task, and participants were

[21] The observation of the data obtained from the transcription task suggests that 20 seconds were, as expected, enough time for participants to transcribe the sentences completely.

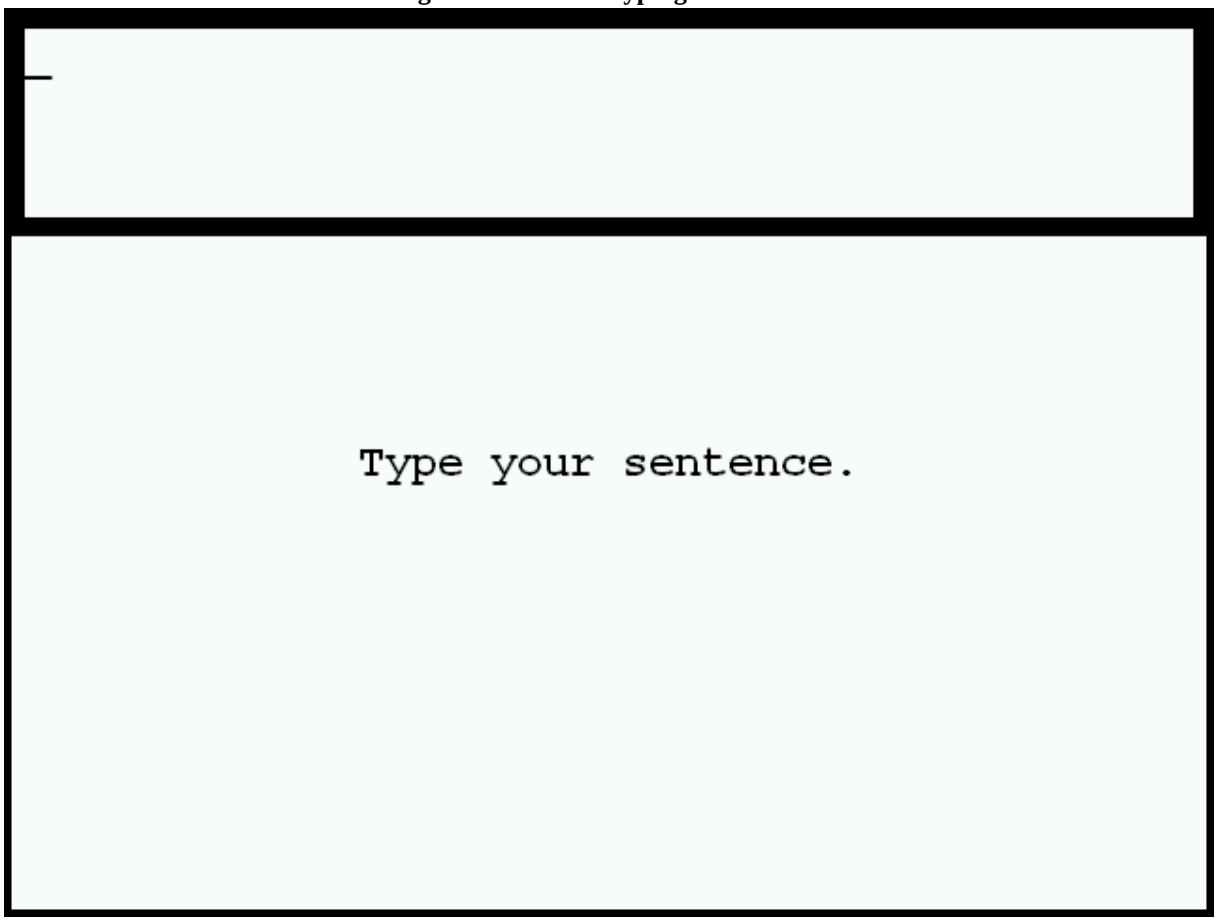
made aware of that. The next *Figure 3* displays the second screen of the task, which provided further instructions:

Figure 3 - E-Prime instructions screen 2



As it is made clear by *Figure* 3, participants were incentivized to take breaks whenever they felt it was necessary. It is interesting to notice, however, that no participant but two chose to take a break longer than a few seconds and stand up. At this point, all participants were also asked by the researcher the following: "Do you have any questions? Please type what you *understood* from the sentences, and notice that not understanding something is part of the task." No further explanation was provided to the participants about the sentences or how to respond to the task. *Figure 3*, which follows, shows the final screen of each trial, which was the screen with the typing field:

Figure 4 - E-Prime typing field screen



In the next subsection, I will detail the process of data analysis conducted with the results obtained from the task.

3.5 Criteria for data analysis

After each participant finalized the experiment on E-Prime, a file was generated, detailing exactly what had been typed in response to each stimulus in the transcription task. This subsection explains in detail how these responses were analyzed.

In order to make an objective analysis of this data, it was important to establish some criteria to assign the transcription to the binary *low-to-no intelligibility* versus *high intelligibility* categories. Before considering these data, some observations must be made.

Firstly, one must consider that the relationship between *understanding* an utterance and being able to transcribe it is not as straightforward as it may seem when we first consider it. In other words, if one is not able to transcribe a sentence, that does not necessarily mean he/she does not understand it. A participant's performance in this kind of task does not rely solely on speech perception or intelligibility. Rather, it depends fundamentally on the participant's ability to recall the auditory information to which he/she was exposed and transposing it into written language – a complex executive function in which participants certainly differ. Moreover, correct spelling depends on more than just recognizing vocabulary and being able to assign meaning to it; it also depends on being able to recall the set of orthographic symbols associated with it correctly. Correct spelling is therefore not a requirement for intelligibility *per se*. In short, not being able to recall an utterance completely may depend on factors which are external to the scope of speech perception or intelligibility, and therefore any generalizations made with respect to this analysis must be limited.

It should also be clear, in this sense, that if we assume that with the task we propose we can actually assess intelligibility, within a dynamic perspective, the number of factors to be considered as causes of intelligibility problems is perhaps not possible to determine – it goes well beyond the contextual information of the sentences and VOT.

These points considered, I now discuss what should be considered as transcriptions that are indicators of high intelligibility in the task conducted in this study and why. The first and obvious exemplar would be the case in which the participant transcribes the identical sentence which was read by the participant who recorded the stimulus, representing each and every word with correct spelling. The few mistakes to be admitted within the conception of binary *low-to-no intelligibility* versus *high intelligibility* categories are slight variations of this "faithful" representation of the stimulus. As an example, we have "The fisherman caught a huge cod." as both the sentence read in stimulus recording and the response provided in the sentence transcription task – although the use of punctuation and capital letters are not necessary.

The next transcription to be accepted as an example of high intelligibility is the one in which there is a small typing mistake. Determining what is a case of typing mistake depends, of course, on a more subjective analysis; still, many examples can reasonably be assumed to be evidence of high intelligibility. For the example in which we have "The fisherman caught a huge cod." as stimulus, the responses "**hte fisherman caught a huge cod"*, *"*the fosherman*

caugt a huge cod" and even *"*the fishermann caught ahuge cod"*[22] are to be accepted as cases of high intelligibility within what is conceived by this study.

Other cases in which transcriptions distinct from the faithful "the fisherman caught a huge cod" are accepted are related to variation in *function words*. As opposed to *content words*, function words are not responsible for conveying the fundamental information in the utterance. Moreover, they might also be produced in their "weak" forms, which might affect their recognition by the listener. Function words in the target sentences selected for this study were 7: *a, an, the, his, her, their,* and *for*. Therefore, some deviant transcriptions are to be accepted as evidence of high intelligibility in the transcription task: a) *a/an* transcribed as *the* and vice-versa; b) *for* is transcribed as *to* and vice-versa; c) one function word is omitted from the transcription, e.g. "**the fisherman caught huge cod"*. In addition, *his, her,* and *there* are to be considered as interchangeable.

One more case in which deviant transcriptions are to be accepted as evidence of high intelligibility is related to grammatical errors. Specifically, this makes reference to errors in the spelling of verb forms. Most of the target sentences in this study contain verbs in the past, which might be a problem for both native speakers and L2 learners of English. In this case, deviant forms of both regular and irregular verbs may be accepted as evidence of high intelligibility, as long as these deviant forms do not stand for different words in the English language. For the targets "The churchman prayed for their God." and "The girl hid her big gut.", the transcriptions "**the churchman praied for their god*" and "**the girl hide her big gut*" are to be accepted as evidence of high intelligibility.

All deviant transcriptions which do not fit the examples discussed above are to be regarded as evidence of *low-to-no intelligibility*. Some cases of these include: a) recognition of a different content word in the sentence, especially of the other member of the minimal pair in the target word (e.g. "*the man found a small pill"* for the target *"the man found a small bill"*); and b) omission of a content word from the transcription (e.g. *"*the man a small bill"*).

In comparison to the criteria proposed by other studies carried out in Brazil, I highlight similarities in terms of acceptance of spelling and grammatical mistakes to the criteria used by Becker (2013). One important difference is found between what is proposed in this study and

[22] The examples of deviant transcriptions presented here are a product of both what was hypothesized before the analysis and the patterns found after the analysis of the transcriptions was conducted. All examples are based on actual deviant transcriptions found in the data analysis.

what is found in Gonçalves (2014). Gonçalves (*op. cit.*) proposes that, if misspelling led to a homophonous word (e.g. *"beet"* for the target *"beat"*), the transcription be regarded as correct (evidence of high intelligibility). This seems to be the reflex of a different conception of intelligibility from the one proposed in this study. As discussed in *subsection 2.4.2*, this investigation is concerned with (a) recognition of all the words in the sentence (*word intelligibility*) and (b) understanding of the interdependence of elements connected by meaning at the sentence level (*sentence intelligibility*). To admit homophonous words as evidence of high intelligibility would be to assess sound rather than sound and meaning, as what is meant by word intelligibility.

In the next subsection, I present the hypotheses to be tested in this study, before I can proceed to the results and analysis section.

3.6 Hypotheses

Once the accuracy levels of all participants were verified in the transcription task, the following hypotheses were tested. These hypotheses were established with basis on findings and discussions conducted in studies presented in the theoretical background section of this dissertation.

When we contrast the two groups of participants, I hypothesize that NSAE will rely more on *Positive VOT* as an acoustic cue in order to achieve sentence intelligibility than PBLE (SCHWARTZHAUPT, ALVES & FONTES, 2013; *in press*). Consequently, they will face more intelligibility problems in the absence of this factor than PBLE. Moreover, with regard to initial plosives, PBLE will rely on acoustic cues other than *Positive VOT* in order to achieve word intelligibility, since this does not seem to be the main cue in the distinction between voiced and voiceless segments in BP (SCHWARTZHAUPT, ALVES & BARATZ, 2012; ALVES & MOTTA, 2014; ALVES & ZIMMER, 2015). As for the contextual information in the sentence, I hypothesize that it will be equally important for achieving intelligibility, as considered by both groups. In the absence of *Positive VOT* as a factor to promote intelligibility, contextual information should account for the voiced x voiceless plosives distinction – therefore, both groups will not have significantly different performances

in this[23]. In order to test the contrast of data between the two groups of participants, the stimuli can be categorized as follows:

(i) sentences with contextual information and a voiced target segment;
(ii) sentences without contextual information and a voiced target segment;
(iii) sentences with both contextual information and *Positive VOT*;
(iv) sentences with contextual information, but without *Positive VOT*;
(v) sentences without contextual information, but with *Positive VOT*;
(vi) sentences with neither contextual information nor *Positive VOT*.

Condition (vi) is therefore central to the analysis conducted in the present study, for, in the lack of contextual information, NSAE are expected to interpret unaspirated plosives (without *Positive VOT*) as voiced segments, which in turn is expected to pose a problem to overall sentence intelligibility.

In this sense, the following hypotheses *H1, H2, H3, H4, H5, H6,* and *H7* were postulated:

H1: The analysis contrasting the two groups of participants as to their overall accuracy levels in the sentence transcription task (ignoring VOT pattern and contextual information as variables) will show that PBLE are significantly more accurate than NSAE.

H2: The analysis contrasting all participants' accuracy levels considering only *Positive VOT* as a variable in the sentence transcription task (ignoring native language and contextual information as factors) will show that (a) there is a significant difference in accuracy in the transcription of sentences with plosives produced with *Negative VOT* and *Zero VOT* - participants are significantly more accurate in the former, for NSAE poor performance in the former has an effect over accuracy levels; (b) there is no significant difference in accuracy in the transcription of sentences with plosives produced with *Negative VOT* and *Positive VOT*; (c) contrasting the transcription of sentences produced with *Zero VOT* and *Positive VOT*, participants are significantly more accurate in the latter, for NSAE poor performance in the former has an effect over accuracy levels.

[23] It is important to notice that this has not been explicitly tested before, and to establish such a judgment of the status of *Positive VOT* as opposed to contextual information for intelligibility may be a contribution of the present study.

H3: The analysis contrasting all participants' accuracy levels considering only contextual information as a variable in the sentence transcription task (ignoring native language and contextual information as factors) will show that participants are significantly more accurate in the transcription of sentences with contextual information than in the transcription of those without such clue.

H4: The analysis comparing PBLE and NSAE as to their accuracy levels considering only VOT as a factor (ignoring contextual information) will show that the groups differ in the transcription of sentences with plosives produced with *Zero VOT*, but not in the one with those produced with *Negative* or *Positive VOT*: PBLE will be significantly more accurate in the transcription of sentences with *Zero VOT*.

H5: The analysis comparing PBLE and NSAE as to their accuracy levels considering only contextual information as a factor (ignoring VOT pattern) will show that the groups do not differ in the transcription of sentences with contextual information, but that they are different in the transcription of those without contextual information – PBLE will be significantly more accurate in this case.

H6: The analysis contrasting all participants' accuracy levels considering both contextual information and VOT as factors will show that participants only differ significantly in the condition 'sentence without contextual information produced with *Zero VOT*', in which accuracy levels will be lower – especially for NSAE. In all other five conditions[24] participants will not be significantly different.

H7: The analysis comparing PBLE and NSAE as to their accuracy levels considering both contextual information and VOT as factors[25] will show a significant difference in transcription of sentences for NSAE only in the condition in which there is no contextual information and *Zero VOT* – because their performance will be poorer –, whereas no significant differences will be found in the transcriptions by PBLE.

[24] There are six conditions overall, as presented in the beginning of this subsection, considering the combinations of two variables in contextual information (*with* and *without*) and three VOT patterns (*Negative*, *Zero*, and *Positive*).

[25] Thus including all twelve conditions, product of the combinations of two variables in participant's L1 (*English* and *Portuguese*), two variables in contextual information (*with* and *without*), and three VOT patterns (*Negative*, *Zero*, and *Positive*).

Summarizing the ideas underlying these hypotheses, the data analysis is expected to reveal that, for NSAE, the absence of *Positive VOT* in target /p/, /t/, and /k/ only poses a problem for sentence intelligibility in the lack of contextual information. As for PBLE, they are not expected to face intelligibility problems: in the lack of contextual information, the absence of *Positive VOT* should be compensated by complementary acoustic cues.

4 RESULTS AND DISCUSSION

This subsection presents the results from the transcription task detailed in *subsection 3.4*. I will begin by presenting an overall descriptive analysis and comments on all the data obtained from the two groups of participants (PBLE and NSAE, as explained in *subsection 3.2*) considered together. Next, I will address the statistical analysis of a comparison of performance between the two groups, testing and commenting on the results of each of the seven hypotheses (see *subsection 3.6*) in a different subsection. A summary of the results obtained from this study and their implications is presented in the final subsection.

4.1 Overall descriptive analysis

The analysis of all the data obtained from the transcription task takes into consideration 1152 transcriptions by PBLE (96 target sentences x 12 participants) and 576 transcriptions by NSAE (96 target sentences x 6 participants). The following *Table 4* displays overall accuracy in the sentence transcription task considering contextual information as a variable:

Table 4 - Overall accuracy in the transcription task considering contextual information

	sentences with contextual info	sentences without contextual info	overall accuracy
PBLE	72.91% (420/576)	72.04% (415/576)	72.48% (835/1152)
NSAE	86.80% (250/288)	77.43% (223/288)	82.11% (473/576)

As we observe the overall descriptive analysis presented in *Table 4*, the suggestion is that NSAE were more accurate in transcribing the sentences. This result raises several questions. For one thing, this might contradict the expectations that intelligibility would be high in the interaction between learners and speakers with a common L1 – it would be reasonable to assume that intelligibility would come to be higher between Brazilians, since they share a common L1 and, therefore, the same phonological system. One might suggest that NSAE performed better because of the target words involved in the sentences transcribed

in the task (see *subsection 3.3.1*): some of the words might be considered not frequent enough for learners to recognize them.

It would also be reasonable to suggest that the auditory information of the stimuli would be more rapidly assimilated by NSAE than PBLE, for the former tend to be more familiar with both vocabulary and the sentences presented in the task. Moreover, such an expected familiarity with vocabulary might cause NSAE to face less difficulty in transcription because of spelling. In this sense, it should be made clear that cases in which transcriptions were not finalized – that is, cases in which 20 seconds were not enough for participants to transcribe the sentences – were very rarely observed in data analysis.

When considering that accuracy levels are virtually the same for PBLE, whereas NSAE have a better performance in the cases of stimuli with contextual information, it seems possible that native speakers direct more of their attention to meaning rather than sound, while L2 learners tend not to do so[26]. In that case, we might explain why performance of NSAE was more affected by the lack of contextual information. It is important to consider that, without contextual information, sentences may seem too "artificial", in the sense that they are not frequently found in the language, except for very specific and contextualized situations. In this sense, we might be talking about the possibility that native speakers expect to find meaningful sentences in all contexts and, when that does not happen, the sentence is thought to be incorrect or pronounced wrongly.

As an example of this situation, I would like to mention the case in which one of the NSAE participating in this study commented on the task after having finished it. The participant said to find it funny that *"there was one sentence in which he kept saying 'gill', when I know he meant 'girl'"*. The sentence to which the participant was referring to, I assume, was "The boy saw an ugly gill", which appeared twice in the transcription task. It is interesting to observe how (a) "The boy saw an ugly girl" is a much more meaningful – and certainly more frequent – sentence, especially for the semantic approximation of the words *boy* and *girl*[27]; and (b) in this case, having noticed that the interlocutor was not a native

[26] Van Patten (2002) suggested that advanced learners tend to direct their attention to finer aspects of speech, as opposed to basic learners, who focus on more global aspects of communication. This suggestion might meet what is suggested by the author in relation to advanced learners: since performance of PBLE was virtually the same in transcription of sentences with or without contextual information, it is possible that NSAE direct their attention to meaning, but PBLE make use of a mixed criterion, looking for both meaning and sound.

[27] More evidence of this case could be found in the fact that the sentence "The lady got a new duck" was transcribed as "The lady got a new jug" by another speaker.

speaker, the listener was expecting the L2 learner to make pronunciation mistakes – we should notice that "girl" is one of the first words to pose a problem for Brazilian learners of L2 English because of the production of the retroflex [r] followed by the velarized [ɬ], both not common in most BP dialects.

The following *Table 5* presents overall accuracy in the transcription task, considering VOT as a variable:

Table 5 - Overall accuracy in the transcription task considering VOT

	sentences with voiced segments	**sentences with unaspirated voiceless segments**	**sentences with aspirated voiceless segments**
PBLE	75.86% (437/576)	71.87% (207/288)	72.48% (835/1152)
NSAE	84.02% (242/288)	80.55% (116/144)	79.86% (115/144)

The accuracy levels presented in *Table 5* might go against expectations, if we consider that *Positive VOT* was not relevant for promoting higher intelligibility by NSAE (as discussed earlier in *subsection 3.6,* that is not the case for PBLE, who very likely direct their attention to complementary cues). It might be the case that contextual information is actually accounting for the intelligibility problems NSAE may face for the absence of *Positive VOT*. It is also possible that the phonetic-phonological material in the sentence – i.e. deviant pronunciation forms in words other than the target one – be responsible for that effect. Moreover, it is interesting to observe that the sentences with voiced segments, in which contextual information is the only variable at stake between the two groups' performances, reinforce the suggestion that NSAE performed better in the transcription task.

In order to discuss these suggestions further, I now proceed to the testing of the hypotheses presented in the previous subsection, through which I will attempt to explain the roles of contextual information and *Positive VOT* as cues for achieving intelligibility.

4.2 Hypotheses tested

In this subsection I present the statistical analyses used to test the hypotheses proposed in *subsection 3.6*. In order to report results more objectively, stimuli are divided into the six categories that follow:

(i) sentences with contextual information and a voiced segment;

(ii) sentences without contextual information and a voiced segment;

(iii) sentences with both contextual information and *Positive VOT*;

(iv) sentences with contextual information, but without *Positive VOT*;

(v) sentences without contextual information, but with *Positive VOT*;

(vi) sentences with neither contextual information nor *Positive VOT*.

The means to be tested in the hypotheses (on 1152 tokens of transcriptions of PBLE and 576 of NSAE) are related to the accuracy levels found as organized according to these six categories, as displayed in the following *Table 6.* Tokens for conditions i and ii are 288 for PBLE and 144 for NSAE, but 144 for PBLE and 72 for NSAE in all other conditions[28]:

Table 6 – Mean accuracy in the transcription task according to six types of stimulus. SD = standard deviation.

	condition i	**condition ii**	**condition iii**	**condition iv**	**condition v**	**condition vi**
PBLE (N=12)	77.79% (SD: 12.02%)	73.95% (SD: 13.75%)	67.33% (SD: 11.41%)	68.75% (SD: 14.25%)	65.25% (SD: 10.50%)	75% (17.33%)
NSAE (N=6)	89.50% (SD: 6.79%)	78.45% (SD: 9.62%)	77.75% (SD: 18.0%)	90.25% (SD: 6.25%)	81.91% (SD: 17.75%)	70.83% (8.66%)

Before proceeding to the statistical tests considering these twelve conditions, it should be mentioned that the *Shapiro-Wilk Test*[29] was conducted, in order to verify whether the data presented a normal distribution. The assumption of normality was met for this sample for all the conditions tested ($p > .05$). Hypotheses are approached individually in each of the next subsections.

[28] As explained in *subsection 3.3*, the study has the same number of sentences with a target voiced plosive and a target voiceless plosive. Voiceless plosives, however, are divided into aspirated and unaspirated segments.

[29] The Shapyro-Wilk Test makes use of the null-hypothesis principle in order to verify whether a sample came from a normally distributed population. If the null-hypothesis is confirmed, the data is in normal distribution, which means that a parametric statistical analysis should have better statistical power.

All of the hypotheses were tested through a *Repeated Measures Analysis of Variance* (hereafter rANOVA)[30], considering the three independent variables (*participant's L1* as a between-subjects factor, and *contextual information* and VOT pattern as within-subjects factors). The test considering all variables and their interactions pointed out that significant differences were found only in the contrasts of variables 'VOT pattern versus participant's L1' [$F(2, 15) = 7.904$; $p = .005$] and 'VOT pattern versus contextual information' [$F(2, 15) = 5.623$; $p = .019$]. The next subsections present the aNOVA results and follow-up Bonferroni's *pairwise comparisons* and T-tests used (when necessary) in order to test each of the seven hypotheses individually.

4.2.1 Testing Hypothesis 1

In *H1 (subsection 3.6),* I hypothesize that the analysis contrasting the two groups of participants as to their overall accuracy levels in the sentence transcription task will show that PBLE are significantly more accurate than NSAE.

An rANOVA was conducted (ignoring *Positive VOT* and contextual information as variables) and a main effect was verified [$F(1, 16) = 6.285$; $p = .023$]. A follow-up test (a Bonferroni's *pairwise comparison*) showed that there is a significant difference between the accuracy of NSAE (mean = 81.48; standard error = 3.29)[31] and PBLE (mean = 71.35; standard error = 2.33) – meaning that native speakers actually outperformed Brazilian learners in the transcription task.

Hypothesis 1 was therefore not corroborated: while, as expected, NSAE and PBLE do differ significantly in their overall accuracy in the sentence transcription task, inversely to what was hypothesized, NSAE had a better performance than PBLE.

Such a result was rather unexpected. *Hypothesis 1* was motivated by the assumption that a shared L1 should promote high levels of intelligibility: if the interlocutors have the same L1, they are expected to behave similarly in terms of the cues to which they direct their attention in L2 speech perception and production. This should, therefore, facilitate oral communication between them. The results from this test indicate, however, that NSAE were

[30] 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.

[31] Such values refer to a percentage of accuracy in the transcription task.

the ones who actually *understood* the Brazilian speakers who recorded the target sentences better.

As I addressed in the previous subsection, it is possible that NSAE performed better because of the target words involved in the sentences transcribed in the task, for some of the words might be considered not frequent enough for learners to recognize them. A more rapid assimilation of the auditory information presented in the task by NSAE, who are more familiar with the vocabulary and possibly with real productions of the sentences in question, is also a reasonable assumption.

4.2.2 Testing Hypothesis 2

Hypothesis 2 (see *subsection 3.6*) predicted that the analysis contrasting all participants' accuracy levels considering only *Positive VOT* as a variable in the sentence transcription task would show that (a) there is a significant difference in accuracy in the transcription of sentences with plosives produced with *Negative VOT* and *Zero VOT* – participants are significantly more accurate in the former, for NSAE poor performance in the latter has an effect over accuracy levels; (b) there is no significant difference in accuracy in the transcription of sentences with plosives produced with *Negative VOT* and *Positive VOT*; (c) contrasting the transcription of sentences produced with *Zero VOT* and *Positive VOT,* participants are significantly more accurate in the latter, for NSAE poor performance in the former has an effect over accuracy levels.

The rANOVA (ignoring participant's L1 and contextual information as variables) indicated that there was no significant difference between the accuracy of participants considering the three VOT patterns: *Negative VOT* (mean = 74.56; standard error = 2.69), *Zero VOT* (mean = 78.64; standard error = 2.46) or *Positive VOT* (mean = 76.04; standard error = 2.40) [$F(2, 15) = 1.516$; $p = .251$] – meaning that the VOT pattern did not have an effect on participant's performance, when they are considered together.

Hypothesis 2 was thus not corroborated: the participants did not have significantly different accuracy levels in the transcription of sentences while we contrast the aforementioned cases (a), (b) or (c).

As I stated earlier (subsections *2.3* and *3.6*), studies have shown that Brazilian leaners tend to identify English voiceless plosives produced with *Negative VOT* and *Positive VOT*

easily, reaching ceiling effects (ALVES *et al.*, 2011, 2012; SCHWARTZHAUPT *et al.*, 2013; *in press*; ALVES & MOTTA, 2014; ALVES & ZIMMER, 2015). In the present study, this led to the assumption that PBLE would perform similarly to NSAE in the transcription of sentences with these VOT patterns. Nonetheless, it was predicted that the absence of *Positive VOT* as an acoustic cue would cause NSAE to have intelligibility problems with sentences in which there was no contextual information, since they would have poor identification of target words without the necessary aspiration – that does not seem to have been the case, at least not enough to cause accuracy levels in the transcription of sentences according to VOT to be significantly different. In this sense, it is important to notice that the methodology adopted in the present study differs from the ones found in the literature, for the target words are not presented in isolation and, more importantly, the "identification"[32] of the plosive segments is being assessed through sentence transcription, and therefore taking into account many other factors. This may, therefore, be the source of divergence between what is found in previous studies and in the present one.

Perhaps one way of approximating these results to the ones of previous studies would be to examine the transcriptions analyzing accuracy in transcription of the target word only – not presenting this data might be regarded as a limitation of the present study. As impressionistic data, however, it is rather important to mention that, as the analyses were conducted, very few cases (less than 5) were observed in which the lack of Positive VOT led to a problem of word intelligibility related to the identification of a voiceless segment as a voiced one by PBLE or NSAE. This goes against what was expected with the methodology proposed by this study, and might be a suggestion of the action of factors other than Positive VOT or contextual information (such as discussed in subsection 3.5) determining accuracy in the transcription task.

4.2.3 Testing Hypothesis 3

In *Hypothesis 3* (*subsection 3.6*), it was predicted that the analysis contrasting all participants' accuracy levels considering only contextual information as a variable in the sentence transcription task would show that participants are significantly more accurate in the

[32] It is essential to clarify that the aforementioned studies tend to address this phenomenon as a matter of speech perception, whereas I discuss recognition within the scope of word intelligibility (subsection 2.4.2).

transcription of sentences with contextual information than in the transcription of those without such a clue.

In order to test the hypothesis, an rANOVA was conducted (ignoring participant's L1 and VOT pattern as variables), but no main effect was verified [$F(1, 16) = .001$; $p = .982$]. Results showed, therefore, that there are no significant differences between the performances of the participants in the transcription of sentences with contextual information (mean = 76.38; standard error = 2.35) and without contextual information (mean = 76.44; standard error = 2.44).

Hypothesis 3 was therefore not corroborated: participants' accuracy levels in the transcription of sentences with and without contextual information are not different.

Against expectations, the lack of contextual information did not pose a difficulty for participants to transcribe sentences accurately. The next hypotheses are expected to provide more relevant information on how contextual information was assimilated in the transcription task – especially through the comparison between groups – so that this question can be discussed further.

4.2.4 Testing Hypothesis 4

In *H4*, I hypothesized that the analysis comparing PBLE and NSAE as to their accuracy levels considering only VOT as a factor would show that the groups differ in the transcription of sentences with plosives produced with *Zero VOT*, but not in the one with *Negative* or *Positive VOT*, for PBLE would be significantly more accurate in the transcription of sentences with *Zero VOT*.

An *Independent Samples T-Test*[33] was conducted in order to test this hypothesis. The T-Test indicated that the two groups did not differ significantly in the transcription of sentences produced with *Negative VOT* (NSAE = mean 84.02, SD 6.67; PBLE = mean 75.86, SD 10.60) [$t(16) = -1.709$; $p = .107$] or *Zero VOT* (NSAE = mean 80.55, SD 5.04; PBLE = mean 71.87, SD 12.45) [$t(16) = -1.622$; $p = .124$]. As regards the transcription of sentences produced with *Positive VOT*, however, results showed that NSAE (mean 79.86; SD 10.67) and PBLE (mean 66.31; SD 9.13) have significantly different performances [$t(16) = -2.808$; p

[33] The *Independent Samples T-Test* is a statistical approach to determine significant differences between two sets of independent and identically distributed samples.

= .013]. Such results suggest that, while groups differ in their overall accuracy in the sentence transcription task (see *4.2.1 Testing Hypothesis 1*), NSAE only actually outperform PBLE in their transcriptions of sentences produced with *Positive VOT.*

Considering this result, *Hypothesis 4* was not corroborated: groups are significantly different in the transcription of sentences with *Positive VOT*, not in the one of sentences with *Zero VOT*.

NSAE were expected to perform poorly in the transcription of sentences with *Zero VOT* because, in sentences without contextual information, in which both members of the minimal pair initiated by a plosive are good candidates for the target word, the absence of *Positive VOT* would lead participants into transcribing target words initiated by voiceless segments as voiced ones (e.g. “the boy found a large bill” for the target “the boy found a large pill”). Such a difficulty faced by NSAE should be enough to make their performance significantly poorer than that of PBLE, who would make use of complementary acoustic cue to account for the distinction between voiced and voiceless segments (ALVES & MOTTA, 2014; ALVES & ZIMMER, 2015). That was not, however, the case (significant differences were not found).

Nonetheless, we must also consider how *unaspirated* the stimuli in the present study actually were (see *subsection 3.3*). As I pointed out , although the VOT levels produced in the sentences selected as stimuli were lower than those considered as standard of voiceless plosives in English, they were also higher than what is commonly regarded as *Zero VOT*. Simply put, it is possible that, if the VOT in those stimuli were lower, the effects expected to be found with *H3* and *H4* would have actually been verified.

The fact that NSAE outperform PBLE in the transcription of sentences with *Positive VOT* is surprising. For one thing, as I pointed out earlier (see *subsections 2.3.2* and *3.6*), studies have suggested that Brazilian learners and native speakers of American English do not differ as to their identification of segments with *Positive VOT*, for both seem to identify the pattern as the one of a voiceless plosive categorically – however, it is important to consider, as addressed in the previous subsection, that these studies tested such a categorization with words in isolation.

What seems to cause the significant difference found in *H4* is a poorer performance of PBLE in the case of sentences with *Positive VOT* (we can see that their mean accuracy drops in this case in relation to the other two). It might be possible to suggest an explanation for this result if it were possible to show that the target words initiated by voiceless segments were

more difficult to be recognized by Brazilian learners than those initiated by voiced ones by a matter of frequency, for example. That being the case, PBLE performance in relation to the transcription of all sentences initiated by voiceless segments (that includes those with *Zero VOT* and *Positive VOT*) would be poorer in comparison to sentences with voiced segments (*Negative VOT*). Since a poorer performance of PBLE in the transcription of sentences with *Zero VOT* was not observed, that does not seem to be the case.

Therefore, perhaps the only possible means to explain these results is by the action of other factors that are external to those controlled in the task (participant's L1, contextual information and VOT pattern), as addressed in *subsection 3.5*. Once again, it is also interesting to notice that an analysis of the transcription of target words alone – not the whole sentence – might be a valuable resource to identify the reason for the divergences found in these results.

4.2.5 Testing Hypothesis 5

In *Hypothesis 5* (*subsection 3.6*), it was postulated that the analysis comparing PBLE and NSAE as to their accuracy levels, considering only contextual information as a factor, would show that the groups do not differ in the transcription of sentences with contextual information, but that they are different in the transcription of those without contextual information, in which case PBLE would be significantly more accurate.

Once, as pointed out in *subsection 4.2*, no main effect was verified in the interaction between the variables of participant's L1 and contextual information in the rANOVA [$F(1, 16) = 2.526$; $p = .132$], no follow-up test was required and, therefore, the hypothesis was not corroborated: there is no significant difference between the performances of NSAE and PBLE in the transcription of sentences with contextual information and sentences without contextual information.

The results obtained from the testing of this hypothesis, along with the results from *H3* and the descriptive comparison of means (see *Table 6*), suggest that contextual information as a factor did not have an effect on participants' performance, for their accuracy levels are very similar in the conditions with contextual information and without contextual information.

An explanation may reside in the possibility that the contextual information presented in the sentences elaborated for this study was not enough to make sentences sufficiently different as to cause divergent performances in the transcription task – meaning that longer

sentences, with more contextual information, might have caused different effects. It might be the case that a new study with different sentences could shed light on this matter; the challenge would be, however, to present the participants with sentences that are similar in number of words or syllables and stress pattern regardless of having what is considered as contextual information or not – observing the criteria I explained in *subsection 3.5*.

4.2.6 Testing Hypothesis 6

Hypothesis 6 in *subsection 3.6* predicted that the analysis contrasting all participants' accuracy levels considering both contextual information and VOT as factors (but ignoring participant's L1) would show that participants only differ significantly in the condition 'sentence without contextual information produced with *Zero VOT*', in which accuracy levels would be lower – especially for NSAE.

The follow-up *Paired Samples T-Test*[34] indicated that there is no significant difference in any of the cases: sentence with *Negative VOT* with (mean = 81.71; SD = 11.96) and without contextual information (mean = 75.46; SD = 12.44) [t(17) = -1.921; p = .072]; sentence with *Zero VOT* with (mean = 75.92; SD = 15.88) and without contextual information (mean = 73.61; SD = 14.92) [t(17) = -.464; p = .648]; and sentence with *Positive VOT* with (mean = 70.83; SD = 14.36) and without contextual information (mean = 70.83; SD = 15.19) [t(17) = .000; p = 1.0].

Hypothesis 6 was therefore not corroborated: considering the performances of all participants together, there is no significant difference between the accuracy levels of the transcriptions of sentences without contextual information and *Zero VOT*, and no effect is found through other correlations between VOT pattern and contextual information as grouping variables.

It may be interesting to observe, however, that in the case of sentences with *Negative VOT*, the results show an *approaching significance*: one can reasonably assume that, if a larger sample were analyzed (with data from more participants), an effect of contextual information could have been found in this test. While these results do not offer the expected outcome for the effects of *Zero VOT*, we may find relevance in them considering that the tokens with *Negative VOT* stand for 50% of the task sentences alone (see *subsections 3.3* and

[34] The *Paired Samples T-Test* is a statistical approach to determine significant differences between two sets of dependent and identically distributed samples, approached in a sample of matched pairs of similar units.

3.4). In this context, to assert that participants were *almost* significantly more accurate in the transcriptions of sentences with *Negative VOT* and contextual information as compared to the transcription of those without contextual information is to say that the analysis was close to finding an effect of contextual information towards promoting higher intelligibility levels in the sentence transcription task. In that case, research may find relevance in further analyses with more participants' data.

4.2.7 Testing Hypothesis 7

In *subsection 3.6*, *Hypothesis 7* predicted that the analysis comparing PBLE and NSAE as to their accuracy levels considering both contextual information and VOT as factors would show a significant difference in transcription of sentences for NSAE only in the condition in which there is no contextual information and *Zero VOT* – because their performance will be poorer –, whereas no significant differences would be found in the transcriptions by PBLE.

Since, as I pointed out in *subsection 4.2*, no main effect was verified in the three-way interaction between the variables participant's L1, VOT pattern, and contextual information in the rANOVA [$F(2, 15) = .154$; $p = .858$], no follow-up test was required and, therefore, *Hypothesis 7* was not corroborated: no significant differences were found in any of the twelve conditions, considering the interaction of the three variables controlled in this study, which means that the negative effect of *Zero VOT* on the intelligibility of sentences without contextual information for NSAE was not verified.

Among all of the factors pointed out as possible explanations for the lack of the expected effects of *Positive VOT* and contextual information as variables promoting intelligibility throughout this section, I would like to highlight those discussed in *subsection 3.5*: factors other than those controlled in this study (such as participant's ability to recall auditory information or accent) may play fundamental roles in the intelligibility of the utterances that were tested. In trying to assess intelligibility, we are dealing with an intrinsically complex phenomenon, and therefore we must bear in mind the ever possible unpredictable action of factors other than the ones we control.

4.3 Summary of findings

The analysis of the data from the transcription task made it possible to test the seven hypotheses presented in *subsection 3.6*, which are once again transcribed in what follows*:*

H1: The analysis contrasting the two groups of participants as to their overall accuracy levels in the sentence transcription task (ignoring VOT pattern and contextual information as variables) will show that PBLE are significantly more accurate than NSAE.

H2: The analysis contrasting all participants' accuracy levels considering only *Positive VOT* as a variable in the sentence transcription task will show that (a) there is a significant difference in accuracy in the transcription of sentences with plosives produced with *Negative VOT* and *Zero VOT* - participants are significantly more accurate in the former, for NSAE poor performance in the former has an effect over accuracy levels; (b) there is no significant difference in accuracy in the transcription of sentences with plosives produced with *Negative VOT* and *Positive VOT*; (c) contrasting the transcription of sentences produced with *Zero VOT* and *Positive VOT*, participants are significantly more accurate in the latter, for NSAE poor performance in the former has an effect over accuracy levels.

H3: The analysis contrasting all participants' accuracy levels considering only contextual information as a variable in the sentence transcription task will show that participants are significantly more accurate in the transcription of sentences with contextual information than in the transcription of those without such clue.

H4: The analysis comparing PBLE and NSAE as to their accuracy levels considering only VOT as a factor will show that the groups differ in the transcription of sentences with plosives produced with *Zero VOT*, but not in the one with those produced with *Negative* or *Positive VOT*: PBLE will be significantly more accurate in the transcription of sentences with *Zero VOT*.

H5: The analysis comparing PBLE and NSAE as to their accuracy levels considering only contextual information as a factor will show that the groups do not differ in the transcription of sentences with contextual information, but that they are different in the transcription of those without contextual information – PBLE will be significantly more accurate in this case.

H6: The analysis contrasting all participants' accuracy levels considering both contextual information and VOT as factors will show that participants only differ significantly in the condition 'sentence without contextual information produced with *Zero VOT*', in which accuracy levels will be lower – especially for NSAE. In all other five conditions participants will not be significantly different.

H7: The analysis comparing PBLE and NSAE as to their accuracy levels considering both contextual information and VOT as factors will show a significant difference in transcription of sentences for NSAE only in the condition in which there is no contextual information and *Zero VOT* – because their performance will be poorer –, whereas no significant differences will be found in the transcriptions by PBLE.

As approached in detail in the previous subsections, none of the hypotheses was corroborated. Nonetheless, the results might allow us to make some important considerations. The lack of significant difference between the performances of native speakers of American English and proficient Brazilian learners in the transcription task seems to endorse the assumption that these groups might have fairly similar *understandings* of the utterances in question in what comes to the attention they direct to *Positive VOT* and contextual information as cues for promoting intelligibility. While it might have been possible that a high proficiency level had led these learners to behave similarly to NSAE towards these cues, the lack of significant differences within subject performances (considering the manipulation of contextual information and VOT pattern as variables) suggests that factors other than those controlled by this study can be at play – as I pointed out earlier, the participants' ability to recall auditory information or accent might be among these factors.

Another important observation concerning the lack of statistical confirmation of the hypotheses must be made relating the results to the number of participants in the present study (6 NSAE and 12 PBLE): it seems reasonable to assume that a larger sample could have greater statistical power and, therefore, lead to the confirmation of differences in performances, which can be suggested by the *approaching significance* found in the case of the difference between participants' accuracy in transcription of sentences with *Negative VOT,* with and without contextual information.

Notwithstanding the lack of significant differences found within participants' performances, it is important to highlight the difference found between the groups. Against the assumption that a shared L1 would promote intelligibility, overall, NSAE had a better

performance in the sentence transcription task – and such a difference found statistical support. Possible explanations may reside in the fact that NSAE find ease in a greater familiarity with the target words used in the transcription task, especially because some of these words may be considered infrequent for L2/Brazilian learners, whose limitation in vocabulary may have been determinant in sentence-transcription performance.

Moreover, even though the differences are not supported through statistical confirmation, it should be mentioned that, when we analyze transcriptions by the differences in stimulus-sentences, NSAE seem to have a significantly better performance than PBLE in the following conditions: (a) sentence with contextual information and a voiced plosive; and (b) sentence without contextual information and *Positive VOT*. In all other conditions (find them in *subsection 3.6*), the two groups do not have significantly different performances. In condition (a), this result might suggest that contextual information promotes intelligibility for NSAE in overall performance – we must consider that, as explained in *subsection 3.3*, sentences with voiced target plosives corresponded to 50% of the target sentences in the task. In condition (b), we might have evidence that, not finding meaningful contextual information in the sentence, NSAE direct their attention to *Positive VOT* as a cue for achieving word intelligibility. In this sense, there was also a suggestion that NSAE might have directed their attention to sentence meaning and frequency over sound – which, again, suggests that this group looked for meaningful contextual information in the sentences –, while PBLE do not seem to have done the same.

The main limitation of the present study may reside in the lack of an objective analysis to determine whether the source of inaccuracy in transcription was the target word or the carrier sentence, comparing these sources between groups. Nevertheless, I would like once again to highlight the fact that an impressionistic analysis (conducted by counting the number of occurrences of these manually) reveals that, considering the performances of NSAE and PBLE altogether, out of 420 wrong transcriptions found in the test (see *Table 4*), the number of cases in which the source of the error was in the target word was below 10%.

Future studies could test these variables with larger samples and look into the transcription of target words solely, in order to conduct a closer verification of the source of the small divergences in groups' performances found in this study, as to associate them with intelligibility of the target word in question or intelligibility of the whole sentence uttered.

5 CONCLUSION

This study aimed to investigate the effects of *Positive* VOT (with and without contextual information) on the intelligibility of short English sentences produced by Brazilians to native speakers of American English (NSAE) and proficient Brazilian learners of English (PBLE).

Sentences without contextual information were meant not to provide participants with informational clues as to which member of a minimal pair (initiated by either a voiced or a voiceless stop consonant) was its target word. Sentences with contextual information, however, should indicate which member of the minimal pair was the target sentence through association of meaning with the other words of the sentence. Four Brazilian speakers recorded the stimulus-sentences: two of them produced sentences with *Positive VOT;* two of them did not do so.

Twelve proficient Brazilian learners and six native speakers of American English participated in the sentence transcription task conducted in software *E-Prime 2.0* (SCHNEIDER, ESCHMAN & ZUCCOLOTTO, 2012). The data obtained from this task was analyzed objectively, assigning transcriptions to binary accurate (high intelligibility) or inaccurate (low-to-no intelligibility) categories.

However, it is essential to observe that, as discussed in *subsection 2.4*, intelligibility goes far beyond binary categories. An objective analysis to assess intelligibility will require methodological choices like the ones from this study, but we must bear in mind the dynamicity of intelligibility – or even *understanding*, if we address it as such. Intelligibility is a construct comprising several complementary levels in a continuum, whose boundaries are very much open to discussion. Perhaps one valid assumption, that I tried to propose throughout this study, is that the construct put forth by models of speech perception (e.g. FLEGE, 1995; BEST & TYLER, 2007) may be regarded as a more basic or lower level for intelligibility, which comes to contribute to higher levels in context. Higher levels will progressively include more variables (going from speech perception to *interpretability*, for example), accounting for overall *understanding* in communication. For that reason, assessing intelligibility with different levels of context and trying to establish the roles these levels play in communication, as it was the purpose of this study, consists of a very interesting gap to be filled by research.

As for the performance of participants in the transcription task, it can be stated that both groups of participants were reasonably accurate in transcribing the sentences – the performance of NSAE was, however, fairly better. The testing of hypotheses comparing mean accuracies in different conditions of stimuli was inconclusive as to how contextual information and *Positive VOT* may actually affect sentence intelligibility to native speakers of American English and Proficient Brazilian learners, since no effect from the manipulation of these variables found statistical confirmation. Notwithstanding, some interesting observations may be made with respect to the results obtained from the task.

The fact that NSAE had a better performance was unexpected. Since PBLE shared a common L1 – more specifically a common phonological system – with the speakers who recorded the stimuli, it was reasonable to assume that intelligibility would be higher for these participants. Moreover, since the stimuli recorded for the transcription task also had variables supposed to pose a problem for intelligibility (that is, the stimuli without *Positive VOT*) by native listeners, the performance of NSAE was expected to be poorer: sentences without contextual information produced with unaspirated voiceless plosives were expected to lead NSAE into assimilating target words as their voiced counterparts. This hypothesis did not find confirmation in the tests conducted in the study. In this sense, we might be looking either at a scenario in which *Positive VOT* does not play an important role to intelligibility in the terms of this study, or at one in which variables external to those manipulated in the study played a determinant role in the overall performance of participants in the task, causing unexpected effects to be verified.

As regards the lack of statistical confirmation of the hypotheses, it is also important to observe that the number of participants in the present study might not have been a large enough sample in terms of statistical power. The *approaching significance* found in the case of the difference between participants' accuracy in transcription of sentences with *Negative VOT* with and without contextual information, leads us to the assumption that larger samples might support the confirmation of differences in performances considering the manipulations of VOT pattern and contextual information as grouping variables.

Some factors must be pointed out as limitations of this study. First, the fact that the frequency of the target words was not measured. Especially because we are working under a dynamic perspective of language acquisition, we must assume that the frequency of the words tested may have played an important role in the results, since less frequent words will tend not to be recognized, or to be assimilated or transcribed differently. Moreover, another limitation

lies in the lack of control over the VOT of the stimuli of target words – as for unaspirated target words, lower VOT levels might have caused the perception of these segments to be different. Another limitation of this study is the lack of an analysis to determine whether the source of inaccurate transcriptions was in the target words or in the carrier sentences. As discussed in *subsection 4.2.4*, this would provide a better understanding of how *Positive VOT* affected the performance in the task. Finally, the small number of participants was also a limitation, especially because it has a direct influence over the power of statistical analysis.

Future directions may include the replication of this study with observations of the limitations mentioned above: controlling frequency of target words and VOT of stimuli, gathering data from more participants. Also, to test overall accuracy of transcription of target words in contrast to sentence transcription might provide interesting insights on the questions raised by this study. The fact that an approaching significance was found in the comparison of means of accuracy in transcription of sentences with and without contextual information, produced with *Negative VOT*, suggests that the same analysis with a larger sample might bring interesting insights.

In conclusion, as to the relevance of *Positive VOT* for pronunciation instruction, it can be said that the present study brings to question at least one important factor to be taken into account: the presence of contextual information – such as presented in the sentences selected for this study – may, as suggested, account for the distinction between voiceless and voiced target stops in a sentence. We must also consider that, in real-life communicative situations, the tendency is for other factors (of linguistic and extra-linguistic nature) to present even more contextual information which should minimize the importance of segmental aspects such as *Positive VOT*. In spite of that, it should not be the case that we dispense with *Positive VOT* from SL classrooms. Aside from the fact that some very specific contexts may necessarily require leaners to produce *Positive VOT* (such as spelling words aloud), the process of driving attention towards acoustic cues other than those from our L1 phonological system is an important one within language acquisition, which may raise learner's awareness of other fine details of speech in the L2. In the case in which we actually choose to address this aspect in language classroom, what may be interesting is to discuss the role of contextual information along, always making it clear that the purpose of learning to produce that aspect should not be to eradicate one's accent.

REFERENCES

ALBANO, Eleonora Cavalcante. *O gesto e suas bordas: esboço da fonologia acústico-articulatória do português brasileiro.* Campinas: Mercado de Letras, ALB, FAPESP, 2001.

_____. *Uma Introdução à Dinâmica em Fonologia, com foco nos trabalhos desta coletânea.* Revista da Abralin, número especial 2, 2012.

ALLAN, Dave**.** *Oxford Placement Test 1*. Oxford University Press, 2004.

ALVES, Mariane Antero. *Production of English and Portuguese Voiceless stops by Brazilian EFL Speakers.* Dissertação (Mestrado em Letras) - Universidade Federal de Santa Catarina, Florianópolis, SC, 2011.

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.* Projeto de pesquisa. Porto Alegre: Universidade Federal do Rio Grande do Sul (UFRGS), 2010.

_____; 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.

_____; MOTTA, Camila Saviczky. *Focusing on the right cue: perception of voiceless and voiced stops in English by Brazilian learners.* Phrasis, vol. 2009-2014 (2).

_____; SCHWARTZHAUPT, Bruno M; BARATZ, Ana Hemmons. Percepção e Produção dos Padrões de VOT do Inglês (L2) por Aprendizes Brasileiros. *Estudos em Aquisição Fonológica, vol. 4.* Pelotas: Editora Universitária/UFPEL, 2011.

_____; ZIMMER, Márcia Cristina. *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, 59 (1): 157-180, 2015.

BECKER, Márcia Regina. *Inteligibilidade da Língua Inglesa sob o Paradigma de Língua Franca: Percepção de Discursos de Falantes de Diferentes L1s por Brasileiros*. Tese (Doutorado em Letras). Universidade Federal do Paraná, Curitiba, 2013.

BECKNER, C; ELLIS, N; BLYTHE, R; HOLLAND, J; BYBEE, J; KE, J; CHRISTIANSEN, M; LARSEN-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; 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: In honor of James Emil Flege.* Amsterdam/Philadelphia: John Benjamins Publishing Company, 2007, p. 13-34.

BOERSMA, P; WEENINK, D. *PRAAT: Doing Phonetics by Computer*. Amsterdam, 2014. Available at <http://www.fon.hum.uva.nl/praat/>. Access in March 3rd 2015.

CELCE-MURCIA, Marianne; BRINTON, Donna M.; GOODWIN, Janet M; GRINER, Barry. *Teaching pronunciation: a coursebook and reference guide.* Cambridge University Press, 2010.

COHEN, Gustavo Vargas. *The VOT Dimension: a bi-directional experiment with English Brazilian Portuguese stops.* Dissertação – Mestrado em Letras. UFSC, Florianópolis: 2004.

CRUZ, Neide Cesar. *Vowel insertion in the speech of Brazilian learners of English: a source of unintelligibility?* Ilha do desterro, UFSC, Florianópolis, nº55, p. 133-152, 2008.

DAVIES, Alan. Assessing World Englishes. *Annual Review of Applied Linguistics*, 29, p. 80-89, 2009.

DE BOT, Kes; LOWIE, Wander; VERSPOOR, Marjolijn. *A Dynamic Systems Theory approach to second language acquisition.* Bilingualism: Language and Cognition, v. 10, n. 1, p. 7-21, 2007.

DERWING, Tracey M.; MUNRO, Murray J. Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly*, v. 39, n.3, p. 379-397, 2005.

_____; _____. Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, v. 421, n. 4, p. 476-490, 2009.

ELLIS, Nick. 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, Karoline. *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.* Dissertação (Mestrado em Letras) - Universidade Católica de Pelotas, Pelotas, RS, 2011.

GEWEHR-BORELLA, Sabrina. *A influência da fala bilíngüe Hunsrückish-Português Brasileiro na escrita de crianças brasileiras em séries iniciais*. Dissertação de Mestrado. Universidade Católica de Pelotas, Pelotas, RS, 2010.

GONÇALVES, Alison Roberto. *In search of speech intelligibility: the case of English high front vowels.* Dissertação (Mestrado em Letras) - Universidade Federal de Santa Catarina, Florianópolis, SC, 2014.

JENKINS, J. *The Phonology of English as an International Language.* Oxford, Oxford University Press, 2000.

KACHRU, Y; SMITH, Larry E. *Cultures, Contexts, and World Englishes*. New York: Routledge, 2008.

LARSEN-FREEMAN, Diane. Chaos/Complexity Science and Second Language Acquisition. Applied Linguistics, v. 18, n. 2, 1997. p. 141-165.

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.

MAGRO, Vivian. *Effects of different types of instruction on the acquisition of aspiration in word-initial /p/ by Brazilian Portuguese speakers.* Monografia de conclusão (Especialização em Ensino de Língua Inglesa) - Centro Universitário Ritter dos Reis, Porto Alegre, RS, 2010.

MUNRO, Murray J; DERWING, TraceyM; MORTON, Susan L. *The mutual intelligibility of L2 speech. Studies in Second Language Acquisition*, 28, 111–31. 2006.

NELSON, Cecil. *Intelligibility in World Englishes*: *theory and application.* New York: Routledge, 2011.

PRESTES, Susana Pinheiro da Cruz. *Produção de Consoantes Oclusivas Iniciais do Inglês por Falantes Nativos de PB*. Dissertação – Mestrado em Letras. UFPR, Curitiba, 2013.

REIS, Mara; NOBRE-OLIVEIRA, Denize**.** Effects of perceptual training on the identification and production of English voiceless plosives aspiration by Brazilian EFL learners. *Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech.* Florianopolis: UFSC, 2008. p. 372-381.

SCHWARTZHAUPT, B. *Factors influencing Voice Onset Time: analyzing Brazilian Portuguese, English and interlanguage data.* Trabalho de Conclusão de Curso. Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2012.

_____. *Conceiving Language as a Complex Adaptive System: a problem for SLA researchers?* Signo [ISSN 1982-2014]. Santa Cruz do Sul, v. 38, n. 65, p. 261-272, jul. dez. 2013.

_____; ALVES, Ubiratã Kickhöfel; BARATZ, Ana Hemmons. Percepção e Produção das Plosivas do Inglês (L2) por Aprendizes Brasileiros. *E-book do I Seminário Internacional de Aquisição da Linguagem.* Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2012.

_____; _____; FONTES, Ana Beatriz Areas da Luz. *O Papel do Conhecimento da L1 na Percepção da Fala em L2: Investigando o Status Atribuído por Falantes Nativos e Aprendizes ao VOT Positivo no Inglês Americano.* Comunicação apresentada no IV Seminário de Aquisição de Linguagem. UFSM, Santa Maria – RS. 2013.

_____; _____; _____. *The role of L1 knowledge on L2 speech perception: investigating how native speakers and Brazilian leaners categorize different VOT patterns in English*. 2015 (in press)

SIMON, Ellen. *Acquiring New Voice Contrasts: an empirical study on the laryngeal phonology of Dutch, English and Dutch Learner English.* Doctoral Dissertation, Ghent, Ghent University, 2006.

SMITH, Larry E. Spread of English and issues of intelligibility. *In:* Braj B. Kachru (ed.) *The Other Tongue: English across Cultures.* 2nd edn. Urbana, IL: University of Illinois Press. 1992. p. 75–90.

WALKER, Robin. *Teaching the Pronunciation of English as a Lingua Franca.* New York, Oxford University Press, 2010.

YAVAS, Mehmet**.** Factors influencing the VOT of English long lag stops and interlanguage phonology. In: RAUBER, Andrea S.; WATKINS, Michael A.; BAPTISTA, Barbara O. (eds.). *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech.* Florianópolis: Universidade Federal de Santa Catarina, 2008, p. 492-498.

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

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

_____; SILVEIRA, Rosane; ALVES, Ubiratã Kickhöfel. *Pronunciation instruction for Brazilians: bringing theory and practice together.* Newcastle upon Tyne: Cambridge Scholars Publishing, 2009.

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

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

APPENDICES

Appendix 1 – Termo de Consentimento Livre e Esclarecido (Group 1)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
Projeto de Pesquisa: ***A Relevância do VOT Positivo Para a Comunicação Oral em Inglês (L2): Testando o Efeito da Informação Contextual sobre a Inteligibilidade***
Mestrando: Bruno Moraes Schwartzhaupt
Orientador: Prof. Dr. Ubiratã Kickhöfel Alves

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Grupo 1: Gravação de Estímulos

Prezado participante,

A presente investigação tem como objetivo a produção de conhecimento sobre as dificuldades de percepção dos sons e da pronúncia em língua inglesa por parte de estudantes brasileiros. A investigação resultará em uma dissertação de Mestrado. Esperamos contar com a sua valiosa participação.

Ao participar da pesquisa, você estará integrando um dos seguintes grupos: o Grupo 1 realizará a gravação de sua leitura oral de frases em Inglês; por sua vez, o Grupo 2 participará de testes de percepção de sons, elaborados a partir do áudio das gravações das produções orais dos participantes do Grupo 1. Este termo é referente ao Grupo 1, que é o grupo do qual o convidamos a fazer parte.

Como parte da investigação, você será levado a um estúdio profissional com isolamento acústico, para que possamos efetuar sua leitura de frases curtas em Inglês. Nessa oportunidade, você receberá uma lista com as frases, que pedimos que você leia na ordem fornecida. Sua voz será utilizada, posteriormente, como estímulo em testes de percepção, a serem elaborados pelo proponente deste estudo, os quais nos ajudarão a entender como são percebidos determinados sons na Língua Inglesa. Em nenhum momento sua identidade será revelada para aqueles que posteriormente fizerem os testes de percepção, ou mesmo em apresentações orais ou relatos escritos do presente estudo. O procedimento de gravação dos estímulos no estúdio não tomará mais do que 20 minutos do seu tempo. Caso você se sinta desconfortável e queira interromper ou encerrar a gravação, fique à vontade para fazê-lo a qualquer momento.

Não há benefício direto para você ao participar do estudo. As descobertas poderão servir como fonte de consulta para estudiosos do processo de aquisição da pronúncia em língua estrangeira, bem como de metodologia de ensino e aprendizagem de línguas, além de formadores de professores de línguas estrangeiras. Salientamos que a tarefa pode causar cansaço ou ansiedade de sua parte, o que justifica intervalos de tempo para descanso conforme sua vontade.

Os resultados da pesquisa serão divulgados à comunidade acadêmica e à comunidade de educadores por meio de publicações, apresentações em eventos acadêmicos, oficinas de formação de

professores, entre outras formas de divulgação. Reafirmamos que em nenhuma dessas ocasiões sua identidade será revelada.

Sua participação é essencial para a realização do trabalho de pesquisa, mas você tem a liberdade para se recusar a participar ou retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo.

Em caso de dúvida ou necessidade de esclarecimentos sobre o estudo, por favor, entre em contato com o Professor Orientador deste trabalho:

Prof. Ubiratã Kickhöfel Alves

Prédio Administrativo do Instituto de Letras – Sala 220 – Campus do Vale

Av. Bento Gonçalves, 9500 – 91501-000 – Porto Alegre, RS

Telefone: (51)3308-7081

E-mail: ukalves@pq.cnpq.br

Em caso de dúvida relacionada a seus direitos ou sobre sua participação nesta pesquisa, por favor, entre em contato com o Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Sul (UFRGS):

Comitê de Ética em Pesquisa/UFRGS
Prédio da Reitoria – 2º andar – Câmpus Central
Av. Paulo Gama, 110 – 90040-060 – Porto Alegre, RS
Telefone: (51) 3308- 3738
E-mail: etica@propesq.ufrgs.br

Data: _____

Participante: _____

_____
Bruno Moraes Schwartzhaupt
(Mestrando responsável)

_____
Prof. Dr. Ubiratã Kickhöfel Alves
(Orientador)

(PPG-LETRAS/UFRGS)

Appendix 2 – Ficha de Informações do Participante

FICHA DE INFORMAÇÕES DO PARTICIPANTE

Por favor, preencha o formulário abaixo. Sua participação neste estudo é muito importante. Muito obrigado!

Nome: _____

Data de nascimento e idade: _____

Cidade natal: _____

Cidade natal do pai: _____

Cidade natal da mãe: _____

Línguas adquiridas até os 6 anos de idade: _____

Idade com que iniciou a estudar a Língua Inglesa: _____

Instituição de ensino em que estuda a Língua Inglesa atualmente: _____

Nível de adiantamento (__ semestre / básico / intermediário, etc.): _____

Outras línguas que você pode falar, além do português e do inglês (para cada uma, determine um nível de proficiência entre básico, intermediário e avançado):

_____

Cidade de residência e período de tempo em que reside nesta cidade: _____

Países de língua inglesa que já visitou/residiu:

_____

Período de tempo da visita/residência:

_____

Data: _____

Informante número: _____

Appendix 3 – Termo de Consentimento Livre e Esclarecido (Group 2)

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
Projeto de Pesquisa: ***A Relevância do VOT Positivo Para a Comunicação Oral em Inglês (L2): Testando o Efeito da Informação Contextual sobre a Inteligibilidade***
Mestrando: Bruno Moraes Schwartzhaupt
Orientador: Prof. Dr. Ubiratã Kickhöfel Alves

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Grupo 2: Realização de Tarefas Perceptuais

Prezado participante,

A presente investigação tem como objetivo a produção de conhecimento sobre as dificuldades de percepção dos sons e da pronúncia em língua inglesa por parte de estudantes brasileiros. A investigação resultará em uma dissertação de Mestrado. Esperamos contar com a sua valiosa participação.

Ao participar da pesquisa, você estará integrando um dos seguintes grupos: o Grupo 1 realizará a gravação de sua leitura oral de frases em Inglês; o Grupo 2, por sua vez, participará de testes de percepção elaborados a partir do áudio das gravações das produções orais dos participantes do Grupo 1. Este termo é referente ao Grupo 2, que é o grupo do qual o convidamos a fazer parte.

Primeiramente, solicitamos não mais do que 30 minutos do seu tempo, para que você faça o teste de nivelamento *Oxford Placement Test 1*. Para a condução deste estudo, é importante que saibamos qual é o seu grau de adiantamento em língua inglesa, o que será revelado a partir de sua classificação nesse teste. O *Oxford Placement Test 1* é composto por um *Teste de Gramática* (50 questões de múltipla escolha) e um *Teste de Compreensão Auditiva* (100 questões de múltipla escolha), e classifica aprendizes em 1 de 10 níves, de *Beginner* a *Functionally Bilingual*. Sua identidade e pontuação não serão reveladas a ninguém além de você.

Em um segundo momento, você utilizará fones de ouvido e um programa de computador para realizar uma tarefa de transcrição de frases. Nessa tarefa, você ouvirá frases (todas em Inglês) e sua tarefa será transcrever a frase que você ouviu – isto é, digitá-la imediatamente no computador, com um tempo limite conforme explicado no momento da realização da tarefa. Os resultados do teste serão de conhecimento da comunidade científica; no entanto, sua identidade permanecerá preservada. O procedimento não tomará mais do que 30 minutos do seu tempo. Caso você se sinta desconfortável e queira interromper ou encerrar o teste, fique à vontade para fazê-lo a qualquer momento.

Não há benefício direto para você ao participar do estudo. As descobertas poderão servir como fonte de consulta para estudiosos do processo de aquisição da pronúncia em língua estrangeira, bem como de metodologia de ensino e aprendizagem de línguas, além de formadores de professores de línguas estrangeiras. Salientamos que a tarefa pode causar cansaço ou ansiedade de sua parte, o que justifica intervalos de tempo para descanso conforme sua vontade.

Os resultados da pesquisa serão divulgados à comunidade acadêmica e à comunidade de educadores por meio de publicações, apresentações em eventos acadêmicos, oficinas de formação de

professores, entre outras formas de divulgação. Reafirmamos que em nenhuma dessas ocasiões sua identidade será revelada.

Sua participação é essencial para a realização do trabalho de pesquisa, mas você tem a liberdade para se recusar a participar ou retirar seu consentimento em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo.

Em caso de dúvida ou necessidade de esclarecimentos sobre o estudo, por favor, entre em contato com o Professor Orientador deste trabalho:

Prof. Ubiratã Kickhöfel Alves

Prédio Administrativo do Instituto de Letras – Sala 220 – Campus do Vale

Av. Bento Gonçalves, 9500 – 91501-000 – Porto Alegre, RS

Telefone: (51)3308-7081

E-mail: ukalves@pq.cnpq.br

Em caso de dúvida relacionada a seus direitos ou sobre sua participação nesta pesquisa, por favor, entre em contato com o Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Sul (UFRGS):

Comitê de Ética em Pesquisa/UFRGS
Prédio da Reitoria – **2º andar** – **Câmpus Central**
Av. Paulo Gama, 110 – 90040-060 – Porto Alegre, RS
Telefone: (51) 3308- 3738
E-mail: etica@propesq.ufrgs.br

Data: _____

Participante: _____

_____
Bruno Moraes Schwartzhaupt
(Mestrando responsável)

_____
Prof. Dr. Ubiratã Kickhöfel Alves
(Orientador)

(PPG-LETRAS/UFRGS)

Appendix 4 – Consent Form

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
Research Project: ***The Relevance of Positive VOT to Oral Communication in English (L2): Testing the Effect of Contextual Information upon Intelligibility***
M.A. candidate: Bruno Moraes Schwartzhaupt
Advisor: Professor Ubiratã Kickhöfel Alves, PhD

CONSENT FORM

Group 2: Participation in Perceptual Tasks

Dear participant,

This investigation aims to obtain knowledge on the difficulties concerning the perception of the English Language sounds by Brazilian learners. The results of the investigation shall be used for the writing of a Master's Degree Thesis. We hope to count on your valuable participation.

If you agree to participate in this research, you will integrate one of the following groups: Group 1 will perform a recording of their reading of sentences in English, while Group 2 will take part in percteptual tasks elaborated with the recordings from Group 1. The present Consent Form is meant for those participants assigned to Group 2, in which we invite you to participate.

In case you choose to participate, you will wear a headset and a computer program in order to perform a sentence transcription task. In this task, you will hear sentences in English and your task will be to type the sentences in the computer within a limited time, as it shall be explained in the beginning of the task. Your identity shall remain unrevealed, and only the M.A. candidate in question and his advisor will know about your identity in this task. The procedure lasts for no longer than 30 minutes. In case you feel uncomfortable and you want to interrupt or end the test, feel free to do so at any time.

There is no direct benefit to you in taking part in this study. The findings may serve as a reference source for those who study Foreign Language pronunciation acquisition and teaching methodologies, as well as for those who are studying to become Foreign Language teachers. It is important to make it clear that you might feel tired or anxious while performing this task, which is why you should feel free to pause and resume the task as you wish.

The results of this research study will be at the disposal of the scientific community and educators through publications, presentations in academic events, teaching formation workshops, among others. Once again we assure you that in none of these occasions your identity will be revealed.

Your participation is essential for this study, but you are free to refuse participating or deny your consent to the use of your data at any time in the conduction of the study without any penalties.

In case you have doubts with regard to the conduction of this research study, please contact the Professor in charge of supervising this work:

Prof. Ubiratã Kickhöfel Alves

Instituto de Letras – Room 220 – Campus do Vale
9500 Bento Gonçalves Avenue – 91501-000 – Porto Alegre, RS

Phone: (51)3308-7081

Email: ukalves@pq.cnpq.br

In case you have doubts concerning your rights as a participant of this research study, please contact the *Comitê de Ética em Pesquisa da Universidade Federal do Rio Grande do Sul* (UFRGS' Ethics in Research Committee):

Comitê de Ética em Pesquisa/UFRGS
Prédio da Reitoria – 2nd floor – Central Campus
110 Paulo Gama Avenue – 90040-060 – Porto Alegre, RS
Phone: (51) 3308- 3738
Email: etica@propesq.ufrgs.br

Date: _____

Participant: _____

_____
Bruno Moraes Schwartzhaupt
(Graduate student responsible)

_____
Professor Ubiratã Kickhöfel Alves, PhD
(Advisor)

(PPG-LETRAS/UFRGS)

Appendix 5 – Participant Information Form

PARTICIPANT INFORMATION FORM

Please fill out the form below. Your participation in this study is very important. Thank you!

Name: _____

Date of birth and age: _____

Hometown: _____

Father's hometown: _____

Mother's hometown: _____

Languages acquired until six years of age: _____

Other languages you are able to speak, apart from Portuguese and English (please state a proficiency level between basic, intermediate and advanced for each of them):

_____

City of residence and time living in the city: _____

Foreign countries you have lived in for more than six months:

_____

Time living in those countries:

_____

Date: _____

Participant number: _____

Appendix 6 – List of Target and Distractor Sentences

	Target word	Sentence without contextual information	Sentence with contextual information
1	pill	The boy found a large pill.	The patient took his night pill.
2	bill	The man found a small bill.	The man lost a dollar bill.
3	pan	The man got a huge pan.	The chef used the largest pan.
4	ban	The boy got a light ban.	The runner got a permanent ban.
5	pug	The guy patted the small pug.	The girl trained her fluffy pug.
6	bug	The girl patted the ugly bug.	The notebook has a minor bug.
7	pox	The guy saw the ugly pox.	The doctor cured his awful pox.
8	box	The girl saw the red box.	The baby opened the white box.
9	tip	The lady had a great tip.	The waiter got a generous tip.
10	dip	The man had a good dip.	The swimmer tried a risky dip.
11	tab	The child needed a little tab.	The can has a silver tab.
12	dab	The man needed a quick dab.	The maid gave it a quick dab.
13	tuck	The woman got another tuck.	The doctor watched a complex tuck.
14	duck	The lady got a new duck.	The farmer fed the black duck.
15	tot	The woman drew a large tot.	The mother carried her cute tot.
16	dot	The child drew a small dot.	The sentence has a blurry dot.
17	kill	The people saw an actual kill.	The witness saw the cold kill.
18	gill	The boy saw an ugly gill.	The shark hurt his large gill.
19	cap	The boy found a similar cap.	The store sold a jeans cap.
20	gap	The people found a different gap.	The study showed a financial gap.
21	cut	The kid hid her small cut.	The knife has a sharp cut.
22	gut	The girl hid her big gut.	The fries inflated his fat gut.
23	cod	The girl drew a strange cod.	The fisherman caught a huge cod.
24	god	The kid drew her own God.	The churchmen prayed for their God.

	Distractor word	Sentence without contextual information	Sentence with contextual information
1	fan	The man saw his old fan.	*The star thanked her huge fan.*
2	whip/lip	The woman held her own whip.	*The victim had a swollen lip.*
3	ring/wing	The animal carried a tiny ring.	*The bird broke his left wing.*
4	ship	The boss sold his expensive ship.	*The sailor cleaned the large ship.*
5	nest	The man made his own nest.	*The birds built a cozy nest.*
6	train/brain	The people saw the heavy train.	*The girl has a smart brain.*
7	house	The parents bought a new house.	*The man entered the pink house.*
8	reef/leaf	The painting showed a beautiful reef.	*The tree has a long leaf.*