UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

DÉBORAH SILVA ALVES KLUG

THE LOCAL INTELLIGIBILITY OF BRAZILIAN LEARNERS' SPEECH IN ENGLISH (L2) TO ARGENTINIAN AND GERMAN LISTENERS: A DISCUSSION ON NON-NATIVE PERCEPTION FROM A COMPLEX, DYNAMIC PERSPECTIVE

Porto Alegre
2021

DÉBORAH SILVA ALVES KLUG

THE LOCAL INTELLIGIBILITY OF BRAZILIAN LEARNERS' SPEECH IN ENGLISH (L2) TO ARGENTINIAN AND GERMAN LISTENERS: A DISCUSSION ON NON-NATIVE PERCEPTION FROM A COMPLEX, DYNAMIC PERSPECTIVE

Dissertação de Mestrado em Letras, vinculado à área de Estudos da Linguagem e à linha de pesquisa Psicolinguística, 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

2021

CIP - Catalogação na Publicação

Klug, Déborah Silva Alves
The local intelligibility of Brazilian learners' speech in English (L2) to Argentinian and German listeners: a discussion on non-native perception from a complex, dynamic perspective / Déborah Silva Alves Klug. -- 2021.
271 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, 2021.

1. inteligibilidade local. 2. fala não nativa. 3. vogais do inglês. 4. aprendizes brasileiros. 5. Teoria dos Sistemas Dinâmicos Complexos. 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 autoria de Déborah Silva Alves Klug, sob o título **The local intelligibility of Brazilian learners' speech in English (L2) to Argentinean and German listeners: a discussion on non-native perception from a complex, dynamic perspective**, 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 e à linha de pesquisa Psicolinguística, tendo sido devidamente defendido e aprovado no dia **27 de abril de 2021**, pela banca examinadora abaixo elencada.

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

Prof. Dr. Pedro Luis Luchini
Universidad Nacional de Mar del Plata - Argentina (UNMdP)

Prof. Dr. Clerton Luiz Felix Barboza
Universidade do Estado do Rio Grande do Norte (UERN)

Prof. Dr. Reiner Vinicius Perozzo
Universidade Federal do Rio Grande do Sul (UFRGS)

ACKNOWLEDGMENTS

First and foremost, I would like to thank my adviser, Prof. Dr. Ubiratã Kickhöfel Alves, for his unwavering guidance and support. From the very first moment I approached him with the intention of becoming a researcher, he has pointed me in the right direction, he has been by my side every step of the way, and all of that he did while encouraging me to trail my own path. Not only is professor Alves an outstanding teacher and adviser, he is *über* sensitive and patient, and he has time and time again given me personal support, in ways that I could never thank him enough for. As the whole is never just the sum of its parts, I praise him with all my heart for the professional and the individual that he is.

I would also like to thank my wife, Luiza Klug, who has never let go of my hand. In the most exciting moments, and in the darkest of them, her presence has kept me sane, strong and grounded. She has never failed to incite me to step into my power. It is a cliché, but having her by my side has made every experience better, sweeter, brighter. And though she says she has done nothing for my thesis, I know that no matter which way my path led me, she was there as my utmost attractor state.

My parents will also say that they did nothing special, but I also know that throughout my life they have been nothing but wonderful. Beyond providing me with worldly privileges, they have always been the best, most loving parents ever, and I can still feel their guiding hand on my shoulder in every new track I trail. My brother, in as many words, has always been and will always be the best companion I could have asked for. And if the family that god has given me was not awesome enough, I am blessed with pretty great in-laws. Many relevant effects in my trajectory have emerged from the interactions of this perfect little complex system that I am proud to call my family.

I would like to thank all of my friends, for their presence in my life has always been a source of joy, of warmth and of motivation to be as special as they are. I must name two friends in particular, Thayse Madella and Renata Dalmaso, for providing me the first spark that led me to become a researcher. They have given me a lot of academic input, they have reviewed my works, endured my personal crises and helped me overcome my insecurities. They are the epitome of the small community of friends that has been an essential part of my embodied experiences.

Finally, I would like to thank all the participants in the present study. For their time and effort, for their willingness to help and their patience, many many *many* thanks. Without each and every one of you, this thesis would not be possible.

I hope my work has done justice to all those people. And once again I thank them for being part of my ongoing development as a researcher and as a person.

Show yourself

Step into your power

Throw yourself

Into something new

(WOOD E. R.; MENZEL, 2019)

ABSTRACT

This thesis aims to investigate the intelligibility of the L2 English spoken by Brazilian learners when perceived by other non-native learners. We examine the local intelligibility (MUNRO; DEWING, 2015) of words with the vowels [æ] – [ɛ] and [i] – [ɪ]. We adopt a Complex, Dynamic view of language (DE BOT; LOWIE; THORNE, 2013, BECKNER *et al.*, 2009; LARSEN-FREEMAN, 2017) and apply two complementary analyses, inferential and exploratory, in order to observe (a) the intelligibility rates of the target words produced by our participants; (b) the variables that have effects on the identification of the target vowels; and (c) the acoustic and participant-related characteristics that play a role in that identification. We collected samples produced by six Brazilian learners of English, at three levels of proficiency, all native speakers of Porto-Alegrense Brazilian Portuguese. Additionally, two native speakers of Canadian English provided baseline tokens. We selected 128 sentences as stimuli for the forced-choice perception task. Our 46 listeners were organised in two groups: Argentinian native speakers of Riverplate Spanish, and German native speakers of Central German. As for our first goal, our mixed-effects logistic models show an effect of the listener's L1 on the intelligibility rates of the words with the [æ] – [ɛ] vowels. Tokens with [æ] were more intelligible to Germans, and the ones with [ɛ] were more intelligible to Argentinians. L1 was not significant for the accurate identification of words with [ɪ], but it was for tokens with [i]; Germans showed higher accuracy rates than Argentinians. As for our second goal, our model calculated L1 as a significant predictor variable of vowel identification for the mid/low pair, but not for the tense/lax one. F1 was not a significant predictor for the identification of [æ] and [ɛ], but it was for [i] and [ɪ]. F2 was significant for both pairs. The inferential statistics was complemented by an exploratory analysis, which took into account those statistically significant variables, as well as speakers' proficiency levels and the length of the vowels in the stimuli. As for our third goal, our stimulus-by-stimulus analyses suggest that participant-related characteristics and acoustic cues are combined by the listeners in different ways, leading to emerging phenomena. We found that the temporal cue seems to play a strong role in the perception of the four vowels by Germans, a role that is not so clear in the perception of [ɛ] or [i] by Argentinians. Our exploratory investigation also suggests that both groups of listeners take the temporal cue in combination with F1 and F2, though the decisive status of those spectral cues appears more clearly in Argentinian identifications. Overall, our analyses suggest that the hybrid nature of both the non-native speakers' and listeners' systems will lead to emerging phenomena, as a result of the Complex, Dynamic way in which individual systems have developed. We understand that our findings highlight the need to take both speaker and listener into account when investigating L2 speech intelligibility, thus confirming the dynamic and complex nature of this process.

Key-words: local intelligibility, non-native speech, English vowels, Brazilian learners, Complex Dynamic Systems Theory.

RESUMO

O presente trabalho investiga a inteligibilidade do inglês (L2) falado por aprendizes brasileiros quando percebido por outros aprendizes não-nativos. Avalia-se a inteligibilidade local (MUNRO; DEWING, 2015) de palavras com as vogais [æ] – [ɛ] e [i] – [ɪ]. O trabalho adota uma visão de língua como Sistema Dinâmico e Complexo (DE BOT *et al.*, 2013, BECKNER *et al.*, 2009; LARSEN-FREEMAN, 2017) e utiliza-se de métodos complementares de análise, inferenciais e exploratórios, para investigar (a) a inteligibilidade das palavras-alvo produzidas pelos falantes no presente estudo; (b) as variáveis que têm efeitos sobre a identificação das vogais-alvo; e (c) o papel de características acústicas e relacionadas aos participantes nessa identificação. Foram coletadas produções de seis brasileiros aprendizes de inglês, em três níveis de proficiência, nativos da variedade porto-alegrense de português brasileiro. Dois falantes nativos de inglês canadense forneceram *tokens* de controle. Foram selecionados 128 estímulos para a tarefa de percepção de escolha forçada. Os 46 ouvintes compunham dois grupos: argentinos nativos de espanhol rio-platense, e alemães nativos de alemão central. Em relação ao primeiro objetivo da presente dissertação, os modelos logísticos de efeitos mistos mostraram efeito da L1 do ouvinte na inteligibilidade de palavras com as vogais [æ] – [ɛ]. Palavras com [æ] foram mais inteligíveis para alemães, e aquelas com [ɛ], para argentinos. A L1 não foi significativa para a identificação correta de palavras com [ɪ], mas o foi para palavras com [i]; alemães tiveram maior acuidade do que argentinos. Em relação ao segundo objetivo, o modelo apontou que L1 constitui variável preditora significativa na identificação de vogais do par médio/baixo, mas não do tenso/frouxo. A variável F1 não foi significativa para a identificação de [æ] e [ɛ], mas o foi para a de [i] e [ɪ]. F2 foi significativa para ambos os pares. A análise estatística foi complementada pela exploratória. Esta última considerou as variáveis estatisticamente significativas, bem como o nível de proficiência dos falantes e a duração das vogais nos estímulos. A análise estatística foi complementada, em relação ao terceiro objetivo, com a análise estímulo-por-estímulo, que sugere que tanto características relacionadas aos participantes quanto pistas acústicas são combinadas de diferentes maneiras pelos ouvintes, levando a fenômenos emergentes. Os resultados indicam que a pista temporal tem um papel forte na percepção das quatro vogais do inglês pelos alemães, papel esse menos claro na percepção de [ɛ] or [i] pelos argentinos. A análise exploratória sugere, ainda, que ambos os grupos de ouvintes tomaram a pista temporal em conjunto com F1 e F2, embora o *status* decisivo dessas pistas espectrais pareça mais claro nas identificações por argentinos. De modo geral, os resultados do presente trabalho sugerem que a natureza híbrida dos sistemas dos falantes e dos ouvintes não nativos permitiram a emergência de fenômenos, em função da maneira dinâmica e complexa com que cada sistema individual se desenvolve. Entende-se que os resultados apresentados evidenciam a necessidade de considerar tanto ouvinte quanto falante em investigações acerca da inteligibilidade da fala em L2, confirmando, assim, a natureza dinâmica e complexa desse processo.

Key-words: inteligibilidade local, fala não nativa, vogais do inglês, aprendizes brasileiros, Teoria dos Sistemas Dinâmicos Complexos.

LIST OF FIGURES

Figure 2.1 – Plot (F1 x absolute duration) of L1 vowel productions by monolingual native Canadian French (left) and monolingual native Canadian English (right) speakers 44

Figure 2.2 – Plot (F1 x absolute duration) of L1 vowel optimal perception by monolingual native Canadian French (left) and monolingual native Canadian English (right) speakers 45

Figure 2.3 – Plot (F1 x absolute duration) of L2 vowel perception by monolingual native Canadian French speakers (left) listening to English vowel productions and by monolingual native Canadian English speakers (right) listening to French vowel productions 46

Figure 2.4 – Acoustic space of stressed vowels in the productions by speakers of the Porto-Alegrense variety of Brazilian Portuguese, as reported by Pereyron (2007) 53

Figure 2.5 – Acoustic space of stressed vowels in the productions by speakers of the Argentinian Riverplate variety of Spanish, as reported by Pereyron (2007) 55

Figure 2.6 – Acoustic space of stressed vowels in the productions by speakers of the standard variety of German, as reported by Sendlmeier and Seebode (2010) 57

Figure 2.7 – Acoustic space of stressed long (left) and short (right) vowels in the productions by speakers of the standard variety of German, as reported by Sendlmeier and Seebode (2010) 58

Figure 2.8 – Acoustic space of stressed vowels in the productions by male (left) and female (right) speakers of the Californian variety of North American English, as reported by Rauber (2006) 59

Figure 3.1 – Screenshot of a slide of the MS PowerPoint presentation file sent to participants in the production task. The target word in this example is 'fit', displayed in bold font-face 93

Figure 3.2 – Order and randomisation process of carrier sentences 95

Figure 3.3 – Desktop version of the perception task, before participants had played the audio file or selected an answer 101

Figure 4.1 – Relative durations of the productions of [æ] and [ɛ], according to proficiency level 106

Figure 4.2 – Individual F1 x F2 plots of the vowels [æ] and [ɛ], according to the participants' proficiency level 110

Figure 4.3 – Proportion of accurate identifications of words with [æ] and [ɛ], according to the listener's L1 groups 114

Figure 4.4 – Proportion of vowel [æ] and [ɛ] identifications, regardless of accuracy, by L1 listener group 115

4.1.2 Identification accuracy of words with [æ] and with [ɛ] (RQ1) 115

Figure 4.5 – Log-odd estimates of the mixed-effects logistic model for effects of interacting L1 and target vowel on accurate identification rates of [æ] and [ɛ] 118

Figure 4.6 – Mixed-effects logistic model's predicted effects of interacting L1 and target vowel on accurate identification rates of words with [æ] and [ɛ] 118

Figure 4.7 – Vowel dispersion of [æ] and [ɛ] productions in English according to speaker's proficiency level 119

Figure 4.8 – Log-odd estimates of the mixed-effects logistic model for the effects of L1, target vowel, F1 and F2 on vowel identifications of [ɛ] 123

Figure 4.9 – Mixed-effects logistic model's predicted effect on [ɛ] vowel identification considering L1 (left) and target vowel (right) as predictor variables 123

Figure 4.10 – Mixed-effects logistic model's predicted effect on [ɛ] vowel identification considering F2 as a predictor variable 124

Figure 4.11 – F1 x relative duration plots of [æ] and [ɛ] tokens with above-chance correct and incorrect identification, by target vowel 130

Figure 4.12 – F1 x F2 plots of [æ] and [ɛ] tokens with above-chance correct and incorrect identification, by target vowel 135

Figure 4.13 – F1 x relative duration plots of [æ] and [ɛ] tokens with above-chance vowel identification, by L1 138

Figure 4.14 – F1 x F2 plots of [æ] and [ɛ] tokens with above-chance vowel identification, by L1 140

Figure 4.15 – F1 x relative duration and F1 x F2 dispersion plots of [æ] and [ɛ] tokens with above-chance vowel identification 142

Figure 4.16 – Relative durations of the productions of [i] and [ɪ], according to proficiency level 146

Figure 4.17 – Individual F1 x F2 plots of the vowels [i] and [ɪ], according to the participants' proficiency level 150

Figure 4.18 – Proportion of accurate identifications of words with [i] and [ɪ], according to the listener's L1 groups 153

Figure 4.19 – Proportion of vowel [i] and [ɪ] identifications, regardless of accuracy, by L1 listener group 154

Figure 4.20 – Log-odd estimates of the mixed-effects logistic model for effects of interacting L1 and target vowel on accurate identification rates of [i] and [ɪ] 157

Figure 4.21 – Mixed-effects logistic model's predicted effects of interacting L1 and target vowel on accurate identification rates of words with [i] and [ɪ] 158

Figure 4.22 – Vowel dispersion of [i] and [ɪ] productions in English according to speaker's proficiency level 161

Figure 4.23 – Log-odd estimates of the mixed-effects logistic model for the effects of L1, target vowel, F1 and F2 on vowel identifications of [i] 164

Figure 4.24 – Mixed-effects logistic model's predicted effect on [i] vowel identification considering target vowel as a predictor variable 165

Figure 4.25 – Mixed-effects logistic model's predicted effect on [i] vowel identification considering F1 (left) and F2 (right) as predictor variables 165

Figure 4.26 – F1 x relative duration plots of [i] and [ɪ] tokens with above-chance correct and incorrect identification, by target vowel 170

Figure 4.27 – F1 x F2 plots of [i] and [ɪ] tokens with above-chance correct and incorrect identification, by target vowel 173

Figure 4.28 – F1 x relative duration plots of [i] and [ɪ] tokens with above-chance vowel identification, by L1 176

Figure 4.29 – F1 x F2 plots of [i] and [ɪ] tokens with above-chance vowel identification, by L1 179

Figure 4.30 – F1 x relative duration and F1 x F2 dispersion plots of [i] and [ɪ] tokens with above-chance vowel identification 181

LIST OF TABLES

Table 2.1 – F1 (Hz), F2 (Hz) and absolue duration (ms) values for [æ, ɛ, i, ɪ] in native productions of Porto-Alegrense Brazilian Portuguese (BP), Riverplate Spanish (ES), Central German (DE) and Californian North-American English (EN) 61

Table 3.1 – Self-assessed proficiency levels of Brazilian participants in their L1, L2, L3 and L4 80

Table 3.2 – Self-assessed proficiency levels of Canadian participants in their L2, L3 and L4 83

Table 4.1 – Average measures of absolute duration (ms) of native and L2 vowels produced by native speakers of Brazilian Portuguese, Riverplate Spanish, German, Canadian English and North-American English 108

Table 4.2 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of interacting L1 and target vowel on accurate identification rates of [æ] and [ɛ] 117

Table 4.3 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of L1, target vowels, F1 and F2 on vowel identification (regardless of accuracy) of response variable [ɛ] 122

Table 4.4 – Average measures of absolute duration (ms) of native and L2 vowels produced by native speakers of Brazilian Portuguese, Riverplate Spanish, German, Canadian English and North-American English 148

Table 4.5 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of interacting L1 and target vowel on accurate identification rates of [i] and [ɪ] 156

Table 4.6 – Average measures of F1 (Hz), F2 (Hz) and absolute duration (ms) of the native [i] vowel produced by native speakers of Porto-Alegrense Brazilian Portuguese, Riverplate Spanish, German, Ontario Canadian English and North-American English 159

Table 4.7 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of L1, target vowels, F1 and F2 on vowel identification (regardless of accuracy) of response variable [i] 163

Table 7.1 – Brazilian speaker participants’ responses to QuExPli’s Linguistic section 249

Table 7.2 – Brazilian speaker participants’ responses to QuExPli’s Demographic section 250

Table 7.3 – Canadian speaker participants' responses to QuExPli's Linguistic section 251

Table 7.4 – Canadian speaker participants' responses to QuExPli's Demographic section 252

Table 7.5 – Argentinian and German listener participants' responses to QuExPli's Linguistics sections 253

Table 7.6 – Argentinian and German listener participants' responses to QuExPli's Demographic sections 265

Table 7.7 – Brazilian speaker participants' responses to the Vocabulary Questionnaire 271

LIST OF BOXES

Box 3.1 – Target and distractor words (CVC monosyllables) 89

Box 3.2 – Word frequency per million words of target and distractor words 90

LIST OF SYMBOLS, ACRONYMS AND ABBREVIATIONS

App – Application

BNC – British National Corpus

BP – Brazilian Portuguese

BR – Brazilian

CA – Canadian

CDS – Complex Dynamic System

CDST – Complex Dynamic Systems Theory

CE – Canadian English

CF – Canadian French

COCA – Corpus of Contemporary American English

CSV – Comma-separated value (file extension)

CVC – Consonant-Vowel-Consonant

DE – German

EN – English

ES – Spanish

F1 – First Formant

F2 – Second Formant

FR – French

IT – Italian

KO – Korean

L1 – Native Language

L2 – Second Language

M – Mean

MS – Microsoft

PDF – Portable Document Format (file extension)

PL – Polish

POA – Porto Alegre

PT – Portuguese

QuExPli – Language History Questionnaire

RoF – Rate of Frequency

RQ – Research Question

RU – Russian

SD – Standard Deviation

SE – Swedish

SLM – Speech Learning Model

SMSPA – Secretaria Municipal de Saúde de Porto Alegre

TCLE – Termo de Consentimento Livre e Esclarecido

UFRGS – Universidade Federal do Rio Grande do Sul

URL – Uniform Resource Locator

VOT – Voice Onset Time

TABLE OF CONTENTS

1 INTRODUCTION 23

2 THEORETICAL FRAMEWORK 28

2.1 A COMPLEX, DYNAMIC VIEW OF LANGUAGE 28

2.2 PERCEPTION AND PRODUCTION OF L2 SPEECH SOUNDS 36

2.3 INTELLIGIBILITY 48

2.4 VOWEL SYSTEMS 52

2.4.1 The vowels in Porto-Alegrense Brazilian Portuguese 53

2.4.2 The vowels in Argentinian Riverplate Spanish 54

2.4.3 The vowels in Central German 55

2.4.4 The vowels in English 58

2.4.5 Contrasts amongst the vowel systems 61

2.5 FINAL CONSIDERATIONS 62

3 METHOD 66

3.1 GOALS 68

3.1.1 Main goal 69

3.1.2 Specific goals 69

3.2 RESEARCH QUESTIONS 70

3.2.1 Research Question 1: Does the L1 have an effect on L2 word intelligibility? 71

3.2.2 Research Question 2: Which linguistic, stimulus-related or participant-related variables might have an effect on vowel identification? 73

3.2.3 Research Question 3: What characteristics of listeners' L2 phonetic categories can be inferred from their identification patterns in the perception task? 74

3.3 PARTICIPANTS 77

3.3.1 Non-native speakers of English (native speakers of Brazilian Portuguese) – Production task 77

3.3.2 Native speakers of English (native speakers of Canadian English) – Production task 81

3.3.3 Non-native listeners of English (native speakers of Riverplate Spanish and Central German) – Perception task 84

3.4 INSTRUMENTS 85

3.4.1 Informed Consent Forms – prior to participation in any task 86

3.4.2 Language history questionnaire – Production and perception tasks 87

3.4.3 Production task 88

3.4.3.1 Selection of target and distractor words 88

3.4.3.1.1 Production samples from native speakers of BP 91

3.4.3.1.2 Production samples from native speakers of Canadian English 92

3.4.3.2 MS PowerPoint presentation file for the recording of stimuli by Brazilian and Canadian participants 92

3.4.4 Perception task 95

3.4.4.1 Samples 95

3.4.4.2 Stimuli 97

3.4.4.3 Task sequence 99

3.5 DATA ANALYSIS PROCEDURES 102

3.6 FINAL CONSIDERATIONS 102

4 RESULTS AND DISCUSSION 104

4.1 MINIMAL PAIRS ‘PAT’ – ‘PET’ AND ‘SAT’ – ‘SET’ 104

4.1.1 Descriptive Analyses 105

4.1.1.1 Stimuli production 105

4.1.1.2 Stimuli perception 113

4.1.2 Identification accuracy of words with [æ] and with [ɛ] (RQ1) 115

4.1.3 Effects of predictor variables on the identification of [æ] and [ɛ] (RQ2) 120

4.1.4 Exploratory analysis of listeners’ composite L1-L2 categories for [æ] and [ɛ] (RQ3) 125

4.2 MINIMAL PAIRS ‘FEET’ – ‘FIT’ AND ‘SEAT’ – ‘SIT’ 144

4.2.1 Descriptive analyses 144

4.2.1.1 Stimuli production 144

4.2.1.2 Stimuli perception 152

4.2.2 Identification accuracy of words with [i] and with [ɪ] (RQ1) 154

4.2.3 Effects of predictor variables on the identification of [i] and [ɪ] (RQ2) 162

4.2.4 Exploratory analysis of listeners’ composite L1-L2 categories for [i] and [ɪ] (RQ3) 167

4.3 FINAL CONSIDERATIONS 183

4.3.1 Research Question 1: does the L1 have an effect on L2 word intelligibility? 184

4.3.2 Research Question 2: Which predictor variables can explain the identification of a token as having [æ], [ɛ], [i] or [ɪ]? 185

4.3.3 Research Question 3: how can the identification patterns in the perception task shed light on the listeners' composite L1-L2 categories? 186

5 CONCLUSION 190

5.1 SUMMARY OF RESULTS 190

5.1.1 Minimal pairs with [æ] and [ɛ] 191

5.1.2 Minimal pairs with [i] and [ɪ] 194

5.2 LIMITATIONS OF THE STUDY 196

5.3 CONTRIBUTIONS TO THE FIELD AND FURTHER STUDIES 199

REFERENCES 202

APPENDICES 210

1 INTRODUCTION

Contemporary research in the field of L2 phonetic-phonological development[1] has focused on the perception and production of speech sounds, seeking to understand how language processing occurs (FLEGE; BOHN, 2021; FLEGE, 1995; BEST; TYLER, 2007; BEST, 1995). Research related to the intelligibility construct (ALBUQUERQUE, 2019; MUNRO; DERWING, 1995; SMITH; NELSON, 1985), in this scenario, has also guided investigations into which L2 production patterns can result in perception identifications with a greater or lesser degree of success in relation to the speakers' intentions. Moreover, studies have shown the relevance of the speaker-listener pair in the analysis of communication success. That is, the interaction between different groups of speakers and different groups of listeners results in different rates of success, in which there is a combination of phonetic-phonological characteristics of all languages at play (ALBUQUERQUE, 2019; ALBUQUERQUE; ALVES, 2017; DERWING; MUNRO, 2015; VAN LEUSSEN; ESCUDERO, 2015; ESCUDERO; POLKA, 2003).

In this context, the present research study seeks to investigate the local intelligibility (MUNRO; DERWING, 2015) of the L2 English spoken by native speakers of Brazilian Portuguese (BP). It is of particular interest to us to explore the communication success of the interaction of those speakers with other non-native learners. Non-native speakers of English outnumber native speakers (CRYSTAL, 2003), so research and pedagogical practices, in our view, should look more closely on how those interactions in the L2 play out when speakers are from distinct L1 backgrounds. We understand that different speaker-listener pairs will lead to emerging phenomena that will have an impact on the success of communication, according to the theoretical frameworks we adopt.

The present study assumes the Complex Dynamic Systems Theory (CDST) as the study's view of language. That means that we understand language as a system that varies with time (dynamic), adapting as a function of the speaker's embodied experiences (adaptive). Additionally, we understand that language is a process, rather than a static product. Accordingly, each stage of the process would be the result of the interaction of various factors (complex), in which the whole is not just the sum of its parts (DE BOT; LOWIE; THORNE, 2013; BECKNER *et al.*, 2009). Hence, we follow the premise that a speaker's productions are not restricted to phonological distinctive features being either present or absent, categorically.

[1] We will make no distinction between the terms 'additional language', 'foreign language' or 'second language' (L2). These terms will be used interchangeably in this thesis.

Rather, the acoustic cues of speech samples are presented in a gradient fashion. Likewise, we expect that a listener's perception would also be affected by aspects such as listener's linguistic experience and its interaction with other subsystems (morphological, syntactic, semantic), also showing gradient and emerging characteristics. Finally, we understand that the interaction of the speaker's and the listener's hybrid systems will also lead to distinct outcomes, depending on the speaker-listener pair.

In this present study, the CDS view is presented as a meta-theory (LARSEN-FREEMAN, 2017). Though dynamism permeates the present work as the concept of language, the method adopted does not follow variability through time. That is, instead of a longitudinal design, we adopt a cross-sectional one. Therefore, our analysis consists of a 'product' approach (YU; LOWIE, 2019). In other words, it is understood that there is a trajectory before and after the moments of data collection and analysis, but the present study will focus only on one static point on this assumed timeline, the moment-in-time of our data collection. Without a longitudinal design, we cannot observe language development through time, nor can we analyse a learner's language trajectory. We do, however, assume that such a trajectory is in course, and we understand that a cross-sectional look at our learners' data can bring to light useful information regarding the individual developmental paths that are bound to be present in any group of participants in a study such as ours.

Our main grouping factor was the learners' native language, as our study collected data from non-native speakers of L2 English. That is because some characteristics in non-native speech are considered to deviate from versions considered as standard in that language. In addition to the acoustic characteristics of non-native speech, specific communities of speakers, as well as of listeners, have relatively standardised characteristics in the production and perception of speech sounds in a given language. In order to understand which factors may have an effect on communication between speakers and listeners, therefore, it is necessary to understand the characteristics of both speech production and perception in the interaction of these communities.

Previous studies point out that acoustic cues produced by non-native speakers with different L1 backgrounds affect the perception by native listeners (ALVES, 2018; ESCUDERO, 2009; ESCUDERO; POLKA, 2003; FLEGE, 1995). Other studies reveal that non-native perception is also affected by the listener's L1 (ESCUDERO; POLKA, 2003; BOHN; FLEGE, 1992). Hence, we understand that it is necessary to investigate the supposed "deviations" on the part of the non-native speakers, as well as evaluate if and how those non-target forms affect the perception by non-native listeners.

In this thesis, we will adopt the Speech Learning Model (FLEGE; BOHN, 2021; FLEGE, 1995) as our framework for the perception and production analyses. Accordingly, we will use the Cue Weighting proposal (HOLT; LOTTO, 2006) – incorporated in the revised version of the SLM (FLEGE; BOHN, 2021) – to analyse the effect of the stimuli's acoustic cues in our listeners' perception. The SLM predicts that new speech sounds will be categorised according to their perceived similarity to known speech sounds. Acoustic similar sounds, thus, are assumed to go through a process of assimilation, in which they will be categorised as a sound the speaker already has in their L1 inventory. This will impact L2 learning, as new L2 sounds might be assimilated into L1 categories, if those sounds are not perceived as dissimilar – that is, as tokens of a different category. The SLM also assumes that speech categories share a common phonetic space, in which L1-L2 composite categories are organised. Though the terms 'similarity' and 'dissimilarity' still lack a more precise definition, acoustic cues presented by productions are known to play a role in how listeners perceive speech sounds, as they work as the unit of speech perception in the model. Those cues are weighed in a language-specific process depending on a number of factors, such as their informativeness, among others. This assumption is what led us to analyse the effects of different acoustic cues in our listeners' perception, within the SLM framework.

Finally, we also adopt 'intelligibility' as a measure of communication success among our non-native speaker participants and our non-native listener participants. Munro and Derwing (1995) define 'intelligibility' as a construct that evaluates if the interlocutor was able to understand the speaker's intention[2]. In the present thesis, we focus on local intelligibility at the word level (MUNRO; DERWING, 2015), that is, the listener's ability to understand the word the speaker intended to pronounce. Moreover, we follow Albuquerque's (2019) account of intelligibility within a Complex, Dynamic framework.

Therefore, our study design included Brazilian speakers as well as Argentinian and German listeners, all of whom are non-native learners of L2 English. We collected data from three groups of Brazilian speakers at different levels of proficiency in English. We expected that this would yield different patterns of production, reflecting different developmental stages in the L2. Brazilian participants were also speakers of the same L1 variety, for we expected that an 'L1 filter' (FLEGE, 1995; FLEGE 2021) would play a role in those participants' productions. The identification task was applied to listeners of two L1 groups: native speakers of Spanish, more specifically of the Riverplate variety from the Buenos Aires state in Argentina

[2] The lack of specificity of this definition will be further discussed in section 2.3.

(Argentinian listeners), and native speakers of German, of the Central variety from the Central area in Germany (German listeners). Our choice of having non-native learners perform the identifications was made so we could investigate the role played by the listener's L1 in their perception. The listener participants were asked to identify Brazilian productions in a forced-choice identification task, in which the options were the target words 'pat', 'pet', 'sat', 'set', 'feet', 'fit', 'seat', 'sit'.

In this thesis, we operationalise 'intelligibility' as the accurate identification of the target words produced by Brazilian learners and presented to our two groups of listeners. We also want to investigate what factors might play a role in higher or lower intelligibility rates. Accordingly, we will look at those aspects considering the speaker-listener pair.

For the analyses, we labelled productions in terms of target word and vowel, as well as acoustic measures (relative duration, F1 and F2). We also took the speakers' proficiency level into account. We observed the listeners' perception in regards to accurate word identification and general vowel identification. Finally, based on the production and identification patterns of the vowels [æ, ɛ, i, ɪ], we exploratorily analysed the role played by acoustic cues in the listeners' perception, attempting to understand how those cues interact and outline vowel categories in our participants' common phonetic space (FLEGE; BOHN, 2021).

Departing from the aforementioned background and experimental design, the main goal[3] of the present study is to investigate the word-level local intelligibility of the L2 English spoken by Brazilian learners, when those speakers are listened by other non-native learners of English – specifically, Argentinian and German learners of L2 English. More specifically, we intend to look at CVC monosyllabic words with the minimal pairs of vowels [æ] – [ɛ] and [i] – [ɪ]. In L1 English, the aforementioned vowels have distinctive temporal and spectral acoustic characteristics. However, previous studies have shown that Brazilian learners tend to not distinctively produce those acoustic cues, or at least not in the same way that native speakers do (GONÇALVES, 2014; RAUBER, 2006; NOBRE-OLIVEIRA, 2003). Therefore, we understand it might prove useful to investigate the local intelligibility of the non-native productions of these words by Brazilians to non-native listeners of L2 English.

In order to achieve our goal, we posed three research questions. The first one enquires about the possible effect of the listeners' L1 on L2 word intelligibility in our stimuli. We also want to check if the target vowel plays a role in those rates. Our second research question disregards 'word identification accuracy' and enquires about the listeners' vowel identification

[3] We will detail our main and specific goals in section 3.1.

patterns. As we understand it, word intelligibility entails the identification of the segments that compose a word, among other processes. Thus, we question which linguistic, stimulus-related or participant-related variables might have an effect on vowel identification. Finally, as we adopt a Complex, Dynamic view of language, we intend to look at what characteristics of the listeners' L2 phonetic categories can be inferred from their identification patterns in the perception task. We expect that our data will be able to outline how L1-L2 composite categories are organised within the common phonetic space and what effects can emerge when they interact in a communicative situation[4].

We applied two distinct methodologies in our analyses. With inferential statistics, we estimated the effects of predictor variables on word accurate identification (RQ1) and on general vowel identification (regardless of accuracy in relation to the target word/vowel – RQ2). Complementarily, we adopted an exploratory method to try and understand the role played by acoustic cues in accurate and in general vowel identifications (RQ3). Our choice for both the statistical and the qualitative analyses is in line with our view of language, as we assumed that interaction and gradience in data would have effects that only a stimulus-by-stimulus observation would be able to pick up on[5] (LOWIE, 2017).

The present thesis is composed of five chapters. Besides this first, introductory one, we will describe the theoretical background we adopt (chapter 2), as well as our goals and the design of the experiment in which we collected our data (chapter 3). We will then move forward to report our results and discuss how they relate to our goals and research questions (chapter 4). The last chapter summarises the contributions we believe our study can provide to the fields of both Laboratory Phonology and Applied Linguistics. We outline the limitations of the present study and indicate further studies that we believe can contribute to the discussions we will present throughout this thesis.

We hope our study can provide a contribution to both the fields of Laboratory Phonology and Applied Linguistics. We have designed this experiment expecting it can yield relevant information to the studies on L2 phonetic-phonologic development. These data could in turn prove useful in future studies that aim to further our understanding of language learning. Likewise, we hope to shed light on L2 production and perception processes, which will hopefully help teachers in the pronunciation instruction they provide to students, leading to higher success rates in L2 communication.

---

[4] We will detail the motivation for our research questions in section 3.2.
[5] This will be further discussed in sections 3.2.3, 4.1.4 and 4.2.4.

2 THEORETICAL FRAMEWORK

In this chapter, we will describe the theoretical framework we adopt, as well as point out how it relates to our methodological choices. We will firstly describe our Complex, Dynamic view of language, as it shapes both our assumptions prior to the study and the adoption of other theoretical models. This view entices change through time, gradience, interaction and adaptation of systems, be them linguistic or otherwise. It conceives language learning as a process, instead of looking at language as a finished product.

Those key points of the Complex, Dynamic view of language led us to adopt the Speech Learning Model (SLM) by Flege (1995), and its revised version by Flege and Bohn (2021) as our perception background theory. We will detail their proposal, as well as how it applies to the present thesis, in the second section of this chapter. Holt and Lotto's (2006) Cue Weighting proposal, incorporated in the revised SLM, will also be detailed in the same section, as it is the basis of the analysis we set out to do in our study. We will also connect both theories to Escudero's (2009) experiment, which was the first inspiration for our research study, though we do not adopt her L2LP (ESCUDERO, 2009) perception model in this thesis.

Additionally, we will describe the construct of intelligibility, adopting Munro and Derwing (1995, 2015) definitions. As described in the Introduction, our main goal with our experiment is to investigate the intelligibility of accented non-native speech, specifically the intelligibility of the English spoken by native speakers of Brazilian Portuguese. We leave the description of intelligibility for the third section, though this construct is intrinsically connected to our main goal, because we understand that it somewhat connects the Laboratory Phonology and the Applied Linguistics dimensions of our work. This two-fold approach, in turn, is closely related to the theories we adopt and that we explain prior to the intelligibility section.

Finally, we provide a brief description of the vowel systems of the four languages that play a role in our study – namely, L1 Brazilian Portuguese, L1 Spanish and L1 German, the native languages of our speaker and listener participants, as well as L2 English, the common, foreign language they have all learned. That description aims to provide an outline of the aspects that were taken into consideration as we designed the experiment and analysed its results.

2.1 A COMPLEX, DYNAMIC VIEW OF LANGUAGE

Complex Dynamic Systems Theory (CDST) understands that language development is a process, in constant change and with gradual steps. Understood as domain-general, speech is an ever-evolving process, rather than a product that reaches a final stage. In this paradigm, both the production and the perception of speech sounds are phenomena that vary over time due to linguistic and socio-cognitive experiences, among many other factors (ALVES, 2018; LOWIE, 2017; LOWIE; VERSPOOR, 2015; DE BOT; LOWIE; VERSPOOR, 2007; DE BOT; LOWIE; THORNE, 2013; BECKNER *et al*., 2009).

The Dynamic view posits that learning a language, be it native or additional, is a process in constant development, in which there are always new (interconnected) items to be learned – vocabulary, structure, intonation, etc. Therefore, it makes little sense to think of 'acquisition' as if knowledge were a finished product, or as if an end-state could be attained. In a process, it is precisely the changes through time that deserve the researchers' attention.

Moreover, such a process will go through different stages depending on the different factors that can affect the subsystems and elements of the language system. Thus, in the case of speech sounds, the theory predicts that linguistic factors as well as social experience will play a role in an individual's language development process (LOWIE; VERSPOOR, 2015). Living in a new area where a local accent has a different colouring from the one in the previous residence, for example, may cause changes in an individual's system. This holds true for both a native language or an additional one. The new characteristics of the speech continuum met in this new speech community will require adjustments in the phonetic-phonological[6] categories so that the communication between the new-comer with their peers can be successful.

It is this view of learning as a process in particular that modulates the theoretical conception of the present study. Language is not a product, capable of reaching a fixed (fossilised) final stage. The shift from 'acquisition' to 'ongoing development' brings implications to the research carried out in our field. A Complex, Dynamic design (YU; LOWIE, 2019) presupposes observation of change over time. Though this is not within the scope of the present analysis, we adopt CDST as a metatheoretical framework. Larsen-Freeman (2017) argues that a Complex system account allows us to understand events as portraits of a moment in time. However, we are still aware that in the moments immediately before and after the timeframe under analysis, the system is different from the one we are investigating. That is, it was not and will not be the same as it was when data were collected. In this sense, this metatheoretical conception allows us to understand that when completing the tasks proposed in the experiments carried out in this thesis, the linguistic systems of our listeners have potentially undergone changes caused by the linguistic experience with the task itself. This hypothesis,

[6] Following Alves (2018), we adopt the term 'phonetic-phonological', as we understand that assuming a continuum from the physical and the more abstract, representational aspects is more in line with the Complex, Dynamic view of language.

however, is not considered in this thesis, which, on the contrary, uses a cross-sectional methodology (single test, without post-test or delayed post-test, or any type of longitudinal collection). Thus, our design is intended to observe what could be considered a single moment of the trajectory under analysis.

This product-based design (LOWIE, 2017) is still capable of yielding generalizable results about factors that may have effects on the learning trajectory. Within the design of our study, this metatheory posits that each proficiency level, with its group averages, will showcase the state of systems after a given amount of experience. We have attempted to equate this experience via the inclusion criteria. We also expect that the questionnaires we applied (see section 3.4.2) could enlighten us as to some indicators of the quality of that experience. At the same time, the fact that each participant's trajectory is inherently different does not elude us. Moreover, we expect that those differences might colour each trajectory in a distinct way. Therefore, as we adopt CDST as a meta-theory, we assume that the learners' trajectories prior to data collection will reflect effects of the variable 'time' with regard to the level of proficiency. In other words, we presuppose that a higher proficiency level is somewhat the result of more experience with the L2 (FLEGE; BOHN, 2021). It is expected, therefore, that less proficient speakers, participants in the production task (see section 3.3), will have had less experience with the target language, while the more proficient learners, both speakers (production) and listeners (perception task), will have had more experience. By the predictions of a Complex, Dynamic view of language, this experience prior to the data collection will have shaped the linguistic systems of the participants up to that point. Thus, we could expect that their proficiency level can reflect a greater or lesser proximity to native patterns[7]. Moreover, in terms of the intelligibility construct that interests us here, proficiency levels could also be reflected in a greater or lesser percentage of accuracy in the identification tests by the listeners, translating into greater local intelligibility (see section 2.3).

Proficiency level is not the only aspect that can be seen under a dynamic light. In addition to constantly changing over time, Complex, Dynamic systems are composed of non-hierarchical, non-linearly organised subsystems. This means that a change in one subsystem can lead to changes in other subsystems. It also implies that a small change at one point can lead to great effects on the system as a whole (DE BOT; LOWIE; THORNE, 2013). The effect is by no means proportional to the size of its cause (DE BOT; LOWIE; VERSPOOR, 2007). In this sense, phonology, as well as morphology and syntax, can be considered to be subsystems

[7] This 'native pattern' could be assumed to be the 'target' of language development, not in a nativeness sense, but as the better 'role model' for a more intelligible speech.

of a larger language system. In the same fashion, each member of a speech community can also be seen in this way, as language itself "operates as a subsystem of embodied cognition" (LOWIE, 2017, p. 124). Thus, interactional situations between the speaker-listener pairs can play a role in the systems' (re)organisation. In sum, the effects of those interactions, over time and in a complex fashion, may lead to system changes.

In a traditional sense, therefore, Complex, Dynamic investigations involve longitudinal studies, in which the change is observed for a period of time with multiple data collections from the same individual. This is due to the fact that, in the interaction of its components, a system tends to leave its initial stability stage and go through a period of destabilisation. Over that time, a new attractor state will take the system to a new point of stability. That is to say that systems are in a constant process of balance/imbalance that results from the constant interactions they are involved in; hence the continuous development of the system (LOWIE, 2017). This process will be particular to each learner, since there is no way two people can have the exact same experience to yield the exact same trajectory. Lowie highlights the relevance of an individual look, while also validating complementary methods of analysis with groups that can be profitable for the field. That is the strategy we adopt, as we will detail further in section 3.5.

As we adopt the CDST framework, it is also relevant that we highlight the adaptive characteristic of Complex, Dynamic systems. That is, new experiences, when bringing new information into the system, lead to a reorganisation of the system as a whole. This process means to incorporate the newly developed knowledge into the pre-existing repertoire. Adaptiveness, therefore, is very much connected with change through time: embodied experiences will play a role in destabilizing the system, as well as bring it back to a new stable state. Adaption is one of the reasons change occurs. In the previous example, when a listener moves to a region where there is a new accent, we expect that an adaptation will be necessary. The new acoustic characteristics, lexical items, prosodic patterns or even pragmatic scenarios presented by the new speech community would require those changes in the listener's system.

For the present study, this adaptive dimension presents itself as a *sine qua non* condition for the interactions between non-native speakers and non-native listeners. We make this claim based on a number of factors. On the one hand, both members of an interaction are dealing with systems that have developed under different conditions and with presumed distinct trajectories. We also expect that those different paths will have suffered effects of the learners' native languages. Additionally, the very interaction with a new speaker/listener may bring about novelties. All of these aspects lead us to expect that some sort of adaptation will be necessary in the communication between our non-native speaker-listener pairs. In fact, the adaptive characteristic of Dynamic systems is one of the factors that can lead to emerging effects in that communication. Within a Complex, Dynamic view, we expect that the elements of the speaker's experience somehow 'show up' in their speech. We also assume it may affect the listener's perception and interaction in a broader sense.

Throughout a learner's language developmental trajectory, the characteristics of the L2 need to be perceived and accordingly produced. Each new piece of information that reaches the learner might lead to system adaptations, in order for that piece of information to be incorporated in the learner's repertoire. Again, we highlight that such a process is not linear. Thus, when we consider two groups of non-native learners, we ought to expect that their trajectories are bound to be distinct – even when we try to control this variable in laboratory experiments. In our study, for instance, we have three groups of non-native learners (see section 3.3). Members of each group (speakers and listeners) would have needed to adapt in different ways, throughout each of their learning trajectories. Firstly, a first point to consider is the participants' native languages. That is, each native language system (Brazilian Portuguese, Spanish and German) differs from that of the common additional language (English). But those differences might not be the same for all three systems: there could be a cognate word in one language that is not present in the other(s), a sound that is similar, a morphological particle that works the same way in both systems, and so on and so forth. These possibilities have been raised by considering only the L1, though we know that all embodied experiences of each individual might affect their language system. As the adaptations that each learner has undergone is different, the state of each learner's system (at the time of our data collection) is also expected to be different. That is highly relevant for our study, as we set out to investigate how those different, hybrid, non-native systems may interact and lead to (lack of) success in communication in L2 English.

In sum, as a meta-theory, the CDST allows us to assume that learning an additional language entices adapting both perception and production to the L2 systems. Moreover, learners from different L1 backgrounds can be expected to perform different adaptations upon learning the same L2. Though we do not investigate those adaptations through time, we highlight the adaptive characteristic of the Complex, Dynamic Systems because they are crucial to the exploratory analyses we will perform (see section 3.4). We chose to work with non-native learners in large part due to the role we assume that each 'L1 filter' will play in our learners' developmental trajectories. As we will explain further on, success in communication does not assume a 'nativeness' parameter anymore (LEVIS, 2005, 2018). That is, the goal of language learning is not to sound like a native speaker of that language, but to be able to successfully communicate using that language – regardless of whether the interlocutor is a native or another non-native speaker. Therefore, as we set out to investigate the intelligibility of productions by Brazilian learners of English, we decided to look into how other non-native speakers would understand those productions. As we have mentioned, we designed our experiment in a way that would allow us to somewhat connect Laboratory Phonology findings with the tenets of Applied Linguistics. That is, we expect that our results will contribute to the field of Laboratory Phonology, but also help teachers in guiding their students towards an intelligible L2 speech.

We reiterate that our proposed methodology combines group-analyses as well as stimuli-by-stimuli observations, each in view of a different goal. These complementary approaches are based on the fact that, on the one hand, trajectories are somewhat unique; on the other hand, this uniqueness does not mean that patterns cannot be found if we look at large-enough time windows. This way, though we do group listeners by their L1 in our study, we cannot assume that there is any strict predictability of trajectories in view of the role played by the native language. Likewise, we cannot expect that, because of idiosyncratic differences, it is not possible to generalise results (LOWIE, 2017). Thus, looking at both ‘levels’ of analysis and discussion can yield complementary results. De Bot, Lowie and Thorne (2013) explain that grouping is possible because Complex, Dynamic Systems present “recurrent patterns at different scales in space and time” (p. 201). In other words,

> "we expect language development over a full lifespan, decades, or years to show similar developmental variability and patterns as at shorter periods such as months, weeks, days, hours, seconds and milliseconds." (*op. cit.*, p. 203)

The authors also highlight that timescales, as well as anything else in a Complex, Dynamic System, also interact. Additionally, since the linguage system is also complex, the changes suffered by one part of the system can result in effects in other parts, as they are all interconnected (BECKNER *et al.*, 2009). On the one hand, these interactions become specific to the point of giving rise to a unique trajectory for each person. On the other, though, the fractality of the system allows us to find points of convergence that are common to different trajectories (DE BOT; LOWIE; THORNE, 2013). In language learning, larger time windows can reveal recurrent patterns in group studies. The the idiosyncrasies of the group members then appear when approaching the eyeglass to smaller time windows – in which changes are occurring, each in its own way, for each speaker. For the present study, this is relevant, once again, in the choice of proficiency groups. As mentioned, we assume that each learner is a Complex, Dynamic system (and a subsystem of the speech community in which they are inserted). However, when grouping participants in view of their level of proficiency, we could be able to find similarities in the developing linguistic systems. In other words, the fractality principle allows us to understand that a given level of proficiency is a large-enough time window to provide evidence of a state of the system. As we have explained, we assume that the proficiency level somewhat reflects the amount of experience (through time) that a learner has had in their language development trajectory. We reiterate that grouping participants by proficiency level may even out individual differences that might become readily apparent when approaching each learner individually.

In view of what has been exposed, it is expected that native speakers of Brazilian Portuguese (BP) – a system in which the vowel [æ][8] does not occur – will need to perceive the distinctive role that this sound has in the English system (in pairs like [æ] 'pat' – 'pet' [ɛ], for example). In other words, the leaner would need to realise [æ] and [ɛ] as two different sounds in order to be able to produce and understand that distinction (FLEGE, 1995; FLEGE; BOHN, 2021). In the same fashion, as BP does not have [ɪ] in its inventory, Brazilian learners would need to perceive the distinctiveness between [i] and [ɪ].

When analyzing productions from three different groups of proficiency, as proposed in the present study, it is not possible to point out the moment when each development stage takes place, as we do not collect data longitudinally. However, by analysing the learners' productions acoustically, we should be able to indicate whether this process of adjustment has already started or not. Additionally, we would be able to find evidence as to what degree this distinctiveness gradually appears in the speech of the Brazilian participants. We should note that the change referred to here would occur guided by the need for intelligibility (amongst other aspects). However, we reiterate that this does not necessarily imply an adjustment towards a given native pattern, but rather towards one equally relevant pattern in functional terms. In other words, when interacting with different listeners, and in order to make themselves understood by such listeners, learners can change the way they produce a sound, by adapting their production in order to allow for their interlocutors to understand it. This, of course, may also have an impact on the way these learners perceive sounds, as both perception and production develop in a complex fashion.

[8] We adopt Flege and Bohn's (2021) SLM proposal (see section 2.2), and their level of analysis is the positional allophone. Thus, we will congruently refer to all sound categories using brackets. Nevertheless, we once again highlight that we assume a continuum from the physical phone to its phonological representation (Cf. ALVES, 2018).

These aspects are of particular relevance to the present study, since interactions between non-native speakers and listeners can lead to new acoustic thresholds that allow for intelligibility without necessarily being equal to the thresholds of native speakers'. Flege and Bohn (2021) discuss how learners in a non-immersion context tend to learn a foreign language in formal contexts such as language schools, in which teachers are non-native speakers, just like their students. This likely means that the input offered by these teachers is produced with different thresholds from those used by native speakers. Likewise, the authors point out that adult learners in an immersion context acculturate more slowly than children, and tend to interact with more non-native speakers of the immersion language. Both these scenarios would mean that the quality of input they are basing their learning on is also skewed, when compared to native patterns. In the present study, the first case would comprise most participants, who learned L2 English in a non-naturalistic way, in both production and perception tasks. As for the quality of input received in communicative situations with other non-native speakers, specifically from different L1 backgrounds, both speakers and listeners can also be assumed to differ greatly. This, in turn, should mean different language system developmental states, according to a Complex, Dynamic framework.

We have mentioned that the native BP system does not have the [æ] vowel. Considering our groups of listeners, the central vowel is also not part of the native Spanish or German inventories. Therefore, we would expect that our Argentinian and German learners of L2 English would also need to go through a process of realizing that the [æ] sound exists in English and that it has a distinctive function from [ɛ] in pairs like 'pet' – 'pat'. That is, a similar process to that of Brazilian learners – though not necessarily in the same way for all groups. Conversely, the [ɛ] vowel is present in both BP and German, but not in Spanish, which would lead us to expect that the latter group of learners would need to realise the functional role of this vowel. However, we highlight that the native BP category of [ɛ], the native German category of [ɛ] and the native English [ɛ] might not share the same acoustic specifications, which in turn might mean that Brazilian and German learners would also have to adapt to that L2 category. As our study is cross-sectional, we do not investigate when or how these processes might have happened, and neither is it our goal. Our perception task is designed to allow us to investigate what role the temporal and vowel quality dimensions play in the vowel identification performed by our two L1 groups of listeners. That is, which (perhaps different) acoustic cue(s) has(ve) effects on each hybrid system, at this point of development, in discriminating between [æ] – [ɛ] and [i] – [ɪ].

In sum, the CDST is adopted in this study as a metatheory, in the sense that it understands language learning as a process, rather than as a product. Language learning is also an ever-changing, ever-adapting process, in which linguistic and extra-linguistic experiences will play a role in each stage of the system at any given moment in time. It is a metatheory in the sense that the study is not longitudinal, but understands that data collection in a single point in time reflects the system status at that point, while still being aware that group results average individual differences[9] in trajectories and experiences. A group-based design was chosen understanding that, despite idiosyncrasies, L2 proficiency level can be a good predictor of a system's stage of development, in which a given set of characteristics has already reached a certain level of stability that can be meaningful to the present research questions. Moreover, results can be interpreted to yield generalizable aspects of the learning trajectory. This way, our findings might also be useful for teachers when instructing Brazilian learners on the production (and perception) of L2 English vowel sounds.

2.2 PERCEPTION AND PRODUCTION OF L2 SPEECH SOUNDS

As language is a dynamic, complex, adaptive system, composed of other subsystems, we now turn our attention to the phonetic-phonological subsystem. Different theoretical contributions seek to explain how the psycholinguistic processing of multiple languages occurs, each of which having implications for the analysis and interpretation of experimental results. In this project, we adopted the proposal of the Speech Learning Model (SLM) by Flege (1995) and Flege and Bohn (2021)[10], which assumes an acoustic representation unit. Incorporated in the revised SLM (FLEGE; BOHN, 2021), we also adopt Holt and Lotto's (2006) Cue Weighting proposal, which provides a framework to analyse the prioritisation of acoustic cues in the perception of the speech signal.

The acoustic signal of oral speech is understood as a set of physical characteristics that the human brain is able to decode into representational units of meaning. Back in the generative view (CHOMSKY; HALLE, 1968), these units are the distinctive features, which gain categorical status of presence [+] or absence [-] and, when gathered in a matrix, compose a phoneme. This perspective, however, addresses the distinctive features in a binary way, one which does not account for the gradience observed in the speech continuum. This, in turn, poses

---

[9] Following Lowie (2017), participants are calculated as random intercept effects in our inferential models. See sections 4.1 and 4.2.

[10] This revised version is called 'SLM-r' by the authors. In the present study, we will refer to both versions as simply 'SLM'.

some difficulties in describing both the production and perception phenomena of L2 speech sounds, especially when temporal characteristics of speech production and development are taken into account.

The gradience in the speech continuum, as well as the way in which it is perceived, are especially relevant in the interaction of non-native speakers. Firstly, we expect that the new developing language is a system that will operate, in different instances, under an organisation different from the L1 system. Besides, the L1 is also a system that presents some gradience and that also changes in this gradience over time. The very fact that it is possible to plot vowel productions by their F1 (y axis) and duration (x axis) measures, as proposed here, and not as a binary feature bundle, is an indication that this gradience exists and has thresholds that characterise each sound category. In the generative view, for example, the phone [ɛ] has the same features in Brazilian Portuguese and in English. However, acoustic analyses indicate that the production of these segments is not carried out in the same way in each language. Besides, previous studies have shown that speech variability in both languages, whether in oral productions by native or non-native speakers, also implies different phonemic perceptions (ESCUDERO, 2009; FLEGE, 1995; FLEGE; BOHN, 2021).

The presence or absence ([+] or [-]) of distinctive features is categorical, and for this reason they are unable to explain gradient distinctions found in the production and perception of speech sounds in different languages and dialects (ALVES *et al*., 2018). The adoption of the SLM proposal allows us to analyse acoustic aspects of speech production and perception that are gradient, and therefore more in line with a change-over-time perspective.

The SLM is an L2 perception model that seeks to explain the individual differences in accented speech that learners present in an additional language[11]. It also aims to understand the aspects that have an effect on the development of that language. Using the positional allophone as its level of analysis, the SLM predicts that category assimilation in the perception process depends on the perceived phonetic similarity between the L1 and the L2 sounds. If the sound of the additional language is acoustically dissimilar to any other option in the native language inventory, the learner is likely to create a new category. The more an L2 sound is perceived as similar to an L1 sound, the greater the chances that phonetic distinctions between them will not be perceived, with a consequent assimilation of the L2 sound into a pre-existing L1 category. Additionally, the SLM highlights that two foreign sounds with a distinctive status in the L2

[11] Though Flege (1995) and Flege and Bohn (2021) specifically state that the SLM aims at a naturalistic (immersion) learning, it is our understanding that research in non-immersion contexts also provide evidence that supports the proposal, thus justifying our choice for it.

might both be perceived as having similar acoustic characteristics to that of a single L1 sound. This would be the situation that poses the greatest difficulty to the learner. In this scenario, the assumption of the model is that the listener will fail to perceive the dissimilarity of the two acoustically similar, yet distinct sounds, and will assimilate both to a single L1 category. Within the scope of the present study, it could be argued that both [i] and [ɪ] in English are perceived by Brazilian learners as similar to BP [i]. This would mean both vowels tend to be assimilated into the L1 category of [i]. The result, then, is a 'neutralisation' of the distinctive status the vowels have in L2 English, for instance in a minimal pair such as 'feet' – 'fit'. The same could be said about both L2 English [æ] and [ɛ], which may be perceived as similar, and consequently be assimilated into the single category of L1 BP [ɛ].

As to why a listener may fail to distinguish between the target sound and their L1 category, Flege (1995) summarises that

> [l]earners of an L2 may fail to discern the phonetic differences between pairs of sounds in the L2, or between L2 and L1 sounds, either because phonetically distinct sounds in the L2 are "assimilated" to a single category (...), because the L1 phonology filters out features (or properties) of L2 sounds that are important phonetically but not phonologically, or both. (p. 238)

In regard to this 'L1 filter', two aspects are highlighted by Flege (1995, 2003) and Flege and Bohn in particular (2021). First and foremost, this 'filter' is not static, to the contrary: the more experience the learner has with the L2, the more likely they are to "become better able to discern L1-L2 phonetic differences which will, in turn, increase the likelihood of a new L2 category being formed" (*op. cit*., p.31). This process, as well, will not happen all at once, as it will rather show some gradience as learning progresses. Moreover, the formation of a new L2 category might lead to changes in the L1 pre-existing categories, because those categories are all in a 'common phonetic space' (FLEGE; BOHN, 2021). In that space, categories will be distributed in a way that allows the speaker to maintain the distinctiveness among them.

As for what will mark up the distinction among categories, the authors (*op.cit.*) point to the acoustic cues present in the speech continuum. Flege and Bohn (2021) highlight that, when learning the native language, a speaker will create categories based on the integration of multiple cues present in the input this learner is exposed to. Among all acoustic aspects, a stronger weight will be put in the cues that are perceived as more salient and/or informative. Congruently, features that are not perceived to carry much relevant information will tend to be overlooked. An acoustic cue that presents too much variability, for instance, might be perceived as 'less informative', regardless of its salience. Furthermore, the acoustic boundaries of these native categories may be affected, over the speaker's lifespan, depending on the distributional patterns perceived in the input to which this speaker is exposed. Once again, let us go back to our earlier example of a person moving to a new city. If that community has a different accent than this speaker was used to, the new input might cause the speaker to adapt to those new acoustic characteristics, according to the distributional patterns they are now being exposed to. Moreover, as we think about 'boundaries', Flege and Bohn (2021) highlight that the native categories have a "narrow range of good exemplars of a phonetic category", around which there is a "tolerance region" (*op. cit*., p. 32). This way, listeners are still able to perceive a token as belonging to a category, even when some of its acoustic characteristics may not be "optimal".

All of these characterisations of how L1 develops will hold true for the L2 learning process. Importantly, inasmuch as the L1 may change over time depending on the input the speaker is exposed to, so will the L2 throughout its development. In that sense, Flege and Bohn (2021) posit that "[t]he influence of L1 cue weighting patterns will be stronger for L2 sounds which remain perceptually linked to an L1 category than for L2 sounds for which a new L2 phonetic category has been formed" (*op. cit*., p.44). With time, the L2 input will allow the learner to develop L2-specific cue weighting patterns. This process, in turn, will also be informed by the L2 tokens this learner is exposed to.

The adoption of the SLM has some important implications for the present thesis. First and foremost, the SLM assumes that language develops through time and that new L2 categories can be created. Flege and Bohn's (2021) proposal also leads us to believe that the 'L1 filter' will have a stronger or weaker effect on learners' L2 production/perception depending on how much experience they have had with the L2[12]. This, in part, is what justifies our choice for speakers at different proficiency levels, while all our listeners are at an advanced level. Moreover, as categories are established based on the integration of multiple cues, we have chosen to analyse the effect of more than one acoustic cue – namely, vowel duration and formant frequencies (F1 and F2), as described in section 3.4.4.1. Lastly, as the boundaries of those categories allow for 'good' and 'tolerable' exemplars, we understand that the SLM proposal is fit for an investigation of the role of speaker-listener interactions in communication success.

That being said, we would like to provide some further details on how Flege and Bohn (2021) understand the common phonetic space, as well as the possible effects that 'sharing' this space may bring upon L1 and L2 categories. As we have mentioned, all categories in the

[12] The quality of the input is also highly relevant. See Flege and Bohn (2021) for a discussion.

common phonetic space are organised in a way to allow for the maximum contrast among those categories – so each of them remains separate from the others. This organisation is the result of the perceived dissimilarities among those categories. Though the authors admit there is still debate as to the best way of measuring this degree of dissimilarity, they also highlight that its perception is language-specific. That is, the same acoustic aspect in a given input can be perceived differently by different listeners from different L1 backgrounds[13]. In this sense, an acoustic cue may be more relevant for a group of non-native listeners, while native listeners might make use of another main cue. Thus, in our study, the two non-native groups (Argentinian and German listeners) can categorise the audio stimuli in this experiment by making use of different acoustic cues – as well as of different combinations of more than one cue. This is highly relevant, as we are dealing with hybrid systems (that is, non-native learners whose common phonetic spaces are shared by different L1s and the same L2[14]).

As we reiterate the hybrid character of our participants' systems, it is also relevant to mention Flege and Bohn's (2021) concept of a 'L1-L2 composite category'. The authors propose this definition for those L2 sounds which are not yet part of a new, totally dissimilated category.

> "Crucially, however, learners do not discard audible phonetic information in such cases. By hypothesis, a perceptual link between the L2 sound and the closest L1 sound will continue to exist and a composite L1-L2 phonetic category will develop, defined by the statistical regularities present in the combined distributions of the perceptually linked L1 and L2 sounds." (*op. cit.* , p. 38)

In combination with the notion that categories will develop gradually, depending on the learner's experience with the L2, the SLM allows us to investigate the gradience in production, as well as in perception. We highlight that Flege and Bohn (2021) posit that perception and production co-evolve. Once again, given the hybrid nature of our participants' systems, we understand that the SLM proposal is a good fit for our Complex, Dynamic view of language.

In the present study, we did not set out to investigate oral productions by Argentinian and German learners. We will, instead, focus on the L2 mapping patterns we can observe for listeners. The SLM defines cross-language mappings as the way the L2 sound is perceived as mapping onto an L1 category. As learning develops, L2 sounds tend to be mapped onto L2

[13] Likewise, the same acoustic cues in tokens of two different languages can have distinct levels of informativeness in each of those languages.

[14] As we will detail in section 3.3, most of our participants had learned more than two languages. Within our Complex, Dynamic view of language, we understand that this fact might have had effects on our results. See section 5.2 for a discussion.

categories. Thus, we will use this definition to try and observe how our stimuli are mapped onto our listeners' L2 or L1-L2 composite categories. Again, we chose complementary methods to investigate how our listeners' categories are established, and what effects may emerge from the differences in the cues weighted by different non-native groups of learners.

In this sense, Flege and Bohn's (2021) model

> "proposes that the influence of L1 cue weighting patterns will be stronger for L2 sounds which remain perceptually linked to an L1 category than for L2 sounds for which a new L2 phonetic category has been formed. Cue weighing patterns for newly formed L2 phonetic categories are expected to develop as in monolingual L1 acquisition, that is, to be based on the reliability of multiple cues to correct categorization found in input distributions." (p. 44)

Thus, the present study focuses on how our listeners map the oral productions of Brazilian learners of English. We also investigate the effects of the temporal (duration) and frequency (spectral) cues on those mappings. This design stems from our main goal to evaluate the intelligibility of our speakers' productions. We expect that 'deviations' in production can lead, in their turn, to failure in communication with other speakers, being them native or non-native (MUNRO; DERWING, 2015). Likewise, we expect that listeners' experiences might also affect the perception process, in a way that they could counterbalance or further hinder those listeners' perception – and not necessarily always following one or the other tendency. Once again, both speaker and listener, when communicating to one another, will have a role in the (in)success of communication.

In view of this, we also adopt Holt and Lotto's (2006) Cue Weighting proposal. The authors highlight that the perception of speech sounds occurs using multiple cues. From this perspective, "[a]coustic dimensions appear to be perceptually weighted in the sense that some are strongly correlated to categorisation responses whereas others, although present, weakly determine perceived category membership" (*op. cit.*, p. 3059). The weight that each cue receives in a listener's linguistic system would be determined by the previous experience with a given "acoustic environment" (*op. cit.*, p. 3059), whose input is recurrent and relevant to the categorisation in question. The weighing of those cues is also language-specific, as well as dialect and even listener-specific. As the Cue Weighting proposal was incorporated by Flege and Bohn's (2021) revised version of the SLM, we highlight that the outcome of these weighting processes is that "the phonetic categories making up the L1 and L2 phonetic subsystems interact with one another dynamically and are updated whenever the statistical properties of the input distributions defining L1, L2, and composite L1-L2 categories (diaphones) change" (*op. cit*., p.1).

Holt and Lotto (2006) point out four variables in determining how cue weighting occurs: cue informativeness, distribution variance, robustness of the acoustic dimension and type of task. This latter variable relates to task aspects that can shape the listener's attention, while the robustness of the acoustic dimension has to do with the ease of the auditory system to perceive sound characteristics. We are particularly interested, however, in the first factor. Cue informativeness has to do with competition among categories – that is, dimensions whose cues overlap in many categories will be less informative than those restricted to fewer categories, in which the presence of the cue becomes a strong indication that the token belongs to one of the categories in question. Holt and Lotto (2006) exemplify this aspect with the Voice Onset Time (VOT) pattern in English, used as a main cue in the distinction between voiced stops / b, d, g / and voiceless stops / p, t, k /. As VOT changes its length in view of place of articulation, VOT is the main cue in the categorisation of these sounds, with closure voicing (Negative VOT) being less preponderant in English. Also relating to VOT, we can mention that native speakers of English use Positive VOT as a highly informative cue to identify voiceless plosives in onset position. In Brazilian Portuguese, on the other hand, the Negative VOT pattern is informative in the identification of voiced plosives in the same context. Therefore, voicing takes precedence in that sense as a more informative cue. Therefore, Positive VOT is not strongly weighted by Brazilian listeners in native BP. As a consequence, it tends not to be heavily weighted when perceiving L2 English speech either. Brazilian learners of English, therefore, have to learn how to focus on Positive VOT in order to discriminate between word-initial voiceless and voiced stops in English (ALVES; LUCHINI, 2020; ALVES; ZIMMER, 2015; SCHWARTZHAUPT; ALVES; FONTES, 2015; ALVES; MOTTA, 2014).

Like Flege (1995), but focusing specifically on what concerns the effects of the different weights assumed by acoustic cues in speech perception, Holt and Lotto (2006) understand the prioritisation of certain cues in relation to others as a possible origin of mismatches between the perception of native listeners and the perception of non-native listeners. As a result of greater previous experience with the L1 system, L2 learners might apply their L1 cue weighting pattern to perceive L2 tokens (what we have been calling the 'L1 filter'). The use of the L1 cue weighting system may not, however, be effective when listening to L2 sounds. In other words, though the acoustic cues may be present in the speech signal, that does not entail that the listener is attuning to them. In not attending to a characteristic that should work as a decisive source of information in L2 processing, it is possible that the non-native listener will not identify the token as the speaker intended it. Accordingly, when the learner assumes the role of speaker in the interlocution, they may not produce this ignored (or less weighted) cue. This mismatch in cue weighting, in turn, can result in the opposite situation. That is, a listener seeks some information that is decisive, but does not find it, and thus fails to identify the target that the non-native speaker intended to produce. Nevertheless, these cues can come to be perceived and weighted by the learner as informative cues, regardless of age of onset of learning (FLEGE; BOHN, 2021) – depending on the quantity and quality of input this learner is exposed to.

Considering this scenario, this thesis maps the acoustic characteristics of the vowels [æ, ɛ, i, ɪ] in the L2 English tokens produced by L1 Brazilian Portuguese learners. In BP, the vowels [ɛ, i] may occur in stressed position, just as they do in English. The traditional notation labels both sounds as [ɛ, i] in both BP and in English, that is, as if they were produced with the same acoustic characteristics in both languages. However, the phonetic-phonological status of each sound can be different in each language system. Moreover, previous research has shown that these 'same sounds' can be produced in different areas of the common phonetic space (see section 2.4), as well as with distinct acoustic characteristics. Those differences in production, in turn, can have effects on perception. Furthermore, the SLM highlights that vowels will maintain a certain acoustic distance from one another in the common phonetic space in order to maintain distinctiveness among categories.

Thus, considering that BP has seven vowels in stressed position, whereas English has eleven (Yavas, 2011)[15], it is possible to predict that the acoustic boundaries of those vowels will be different in each language system. Pereyron (2017) found that the vowel spaces of L1 Brazilian Portuguese are larger for each sound category than they are in L1 English, given that in the BP system the phonetic space needs to be divided into a smaller number of categories. As the Cue Weighting theory puts it, the reallocation of categories in the acoustic space would also depend on the prioritisation of certain acoustic cues (such as F1 or duration) over others. Holt and Lotto (2006) indicate, for example, that in the pair [i-ɪ], native English listeners prioritise spectral cues (characteristics of F1 and F2) over temporal cues (vowel length). When the spectral cues are ambiguous, however, vowel duration is used to distinguish between the tense or lax members of the pair (ESCUDERO; POLKA, 2003). Considering this scenario, for a sound to be perceived as [i] or [ɪ] by a native English listener, then, it appears that the stimulus should bring both dimensions (spectral in F1/F2 and duration). That is, in case one dimension is not informative enough, the other can render the decisive information needed for accurate

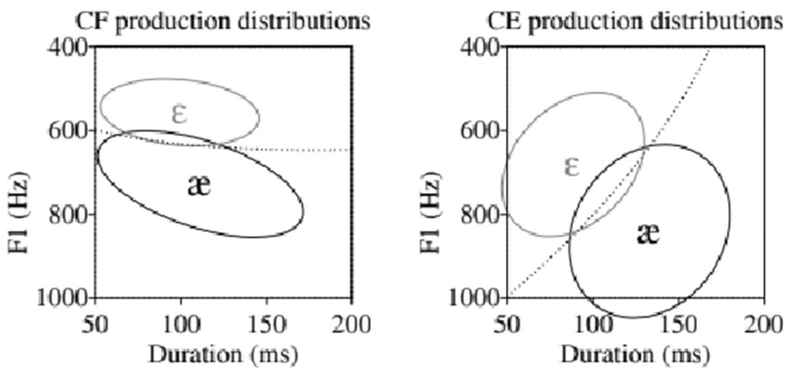
[15] Yavas (2011) lists 11 vowels, as the author takes [e, o] into account. Ladefoged (2010), however, treats [eɪ] and [oʊ] as diphthongs.

perception. However, a native speaker of an L1 (which prioritises one of the cues, but not the other) might produce the vowels in question in L2 English without prioritizing these cues. This, in turn, could make it difficult for native listeners to correctly categorise the tokens – ultimately, hindering speech intelligibility.

In view of this, it is important that we connect our theoretical background to the study conducted[16] by Escudero (2009), and which has inspired the present thesis. Escudero investigated the main cue prioritised by two groups with different L1 backgrounds. She analysed how F1 and temporal cues are weighted by monolingual native speakers of Canadian French (CF) and of Canadian English (CE). Those participants provided the speech samples in each of their native languages, and those productions served as stimuli for a vowel identification task – in which listeners were also monolingual native speakers of CE and CF, as we will detail further on.

Figure 2.1 shows the mappings of vowel production patterns by CF (left) and CE (right) participants, in relation to duration by F1. As can be seen, spectral information (F1) appears as a primary cue in the CF productions. That is, regardless of a shorter or longer duration, the acoustic boundary between [æ] and [ɛ] is established in relation to the first formant (F1), at a value close to 700Hz (*op. cit.*, p.6).

Figure 2.1 – Plot (F1 x absolute duration) of L1 vowel productions by monolingual native Canadian French (left) and monolingual native Canadian English (right) speakers



Source: Escudero (2009, p.7)

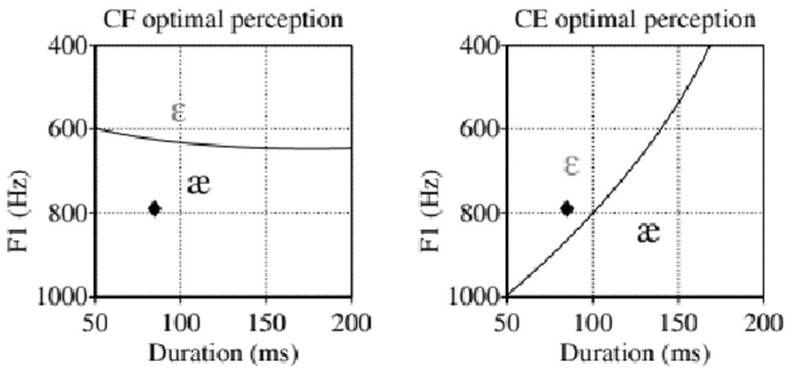
The plot on the right in Figure 2.1, in turn, shows that both duration and F1 are determinant in the acoustic threshold of production by CE participants. For example, a shorter

[16] We highlight that we do not adopt (nor do we concur with) the L2LP perception model, proposed by the author in the study that first inspired the present investigation. To the contrary, we believe that the SLM fits much better with a Complex, Dynamic view of language, which is why we adopt the theoretical frameworks detailed in this chapter.

duration will be coupled with a higher F1 for target vowel [æ], while the same temporal cue will be paired with a low F1 measure for target vowel [ɛ].

In Escudero (2009), the production patterns shown above were mapped in order to evaluate if they would coincide with perception patterns, both in 'native' L1 and in 'accented' L1. We use quotation marks because those are our terms, not Escudero's – she considers what we call 'accented L1' to be 'L2'. In her design, all 120 tokens (60 produced by CE and 60 by CF speakers) were presented to both L1 CF and CE listeners as if all 120 stimuli had been produced by native speakers of the listeners' respective L1s. Those 'accented' tokens produced by the 'non-native' speakers would have acoustic differences in comparison to the 'native' ones, as shown in Figure 2.1. In Escudero's study, this design allowed her to test the workings of the 'L1 filter', operationalised as the way native parameters were used in identifying tokens produced by participants that were native to the other language of the two. Acoustic specifications of L1 categories were taken as the 'L1 filter' expected to be applied in the perception of the 'accented' (L2, in her terms) productions. Figure 2.2 showcases her predictions of how the same token would be "optimally" perceived in each L1 – the same token, as shown in the figure, would be identified as a different vowel by each L1 group of listeners.

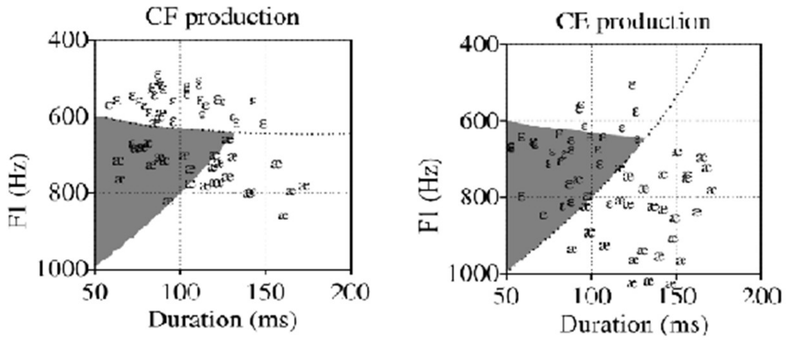
Figure 2.2 – Plot (F1 x absolute duration) of L1 vowel optimal perception by monolingual native Canadian French (left) and monolingual native Canadian English (right) speakers



Source: Escudero (2009, p.8)

Figure 2.2 showcases the expected match between production and perception boundaries for both groups in their own L1. Hence, if L2 speech is indeed perceived through an 'L1 filter', it would be expected that some 'accented' targets would yield lower accurate identifications. That is, in the same acoustic area of the plot where different vowels are categorised in CF and in CE, some 'confusion' was expected. Figure 2.3 plots the experiment results, in which native speakers of CE identified CE and CF tokens altogether, and native speakers of CF also identified CF and CE tokens altogether.

Figure 2.3 – Plot (F1 x absolute duration) of L2 vowel perception by monolingual native Canadian French speakers (left) listening to English vowel productions and by monolingual native Canadian English speakers (right) listening to French vowel productions



Source: Escudero (2009, p.20)

In Figure 2.3, the threshold (horizontal and diagonal lines) confirm that the 'L1 filter' is hindering the intelligibility of the 'accented' productions when these productions have temporal and spectral characteristics that do not match the listeners' L1 categories in terms of the acoustic cues under analysis. This holds true for both groups of listeners in Escudero's study.

Though we were initially inspired by Escudero's experiment, we reiterate that her study was performed by monolingual speakers. We also wanted to investigate the 'L1 filter', but in terms of L2 intelligibility in the interactions involving bi/plurilingual participants. The Complex, Dynamic view of language we adopt leads us to believe that if Escudero's participants were bilingual, those results could be different. We would expect, on the one hand, that bilingual participants would be somewhat attuned to the L2 system[17]. Moreover, we would also assume that those participants' L1 boundaries could have undergone some changes as an effect of L2 development (cf. PEREYRON, 2017). Additionally, the listeners' L2 level of proficiency could have had a further impact on those results, had they not been monolingual participants.

Despite these differences, we understand that Escudero's (2009) results do shed light on how L1 and L2 category mappings interact. As the SLM posits, L1, L2 and L1-L2 composite categories occupy the same common phonetic space, which means they may affect one another. The Complex, Dynamic framework, accordingly, predicts that a bilingual's language systems

[17] Hence our previous choice for the 'accented L1' denomination in describing Escudero's 'L2' tokens.

will interact and may impact one another. Thus, when looking at our dataset, we would expect that the native speakers of Riverplate Spanish in our study would have their acoustic thresholds in L2 English established (to some degree) in view of the decisive cues used in their native language. Congruently, the same would hold true for the native speakers of Central German. Moreover, as already mentioned concerning Brazilian speakers, both groups of listeners presumably would have gone through various stages of L2 development. As our listener participants are all at an advanced proficiency level, we would expect that their learning trajectories were somewhat shaped by their experiences with L2 English.

We once again highlight that the present study is cross-sectional, that is, though we assume a learning trajectory, we do not focus on that. Besides, we did not collect speech data from our listener participants. Thus, we will base our comparisons of native systems ('L1 filter') on data compiled by previous studies (see section 2.4). We consider the 'current' state of systems at the time of data collection. We do so because we want to focus on how those participants' systems interact, as we do not expect their native (nor their L2) system to be solely accountable for the processes we are setting out to investigate. On the contrary, we assume that the 'L1 filter', along with other processes, will lead to emerging outcomes when the hybrid, non-native systems of our speakers and of our listeners 'meet' (encounter) in a communication setting. Therefore, we do not propose to just 'superimpose' one mapping onto the other, but to try and understand the gradience that arises from the interaction of our non-native speaker-listener pairs.

As a goal of L2 learning, it is assumed that any individual learner of any group of individuals is looking to have a successful interaction in the L2 they are learning. This, in turn, involves the ability to communicate in an intelligible way. Hence, what ultimately motivates our study is the wish to understand the extent to which the acoustic characteristics of oral productions by Brazilian learners at different proficiency levels of English can lead to a greater or lesser success rate in communicating with other learners with different L1 backgrounds. Besides, we wish to observe how those characteristics relate to other participant-related factors that might also play a role in the intelligibility of the L2 English spoken by Brazilian learners.

Again, we highlight that, in this thesis, we are particularly interested in the bidirectional interactions of this interlocution relationship. On the one hand, considering that BP has a distribution in which vowel duration does not seem to play a functional/phonological role, it could be expected that this cue will not be produced as sharply. Previous studies, however, have shown that the temporal dimension is produced in a more native-like fashion than the F1 and F2 cues (PEREYRON, 2017), as far as [æ, ɛ, i, ɪ] are concerned. Thus, the predictions of both the SLM and the Cue Weighting models would be that L1 English listeners could have less accuracy in the perception of each vowel of the pair [i-ɪ] produced by L1 BP speakers, for example, since the spectral information is a more informative cue for these native English speakers. We reiterate, though, that the spectral distinction may not be the sole fundamental cue, as all cues play a role in perception. As we have mentioned, if F1 and/or F2 are not decisive, L1 English listeners make use of the temporal cue. The question that emerges concerns what the perception of a native listener of another language (other than BP or English) will be like. The theoretical scope of this study predicts that, in part, this will depend on the weighting of different acoustic cues in the listener's L1. Additionally, we ought to remember that our learners' L1-L2 composite categories could also be developed in a manner in which the 'L1 filter' does not strongly affect perception (any more), especially among advanced learners of English. Those are but some of the scenarios that led us to conduct the present study, as we will describe in sections 3.1 and 3.2.

Before we detail our goals, we must first discuss the intelligibility construct, which is one of the main aspects discussed in our investigation. Now that we have explained our view of language in general and the framework we adopt to investigate speech perception, we are ready to discuss what we understand as 'intelligibility' and how we will operationalise it in our experiment.

2.3 INTELLIGIBILITY

Between the 1950s and the 1980s, the teaching of L2 pronunciation, to a large extent, aimed to make students speak like native speakers, that is, without a foreign accent. From the 1980s on, however, nativeness was no longer the focus of teaching, and success in communication began to occupy the spotlight (DERWING; MUNRO, 2015). In this paradigm shift, nowadays an accented speech (with a foreign accent) can even be regarded as the result of a learner's choice, consisting of an identity element, especially in immigration contexts. As successful communication becomes the goal of teaching, teachers and researchers need to understand what can lead to the success/failure of an interaction in a foreign language.

Intelligibility is defined by Munro and Derwing (1995) as "how much a speaker's message is understood by a listener" (*op. cit.*, p. 289). This construct of intelligibility – other authors define it in different ways (cf. CRUZ, 2007) – is one of the three characterisations that the authors present in an attempt to investigate the aspects involved in the success of communication. In addition to intelligibility, the authors pose two other relevant dimensions.

'Comprehensibility' is defined as a level of ease/difficulty in understanding what is being heard. 'Accentedness' is a measure of the perception of accented speech by the listener. These three dimensions are "partially independent". This means, for example, that a speaker can be highly intelligible, despite having a strong accent. The measurements in each dimension are gradual, since the three aspects are "continuous phenomena" (MUNRO; DERWING, 2015, p. 379). Additionally, to some degree all three depend on the acoustic characteristics of speech production, which justifies the adoption of the framework in the present study.

The intelligibility construct also poses the challenge of integrating linguistic aspects – such as acoustic characteristics of production, phonotactic aspects, prosodic issues, among others – and 'extra-linguistic'[18] ones – linguistic experience, context and registry, etc. In investigating those factors, it is important to bear in mind that they are also dependent on the interacting agents. In that sense, Munro and Derwing (2015) point out that both speaker and listener have an active participation in a dialogue and that a listener can, in general, adopt one of two positions: wanting to understand communication, or refusing to do so. One of the factors that can lead to one or the other attitude is the speaker and their way of speaking. Therefore, on the one hand, the speakers and the characteristics of their speech continuum in a given language must be taken into account, as well as additional factors that have an effect on these characteristics. On the other hand, it is necessary to know the listeners and how their language experience can affect the way in which they understand the sounds of non-native speech – or of a dialect distinct from theirs –, as well as their potential attitude towards the speakers. Thus, understanding what factors affect speech intelligibility encompasses an effort to analyse both the speaker and the listener.

A discussion on the intelligibility construct, as well on the combined role of both speaker and listener in the interaction within the Complex, Dynamic framework, can be found in Albuquerque (2019). The author defines intelligibility as

> a process that entails the perception of linguistic data imbricated with gradient comprehension, whose stages go from recognition/attunement, recovery of lexical and phonic processes,to semantic association and linguistic-cognitive stabilization (not necessarily following a linear order across such gradience.[19] (*op. cit*., p. 121)

[18] Though the authors use this term, we understand that within CDST all systems are interconnected. This way, embodied experiences are not segmented into linguistic or not, seeing that all experiences might lead to changes in the whole system.

[19] "[Inteligibilidade] é um processo que implica a percepção dos dados linguísticos de modo imbricado a um gradiente de compreensão, passando por estágios de reconhecimento/sintonização, recuperação e processamento lexical e fônico até a associação semântica e a acomodação linguístico-cognitiva (não necessariamente seguindo uma ordem linear ao longo do gradiente)."

Albuquerque and Alves (2020) highlight that both the speaker and the listener are a part of a 'comprehension gradient', and that "each speaker-listener pair characterises a locus of analysis" (*op. cit*., p. 216) within a Complex, Dynamic framework.

In their traditional account, Munro and Derwing's intelligibility construct (1995) was further subdivided into two types (MUNRO AND DERWING, 2015): global and local intelligibility. 'Global intelligibility' "entails larger units of language that include rich contextual information" (*op. cit*., p.381). The term "larger" is being compared to (the 'smaller' units of) segments and words, which are the scope of 'local intelligibility'. The acoustic cues investigated in the experiment analysed in this thesis are taken to characterise this 'local intelligibility', namely, the identification of speech sounds in "relatively small units of speech, such as segments and words" (2015, p. 381), that is, isolated from larger contexts. As the authors pose,

> "research on local intelligibility is more useful to our understanding of L2 learning processes and to identifying some of the underlying components of global intelligibility. For instance, a local study might help us determine several speaker errors that lead to problems for the listener; however, only some of those may cause difficulties when contextual information is present. " (*op. cit*., p. 381)

Each level of analysis will have different implications for experiments investigating intelligibility, as well as for the application of the findings in further studies and in teaching. Moreover, the authors point out that the tasks used to operationalise the construct of intelligibility lack greater precision. This perhaps explains the fact that studies on the subject report different or not comparable results (cf. MUNRO; DERWING, 2015). Studies focusing on more specific aspects of intelligibility – at a global or at a local level – might help in that aspect, they suggest.

To try to gather more precise results, therefore, the present study opts for an investigation on local intelligibility. Furthermore, as Munro and Derwing's construct entailed segments and words, we narrowed our study to look at the word level. As we will detail in section 3.4.4.3, our experiment was designed as an eight-option[20] forced-choice word identification task. Because we wanted to investigate the local intelligibility at word level, the task provided the target words as the eight options. We understand that providing just the nuclei target vowel would be a more appropriate design if we were looking at the segment level, which

[20] As we will detail in section 3.4.4, listener participants were given eight words to choose from, which included the four target words with [æ] – [ɛ], as well as the four lexical items with [i] – [ɪ]. In chapter 4, we will detail that, out of 5,888 tokens, 62 (1,05%) consisted of identifications outside of the intended minimal pairs (such as a target word 'sit' identified as 'set', or a target word 'set' identified as 'seat').

we are not. Thus, in our experiment, we operationalised 'accurate identification' of a word as the correct identification of all three segments in our CVC monosyllables. This way, a 'sat' token was considered accurately identified if it was identified as 'sat', but inaccurately identified if it was identified as 'pat' (correct vowel) or as 'set' (correct onset consonant).

Moreover, as we chose target words with plosive codas – which might be unreleased in production –, we chose the forced-choice task over a cloze test to avoid data loss in transcriptions that did not match any of our vowel contexts. For instance, a cloze task could yield a 'see' transcription for the 'seat' token, which would render this response useless for the purposes of our accurate/inaccurate ratings. Moreover, a cloze task would recruit a productive skill (writing). Additionally, it could lead to a 'cross wiring' of subsystems, so to speak, mixing phonological aspects with those relating to orthography, for example. We do conceive that the predefined labels (choices) we provided still allow for other subsystems to interact during the identification process. In other words, as our listeners must select the word they hear, they can still use orthography as an input (passively, in reading), in addition to the acoustic cues under analysis. This, in turn, might have effects on our results (see section 5.2). On the other hand, as Munro and Derwing (2015) highlight, our methodological choice was made based on what we thought best fit our goals, as any method would have some downside. Besides, we highlight that within a Complex, Dynamic framework, different psycholinguistic aspects are assumed to interact in language processing – though we will not discuss this further, as it is not our present goal.

We highlight that our focus on accurate word identifications *per se* was explored as one of our research goals. Our study also investigates the identification of vowel segments, albeit not operationalised as 'accurate' or 'inaccurate'. Word identification accuracy measures are complemented with mappings of the acoustic (spectral and temporal) vowel characteristics, taken as the cues to be weighted by listeners (in regards to those segments) when deciding on their responses in the task. We understand that this design is in line with our combined Laboratory Phonology and Applied Linguistics approach. That is, we expect that our results can contribute to future research on the field of Laboratory Phonology, but also that it could prove useful for language teachers in their pronunciation activities. Again, we refer to chapter 3 for further discussion.

Finally, we reiterate that intelligibility levels may also dynamically affect one another. A segment might not be intelligible by itself, but within a word it can come to be understood. Likewise, a word that was not intelligible by itself can have its meaning grasped within a sentence or utterance context (cf. CRUZ, 2017). We highlight this aspect to once again justify our choice for the complementary analysis (in our third Research Question) that we will discuss in chapter 4. The methodological choices we made aimed to provide us with variables that could be analysed in a somewhat controlled way. This, in turn, would allow us to investigate different aspects of local intelligibility in simulating an interaction between non-native speakers and listeners of English. As we have mentioned, it is in this interaction that the present thesis focuses on, hence our choice for a cross-sectional design (that is, a design which investigates the state of systems at the moment of the interaction – in our laboratory conduction, the moment of data collection).

We will now describe the vowel systems that are at play in our experiment. As we have mentioned, those systems will be relevant when we discuss the significant predictor variables in our inferential results, as well as the patterns we will look at in our exploratory analysis (see chapter 4).

2.4 VOWEL SYSTEMS

As we have mentioned, we will focus on local intelligibility at the word level. Nevertheless, as our CVC monosyllable target words have a nuclear vowel that is distinctive within the minimal pairs we investigate, we will now describe the vowel systems at play in our experiment. This description is particularly relevant when we consider RQ2 and RQ3, which focus on general vowel identification, that is, regardless of accuracy in relation to the target vowel.

In acoustic terms, vowel production is characterised by the presence of fundamental frequency (f0). The acoustic correlate of vowel height is the first formant (F1), whose measure in Hertz yields values that are inversely proportional to the articulatory position. Thus, the F1 values of [i] should show low frequencies, since the vowel is high. The front-back acoustic correlate is the second formant (F2), which is proportional to the frontness of a vowel (LADEFOGED, 2010). Thus, [i] has a high F2 value, because it is the most fronted vowel[21].

We turn now to a brief acoustic description of the vowels in the four languages involved in this project, namely: Brazilian Portuguese (Porto Alegre variety), Argentinian Spanish (Riverplate variety), German (Central variety) and English (non-native/L2 for all our groups of participants).

[21] For a detailed account of F1, F2 and fundamental frequency, see Silva *et al*. (2019).

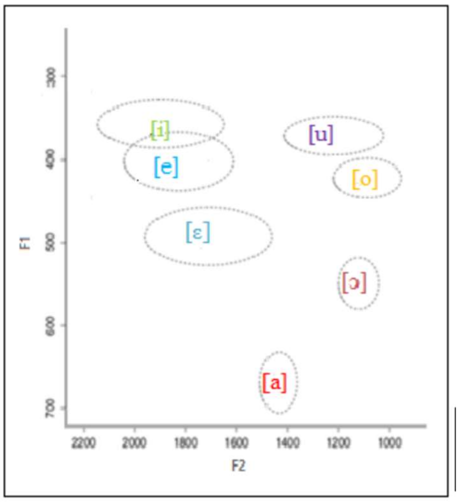
2.4.1 The vowels in Porto-Alegrense Brazilian Portuguese

In this section, we will describe the formant values in the Porto-Alegrense variety of Brazilian Portuguese (BP), which is the native language of the participants in the production task of this study.

Of the four vowels we analyse, only [i] and [ɛ] occur in a tonic position in BP, in words like "piso" (floor) and "ferro" (iron), respectively. We highlight, however, that "the use of the same symbol for a sound in two or more varieties does not mean the sound is identical in different varieties" (Yavas, 2011, p.89). That is, the acoustic characteristics of these vowels in BP are not the same as they are in other varieties/languages, as we will show in this section. The lax vowel [ɪ] of words like 'sit' [sɪt] in English does not occur in stressed positions in the native BP vowel inventory. Likewise, the [æ] of words like "pat" in English, produced as [pæt] in the native pattern, is not found in BP.

Figure 2.4 shows the acoustic space of stressed vowels in the productions by speakers of the Porto-Alegrense variety of Brazilian Portuguese, as reported by Pereyron (2007).

Figure 2.4 – Acoustic space of stressed vowels in the productions by speakers of the Porto-Alegrense variety of Brazilian Portuguese, as reported by Pereyron (2007)



Source: Pereyron (2017, p.103).

Pereyron (2017) analysed the speech of four female and one male monolingual native speakers of the Porto-Alegrense variety of Brazilian Portuguese[22]. The speakers' productions

[22] We refer to Pinto (2017) for another account of the Porto-Alegrense variety of Brazilian Portuguese, elicited by a reading-task, though his work does not investigate vowel duration. Additionally, Callou, Moraes and Leite (1996) also provide a description of spectral acoustic characteristics of spontaneous speech, though again not of vowel duration.

in their native language were elicited by a list-reading task, and acoustic formant measures were normalised by the Lobanov Method[23].

The acoustic characteristics of the vowels produced by monolinguals from Porto Alegre and its metropolitan area described in Pereyron (2017) account for an average of 361.3Hz for the first formant (F1) of [i], with a standard deviation (SD) of 19.08Hz. This is the vowel at the high front corner of the phonetic space, with an average F2 value of 1,936.5Hz (SD = 119.5). The reported relative durations are on average 13.64ms for disyllables (SD = 2.89) and 12.36ms (SD = 1.91) for trisyllables.

For [ɛ], F1 has an average of 492.5Hz (SD = 26.06), while F2 has an average of 1,776.5Hz (SD = 161.5). The average relative durations of [ɛ] were 17.23ms for disyllables and 17.45ms for trisyllables, with SDs of 2.15ms and 2.08ms, respectively (PEREYRON, 2017).

2.4.2 The vowels in Argentinian Riverplate Spanish

From the same study by Pereyron (2017), we also find acoustic descriptions of the stressed vowels of Argentinian Riverplate Spanish[24] – which is the variety of L1 of one of the groups of listeners that participated in the perception task. Of the four English vowels under analysis in the present study, Spanish has only one in tonic position, out of a total of five tonic vowels. The vowel [i] occurs in words like “sito” (“localised”). Again, we highlight that the existence of a high front tense vowel in the native Riverplate Spanish inventory does not mean that it is produced with the same acoustic characteristics that it shows in productions in other languages, like BP or English, for example. The [ɛ][25] and [æ] of words like “pet” and “pat” in English, produced in the native pattern as [pɛt] and [pæt], respectively, are not found in the phonological inventory of Riverplate Spanish.

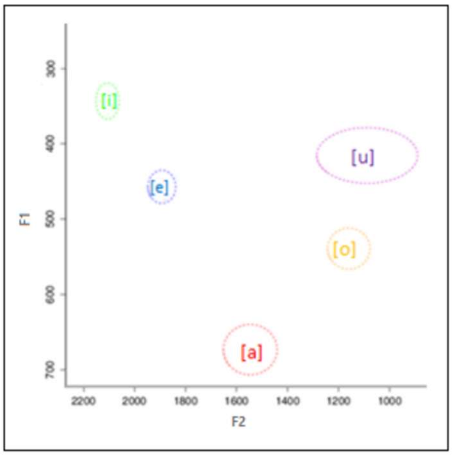
Figure 2.5 shows the acoustic space of stressed vowels in the productions by speakers of the Argentinian Riverplate variety of Spanish, as reported by Pereyron (2007).

[23] As we will detail in chapter 3.4.4.1, our study applied the Watt & Fabricius normalisation method.

[24] We also refer to De Los Santos (2017), Santos (2014) and Santos and Rauber (2016) for accounts of the Riverplate variety of Spanish spoken in Uruguay.

[25] The vowel [ɛ] can be found in stressed position before some segments, like /r/, as in that position it is an allophone of /e/. (BRISOLARA; SEMINO, 2014 apud DE LOS SANTOS, 2017. p.47)

Figure 2.5 – Acoustic space of stressed vowels in the productions by speakers of the Argentinian Riverplate variety of Spanish, as reported by Pereyron (2007)



Source: Pereyron (2017, p.107).

Pereyron (2017) analysed the speech of three female and two male monolingual native speakers of the Riverplate variety of Spanish. The speakers' productions in their native language were elicited by a list-reading task, and acoustic formant measures were normalised by the Lobanov Method.

The acoustic characteristics of the variety of Spanish spoken in the Buenos Aires province found by Pereyron (2017) account for an average of 343.3Hz for the F1 of [i], with a standard deviation (SD) of 14.81Hz. This vowel shows an F2 average of 2,105.7Hz (SD = 24.05). The reported relative durations show averages of 8.69ms for disyllables (SD = 1.16) and 7.80ms (SD = 0.38) for trisyllables.

As Pereyron (2017) has observed that L1 Spanish learners of L2 English tend to assimilate English [æ] as Spanish [a], we report here the average values found by Pereyron (*op. cit.*) for the low central vowel in the Riverplate variety. The average F1 is 673.15 (SD = 11.35), while the average measured F2 is 1,528.1Hz (SD = 30.11). The average relative duration is 11.06ms for disyllables and 9.12ms for trisyllables, with DPs of 0.74ms and 0.63ms, respectively.

2.4.3 The vowels in Central German

Before we detail the acoustic aspects of the German vowels, we ought to make a slight digression. As summarised by Stoeckle (2009), what is commonly referred as 'German' or 'High German' is a standardised language that came into being in the 16th century, arising from the interaction between many dialects in the area. The author details that, during the centuries since, the use of both dialect and standardised language have led to different conformations of this relation in different regions of Germany. We start this section with this highlight because our descriptions will be impacted by this fact. Firstly, because many of the references we were able to find describe the standardised variety of German, sometimes pointing out that the productions had ‘no dialect colouring’. Moreover, among the few descriptions that do mention a given variety, we could not find one describing precisely the Central German variety of the participants in our study. We chose Central German, on the one hand, due to the greater number of participants we were able to recruit in this study (see section 3.3.3). Additionally, according to Stoeckle,

> [g]enerally it can be said that in all parts of central Germany there is a tendency to lose base dialects, and most speakers are able to use a variety of the standard language in any situation (...). Nevertheless, there are regional differences concerning the variability of language use[.] (*op. cit.*, p.6)

The author goes on to describe that in the northern and central parts of central German, dialects have virtually disappeared (cf. STOECKLE, 2009 for details). “The only parts of central Germany where dialects are still alive and in use are the Moselle-[26] and Rheno-Franconian[27] dialect areas” (*op. cit.*, p.7).

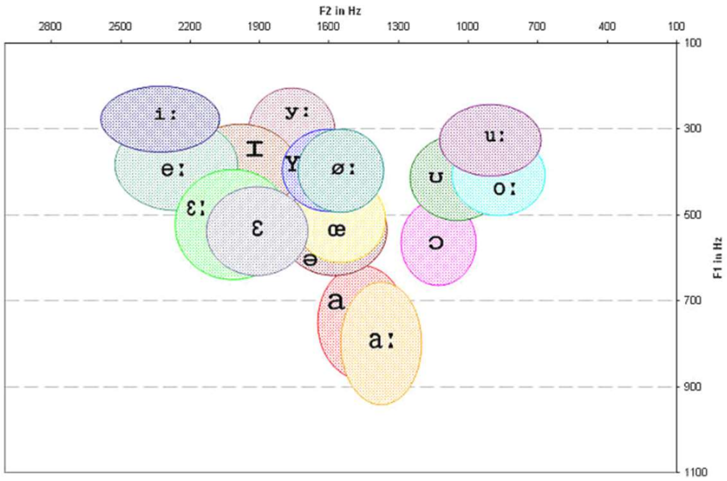
Having pointed out the characteristics of the German language and its varieties, we will now move on to the acoustic descriptions we will follow in our analysis (see chapter 4). We will provide descriptions that relate the closest to the Central German variety.

Like BP and Spanish, the native German vowel inventory has the vowel [i]. Once again, though, we highlight that the acoustic characteristics of the German productions of [i] may not be the same as those in BP, Spanish or English. This, we reiterate, holds true for any speech sound, like the [ɛ] vowel – absent only in the Spanish phonological inventory –, as well as for [ɪ] – which is only present in German and in English inventories. Like BP and Spanish, German does not have [æ] as a native vowel. Moreover, we highlight that extrinsic vowel duration is a distinctive feature in German.

Figure 2.6 shows the acoustic space of stressed vowels in the productions by speakers of the standard variety of German, as reported by Sendlmeier and Seebode (2010).

[26] None of our 18 German participants are from that area.
[27] Two out of our 18 German participants are from that area. They are participants DE21 and DE25.

Figure 2.6 – Acoustic space of stressed vowels in the productions by speakers of the standard variety of German, as reported by Sendlmeier and Seebode (2010)



Source: Sendlmeier and Seebode (2010, p.1).

Sendlmeier and Seebode (2010) analysed the speech of 58 female and 69 male[28] native speakers of German. The speakers' productions were in 'standard' German and were elicited by a list-reading task.

Sendlmeier and Seebode (2010) report data from standard German "without dialectal colouring"[29] (*op. cit*., p.1). They report averaged results, as well as values segmented by male and female speakers. For [i], the F1 averages are 263Hz for male voice and 302Hz for female, with the average F2 at 2,171Hz and 2,533Hz, in the same order. For [ɪ], the mean F1 values are 369Hz and 433Hz, respectively for men and women, while the recorded F2 averages are 1,902Hz and 2,095Hz.

The same authors note, for [ɛ], an average F1 of 489Hz for men and 608Hz for women, with the average F2 of the same groups at 1,817Hz and 2,040Hz.

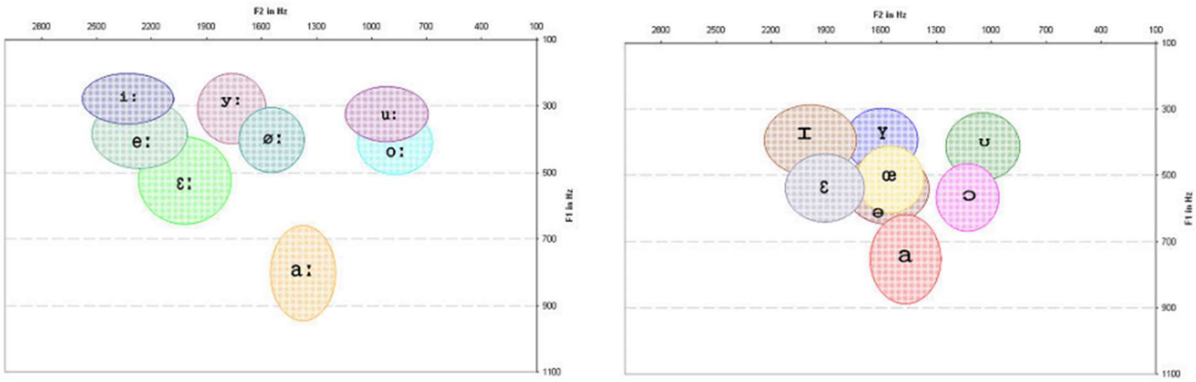
In terms of vowel length, though Sendlmeier and Seebode (2010) do not report duration measures, they do showcase different F1 by F2 acoustic spaces distinguishing long and short vowels. Short vowels are more centralised than long vowels. We reproduce their findings in Figure 2.7, in which long vowels are indicated by the ':' sign on the left plot. Figure 2.7 shows the acoustic space of stressed long (left) and short (right) vowels in the productions by speakers of the standard variety of German, as reported by Sendlmeier and Seebode (2010).

---

[28] Though the authors mention that male and female speakers might produce very different formant frequency values due to their physiological differences. Indeed, they find that female participants tend to produce higher frequencies than male participants. However, the authors do not normalise their values, but rather use production means by speaker and by genders.

[29] "Die Sprecher realisierten die deutsche Standardlautung ohne dialektale Einfärbungen."

Figure 2.7 – Acoustic space of stressed long (left) and short (right) vowels in the productions by speakers of the standard variety of German, as reported by Sendlmeier and Seebode (2010)



Source: Sendlmeier and Seebode (2010).

As Sendlmeier and Seebode (2010) do not report the duration of the vowels, we bring those values as measured by Maack (1949)[30]. This author analysed the speech of three male monolingual native speakers of German, two of the Silesian variety and a third of the Bavarian variety. The speakers' productions in their native language were elicited by text-reading tasks ("narrative" and "conversational"). Maack (*op. cit.*) reports vowel durations in four groups: stressed and unstressed vowels, each group subdivided in long and short vowels. We will recall only the values for the stressed segments, as those are the ones under analysis in our study.

For [i], Maack (1949) reports a short duration of 63ms and a long duration of 111.67ms (no SD is reported[31]). For [ɛ], his averaged values are 82.78ms and 122.50ms, for short and long vowels, respectively. Maack does not report [ɪ] duration values. Thus, we provide the vowel length of the lip-rounded counterpart [ʏ] as a basis for comparison, though we are aware that lip rounding affects vowel duration. For [ʏ], the author reports an average duration of 73.33ms for the short vowel, and 120ms for the long one.

2.4.4 The vowels in English

Yavas (2011) describes 11 monophthongal vowels in English. All four sounds under analysis in the present thesis, [æ, ɛ, i, ɪ], are classified as front vowels in relation to tongue frontness. As for tongue height, [i, ɪ] are classified as high vowels, [ɛ] as a mid vowel, and [æ]

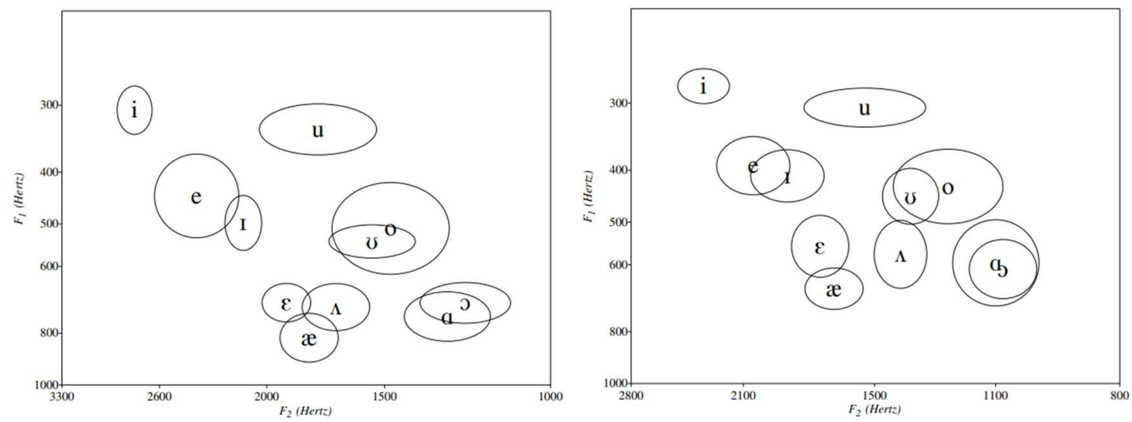
[30] Maack remarks that the samples used in his measures were obtained in different contexts. Two were elicited by two different texts participants had to read out loud, and the third one was a "conversation text" ("Gesprächstext"). Moreover, the linguistic background of the three participants was different. Two of them showed a dialect colouring in their production "only slightly" ("nur leicht"), whereas in the third participants' productions this colour was "considerably stronger" ("erheblich stärker"). We report to Maack (1949, p.192-4) for further details.

[31] Though Maack does not report standard deviations, he does provide the arithmetics for the averaged durations, which includes a comparison only within the vowels in each eliciting task (cf. MAACK, 1949, p. 196-7).

as a low vowel. Regarding tenseness, in the phonetic definition[32], [i] is the only tense vowel in the scope of the present work, whereas [ɪ], as well as [æ, ɛ], are taken as lax vowels. Given that [æ] and [ɛ] are classified in the same manner in regards to tenseness, we will not use this dimension to describe the distinctiveness between the two sounds. We will use it, however, when we mention [i] and [ɪ].

Figure 2.8 shows the acoustic space of stressed vowels in the productions by male (left) and female (right) speakers of the Californian variety of North American English[33], as reported by Rauber (2006).

Figure 2.8 – Acoustic space of stressed vowels in the productions by male (left) and female (right) speakers of the Californian variety of North American English, as reported by Rauber (2006)



Source: Rauber (2006, p.105)

Rauber (2006) analysed the speech of four female and five male monolingual native speakers of the Californian variety of North-American English. The speakers' productions in their native language were elicited by a reading task encompassing both words in isolation and

[32] The author highlights that 'tense' and 'lax' are labels that may be applied in a phonetic or phonological definition. He summarises the discussion: "In some manuals, the tense–lax distinction is present to account for two vowels that are otherwise described identically. For example, vowels /i/ and /ɪ/ (...) will both be described as "high, front, unrounded" vowels; (...). To solve these problems, tense and lax are introduced; the first member in each of these pairs of vowels is called 'tense', because (a) it has a higher tongue position, (b) it has greater duration than its 'lax' counterpart, and (c) it requires a greater muscular effort in production (hence the term 'tense') than the lax vowel. This phonetic definition, however, is not universally adopted. Rather, one finds a phonologically defined 'tense–lax' separation more popular in the literature. This distributionally-based classification is more useful, because it divides the vowels into two groups that are distinguished by the environments in which they occur. Also, (...) this division will play an important role in the stress rules of English" (*op. cit*., p.79). As mentioned, we will only use this description to discriminate between the members of the minimal pair of high front vowels, in which both classifications are congruent. For a discussion on the two dimensions, cf. Yavas, 2011.

[33] As the present study did not focus on analysing production, but perception, we opted to reference standard production frequencies and durations in works that have more thoroughly analysed acoustic aspects of productions. Those studies also have a larger number of participants, as the present thesis only collected samples from two native speakers in order to have a baseline when analysing perception patterns. We will further discuss those productions in section 4.1.1.1.

within a carrier sentence. Acoustic formant measures were normalised by means of a script which accounts minimum and maximum formant values for each gender (cf. RAUBER, *op. cit.*, p.117-9).

Rauber (2006), using data from the Californian[34] variety of North American English[35], describes the F1, F2 and duration values of the tonic vowels of the language in this community, also considering studies with women and men speakers separately. For [i], the author's F1 measures show an average of 308Hz (SD = 35) for female productions, and of 280Hz (SD = 22) for the male ones. The F2 measures are averaged, respectively, around 2,766Hz (SD = 117) and 2,331Hz (SD = 152). The [ɪ] productions show F1 values averaged around 412Hz (SD = 43) for female productions and 501Hz (SD = 55) for male productions. Respectively, F2 values are averaged around 2,121Hz (SD = 95) and 1,884Hz (SD = 172).

Female participants in Rauber's study produce [ɛ] with an average F1 of 704Hz (SD = 58), while male participants show an average of 559Hz (SD = 69). The F2 averages for these same speakers, respectively, are 1,910Hz (SD = 113) and 1,729Hz (SD = 124). For the [æ] vowel, female productions have an average F1 measure of 820Hz (SD = 89), while male participants show an average of 668Hz (SD = 59) for male productions. F2 values are averaged around 1,808Hz (SD = 128) and 1,669Hz (SD = 123), respectively.

As for vowel length[36], the [i] values reported by Rauber (*op. cit.*) are averaged around 140ms and 130ms, for male and female participants, respectively. In the same order, durations of 118ms and 103ms are reported for [ɪ]. Female participants produce an average duration of 116ms for [ɛ], while males show an average of 134ms. The vowel [æ] is produced with 167ms and 179ms on average, respectively.

Finally, we highlight that those measures represent the Californian variety of North American English. Other varieties, as well as non-US Englishes, may present different acoustic characteristics in their productions. Yavas (2011) points out that [æ] and [ɛ] are produced with higher tongue height in British English than they are in North American English. As we will not analyse production or perception by native speakers of English, the measures presented in this section are meant as an outlined reference. We understand that, in a way or another, native

[34] We opted for Rauber's (2006) description precisely because her data focuses on a specific variety, which is one of the criteria we adopted for the data we collected ourselves. Though in Rauber's results having participants of a single variety is considered a downside, in our design the very specificities of the speech community are taken as a relevant factor that we will consider in our analysis.

[35] We also refer to Assmann and Kats (2000) and Peterson and Barney (1952) for accounts of the North-American English vowels in relation to F1 and F2. Lima Junior (2012) describes both frequency and temporal acoustic characteristics of native North-American English productions by participants from different areas of the United States.

[36] Rauber (2006) does not report relative durations, only absolute durations.

varieties are part of the input received by both teachers and learners of an additional language, and should be considered as part of the input used in the distributional process by which the L2 or L1-L2 composite categories are formed.

2.4.5 Contrasts amongst the vowel systems

Having described the vowels we will be analysing in relation to how they are characterised in their native inventories, the purpose of the present section is two-fold. We begin by summarizing the data we presented in a single table, in order to allow for a comparison of native systems. Additionally, this summary is intended to highlight the fact that even if a given vowel is part of the inventories of two or more languages, that does not mean they are produced with the same acoustic characteristics.

Table 2.1 summarises F1, F2 and absolute duration values for [æ, ɛ, i, ɪ] in native productions of Porto-Alegrense Brazilian Portuguese, Riverplate Spanish, Central German and Californian North-American English.

Table 2.1 – F1 (Hz), F2 (Hz) and absolue duration (ms) values for [æ, ɛ, i, ɪ] in native productions of Porto-Alegrense Brazilian Portuguese (BP), Riverplate Spanish (ES), Central German (DE) and Californian North-American English (EN)

L1	**æ**			**ɛ**			**i**			**ɪ**		
	F1	**F2**	**dur**	**F1**	**F2**	**dur**	**F1**	**F2**	**dur**	**F1**	**F2**	**dur**
BP	—	—	—	492	1,776	195	361	1,936	146	—	—	—
ES	—	—	—	—	—	—	343	2,105	70	—	—	—
DE	—	—	—	533[37]	1,978	122[38]	275	2,313	111	396	1,914	– [39]
EN	820	1,808	167	704	1,910	116	308	2,766	130	501	2,121	103

[37] As Sendlmeier and Seedbode (2010) report F1 and F2 values only by gender, we average the results proportionally to the number of participants in order to summarise the data. Detailed descriptions can be found in section 2.4.3.

[38] As Maack (1949) reports absolute duration values of long and short vowels, we chose to display the values of the longer durations in the summary in Table 2.1. That is because, as we will further describe in sections 4.1.1 and 4.2.1, the Porto-Alegrense variety of BP tends to produce long vowel lengths.

[39] Maack does not report duration values for [ɪ]. As we have mentioned, he reports the values of the lip-rounded counterpart [ʏ] as a basis for comparison, though we are aware that lip rounding affects vowel duration. For [ʏ], the author reports an average duration of 120ms for the long one.

Source: elaborated by the author (2021) based on the data from Pereyron (2017); Sendlmeier and Seebode (2010); Maack (1949); Rauber (2006).

Table 2.1 shows, as we have mentioned, that the 'same' categories are just the 'same' in the name, as their acoustic characteristics are not uniform across all languages. For instance, BP [i] is produced with the highest F1, as well as with the lowest F2. That is, Porto-Alengrese BP productions of the 'high front vowel' are lower and less fronted than the 'high front vowel' in the vowel inventories of the other languages described. Additionally, we can see BP productions of [i] are the longest[40] amongst the ones listed. Likewise, the Porto-Alegrense BP [ɛ] is longer than the [ɛ] of both Central German and Californian English.

As we will be analysing the role played by acoustic cues in the perception of the vowels (see chapter 4), we ought to reiterate that each speech community will have their own patterns of productions. This is one of the reasons why we chose participants (speakers and listeners) that were native speakers of the same L1 variety. Moreover, within a Complex, Dynamic framework, we assume that even controlling native variety is a feeble attempt, as the individual experiences of each participant might have shaped their language systems in a different way. Therefore, despite the fact that we classify a speech sound as 'high front', 'tense' and refer to it with the same '[i]' notation, the speech continuum allows for much more gradience than those categorical descriptions allow for. Hence the relevance of our reiteration that any statement regarding a sound being 'found in both languages' must take into account that those category acoustic boundaries are not necessarily the same.

2.5 FINAL CONSIDERATIONS

In this section, we will summarise the theoretical proposals we adopt, as well as how they relate to our goals and research questions, which will be presented in the following chapter.

As we have mentioned in section 2.1, we adopt a Complex, Dynamic view of language, in which language is an ever ongoing process, rather than a product that reaches an 'end-state' at some point (ALVES, 2018; LOWIE, 2017; LOWIE; VERSPOOR, 2015; DE BOT; LOWIE; VERSPOOR, 2007; DE BOT; LOWIE; THORNE, 2013; BECKNER *et al*., 2009). Change happens through time and as a result of the system's adaptation to an individual's embodied experiences. On the one hand, this adaptation means that the systems will show some gradience. On the other hand, it means that new experiences can lead to new system organisations, in a

[40] In our study, we make use of the relative duration values, which allow us to account for individual speech rates. The data in Table 2.1 report absolute durations, as that is what is available in the literature reported in this section.

never-ending changing process. Additionally, as each trajectory is individual, each system will also be unique (in a way), though commonalities can be found across groups of people, when we look at a large-enough time window (in relation to processes that time would have allowed for).

These assumptions impact our study in different ways. First of all, we understand that our speaker participants' systems are changing as a function of their experiences. Thus, our effort to group them into proficiency levels is bound to even out the intragroup differences. We are very much aware of that. Secondly, as ours is a cross-sectional experiment (see chapter 3), we account for this individuality by means of taking participants as random intercept effects, as far as our inferential statistics are calculated, and by looking at specific behaviours in our stimulus-by-stimulus analysis.

Moreover, as we consider that embodied experiences and speech communities are part of what shapes a system, as well as of what leads it to change and adapt, we justify our choice for non-native speakers. We also took that into account when we chose to narrow down the L1 varieties that were taken as inclusion criteria. Finally, our Language Experience Questionnaire (see section 3.4.2) was applied in order to allow for some overview on which experiences might have had an influence on the state of our participants' systems at the time of data collection.

In view of the Complex, Dynamic framework we adopt, we need a perception model that will allow for the analysis of the gradience in the system. We chose the Speech Learning Model (FLEGE; 1995; FLEGE; BOHN, 2021) for that purpose, as this model accounts for such gradient behaviour in speech perception. The SLM predicts that an L2 sound will be perceived as similar or dissimilar to a native sound, leading to category assimilation or dissimilation, respectively. In other words, upon hearing an L2 sound, a learner will perceive as belonging to a category. If there is no category already in place in which that sound belongs to, the learner will either create a new category or assimilate the sound to the closest, already established category. It is important that we highlight that those processes, which we describe in an oversimplified way in this section, will not take place all at once, nor will they be categorical or linear. While learning an L2, a person will take the distributional characteristics of the input they are being exposed to in order to create/polish those categories. Importantly, the SLM assumes that all categories – those of L1, those of L2 and any composite L1-L2 category – will share a common phonetic space.

For the present study, Flege and Bohn's proposal allows us to account for non-native productions and perceptions that may not be native-like, but that showcase the complex relation that arises from the L1 and the L2 systems that interact[41] throughout a learner's language development process. Additionally, as we deal with non-native speakers and listeners, the SLM is a good fit for our analysis, as it puts great stock in the role played by the learner's embodied experience with the L2.

Furthermore, the SLM assumes the positional allophone as its level of analysis. According to the model, the acoustic characteristics of a speech sound will be taken into account when that sound is being judged as similar or dissimilar to other known sounds. The common phonetic space is organised, partially, as a function of those acoustic characteristics. The SLM incorporates Holt and Lotto's (2006) Cue Weighting proposal, which states that acoustic cues will be combined in language-specific ways in order to inform the category to which a speech sounds belongs.

To the present study, the acoustic cues are taken as predictor variables in our statistical analysis (see sections 4.1.2, 4.13, 4.2.2 and 4.2.3). They are also taken as some of the characteristics that might play a role in the way speech is perceived, when we consider that the phonetic-phonological subsystem interacts with other subsystems (morphology, semantics, etc), within a Complex, Dynamic framework. We will explore those roles in our exploratory analysis.

All of the assumptions we have summarised thus far will come together when we consider intelligibility, as defined by Munro and Derwing's (1995) definition as "how much a speaker's message is understood by a listener" (p. 289). As one of three partially independent dimensions (the other two being comprehensibility and accentedness), intelligibility is a construct that attempts to measure the success of communication in the exchange between speaker and listeners. Munro and Dewring (2015) subdivide this 'understanding' of the message in local intelligibility (at segment/word level) and global intelligibility (at sentence/utterance level). The present thesis, as mentioned before, is concerned with the local intelligibility at the word level.

We adopt Albuquerque's (2019) account of intelligibility within a Complex, Dynamic view of language. In her proposal, the speaker-listener pair is taken as the locus of analysis, which allows for other processes to be investigated in their communication.

In the next chapter, we will detail our main goal of investigating the local intelligibility of the L2 English spoken by Brazilian learners, at the word level, when those speakers are listened by other non-native learners of English – specifically, Argentinian and German learners

[41] In this sense, we do not mean the term 'interaction' in a statistical sense. We will pointedly note statistical interactions when we refer to the statistical sense of the word.

of L2 English. We will also detail the research questions we posed to guide our experient design and the analysis of our results.

3 METHOD

Having established our theoretical framework, we will move on to describe our main goal, as well as our specific goals. In this chapter, we will also lay out the research questions (RQs) that guided our experimental design and our discussion of the results. Finally, we will describe the Participants, the Instruments and the Data Analysis procedures adopted in the present study.

Our goals will be specified before our research questions, as the latter directly relate to the former. Participants are described in terms of L1 groups, comprising native and non-native speakers of English in the Production task, as well as non-native listeners who partook in the Perception task. Instruments are also described per task. We start with a description of the stimuli and language history data collected from speakers, and move next to the description of the perception and language history data collected from listener participants for subsequent analysis. The Data analysis procedures are described in terms of which methods were used, in relation to the Research Questions we aim to answer in the present thesis.

As mentioned in the Introduction, we mainly want to investigate the word-level local intelligibility of the L2 English spoken by Brazilian learners, when those speakers are listened by other non-native learners of English – specifically, Argentinian and German learners of L2 English. To achieve this main goal, we have the specific goals of investigating the intelligibility rates of target words in English (RQ1), as well as the general way in which the vowels in those words are identified, regardless of accuracy (RQ2). We also want to investigate how the hybrid, non-native systems of both speakers and listeners may affect intelligibility when those learners interact (RQ3).

Six Brazilian learners participated in our study and recorded the stimuli for the identification test. They were divided into three proficiency groups – beginner, intermediate and advanced –, each with a self-declared female participant and a self-declared male participant. Two native Canadian English speakers also participated as speakers, in order to provide baseline values.

The production task enticed the recording of 12 monosyllabic target words, six composing the minimal pair [æ-ɛ] and six tokens of the pair [i-ɪ]. In addition, participants were asked to record 12 distractor items, whose nuclei vowels were dissimilar from those of the target items. All targets presented the same CVC, monosyllabic lexical structure. All words, the targets and the distractors, were inserted in the carrier sentence “The word is ___ too”. This choice was made so target words could be produced within a prosodic context, avoiding a list reading effect. It also intended to prevent a semantic effect on the production of the speech continuum. The aforementioned Brazilian English learners and the native speakers of Canadian English recorded three repetitions of each carrier sentence (72 sentences total, per participant). There were also four carrier sentences used as practice items, prior to the recording of the samples. The target words were analysed acoustically and mapped in terms of absolute and relative duration, as well as first (F1) and second (F2) formats. After analysis, 16 samples from each participant were selected[42] to be used as stimuli for the perception task. All participants in the production stage voluntarily signed an Informed Consent Form (TCLE[43]) and completed a Self-Assessed Proficiency questionnaire[44] (before recording). The recording sessions should have taken place at the laboratory at UFRGS. However, because of the social isolation period during the Covid-19 pandemic, each participant recorded the samples at their own home, using varied mobile devices and agreeing to follow specific instructions for this scenario. Although the amount of time spent by each participant in their houses cannot be estimated, according to the original project, participation in the study had been estimated to take approximately 37 minutes, considering the recording of samples (20 minutes) and filling the documents (17 minutes).

The perception task was carried out online, by participants born and raised in the state of Buenos Aires (Argentina) and in different states of Central Germany. These listeners were native speakers of Spanish and German, respectively, and with a self-assessed advanced proficiency level in English as an additional language. After the participant inclusion criteria were applied, the Argentinian group totalled 28 participants, and the German group totalled 18 participants. The proficiency of the listeners was also determined by the application of a Self-Assessed Proficiency questionnaire, after the digital signature of the TCLE[45]. The participation of the Argentinian and German learners was estimated in a total time of 40 minutes – 25 minutes for the perception task and 15 minutes for the Self-Assessed Proficiency and the Language History questionnaires. All instruments were presented in digital form, in a URL exclusively meant for the present study and with data recording on the researcher's hired server.

---

[42] Details on methodological criteria for these selections are described in section 3.4.4.2.
[43] The acronym comes from the BP name 'Termo de Consentimento Livre e Esclarecido'.
[44] Canadian participants did not self-assess their proficiency in English, given that it is their native language. They did, however, assess their proficiency in additional languages, and they also filled out the language history questionnaire, as a means of gathering information on their experience with foreign languages. We understand that, from a Complex, Dynamic point of view, there is a distinct possibility that those foreign languages have already had an effect on these participants' native patterns (Cf. HAGIWARA, 2007).
[45] This study was approved by *Comitê de Ética em Pesquisa da Secretaria Municipal de Saúde de Porto Alegre/SMSPA*, under project nº 30099020.2.0000.5338.

The acoustic analysis of the recordings of the Brazilian participants were performed with the *Praat* 6.1.14 software (BOERSMA; WEENINK, 2020). The analysis allowed for the mapping of the regions of the BP-English common phonetic space (FLEGE; BOHN, 2021) occupied by the vowels [æ, ɛ, ɪ, i] in the learners' productions in English. Relative and absolute duration data were also tabulated. After the elaboration of the perceptual task, it was made available online for the listener participants. The data we collected was analysed statistically in the RStudio application (RSTUDIO TEAM, 2021) and plotted in the web-app Visible Vowels and with RStudio packages 'ggplot2' (WICKHAM, 2016) and 'phonR' (MCCLOY, 2013). The responses for each pair of vowels ([æ, ɛ] and [i, ɪ]) were analysed separately. In the statistical analysis, the mixed-effects logistic model considered, as predictor variables, (i) the native language of the listeners (Riverplate Spanish or Central German); (ii) the target vowel intended by speakers; and (iii) vowel quality measures of F1 and F2. The response variables were the accurate identification of the intended vowel (target) by the speaker (RQ1), as well as the identified vowel, regardless of accuracy/target word (RQ2). Participants and lexical items (target words) were taken as random intercept effects[46]. The exploratory analysis also took into account the participants' level of proficiency as well as the produced vowel length in the stimuli (RQ3).

Finally, participants who did not carry out all the tasks required at each stage of participation were excluded from the study. Those who participated in the study but did not take one or more inclusion criteria into account were also excluded. Inclusion criteria were verified with the responses in the questionnaire.

All aspects mentioned above will be further described in this chapter.

3.1 GOALS

In this section, we will state the general and specific goals of the present study. These goals are related to the Research Questions (see section 3.2) that arose from the review of literature and that guided the experiment design, as well as the choice for both quantitative and qualitative analyses.

This section is divided into main goal and specific goals. The main goal is to investigate the word-level local intelligibility of the L2 English spoken by Brazilian learners, when those

[46] Random intercept effects of 'target word', 'listener' and 'speaker' were taken in varied combinations across models, as in some instances the presence of all three effects did not allow the model to converge. Only random intercepts yielded convergence. See chapter 4 for details.

speakers are listened by other non-native learners of English – specifically, Argentinian and German learners of L2 English. The specific goals, as will be shown below, help attain the main goal. Therefore, they aim to analyse the rates of accurate target word identification, as well as how spectral (F1, F2) and temporal (relative duration) cues may play a part in vowel identification by each group of participants. Finally, the vowel systems at play (that of L2 categories for Brazilian, Argentinian and German non-native speakers of English) will be explored to help understand how they interact in regard to communication success in L2 English.

3.1.1 Main goal

Considering the language view that guides the present study, as well as the new perspectives and goals of foreign language pronunciation teaching, our main goal is to investigate the word-level local intelligibility of the L2 English spoken by Brazilian learners, when those speakers are listened by other non-native learners of English – specifically, Argentinian and German learners of L2 English. Our theoretical framework highlights a speaker-listener relation (see section 2.3).

3.1.2 Specific goals

In order to address our main concern, we set specific goals for the present study. Accordingly, we begin by investigating whether the listeners' native language (namely, Spanish or German) has an effect on the identification accuracy of the selected words (RQ1). Our CVC target words have vowels [æ, ɛ, i, ɪ], a choice that will be detailed in section 3.4.2.1. We grouped words in this way because we also mean to investigate if the target vowel is a significant predictor variable of the accuracy rates. Accuracy is understood, in this study, as a perfect match of the word intended by the speaker and the word identified by the listener. Words are later grouped by target vowel for the inferential statistics. As we have mentioned in section 2.3, thus, accuracy is the measure employed to address the construct of local intelligibility – with local meaning “word”, and intelligibility regarded as “being understood” (cf. MUNRO; DERWING, 2015).

Our second specific goal is to verify which linguistic or participant-related variable may play a role in the listeners’ general vowel identification rates, regardless of accuracy in relation to the target. As we understand that identifying the vowel is part of the process of identifying a word, we are interested in verifying which acoustic cues are playing a role in the non-native listeners' identifications (RQ2). We estimated the significance of these acoustic predictors, as well as of the listener's L1 and of the target vowel, by means of a mixed-effects logistic model.

Finally, once we obtained the inferential results about which dimensions have a main effect on the listeners' perception, we performed exploratory analyses to observe how the speaker-listener interactions may shed light on our participants' general system organisation and development. Thus, our third specific goal is to investigate whether there are emerging effects, depending on the group of listeners (namely, Argentinians and Germans), with regard to the weighting of acoustic cues (relative duration, F1 and F2). The exploratory analysis will take a second look at different accuracy rates for each L1 group of listeners. It also includes an analysis of possible thresholds in the listeners' common phonetic space (FLEGE; BOHN, 2021), which was operationalised here by general identifications – regardless of an accurate match to the word/vowel intended by the speaker (RQ3). By mapping vowel quality and temporal acoustic characteristics in the production of the vowels [æ], [ɛ], [i] and [ɪ] by Brazilian English learners at different levels of proficiency, we set out to verify the effects of such characteristics in the perception of Argentinian and German participants.

For the exploratory analyses, we include variables that did not necessarily have a significant main effect in the logistic models carried out in the inferential analysis. We understand that a stimulus-by-stimulus investigation can highlight emerging, hybrid processes that the inferential approach might not pick up on, depending on the dataset size. As already explained in section 2.1, the exploratory approach is complementary to the inferential statistics. Thus, based on acoustic space plots, we also investigate the way in which acoustic thresholds are presented in the common phonetic space of the L1 groups of non-native English participants (Argentinians and Germans). Plottings and analyses in Escudero (2009) and Escudero and Polka (2003), regarding absolute duration (x-axis) by F1 (y-axis), were major guides in the present study, though we do not limit ourselves to the same parameters. We intend to move further and observe emerging patterns in the speaker-listener pair.

3.2 RESEARCH QUESTIONS

Acoustic cues relating to vowel quality (first – F1 and second – F2 formants), as well as the temporal cue (relative duration) are used by listeners to distinguish speech sounds (ESCUDERO; POLKA, 2003; LADEFOGED, 2010). Previous studies have shown that L2 perception can also present effects from the L1 production patterns. This means that, to some extent, listeners perceive L2 sounds with “L1 ears” (FLEGE; BOHN, 2021; MUNRO; DERWING, 2015; BEST; TYLER, 2007; HOLT; LOTTO, 2006; FLEGE, 1995; BEST, 1995).

Thus, the present work sought to answer three research questions in order to achieve our specific goals. This section is also subdivided, in order to explore each one of our research questions. Additionally, each of them is motivated by pedagogical and/or Laboratory Phonology constructs.

3.2.1 Research Question 1: Does the L1 have an effect on L2 word intelligibility?

As we look into local intelligibility, which in our case relates to word intelligibility, does a learner’s L1 have an effect on accurate token identification? That is, will the words in the minimal pairs with the target vowels under analysis be correctly labeled more often by one or the other group of L2 English learners (native speakers of Spanish or German)? How does each L1 group identification pattern compare to each other, with regard to the identification of the Brazilian learners’ productions in English?

The SLM assumes that an ‘L1 filter’ might have an effect on learners' productions and/or perception. Thus, we enquire whether our listeners’ different L1s will show an effect on identification accuracy rates. In other words, we ask if Brazilian productions will be perceived as more (or less) intelligible for one or the other group of listeners.

Derwing and Munro define accent, in any language, as “a particular pattern of pronunciation that is perceived to distinguish members of different speech communities” (2015, p. 5). We see that both speaker and listener need to be considered when looking at accented speech (ALBUQUERQUE, 2019; MUNRO; DERWING, 2015). As we have explained, one of the effects of the ‘L1 filter’ in speech production is that of an accented speech in L2. We also assume that this ‘L1 filter’ will impact perception. We expect, then, that the members of both the Brazilian speaker group and the German/Argentinian listener groups will have a somewhat similar accented pattern within their group, as an accented pattern is specific to each of their speech communities. In other words, as non-native speakers, Brazilian learners will likely have similar patterns in the productions of L2 English sounds, given their native variety of Brazilian Portuguese, the non-native input they receive in their L2 learning trajectory, the common accented speech they are exposed to, and so on. The same can be expected among the members of each one of the listeners’ groups.

Previous studies have shown that an ‘L1 filter’ may somewhat predict identification patterns, more specifically considering assimilation of non-native sounds into native categories

(ESCUDERO, 2009; ESCUDERO; POLKA, 2003; STRANGE, 1995). Acoustic similarities, however, are not the only factors that may impact the way a segment, word or utterance is perceived. Contextual information (CRUZ, 2017; SCHWARTZHAUPT, 2015; STRANGE *et al.*, 2004), word frequency (SILVEIRA, 2004; GONÇALVES, 2014), familiarity with a given accent (SALVES; WANGLON; ALVES, 2020; CRUZ; PEREIRA, 2006) and orthography (SILVEIRA, 2012; ZIMMER; SILVEIRA; ALVES, 2009), to name but a few, are factors that can play a role in the production and/or perception of L2 speech.

As will be described in the next sections of this chapter, some of those factors, such as word frequency and contextual information, can be somewhat controlled by the experimental design. Additionally, though the Complex, Dynamic framework posits that each person's system will be unique – for it is shaped by an individual's embodied experiences –, it also allows us to work with groups that are assumed to be in somewhat similar stages in terms of language development.

Given the geographical proximity of the Buenos Aires state and the Porto Alegre metropolitan area, as well as the cultural exchange that those regions are subjected to, we enquire if the English spoken by Brazilians would be more intelligible for this group than for the German group. Moreover, considering the Latin vs. Gemanic language families at play, we also asked if the (dis)similarity of native Brazilian/Spanish and native German vowel sounds to those of L2 English would have an effect on word intelligibility in the interactions of those non-native learners.

Finally, we ought to point out that we are looking at local intelligibility, a construct that includes, by Munro and Derwing's definition (2015), both word and segment intelligibility. As we ponder about all linguistic and extra-linguistic factors that might play a role in intelligibility, it seems to us that word intelligibility (when compared to segment intelligibility) is a more suited unit of analysis in terms of Applied Linguistics. As we have mentioned in the previous chapter, in this thesis we intend to combine an applied approach with that of Laboratory Phonology. Thus, this first research question was formulated in order to focus on possible teaching and training applications of the findings obtained in the present study. In the next section, we will detail how acoustic characteristics were taken into account to help answer the following research questions.

3.2.2 Research Question 2: Which linguistic, stimulus-related or participant-related variables might have an effect on vowel identification?

As we have mentioned, speech perception will show the effects of a number of factors, such as the acoustic cues in the stimuli, acoustic contextual information and learners' language developmental stage, for example. What, then, has an effect on identification? More specifically, which linguistic, stimulus-related or participant-related variables might have an effect on L2 vowel identification?

To answer this question, we move from local intelligibility (at the word level) to vowel identification. Vowel identification is assumed to be part of the language processing enticed in identifying a given word. Thus, we could expect that a possible lack of success in the communication between our Brazilian speakers and our Argentinian or German listeners may arise from lack of distinctiveness in production and/or perception of the L2 utterances, hence our choice to investigate vowel identification.

Considering that perception is a process that occurs through an 'L1 filter' (FLEGE, 1995; FLEGE; BOHN, 2021; BEST; TYLER, 2007), it is relevant to consider which acoustic cues favour the identification of a given member of the minimal pair. Our theoretical framework states that linguistic experience with a native language is what guides the perceptual patterns of a speaker in the establishment of new perceptual weightings for a given cue in a new language (HOLT; LOTTO, 2006). This process also occurs in relation to the linguistic experience with an additional language. Thus, as cue weighting is language-specific, we consider it worth asking whether different L1 backgrounds will have an effect on the identification of L2 vowels. Moreover, given the perceived informativeness of each acoustic cue, which one(s) will be decisive in the identification of the segments?

As mentioned in the previous chapter, from a Complex, Dynamic perspective, the learning trajectory progresses from an initial 'L1 filter'-only state, to a state in which the L2 is perceived (to some degree) as having its own set of characteristics. In more traditional terms, as the L2 is learned, it becomes different from the L1, in each subsystem involved in language learning. We did not conduct a longitudinal study, but we assume, from our literature review, that both speaker and listener groups of non-native learners of English have started that development process prior to the moment of data collection. Moreover, as language is constantly developing and adapting, we understand that this process has continued to happen after we collected the data analysed here. Such a learning trajectory should include perceiving and reconsidering the prioritisation of acoustic cues when producing and perceiving speech sounds in an additional language. Thus, for instance, if a given cue plays a primary role in the L1, and that is also the case in the L2, it could be expected that the need for perceptual adaptation would be minimal – or at least lower than the need in the case of a cue that is not primary in the L1 in question. On the other hand, if a cue is heavily weighted in the L1, but is not distinctive in an L2, it could exert more adaptation from a learner to distinguish a sound through other characteristics. Lastly, it could also be the case that the same cue is weighted differently in the L1 and in the L2, a scenario in which adaptation would be necessary as well. Therefore, it is worth questioning whether the prioritisation of F1, F2 and relative duration in the perception of each listener group will be different when considering L2 English.

It should also be noted that the present study sought to identify these effects on the communication process between non-native groups (Brazilians with Argentines and Brazilians with Germans), in which the prioritisation of acoustic cues may already be, from the start, different from what they would be for a native English speaker, given their 'L1 filter'.

3.2.3 Research Question 3: What characteristics of listeners' L2 phonetic categories can be inferred from their identification patterns in the perception task?

Our investigation path has so far considered non-native learners of English and their identification accuracy, that is, the level of local intelligibility in the interaction of Brazilian speakers and Argentinian/German listeners (RQ1). From a pedagogical perspective, understanding what may be hindering intelligibility is a relevant aspect in conducting language teaching and pronunciation training activities. We have also observed each group's L2 vowel identification patterns (regardless of accuracy) in terms of linguistic, stimulus-related and participant-related factors (RQ2). Those enquires have also proven relevant for Laboratory Phonology investigations hitherto. Those first two research questions will apply inferential statistics in order to yield the main predictors of each response variable – namely, accuracy rate (RQ1) and identified vowel (RQ2). From the significant predictors yielded by the inferential analysis used to answer those initial questions, thus, we now ask: what can we infer about the listeners' common phonetic space and vowel systems, in terms of L1-L2 composite categories? Moreover, what gradient and/or interactional roles are played by the vowel quality and temporal dimensions that compose those categories? Is there a different, L1-based cue weighting pattern that leads different groups of listeners to identify the stimuli differently? How do those groups compare? Additionally, from a pedagogical perspective, in which way can our data shine some light on relevant learning landmarks, considering the minimal pairs we are investigating?

This third research question intends to work with two complementary aspects. On the one hand, it aims at some interface between laboratory analyses and their possible application in the classroom. As we have mentioned, the investigation proposed in RQ1 was directed towards an applied approach, whereas RQ2 focused on the role played by the acoustic characteristics of stimuli in speech perception. Here, we intend to observe how both types of identifications (accurate and otherwise) relate to the speaker-listener interaction. Furthermore, as we have applied inferential statistics to estimate the effects we deal with in RQ1 and RQ2, we now want to complement that analysis with an exploratory approach in RQ3. We believe that, within a Complex, Dynamic framework, a stimulus-by-stimulus analysis can shed light on processes that are too gradient to be picked up by group statistics, especially given the small dataset we have in the present study (LOWIE, 2017).

According to Flege (1995) and Flege and Bohn's (2021) Speech Learning Model (SLM) predictions, L2 learners will, at first, produce L2 speech sounds based on the characteristics mapped from their native vowel categories. For instance, a Brazilian learner from the Porto Alegre area is expected to assimilate both [æ] and [ɛ] in English as the BP category of [ɛ]; likewise, [i] and [ɪ] in English would be assimilated into the BP [i] category. The SLM predicts that those learners should realise that those sounds are different, that is, belong to different categories, at some point in the learning trajectory. Only after they have noticed this distinctiveness regarding vowel duration, tongue height and tongue frontness (among others) will they be able to start a dissimilation process. As Flege and Bohn (2021) point out, perception and production co-evolve, so it cannot be said, a priori, on which 'side' (perception or production) that distinctiveness will take form. But it is assumed that, wherever it starts, it should affect the other domain when a new (composite) category is established.

One relevant aspect of this process is that the way distinctiveness is processed by an L2 learner may not be the same as it is for a native speaker. Moreover, the gradience that might be present in such (native and non-native) systems will be, from a Complex, Dynamic perspective, in constant change, as a response to a speaker's embodied experiences. That leads us to assume that groups of different levels of proficiency can showcase different moments of the learning trajectory predicted by the model. When mapping how the characteristics of F1, F2 and duration appear in the production of these speakers (grouped by proficiency), it could be possible to describe the acoustic thresholds of each vowel in the speaker's system. Additionally, we could investigate the state of L2 category dissimilation[47]. Likewise, mapping identifications according to the acoustic cues in the stimuli might give us the same (or at least a similar) overview of the listeners' developing systems. Therefore, as a last question to the present study, we sought to understand how the acoustic spaces of speakers and of listeners are organised. We expected that those mappings might allow us to infer the boundaries of these systems' categories, and how they might be reorganizing as a function of the ongoing language development. Furthermore, from the distinct ways in which each group's composite categories are established, we sought to analyse if (lack of) communication success can be predicted – and, of course, if pedagogical techniques can be used to enhance that success.

As we have mentioned, word intelligibility includes (to some degree) the intelligibility of the segments that compose that word. Thus, accurate identifications and general identification patterns (regardless of their accuracy or not) can point to some intelligibility issues that might arise in the speaker-listener interaction. The very concept of intelligibility implies that success in communication is not (necessarily) based on nativeness. Therefore, local intelligibility, as understood in the present study, could be attained depending on the speaker-listener pair, as well as on the role acoustic cues are playing in that interaction – even though, as mentioned, it may not be the same role it would play in the interaction with a native speaker. Investigating the local intelligibility of the L2 English spoken by Brazilian learners, at the word level, when those speakers interact with other non-native learners of English – specifically, Argentinian and German learners of L2 English – is our main goal in the present study. This is why our research questions seek to analyse different processes that might have an effect on intelligibility.

As mentioned in section 2.2, this third and final Research Question is largely inspired by the work of Escudero (2009). The author conducted an experiment in which she analyses if the 'L1 filter' might predict L2 perception. Escudero's analysis looked at acoustic cues that play a role in L2 vowel identification. Given our Complex, Dynamic framework, we wanted to investigate whether other linguistic factors might also play a role in that process – and influence local intelligibility. As previously explained, we wanted to complement the Laboratory Phonology approach with an Applied Linguistics view, providing a sort of interface between findings in both fields. Additionally, as we consider language as a process, which changes in time and constantly adapts to an individual's experience, we also believe it useful to look closely

[47] Though the present thesis does not focus on production, in the Results section we do provide descriptive analyses of those processes as background information for the statistical and exploratory findings we discuss in regard to listeners' perception.

at how hybrid, non-native systems might show some gradience that a single approach might not.

3.3 PARTICIPANTS

In this section, we will describe the four groups of participants we collected data from. There are two groups of speakers, that of native speakers of Brazilian Portuguese, and that of native speakers of Canadian English. They are described in that order. There are also two groups of listeners, that of native speakers of Riverplate Spanish, and that of native speakers of Central German. These are grouped in a single, final subsection of non-native listeners of L2 English.

3.3.1 Non-native speakers of English (native speakers of Brazilian Portuguese) – Production task

Production samples were collected from six Brazilian participants (M_{age}: 26y; SD: 3.41), born and raised in the southern Brazilian city of Porto Alegre (RS) and its Metropolitan Area. This was done to ensure that they spoke the same variety of Brazilian Portuguese as their native language. There were three self-declared male and three self-declared female participants, totaling six participants. The use of one person of each self-declared gender was chosen because it is recommended to use samples from at least one person of each gender in perception tests. This recommendation comes from the fact that an individual's physiology has acoustic effects in their productions and, consequently, this can have effects on perception (LEHET; HOLT, 2017; ESCUDERO, 2009; FLEGE, 2003). There were two speakers in each proficiency group (beginner, intermediate and advanced), according to the information provided in the Self-Assessed Proficiency instrument (detailed in section 3.4.2).

As for our inclusion criteria, beginner and intermediate level learners could not have had an immersion experience in an English-speaking country for more than six months. We also established that beginner learners could not have had phonetic / phonological pronunciation training in English. These criteria aimed to avoid any effects of immersion or training on those speakers' pronunciation patterns, since we expected proficiency levels to somewhat reflect different stages of the speakers' learning trajectories.

Participant recruitment was to take place by invitation in the face-to-face classrooms of the English Language program of the Department of Modern Languages / Instituto de Letras at UFRGS. As there were no classes due to the Covid-19 pandemic, the invitations were sent using messaging applications, such as WhatsApp. Invitations included unofficial student groups of the English Program, as well as individuals that the researcher had prior knowledge with and who might fit the profile. In the invitation message (in Portuguese), it was clarified that participants had to have been born and raised at least until the age of 7 years old in Porto Alegre or its Metropolitan Area. It also stated that all proficiency levels of English could take part in the study. Additionally, the message clearly explained that the participation enticed voice recordings, and that those recordings would be played back to other non-native speakers of English in an intelligibility task carried out online later on. Furthermore, given that the university lab was no longer a viable option, the invitation message stated the ideal conditions for recording at home. We recommended playing the PowerPoint presentation file containing the carrier sentences (see section 3.4.3.2) on a desktop or laptop computer. We also advised participants to record their voices using a mobile device, preferably with data connection turned off, so that no notification sounds/vibrations should disrupt the audio quality of the recording.

Interested participants were asked to provide an email address, in which they received the link to the online Self-Assessed Proficiency questionnaire (see section 3.4.2), as part of the screening process. After the researcher checked the answers in the questionnaire for the inclusion criteria[48], the selected participants were sent the online Informed Consent Form (Appendix A), along with a message that reiterated the instructions already detailed in the invitations. The email was written in Brazilian Portuguese and sent to each selected participant, as follows:

> *Olá, obrigada pelo interesse em participar da pesquisa.*
> *Agora que você já preencheu sua avaliação, vamos à gravação. Em anexo envio uma apresentação em PowerPoint. Por favor, faça o download e abra o arquivo no seu computador ou notebook/laptop. A gravação você pode fazer no celular mesmo. Vou explicar tudo melhor a seguir :)*
> *Antes de mais nada, por favor clique no link a seguir para assinar digitalmente o Termo de Consentimento Livre e Esclarecido, que é o documento oficial de que você aceita participar da pesquisa:*
> *[LINK]*
> *A apresentação em anexo foi feita para rodar sozinha. As primeiras telas têm botões para você clicar, e na hora que as frases forem começar a aparecer, a tela vai passar sozinha. Isso é uma exigência da metodologia do estudo, e como a apresentação já está programada, você não precisa se preocupar em fazer nada.*
> *Primeiro você vai ver algumas telas com instruções. Uma delas vai indicar que você já pode começar a gravar. É só apertar o botão e ler as frases em voz alta de acordo com as instruções.*

[48] Two participants volunteered but were not selected because they did not fit the Inclusion Criteria for their self-assessed proficiency level. They were intermediate level learners, but one had had pronunciation training and the other had had an immersion experience of over 6 months living in a country where the target language (English) was spoken.

> *Se possível, peço que você desligue a internet do seu celular (modo avião) enquanto faz a gravação, para não ter som de notificações ou da vibração do aparelho. A gravação inteira deve levar cerca de 20 minutos. Se for possível na sua casa, faça a gravação sozinha(o) em um cômodo com janelas e portas fechadas, para ficar silencioso, e deixe seu bicho de estimação do lado de fora por esse tempo.*
> *Vamos lá? O arquivo do PowerPoint para você baixar e abrir no seu computador está em anexo. E mando também o manual para você fazer a gravação no seu celular.*
> *Se você tiver qualquer problema, é só responder a este email ou falar comigo no WhatsApp [NUMBER].*
> *Mais uma vez, obrigada por participar da pesquisa!*
> *Abraço,*[49]

All participants also filled out a Language History Questionnaire after completing their recordings. Participants reported that English was the second language they had learned. Spanish was the third language learned by four out of the six participants; French was the third language for another[50], and the last one did not speak a third language. One of the six participants also spoke a fourth language, Italian. None of the participants reported a proficiency level higher than intermediate (3-4) in their third or fourth language in the Self-Assessed Proficiency questionnaire. Details on self-assessed proficiency levels of Brazilian participants are provided in Table 3.1.

[49] Hello, thank you for your interest in taking part in our research study.
Now that you have already filled out your Self-Assessment Proficiency questionnaire, let us move forward to the recording. Attached is a PowerPoint file. Please, download the file and open it in your desktop or notebook/laptop computer. You can make the recording on your mobile phone. I will explain it better below [smile emoticon].
Before we move further, please click the link below in order to digitally sign the Informed Consent Form. That is the official document that states that you wish to partake in our research study:
https://forms.gle/d2Hse3f9zK69knBS6
The presentation file attached to this email is supposed to forward from one slide to the news automatically. The first few screens have buttons for you to click, and when the sentences intended for recording start to be exhibited, the presentation will automatically forward from one to the next. That is a methodological requirement of our study, and since the file is already set to do it automatically, you are not required to take no actions in that regard.
First, you will see some instruction slides. One of them will indicate to you that you can initiate the recording on your recording app. You only need to push the 'start' button and read the sentences out loud, according to the instructions in the previous slides.
If possible, we ask you to turn off the internet connection of your phone (Airplane mode) whilst you record the sentences. That is to avoid any notification sound vibration to be recorded along with your voice. The recording of all sentences should take about 20 minutes. If possible at your household, please be alone in the room and keep doors and windows closed, so that the space can be quiet. We also ask that you leave your pet out of this room during the recording.
Let's go? The PowerPoint file you should download and open on your desktop computer is attached to this email.
I am also sending you a tutorial on how you can make a recording on your mobile phone.
If you should encounter any difficulties, just reply to this email or contact me on WhatApp adding the phone number [author's personal phone number].
Once again, thank you for your interest in taking part in our study.
Cheers,
[signature]

[50] We understand that the fact that some of our participants speak other languages besides BP and English might mean that their systems do not reflect a similar state to those of participants who only speak BP and English. We recognise it as a limitation of our study. However, during the Covid-19 pandemic, we had some constraints with regard to participant recruitment.

Table 3.1 – Self-assessed proficiency levels of Brazilian participants in their L1, L2, L3 and L4
Source: present study

	Participant	Beginner		Intermediate		Advanced	
		BR08	BR01	BR04	BR07	BR02	BR06
Gender		F	M	F	M	F	M
L2		EN	EN	EN	EN	EN	EN
proficiency in written skills	Reading	3	2	5	4	5	6
	Writing	2	2	4	4	4	5
proficiency in oral skills	Listening	2	3	4	4	5	5
	Speaking	1	3	3	4	5	5
L3		ES	ES	FR		ES	ES
proficiency in written skills	Reading	4	4	4		4	2
	Writing	2	2	4		2	1
proficiency in oral skills	Listening	3	3	3		3	1
	Speaking	1	2	3		3	1
L4		IT					
proficiency in written skills	Reading	2					
	Writing	1					
proficiency in oral skills	Listening	1					
	Speaking	1					
Immersion experience (6+ months)		no	no	no	no	no	no
Phonetic/pronunciation training		no	yes	no	yes	yes	no

On average, "Music / radio / podcast" (M: 6, SD: 0) and "Television (movies, shows, YouTube, etc)" (M: 5.83, SD: 0.41) were reported as the experiences that most contributed to the participants' learning of English, accounting for 2h55mins (SD: 16 minutes) and 1h40min (SD: 30 minutes) daily time of usage of the language, respectively. Three of the participants reported living in the city of Porto Alegre up to 7 years of age, and the other three reported living in the Metropolitan Area during that period. Only one of the six participants reported that their mother[51] was born in a different area[52] than Porto Alegre or its Metropolitan area. Detailed information of each participant's responses to the questionnaire are presented in Appendixes K and L.

3.3.2 Native speakers of English (native speakers of Canadian English) – Production task

Two native speakers of Canadian English were invited to provide samples for baseline purposes. Baseline productions helped us control to which degree the listeners were able to correctly identify tokens that were produced without a foreign accent colouring (CRUZ, 2017). This, in turn, allowed us to analyse identifications of Brazilian productions within the speaker-listener pair, discarding a possible listener-only effect on the data.

A male and a female participant (M_{age}: 37.5y; SD: 0.71) from the same Ontario region (variety of Canadian English) were invited to take part in the study, using the same procedure for participant recruitment as we did with Brazilian participants. Both showed interest and provided email addresses, to which the link of the online Informed Consent Form in English (Appendix B) was sent, along with a similar message to that of the invitation. This message (see section 3.3.1) was in English and once again recommended recording conditions. Attached to the same email was the MS PowerPoint presentation file (Appendix I) with the carrier sentences. Both participants received different MS PowerPoint files, as the carrier sentences were randomised for each (see section 3.4.3.2). Their presentation files, as the ones used with Brazilian participants, contained information as to how to send the recordings to the researcher, and both opted for the WhatsApp messaging application.

Canadian participants were also asked to fill out the Language History Questionnaire. The male participant reported learning six additional languages, Portuguese being the fifth in order of languages. This means the male participant provided information about English (native), French, Spanish and Italian, but not about Portuguese. This participant ranked proficiencies in all three reported foreign languages as 'high' or lower, but none as 'fluent'.

[51] Or an equivalent female figure that raised them.
[52] That mother/female figure was born in Santa Maria, a city in the central area of the state of Rio Grande do Sul, in the southern part of Brazil. (Cf. SCHOLL; FINGER; FONTES, 2017, for a discussion on the role of the female parental figure on language learning).

The female Canadian participant is a bilingual English-Portuguese native speaker, having been born in Canada to immigrant Portuguese parents[53]. She also speaks French and Spanish. She assessed her proficiency in (Mainland) Portuguese as very high, and in French (third language) as fluent. The female participant reported learning Portuguese at home, school, language school and in other ways. Family interaction, language school and music/radio/podcasts were reported as having contributed considerably to her development in the second language. Additionally, she lived in a Portuguese-speaking country for a year, and studied at a school where Portuguese was spoken for two years. It is also worth mentioning that the female participant reported having pronunciation training in English (native language), but not in Portuguese (simultaneous second language). Her daily usage of Portuguese was estimated in two hours of family interaction as well as music/radio/podcasts, and in one hour of interaction with friends, media consumption, as well as social networks.

Details on self-assessed proficiency levels of the Canadian participants are detailed in Table 3.2.

[53] Within a Complex, Dynamic framework, we can assume that this participant may not reflect the same system organisation as a monolingual speaker would. This is a limitation of the present study that will be further discussed in section 5.2.

Table 3.2 – Self-assessed proficiency levels of Canadian participants in their L2, L3 and L4

	Participant	Baseline	
		CA09	CA10
Gender		F	M
L2		PT	FR
proficiency in written skills	Reading	5	4
	Writing	5	2
proficiency in oral skills	Listening	6	3
	Speaking	6	2
L3		FR	IT
proficiency in written skills	Reading	4	3
	Writing	4	2
proficiency in oral skills	Listening	4	3
	Speaking	3	3
L4		ES	SP
proficiency in written skills	Reading	5	3
	Writing	5	2
proficiency in oral skills	Listening	5	3
	Speaking	5	3
Immersion experience (6+ months)		no	no
Phonetic/pronunciation training		no	yes

Source: present study

Detailed information on each participant's response to the questionnaire is presented in Appendixes M and N.

3.3.3 Non-native listeners of English (native speakers of Riverplate Spanish and Central German) – Perception task

The samples produced by Brazilian and Canadian speakers were segmented per carrier sentence and a subset of those productions was used as stimuli in the perception task (see section 3.4.4). The perception task was then presented to listeners of L1 Riverplate Spanish (state of Buenos Aires) and L1 Central German (states of Nordrhein-Westfalen, Hessen, Thüringen, Sachsen-Anhalt, Brandenburg und Sachsen, in the Central Germany area).

The recruitment of participants for the perception task took place via posts in social media, in personal profiles and in communities created autonomously (without institutional involvement) by people from Buenos Aires (Argentina) and Germany[54]. Email invitations were also shared with professors from both countries[55], should they want to share the invitations with their students and acquaintances. Invitations were also posted in messaging groups in services like WhatsApp. The invitation contained the link to the experiment, along with information about the volunteer character of the participation. It also explained the general goal of the study and made it clear that the research study was part of the present Master's thesis. Any questions by potential participants were replied to in the same channel (group or private) it had come through.

A total[56] of 28 native speakers of Riverplate Spanish and 18 native speakers of Central German took part in the study. Participation took about 47 minutes[57], 33 minutes for the perception task[58] and 14 minutes for the language history questionnaire[59]. All listener participants reported having an advanced level of proficiency in L2 English (see section 3.3.3). All 28 Argentinian participants reported that English was their second language, whereas 3 of the 18 German participants reported it was their third language (with Russian being the L2 for

[54] See section 5.2 for a detailed account of invitations to German groups.

[55] Invitations were sent to emails listed in Linguistics and Modern Languages department websites from universities in Argentina and Germany, as well as to personal contacts of the adviser to the present study.

[56] This number does not include participants whose data were excluded due to noncompliance with inclusion criteria. A total of 70 participants either never finished the perception task, or did not speak the appointed L1 (both groups) or L1 variety (Argentinian participants). Additionally, 8 German learners were excluded *a posteriori* for not having Central German as their L1 variety (see section 5.2 for details).

[57] Because the task was long, participants were shown, at all times, a notification reminding them that they could take breaks along their participation, so long as they kept the tab/browser open to prevent their data from being lost. Pilot tests done by the researcher and her adviser estimated 25-30 minutes to complete the task without breaks. This estimated time was also informed to participants before the start of the data collection.

[58] This estimate is based on the time participants took from identifying the first carrier-sentence (by clicking the 'next' button) to identifying the last carrier sentence, which led them to the language history questionnaire they were required to fill out.

[59] This estimate is based on the data of the 46 listener participants, considering the time it took from starting the questionnaire and the time registered by Google Workspace as the time the questionnaire data was sent.

two of them and French being the L2 for the other). Portuguese was the L3 of four Argentinians and one German, and the L4 of two Argentinians and one German.

Detailed information on each participant's response to the questionnaires is presented in Appendix M.

3.4 INSTRUMENTS

In this research study, we made use of three versions of the Informed Consent Form, one in Portuguese for Brazilian participants (Appendix A) and two in English – for Canadian speakers (Appendix B) and for Argentinian and German listeners (Appendix C). We also made use of two language history questionnaires, one in Portuguese for Brazilian participants (Appendix E) and another in English for Canadian, Argentinian and German participants (Appendix F). The questionnaires were both adapted from Scholl and Finger (2013) and Scholl, Finger and Fontes (2017). Two Likert scales were used for self-assessment of proficiency, one in Portuguese for Brazilian participants (Appendix D) and another in English for Canadian, Argentinian and German participants (inserted in the questionnaire in Appendix F). Likert scales were presented to participants after the recording had been sent by native speakers of BP[60] and Canadian English (production task). Listener participants, native speakers of Riverplate Spanish or Central German, filled out the questionnaire after the identification task.

For the production task, Brazilian speakers used an MS PowerPoint presentation file, created in the Slides application of the Google Workspace and displayed in the PowerPoint application of the Microsoft Office suite. We used the Randomizer.org website to randomise the selected words used in the presentation files. The audios were recorded in different mobile devices[61], given that each participant made the recording with their own smartphone native[62] voice recording app, due to the Covid-19 pandemic. Audio files were segmented by carrier sentence and were analysed using the *Praat* application (BOERSMA; WEENINK, 2020). Formant frequency normalisation was made using the Norm tools made available by the

[60] Brazilian participants also filled out a vocabulary questionnaire (Appendix J) adapted from Lepage and LaCharité (2015). This was done after the recording and in order to verify the possible influence of a lack of knowledge of the target words on the production pattern, given that BP participants had different levels of proficiency in English. Our original project intended to make use of this information, but due to delimitational issues, we did not approach this question in the final version of the study we report in the present thesis.

[61] Technical specifications on each mobile device were not asked of participants, in order to avoid a possible embarrassment due to whatever perception speakers might have had over what their mobile device model might imply.

[62] A native application is pre-installed by the phone carrier (service provider) in the device prior to purchase. That means that participants did not have to install a recording app in order to record their voice when taking part in the present study.

University of Oregon[63] and graphics were done with the Visible Vowels application[64] and the 'ggplot2' (WICKHAM, 2016) and the 'phonR' (MCCLOY, 2013) packages for the RStudio application (RSTUDIO TEAM, 2021). Statistics were performed with the RStudio application (*op. cit.*).

Argentinian and German listeners in the perception task were provided a dedicated[65] website built by the researcher and made available at a URL on the researcher's personal page (www.deborahsalves.com.br/reserach-study). The answers were recorded by the FormVibes WordPress plugin (WPVIBES, 2021) installed in the website's server and accessible via login only to the researcher. They were later tabulated for use in the statistical analysis with the RStudio application (*op. cit.*). In the subsections that follow, each one of the instruments mentioned in this brief introduction will be described in detail.

3.4.1 Informed Consent Forms – prior to participation in any task

The Informed Consent Form (TCLE[66]) was prepared in accordance with Resolution No. 510/2016 from the Brazilian National Health Council[67]. The form was signed before any other tasks were asked of participants. This ensured that only participants who had previously agreed to the terms of the Informed Consent Form participated in the tasks. As can be seen in Appendixes A to C, there were three versions of the TCLE. Each version specified which tasks each group would perform, as well as the risks and benefits in taking part in the study. That was because different groups (two of speakers and two of listeners) performed different tasks.

For Brazilian and Canadian participants, the TCLE-BR (in Portuguese) and the TCLE-CA (in English) were presented in a Google Workspace form (Appendix A and B). The forms were electronically signed before the start of the collection of speech data (recordings). Participants could click a checkbox[68] if they wanted to receive a copy of the document.

For Argentinian and German participants, the TCLE-EN (in English) was presented on the second screen of the URL dedicated to the present research study (Appendix C). The form was also electronically signed prior to advancing to the survey page *per se*. Argentinian and

---

[63] http://lingtools.uoregon.edu/norm/norm1.php
[64] https://www.visiblevowels.org/
[65] During data collection, the website was used exclusively for this purpose. No other content was made available and no other person used the website but the researcher and the participants.
[66] The acronym comes from the BP name 'Termo de Consentimento Livre e Esclarecido'.
[67] This study was approved by *Comitê de Ética em Pesquisa da Secretaria Municipal de Saúde de Porto Alegre/SMSPA*, under project nº 30099020.2.0000.5338.
[68] In this form, the copy was provided by Google Workspace default tools, with Google Workspace default aesthetics.

German participants could leave their email if they wanted to receive a copy[69] of the TCLE-EN in their email inbox.

3.4.2 Language history questionnaire – Production and perception tasks

The language history questionnaire was applied in order to obtain information on the learning trajectory and pattern of L2 English use by the participants, both Brazilian (Appendix E) and international (Appendix F). As already pointed out, such data can help to understand the participants' current stage of language development. It also allows us to verify possible changes that the overall language system might have gone through due to other additional languages that these participants have learned. Furthermore, these data are important so that the characteristics pertaining to the participants' language developmental trajectories could be more accurately described. These, in turn, could contribute to the verification of the effects of individual participants in relation to the group in which they were inserted. They can also allow for some comparability with future studies. Finally, such information allowed us to confirm the adequacy to the inclusion criteria.

The questionnaire was adapted from Scholl and Finger (2013) and Scholl, Finger and Fontes (2017). Questions from the original questionnaire that did not connect to the goals of the present study were not included. Additional questions asked for more information that could have influenced the trajectory of speakers/listeners and their current stages of development as English learners. Therefore, there was a version written in Brazilian Portuguese (for Brazilian speakers), and another version written in English (for Canadian English speakers and Argentinian/German listeners), both made available online using a Google Workspace form (see Appendices E and F, respectively). The questionnaires were answered after the production task (Brazilian and Canadian speakers), and after the perceptual task (Argentinian and German listeners[70]).

The Language History Questionnaire included questions related to the first four languages learned by respondents. It queried the informant about: (a) where they had learned the reported languages; (b) at which age they started to learn, started to actively use and became

[69] In this form, the copy was manually sent by the researcher using her personal email account. Because the website was developed exclusively for this research, these participants received a PDF file formatted by the researcher. That format was aesthetically different from the Google Workspace default aesthetics.

[70] Data collected from Argentinian and German participants who did not fill in the language history questionnaire were not included in the results presented in this thesis. Without those data, it was not possible to confirm our inclusion criteria, nor to analyse learning development factors which are crucial within a Complex, Dynamic framework.

fluent in those languages; (c) which factors had contributed to their learning; (d) whether they had had any pronunciation training or taken a course on Phonetics/Phonology at college/university; (e) the number of hours (on a daily basis) participants used the language for a variety of activities; (d) their immersion experiences; and (e) some demographic background information. There was also a section for self-assessment of proficiency in the first four languages the informant had learned. That assessment made use of a likert scale ranging from 1 to 6, in which 1 meant a 'very low' proficiency level and 6 meant 'fluent'. Participants were invited to self-assess their proficiency level by skill: reading, writing, listening and speaking.

For Brazilian participants, the Self-Assessment Proficiency section was detached and applied prior to the Language History Questionnaire, during the screening process. This was done in order to only select participants that matched the inclusion criteria in relation to proficiency level. For native speakers of Canadian English, the self-assessment section was applied along with the Language History Questionnaire[71]. The same was done for Argentininan and German listeners. For these non-native participants, we had a checkbox prior to the beginning of the perception task in which they confirmed the match to the 'advanced proficiency level' inclusion criteria. This is why we assumed it was not an issue to leave all further questions relating to their proficiency level in English (as well as in other languages) to be informed after the perception task and along with the Language History Questionnaire.

3.4.3 Production task

In this section, we will describe the steps we took to select the lexical items for the production task. We will also describe how we elicited those words from our speaker participants (Brazilian and Canadian) in the production task.

3.4.3.1 Selection of target and distractor words

The six Brazilian and the two Canadian participants in the production task of the present study recorded a total of 576 carrier sentences. Each individual recorded 72 sentences, which included 12 target and 12 distractor words, in three repetitions each. Box 3.1 shows the final selection of words recorded by participants in the production task.

[71] We assumed that, as they were native speakers of the target language, there was no need to assess their proficiency level as an inclusion criteria, as we had to do with Brazilian participants.

Box 3.1 – Target and distractor words (CVC monosyllables)

Target word	Target vowel	Target word	Target vowel	Distractor word	Target vowel	Distractor word	Target vowel
bat	æ	bet	ɛ	boot	u	book	ʊ
pat	æ	pet	ɛ	foot	u	food	ʊ
sat	æ	set	ɛ	shoot	u	should	ʊ
feet	i	fit	ɪ	but	ʌ	bought	ɔ
heat	i	hit	ɪ	cut	ʌ	caught	ɔ
seat	i	sit	ɪ	shut	ʌ	shot	ɔ

Source: present study.

The 24 words that comprised the production stage of the study were selected based on syllabic and segmental features, and taking their frequency of use into account (GONÇALVES; SILVEIRA, 2015; GONÇALVES, 2014). All words were CVC monosyllables, with voiceless[72] alveolar plosive codas. The voicing of onset and (mainly) coda segments may affect vowel duration (LADEFOGED, 2010; YAVAS, 2011), which is why a voiceless coda was selected as the uniform coda for the target words. Ladefoged (2010) and Yavas (2011) also posit that place of articulation affects vowel quality as well, which is why the uniform coda was alveolar[73].

Target words were selected based on the target vowels that are under analysis in this thesis, namely [æ], [ɛ], [i] and [ɪ]. Distractor words were included, on the one hand, not to hint the focus of study to participants. This is justified as we try to avoid an effect of having participants turn their attention to that particular detail, as we wanted them to behave as naturally as possible – despite the laboratory condition of the data collection. Additionally, distractor words were chosen to have those particular vowels [u, ʊ] in order to allow for vowel formant frequency normalisation (see section 3.4.4.1).

[72] There were two exceptions, 'food' and 'should', chosen because they were the best fit considering all selection criteria together. We highlight, however, that both 'food' and 'should' were distractor words.
[73] Besides the already mentioned 'food' and 'should', the 'book' token was an exception, again because it was the best fit considering all criteria together. Once again, we point out that 'book', as well as 'food' and 'should', was a distractor word.

The target and distractor words had their frequency of use checked in two corpora – The Corpus of Contemporary American English (COCA)[74] and the British National Corpus (BNC)[75] databases. We did so in order to select words that were both frequently used and, considering the original minimal pairs, similarly frequent within each pair. Box 3.2 shows the results of the 24 words selected for the production task. Highlights indicate the words used in the perception task (see section 3.4.4).

Box 3.2 – Word frequency per million words of target and distractor words

Word	COCA*	BNC**	RoF***	Word	COCA	BNC	RoF
bat	15	10	24	boot	13	16	23
bet	51	26	19	book	305	241	4
pat	29	21	21	foot	60	71	16
pet	20	14	22	food	232	184	5
sat	92	110	10	shoot	52	16	20
set	329	434	3	should	927	1078	2
feet	151	132	7	but	4633	4409	1
fit	74	79	14	bought	70	87	13
heat	88	57	15	cut	192	168	6
hit	176	94	8	caught	85	81	12
seat	68	58	18	shut	82	47	17
sit	110	83	11	shot	138	78	9

*Frequency in COCA[76]

**Frequency in BNC[77]

***RoF = Rank of Frequency[78]

Source: present study.

[74] "The Corpus of Contemporary American English (COCA) is the largest freely-available corpus of English, and the only large and balanced corpus of American English. The corpus was created by Mark Davies, from Brigham Young University, and it is used by tens of thousands of users every month (linguists, teachers, translators, and other researchers)." (DAVIES, 2008).

[75] "The British National Corpus (BNC) is a 10-million-word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written. The latest edition is the BNC XML Edition, released in 2007." (THE BRITISH NATIONAL CORPUS, 2007)

[76] https://www.english-corpora.org/coca/

[77] https://www.english-corpora.org/bnc/

[78] The Rank of Frequency (RoF) is established considering only the words used in the present study. RoF was calculated by averaging both COCA and BNC frequencies, though those corpora have different amounts of words.

The target words were inserted in the carrier sentence “The word is ___ too”. The gap represents the target word. We chose to use carrier sentences in order to avoid providing the listeners with contextual information that might help identify the target word (DE LOS SANTOS, 2017; RAUBER, 2006). Additionally, it ensured that all sentences would have a similar prosodic pattern, thus allowing some control over that variable. Finally, we decided that the target word should be the second last in the sentence (cf. BARBOSA; MADUREIRA, 2015), in order to avoid the process of final consonant devoicing, which could affect vowel length (LADEFOGED; JOHNSON, 2015; ZIMMER; SILVEIRA; ALVES), as well as possible alterations in pitch. The last word should be started by a voiceless consonant, in order to avoid connected speech[79] and lengthening of vowel duration[80] (LADEFOGED, 2010; ZIMMER; SILVEIRA; ALVES, 2009).

3.4.3.1.1 Production samples from native speakers of BP

As previously mentioned, the six native speakers of BP recorded 72 carrier sentences each, 36 of which containing the target words (12 target words in 3 repetitions each). The other 36 contained distractor words (12 distractor words in 3 repetitions each). See Table Box 3.2 in section 3.4.3.1 for a list of words. Sentences were elicited with an MS PowerPoint presentation file (see section 3.4.3.2) and read out loud.

The selected participants were emailed the file (see section 3.4.3.2) after filling out the Self-assessed Proficiency questionnaire (see section 3.4.2) in the screening stage. The email also had the same recording instructions given in the invitation message. Each participant received a different MS PowerPoint file, for the carrier sentences were randomised for each person. Along with the presentation, the email had a supporting PDF file attached (see Appendices G and H). This PDF file explained how to use the native[81] voice recording app in the participants’ mobile phones. At the end of the presentation file (see section 3.4.3.2), participants were informed that they could email, upload to a cloud or instant message the recording file to the researcher. All participants chose to send the files using their WhatsApp app.

[79] Although this configuration avoided consonant-vowel resyllabifications in the learners’ productions, connected speech was, nonetheless, present, as the coda [t] segment of the target words was often unreleased and connected to the onset [t] segment in ‘too’.
[80] Since all target words had a final voiceless alveolar stop, the final word should also have a voiceless onset. In case of unreleased productions, a voiced onset might have allowed for vowel lengthening.
[81] Pre-installed, that is, participants were not required to download or install an app in order to partake in the present study.

Upon sending their audio files, participants were sent the link to fill out the Language History Questionnaire (see section 3.4.2).

3.4.3.1.2 Production samples from native speakers of Canadian English

In the same fashion as Brazilian participants, the two native speakers of Canadian English, one male and one female, recorded 72 carrier sentences each. Half of those sentences (36) contained the target words (12 target words in 3 repetitions each), and the other half (36) contained distractor words (12 distractor words in 3 repetitions each). Sentences read out loud were elicited with an MS PowerPoint presentation file (see section 3.4.3.2). The procedure was identical to that with Brazilian participants.

The productions obtained from the Canadian participants were used along with the Brazilian learners' productions in the perception task, and served as baseline stimuli. This methodological choice was made in order to evaluate if listeners were able to identify the vowels when their acoustic characteristics matched the native acoustic patterns, in other words, when listeners were asked to identify productions that ideally had no non-native accent/colouring (CRUZ, 2017).

As suggested in the literature on L2 speech (CRUZ, 2017; STRANGE; BOHN, 1998), non-native listeners might miss out on acoustic cues not because of the stimuli, but because of their own learning stage. Thus, having a baseline helped to identify, within the speaker-listener pair, when participants were failing to identify the words due to their own language development stage (listener's side) or due to production characteristics that hindered stimuli intelligibility (speaker's side). We also expected a likely scenario in which a combination of aspects from the speaker-listener pair would be responsible for any lack of success in communication.

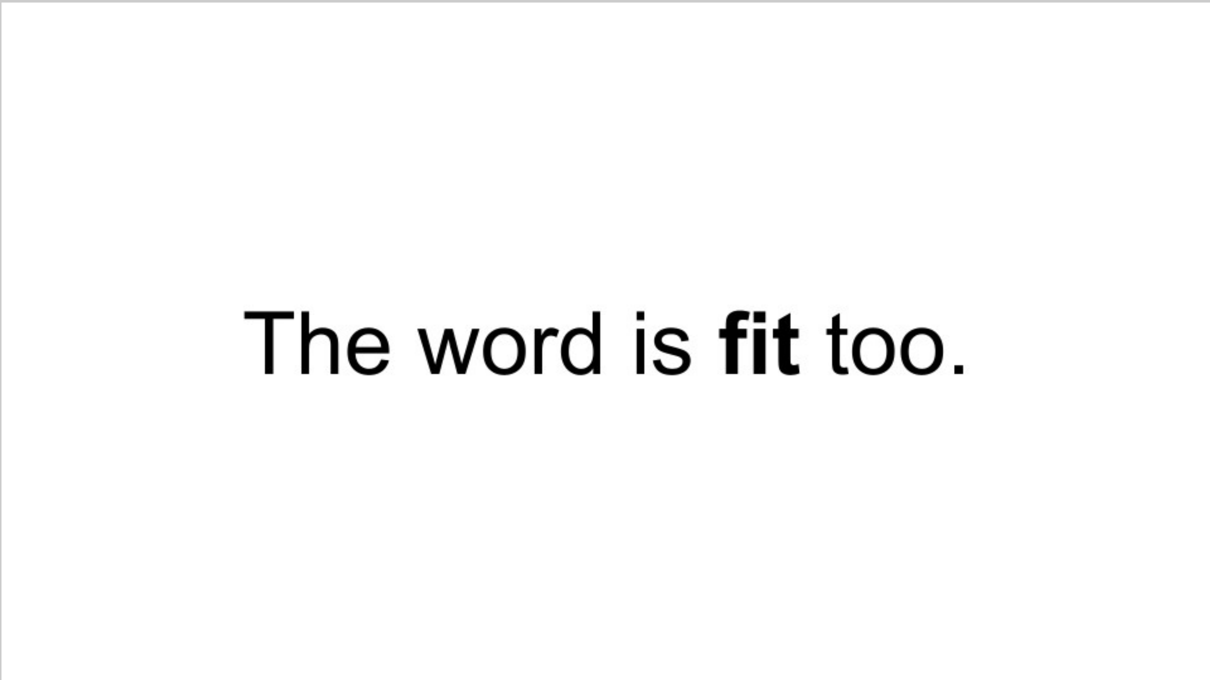
3.4.3.2 MS PowerPoint presentation file for the recording of stimuli by Brazilian and Canadian participants

An MS PowerPoint presentation file was used to elicit speech samples in the production task. Our aim was to standardise the way the carrier sentences containing the target words were displayed to Brazilian and Canadian participants.

In order to avoid distractions from the sentence that had to be spoken aloud for recording in the mobile device, the slide design chosen consisted of a white background. The only constant graphic element was the carrier sentence, which was centered horizontally and vertically on the page. The sentence that was repeated on all slides had a Regular font weight, while the target items were shown in a Bold font weight. The font family used was Arial, and the body of the text was set to 52pt. The ending of each carrier sentence was marked by a period.

Figure 3.1 shows a screenshot of a slide of the MS PowerPoint presentation file sent to participants in the production task. The target word in this example is 'fit', displayed in bold font-face.

Figure 3.1 – Screenshot of a slide of the MS PowerPoint presentation file sent to participants in the production task. The target word in this example is 'fit', displayed in bold font-face



Source: present study

As seen in Figure 3.1, one sentence was presented in each slide. In order to avoid a list effect on reading, there was a 5-second slide-change, with an automatic transition. Because the Google Workspace application does not allow for automatically timed transitions, presentation files were firstly made in Google Workspace, and then exported to Microsoft PowerPoint editing format. In that application, they were edited to contain the 5-second transitions. Finally, they were re-exported to a non-editable, view-only format, in order to be playable on the participants' desktop or laptop computers, regardless of their having the MS Office suite installed[82]. Participants were instructed twice (in the invitation and in the email with the

[82] Unlike the Google Workspace, the MS Office suite is a paid package, as far as editing is concerned. This is why we first attempted to use Google Workspace. Upon finding that the Google Workspace web-app did not allow us

attachment) that they should view the presentation file on a computer, given that mobile devices use different applications to open presentation files. Some of those apps do not allow for the automatic transition and/or allow for the slide to be manually forwarded. This would override the programmed 5-second-inter-stimulus interval.

The presentation file (Appendix I) was divided into four blocks. The first block was initiated by a screen explaining the recording procedure and offered a practice section next. In this section, the same carrier sentence was displayed, but the words were ‘home’ and ‘moon’ – ‘The word is moon too’. Those words maintain the CVC, monosyllabic structure of the target and distractor words. The carrier sentences in the practice section were always presented in the same order. Participants could initiate/retake the practice section as many times as they felt the need for. When they were ready, they pressed a button and moved on to the next block. Blocks 2, 3 and 4 of the presentation file corresponded to the three repetitions of each of the 24 carrier sentences. In those blocks, carrier sentences presented the 12 target words and the 12 distractor words.

A slide indicated the point at which participants should start the recording, at the beginning of block 2. Each one of the carrier sentences was presented in sequence. There was an indicative slide of a break time for resting between blocks 2 and 3. This slide was automatically set to last 60 seconds, but had the option to resume the recording before that, if the speaker felt they had rested enough. When the break was over, block 3 began. The same carrier sentences were displayed again, each on a different slide, in a different, randomised order. There was another indicative slide of a break time, configured as the first one, between blocks 3 and 4. Then, the same carrier sentences were displayed again (third repetition). Again, each sentence was shown on a different slide, and once more in a randomised, distinct order from that of both the previous blocks. A final slide thanked the speaker for their participation and indicated they could end the recording. That last slide displayed the options to send the audio file to the researcher – namely, messaging app (WhatsApp), cloud folder on Google Drive (shared only with each participant, upon filling out the TCLE) or email. As mentioned before, all participants chose to send the files via WhatsApp. The average recording time was 7 minutes and 54 seconds (SD: 0.048 seconds).

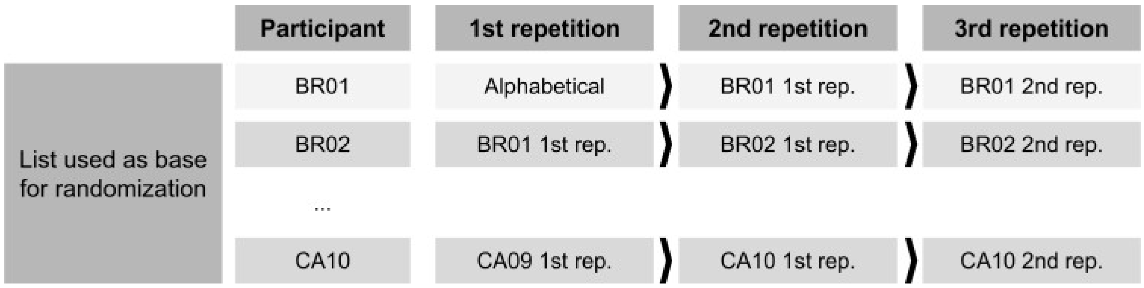
The carrier sentences used in the MS PowerPoint presentation files were randomised for each block, as well as for each of the participants, generating a total of 24 randomised lists with

---

to set up the slide transition as we needed it to be, we switched to the paid application by Microsoft. The MS app, in its turn, has a free, view-only file extension, which we used to allow for all participants to use the presentation file as we intended them to.

72 items each. The first randomisation was done from an alphabetical order of target words, generating the first set of sentences that was sent to Participant BR01. The second repetition was done from the order in the first repetition, and the third repetition was based on the second. The first repetition list of each participant was used to generate the randomisation of the first repetition for the next participant, with second and third repetitions being randomised based on her/his first list. Figure 3.2 illustrates the order and the process of randomisation of carrier sentences for speakers as we have just described them.

Figure 3.2 – Order and randomisation process of carrier sentences



Source: present study.

According to the order of randomisation illustrated in Figure 3.2, for instance, BR02's first list was randomised based on BR01's first list. BR02's second repetition list was based on BR02's first repetition list, and BR02's third repetition list was based on BR02's second repetition list.

3.4.4 Perception task

In this section, we will describe the steps we took to analyse the samples from the production task and select the subset of stimuli for the perception task. We will also describe the sequence of the task as it was presented to our listener participants (Argentinian and German) in the perception task.

3.4.4.1 Samples

As mentioned in section 3.4.3, Brazilian and Canadian participants in the production task recorded the carrier sentences in a single audio file each, resulting in a total of 8 audio files. Each file was recorded by each participant's native voice recording app on their mobile devices. Participants were asked twice (upon invitation and in the email with the attached presentation file) to turn off data on their mobile device, in order to avoid interference from notification sounds/vibration. They were also asked, if possible, to choose a quiet room in their quarantined space, preferably closing all windows and doors and keeping pets and children out.

The audio files we received from participants were segmented into individual carrier sentences, so that their spectrogram could be analysed individually. This segmentation was important as these segmented audios were to later constitute the stimuli of the perceptual task described in this section.

Each of the eight recorded files was segmented into 72 sentence-long files per speaker. Sentences were segmented from the first pulse with a constant waveform of the segment [ð] in "The (word…)" to the last distinguishable pulse in the segment [u] in "(... is) too". This segmentation was done with the Praat 6.1.14 software (BOERSMA; WEENINK, 2020). The 72 sentences per speaker comprised three repetitions of the 12 target words, along with three repetitions of the 12 distractor words. The analyses were annotated[83] so that possible subsequent consultations could be performed, should it be necessary.

All 576 sentences were analysed[84] regarding the vowel in the target/distractor words. Again, the first pulse with a regular waveform was used to mark the beginning of the vowel, the same criteria being used to mark the end of the vowel (GONÇALVES, 2014). Once the beginning and the end of the vowel were flagged, its absolute duration was measured. Then, the central point of the selected vowel – where the waveform tends to be steadier – was used to measure the first and second formants (F1 and F2) values. In some cases, in which productions were not steady throughout the vowel (eg. creaky voice quality), the steadier bit was selected, and the measure took place at the central point of this selected portion of the vowel – duration was measured considering the whole vowel length. Carrier sentence total duration was also measured, as relative duration was calculated as a percentage of the total sentence duration[85]. Relative duration allows us to account for individual differences in speech rate.

The F1 and the F2 values were measured using Praat's Linear Predictive Coding (LPC) script, adjusting maximum formant (default value of 5,500Hz, adjusted in steps of 500Hz) and

[83] The Praat software has an 'annotation to textgrid' function which allows for text comments and marked-down points on the audio file to be recorded in an auxiliary .TextGrid file. With both the audio and the .TextGrid file, it is possible to recall the exact point at which any marks were placed, as well as any comments regarding those marks.
[84] Carrier sentences containing distractor words were also analysed because of the subsequent normalisation process. As mentioned in section 3.4.4.1, Watt and Fabricius normalisation requires values that represent the corners of the vowel space – high front, high back and low central. Vowels [i] and [æ] were present in target words, but [u] and [ʊ] were taken from distractor items. We reiterate that our choice of distractor words already took into account the need of those vowels for the normalisation process.
[85] The formula used was: vowel duration / total duration of carrier sentence.

number of formant settings (default value of 5.0, adjusted in steps of 1.0) individually (SCHOORMANN; HEERINGA; PETERS, 2019). The default settings that were unaltered include a window length of 0.025s and a dynamic range of 30dB.

Due to the social isolation context, recordings were not made in a sound booth. The audio material was recorded in different environments and with distinct devices. Therefore, in some instances the measures were not considered to be reliable. In some others, due to background noise or interference from the microphone of the mobile phone, some samples were not in the same quality levels as the rest of the set. These scenarios led to the discarding of some tokens[86]. We highlight that we did not acoustically normalise audio amplitude in the chosen segmented recordings, in order to avoid artificial enhancements to have any effect on the perceptual task.

For statistical purposes, spectral F1 and F2 values were normalised by the Watt and Fabricius method using the Norm tools made available by the University of Oregon[87]. This was done in order to standardise individual speaker physiological effects on the measures (SCHOORMANN; HEERINGA; PETERS, 2019; LIMA JÚNIOR, 2017). Normalisation was done separately by each proficiency group (male and female beginner, intermediate, advanced and baseline levels). As mentioned, data from [u] and [ʊ] productions were also analysed because they were an integral part of the Watt and Fabricius normalisation method. Normalised data were also used for figures and graphics presented in this thesis and plotted with the Visible Vowels web-app, as well as with the ‘ggplot2’ and the ‘phonR’ packages for the RStudio application (RSTUDIO TEAM, 2021).

The measurements of F1 and F2, as well as those of absolute and relative duration, were also registered in a Google Workspace document. They were later exported to an .CSV (comma-separated value) format in order to be used in the statistical analysis. The statistics were run in the RStudio (*op. cit.*), as we will describe in section 3.5.

3.4.4.2 Stimuli

Of all samples read out loud by participants, we selected a subset to work as stimuli in the perception task. Thus, the productions in the first repetition series of the three were discarded to avoid prosodic and task effect issues that might have aroused from the participant

---

[86] This was expected, which is why we had participants record the samples in three repetitions. We selected the second and third repetitions as a default, but in those cases in which there was an audio issue, the first repetition was selected. This was the case of 15 out of 128 tokens selected as stimuli in the perception task.
[87] http://lingtools.uoregon.edu/norm/norm1.php

getting used to the experiment design (3.4.4.3). Additionally, from the 12 distractor words in each repetition, six were only intended for normalisation – namely the ones containing the vowels [u, ʊ] in the words ‘boot’, ‘book’, ‘foot’, ‘food’, ‘shoot’ and ‘should’ (see section 3.4.4.1)[88]. Thus, in our first proposed design, each speaker set of stimuli would entice six minimal pairs of each target vowel under analysis (that is, 12 words, in two repetitions each, meaning 24 carrier sentences). It would also include six distractors (in two repetitions each, or 12 carrier sentences). As we have eight speakers, this would total 288 carrier sentences.

Considering the experiment would be too long, it was decided to keep just one repetition of the distractors, reducing the total number of stimuli to 240. Upon piloting[89], due in major part to server delays – as the experiment was to take place on a website –, we noticed that the task would take over an hour to complete. We deemed this duration as unfit, for it might mean that potential participants would not or could not take part in/finish the experiment. It also might mean that their concentration would fade throughout the task. A second pilot, removing the distractors altogether (as seen in Alves, 2018), reduced the set of stimuli to 192 samples, with a length of participation varying from 35 to 45 minutes, depending on the listener’s server and personal internet connection speeds.

We still perceived that duration as too long, considering that participants also ought to fill out the language history questionnaire (estimated in about 15 minutes), besides completing the perception task. The questionnaire (see section 3.4.2) was extremely relevant, given that the mapping of the listeners’ developmental trajectories are essential to the comparability across Linguistic studies (SCHOLL; FINGER; FONTES, 2017; LUK; BIALYSTOCK; 2013; MARIAN; BLUMENFELD; KAUSHANSKAYA, 2007). Therefore, we decided to work with two (instead of three) minimal pairs for each contrasting vowels, namely [æ, ɛ] and [i, ɪ]. Of the 12 recorded samples, the ones with the most similar segmental pattern were selected, that is, the ones with voiceless onsets, namely: ‘pat’ – ‘pet’ and ‘sat’ – ‘set’ for vowels [æ, ɛ], and ‘feet’ – ‘fit’ and ‘seat’ – ‘sit’ for high front vowels [i, ɪ]. Following Gonçalves (2014) and Rauber (2006), using voiceless onsets also allowed for more trustworthy measures of vowel duration, since voiced contexts lead to a harder voice quality analysis. An additional factor was a possible VOT variable arising from the ‘bat’ – ‘pat’ / ‘bet’ – ‘pet’ quartet (ALVES; ZIMMER, 2015;

[88] The other six were chosen to appear in the PowerPoint presentation because they fitted other criteria used for the target words, like high frequency of use, CVC monosyllabic structure and were part of a minimal pair, among other factors.

[89] We conducted pilot sections with the researcher and her adviser, to account for design aspects. Some members of the researcher’s family also participated in some pilots intended to test technical aspects, as the experiment was run on a dedicated website fully developed by the researcher and for the purpose of the present study (Cf. section 3.4.4).

SCHWARTZHAUPT; ALVES; FONTES, 2015; ALVES; MOTTA, 2014), whose word-initial VOT patterns could have implications for the word-level intelligibility analysis intended in the study.

Thus, the final perception task comprised a total of 128 stimuli: two minimal pairs of each vowel under analysis (‘pat’ – ‘pet’ and ‘sat’ – ‘set’ for [æ-ɛ]; ‘feet’ – ‘fit’ and ‘seat’ – ‘sit’ for [i-ɪ]). This meant four carrier sentences for each pair, in two repetitions each, using samples from all eight speakers.

Stimuli were randomly presented to each participant, generating a total of 46 randomised lists, 28 for Argentinian listeners and 18 for German listeners. The task sequence will be explained in the following section.

3.4.4.3 Task sequence

As paid services that could provide a platform fit for the experiment design we wanted to apply were unaffordable, the author of the present thesis developed the data collection website herself. It was hosted in a private server and developed based on WordPress. Plugins such as Elementor (ELEMENTOR, 2021), FormVibes (WPVIBES, 2021) and Code Snippets (BUNGE; HEINZ, 2021) were used, along with HTML, CSS and PHP coding, to create the necessary technical conditions for the experiment to run as we had designed it. Conversely, some deficit in our knowledge of the programming languages in use also hindered the development of desired, albeit not required, features[90].

As the experiment was built on a webpage, each step was done at a different URL[91]. We will refer to them as ‘pages’ or ‘screens’. The initial page had a summary of the study’s information: (a) the expected length of participation; (b) the need to sign the TCLE (before starting); (c) the need to answer the Language History Questionnaire (after the perceptual task); and (d) recommendations to wear headphones and switch off browser auto-translation tools[92].

[90] Among those we can name (a) a countdown of how many tokens were already identified and how many were left; (b) a way to use cookies to save the participant’s progress, without asking the participant to sign up to the website; (c) a loading screen that would prevent the participant from trying to click the ‘play’ button before the audio was fully loaded, in case of slow internet connection.
[91] Uniform Resource Locator.
[92] A user can set their browser to automatically translate pages written in a given language. As the target words in the perceptual task must be chosen from a list, and that list contained words written in English, the browser auto-translation function must be turned off so the listener can pick the exact word they heard. In piloting, auto-translation tools showed options such as ‘bastão’ (BP, ‘stick’) ‘Schläger’ (DE, ‘racket’) and ‘raqueta’ (ES, ‘racket’) for ‘bat’. Moreover, a translation implies a single meaning – in this example, the object used to hit –, whereas some words might have more than one meaning – like the nocturnal animal associated with vampires in the case of ‘bat’. It may also have a priming effect, allowing the participant to code switch (BLANK, 2013). This, in turn, might have effects on participants’ responses.

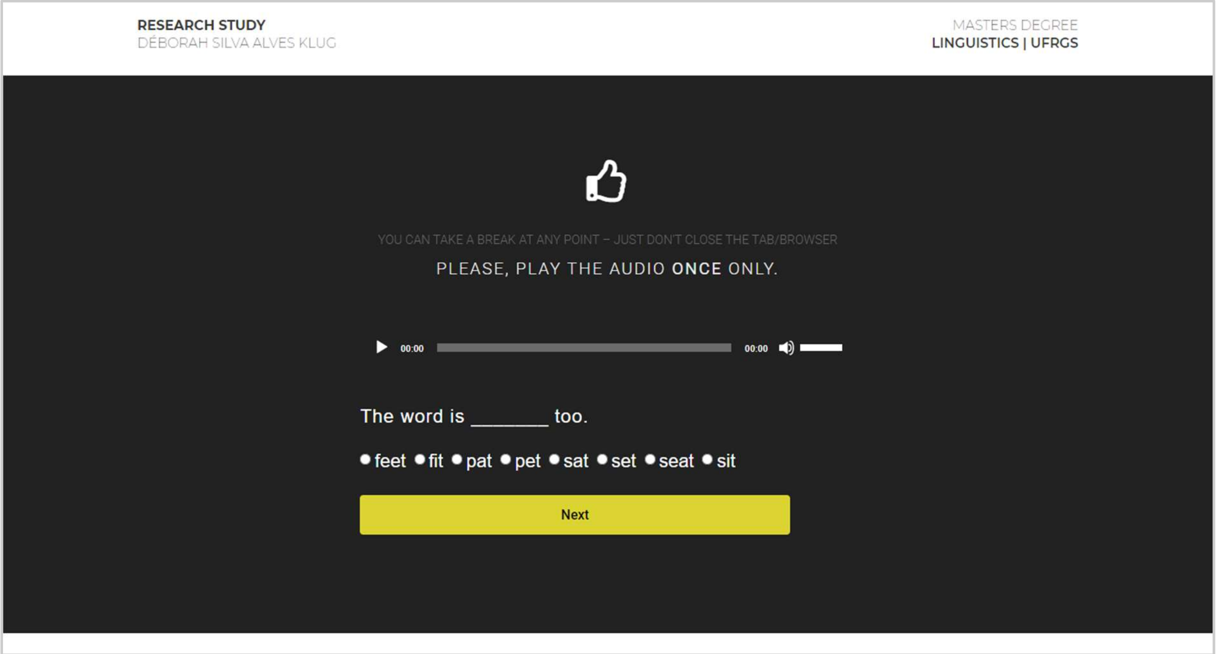
There was also a checkbox confirming the participants' advanced proficiency level in English (inclusion criteria informed in the invitation, see section 3.3.3). After checking the box and advancing to the next screen, participants had to click another checkbox, this time to electronically sign the Informed Consent Form in English (see section 3.4.1 and Appendix C). If the participant wished to receive an electronic copy of the signed Informed Consent Form, they could inform an email address in which to receive it. Participants were then directed to the screen explaining the task *per se* and informing that a practice section would precede the actual experiment.

Upon clicking the 'begin' button, the participant was directed to the practice session screens. This section consisted of four practice carrier sentences – three instances of "The word is home too" and one of "The word is moon too" –, randomly presented. The forced-choice options for selecting the target word in the practice section were 'home', 'rum', 'moon' and 'move'. After going through all four sentences, another page informed participants that they could practice some more if they deemed necessary, or they could start the task. If they wanted to practice some more, they were randomly redirected to one of the practice carriers, and had to identify all four words again, also in a random order. When they chose to begin the perception task, they were directed to a second introduction screen. This time, the introduction page informed that once the participant began the task, they could not close the browser window/tab before finishing the experiment, or their data would be lost. Upon clicking the 'start' button, the data collection began.

All stimuli were presented in an identical screen containing an informational section with: (a) a thumbs-up icon; (b) the message that the tab/browser could not be closed; (c) the information that the participant could take breaks at any point of their participation. There was also, in a larger, heavier font-face, the recommendation to play each audio once only. Below that standard information, the audio player was presented, followed by the eight options of the forced-choice task. Listeners had to choose one of those options as the target word they had listened to. Once the choice was made[93], they were to click the 'next' button. Figure 3.3 shows the desktop version of the perception task, before participants had played the audio file or selected an answer, with the elements we have just described.

[93] The initial instruction screen stated that if the participant was not sure about the word they heard, they should make a "sensible guess". Participants could only move forward to a new sentence once they had chosen an answer to the current one.

Figure 3.3 – Desktop version of the perception task, before participants had played the audio file or selected an answer



Source: present study.

While the participant’s answer was saved, a message reading “Saving. Please wait” was shown, due to the website's server and individual connection speeds. Then the next page loaded, with the next stimulus. There were a total of 128 carrier sentences to be heard and to have a target word identified by each listener.

Upon completing the perception task, the participant was directed to the online version of the Language History Questionnaire (Google Workspace form). After filling in the fields and clicking on the button labeled ‘send’, the participant was given the link to the screen thanking them for their participation. They were not required to click that link. The ‘thank you’ page reiterated the researcher’s contact email. It was also possible to share the research invitation through social media (Facebook and Twitter), messaging apps (Telegram and WhatsApp) or via email.

The researcher was informed via email each time a listener concluded their participation in the study. This allowed us to control the total number of participants already reached on each day. As soon as the data collection phase was completed, the content of the website was taken down. The main URL used for the task had its content changed to a message informing that the study had already completed the data collection phase, in case someone had saved the link posted with the invitation. All data pertaining to the experiment was then downloaded to a physical, local hard drive. We also proceeded with the exclusion of all audio files and responses from the page server.

3.5 DATA ANALYSIS PROCEDURES

The data collected were investigated by means of visual inspection and descriptive statistics, inferential statistics and exploratory analyses.

Descriptive analyses were performed based on tabulated values for vowel absolute and relative duration values, as well as first (F1) and second (F2) formants. Data were also described by visual inspection of plots. In Chapter 4, we describe productions based on proficiency level and individually by speaker, where we find it helpful. These descriptions provide some contextual background information for the discussions we will present in the sections that follow.

Inferential statistics were run in order to provide answers to our first and second research questions. Statistics were calculated with the RStudio application (RSTUDIO TEAM, 2021). Our mixed-effects logistic models were fitted from the .CSV files exported from the data we tabulated using the Google Workspace web-app. The functions we used are from the ‘lme4’ package for RStudio (*op. cit.*). The tables and figures used in Chapter 4, where we report the results of our model fittings, were produced by the ‘ggplot2’ and ‘phonR’ packages for RStudio (op. cit.), as well as by the Visible Vowels web-based app.

The exploratory analyses were performed in order to answer RQ3. These analyses were based on the tabulated values for acoustic cues in the stimuli, the inferential results and the visual inspection of the tables and figures provided in the sections referring to RQ1 and RQ2. Additional figures pertaining to RQ3 were also created using the ‘ggplot2’ and the ‘phonR’ packages for RStudio (*op. cit*.).

3.6 FINAL CONSIDERATIONS

In this chapter, we posed our main goal and our specific goals in the present research study. In order to investigate the word-level local intelligibility of the L2 English spoken by Brazilian learners, when those speakers are listened by other non-native learners of English, we posited three research questions and we explained the motivation behind them.

We have also detailed how we recruited speaker and listener participants, as well what inclusion criteria we adopted to select those participants. The instruments used in the production and in the perception tasks were also minutely described. Lastly, we explained how we intended to perform our data analysis.

We will now move on to presenting the results of those analyses and discussing their findings.

4 RESULTS AND DISCUSSION

In analysing L2 speech perception of accented speech and its impact on intelligibility, the present study investigated the role played by linguistic characteristics (F1, F2 and relative duration), as well as stimulus-related aspects (lexical item and speaker's proficiency) of the audio stimuli. The L1 of our listeners, organised in two groups (native Riverplate Spanish and native Central German), was also taken into account.

We will report results that correspond to each Research Question (see section 3.2), as each of them requires a different analysis to answer those questions. Before we report inferential results, we present a descriptive overview of the dataset. We start by describing the stimuli produced by the speakers, then provide some insight into the perception of these stimuli by non-native listeners. These initial, descriptive sections (4.1.1 and 4.2.1) aim to provide background information to the inferential results of mixed-effects logistic models we explore in the sections that follow. Models were fitted from the data collected in the perception task. Following that, when answering the third research question, we also present qualitative analyses.

Section 4.1.2 reports results regarding the listeners' accuracy in the identification of the words 'pat', 'pet', 'sat' and 'set'. Our discussion attempts to provide an answer to the first Research Question, i.e., "does a listener's L1 have an effect on local intelligibility?". Section 4.1.3 then goes on to report results that estimate which predictor variables might have an effect on the vowel identification of either [æ] or [ɛ] (Research Question 2), regardless of word or vowel identification accuracy. The final section (4.1.4) addresses Research Question 3, which enquires how the identification patterns in the perception task can shed light on the listeners' composite L1-L2 categories.

In sections 4.2.2 and 4.2.3, the same structure is then repeated to report and discuss results concerning the accuracy rates for target words 'feet', 'fit', 'seat' and 'sit' and the vowel identification rates of [i] and [ɪ]. Composite L1-L2 categories for high front tense and lax vowels are also discussed in the final section 4.2.4.

4.1 MINIMAL PAIRS 'PAT' – 'PET' AND 'SAT' – 'SET'

This section will begin by describing the characteristics of the stimuli used in the identification task. We will then carry out another descriptive analysis, this time of the identification results of the two vowels. Those act as background information to the inferential analysis and the qualitative discussion that attempts to provide answers to our Research Questions.

4.1.1 Descriptive Analyses

In this section, we will present descriptive analyses of our production and perception stimuli. These descriptions aim to provide some context for the description and discussion of the inferential statistics and exploratory analyses that will follow later.

4.1.1.1 Stimuli production

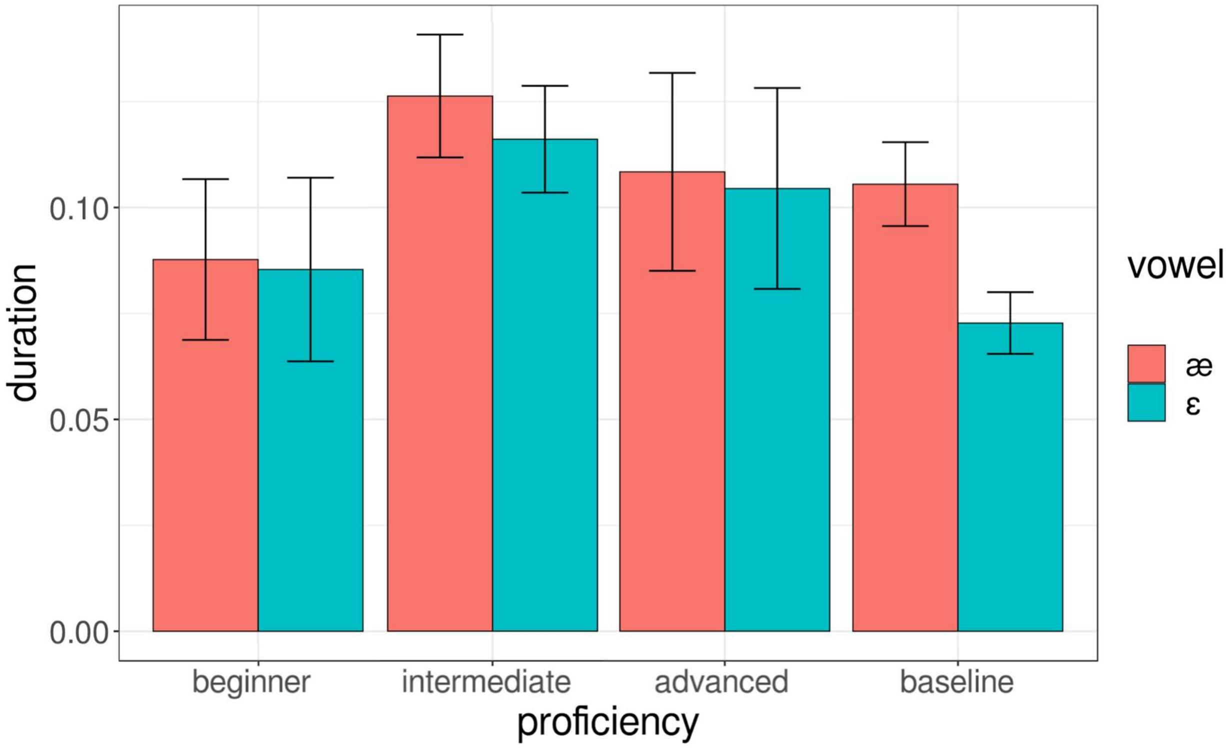
As mentioned before, in the native speech of English, the vowels [æ] and [ɛ] are distinctive, which can be seen in minimal pairs such as 'pat' – 'pet' and 'sat' – 'set'. Because of the acoustic similarity between both vowels (specially in North American English[94]), Brazilian learners of English tend to assimilate these two target vowels as the BP vowel [ɛ]. In line with the SLM, then, a dissimilation process would need to take place in order for a new [æ] category to be created in the learners' composite L1-L2 system.

In the present study, we collected speech samples from Brazilian learners of English in three different proficiency levels, aiming to see such a dissimilation process in different developmental stages (see section 3.3.1 for details). Consistent with Zimmer, Silveira and Alves (2009) and Nobre-Oliveira (2003), our exploratory analyses of the speakers' productions showed evidence of this process.

Throughout this section, we will analyse relative duration, F1 and F2. Relative duration will be observed first, as previous studies have shown that Brazilian learners tend to perceive temporal cues more prominently than formant frequency cues (ZIMMER; SILVEIRA; ALVES, 2009; RAUBER, 2006; NOBRE-OLIVEIRA, 2007). As a result, it could be expected that their productions would also show extrinsic duration[95] as a decisive cue for distinguishing between [æ] and [ɛ]. Figure 4.1 describes the relative durations produced in [æ] and [ɛ] vowels by participants in the production task.

[94] See section 2.4.4 for a description on the vowel systems of English.
[95] For a discussion on intrinsic and extrinsic durations, see section 2.4.

Figure 4.1 – Relative durations of the productions of [æ] and [ɛ], according to proficiency level



Source: present study.

Figure 4.1 compares the relative durations produced by the participants in each proficiency level. A visual analysis shows that the relative durations of [æ] (salmon) and [ɛ] (blue) are really close (with overlapping Standard Error bars) to each other for all proficiency levels of non-native speakers, and the small difference they do present could be taken as due to their intrinsic vowel duration.

Moreover, we can see that relative durations produced in [ɛ] tokens by intermediate and advanced level learners (with an average relative duration above 10%) are much longer than the ones produced by baseline native speakers (with an average relative duration of around 7%). Thus, despite a lack of distinctive extrinsic durations between [æ] and [ɛ], Brazilian learners' productions of [ɛ] seem to be closer to native [æ] productions in terms of relative duration – as baseline speakers produce [æ] tokens with an average relative duration of over 10%.

It is worth mentioning that vowels in the Porto-Alegrense variety of BP tend to be longer than native vowels in German[96] and in Riverplate Spanish. Though the three languages do not have the same vowel inventory, some comparisons are feasible. Maack (1949) reports values for stressed long [ɛ:] and short [ɛ] vowels in German with absolute durations averaged in

[96] As explained in section 2.4.3, we were unable to find references to relative durations of vowels for the Central variety of German. However, as the literature points to North German as the closest variety to Standard German, we have used measures of that variety as a means of comparison. For details on this discussion, see section 2.4.

122.50ms and 82.78ms, respectively. Comparatively, Pereyron (2017) reports monolingual speakers of Porto-Alegrense BP produce [ɛ] with an average absolute duration of 195.20ms. Pereyron (2017) also measured absolute and relative durations of the vowels produced by monolingual native speakers of Riverplate Spanish and by bilingual Spanish-English speakers. She reports average absolute durations of 174.9ms for bilinguals'[97] L2 [ɛ]. Given the 'L1 filter', the fact that Brazilian speakers produce longer vowel durations in their native language, therefore, could lead to the longer absolute durations in L2 productions we report. Finally, the duration of the [ɛ] produced by native speakers of North-American English, measured in previous studies and compiled in Pereyron (2017), vary from average 83ms to average 134ms. Our baseline native speakers of Canadian English produced [ɛ] with an average absolute duration of 116.91ms. Those measures are still shorter than the average shown by Porto-Alegrense BP speakers. Table 4.1 shows a comparison of vowel durations among speakers of the three language groups under analysis in the present study.

[97] Additionally, as [ɛ] is not a native category in Argentinian Spanish, we could compare native [e] absolute durations for monolingual Spanish and Brazilian Portuguese speakers. Pereyron (2017) reports that Argentinians produce [e] with an average absolute duration of 78.03ms, while Brazilian productions of [e] have an average of 175.13ms. We take this comparison as further evidence that native speakers of the Porto-Alegrense variety of BP produce longer vowel durations than native speakers of Riverplate Spanish do.

Table 4.1 – Average measures of absolute duration (ms) of native and L2 vowels produced by native speakers of Brazilian Portuguese, Riverplate Spanish, German, Canadian English and North-American English

Speaker and their languages	Vowel duration (in ms)					
	L2 [æ]	**L2 [ɛ]**	**L1 [ɛ]**	**L1 [e]**	**L1 [a][98]**	**L1 [æ]**
Monolingual Brazilian native speakers of the Porto-Alegrense variety of Portuguese (PEREYRON, 2017)	-	-	195.20	175.13	198.12	-
Monolingual Argentinian native speakers of the Riverplate variety of Spanish (PEREYRON, 2017)	-	-	-	78.03	86.74	-
Monolingual German native speakers of L1 Silesian/Bavarian variety of German[99] (MAACK, 1949)	-	-	82.78* / 122.5**	-* / 149.17**	80.61* / 154.07**	-
Plurilingual[100] Canadian native speakers of the L1 Ontario variety of English (present study)	-	-	116.91	-	-	182.06
Monolingual North-American native speakers of L1 English (compiled by PEREYRON, 2017)			83-134***	-	-	123-179***
Bi/Plurilingual Brazilian speakers of L1 Porto-Alegrense BP and L2 English (present study)	159.31	146.00	-	-	-	-
Bi/Plurilingual Argentinian speakers of L1 Riverplate ES and L2 English (PEREYRON, 2017)	108.97	174.9	-	74.87	84.47	-

*Short vowel[101]
**Long vowel
***Measures compiled from Rauber (2006) and Lima Junior (2013), apud Pereyron (2017).
Source: elaborated by the author (2021) based on the data from the present study; Pereyron (2017); Maack (1949).

[98] As neither BP nor German and Spanish have the low [æ] sound, we list values for their closest L1 low vowel, namely [a] in all three languages.
[99] See section 2.4.3 for details.
[100] We did not set monolingualism as an inclusion criteria for our native speakers of English, due to our difficulty in finding native speakers of English as participants. We understand that the additional languages our participants speak may have had an effect on their vowels systems. See section 5.2 for details.
[101] As mentioned in section 2.4.3, German has distinctive long/short vowel durations. (Cf. KÖNIG, 2004).

Thus, considering the data shown in Table 4.1, a remark ought to be made. As we have mentioned, Brazilian learners tend to perceive temporal cues more acutely than they do for formant frequencies, which would lead us to believe that a temporal distinction might appear more easily than spectral distinctions in the productions of those learners. Our dataset seems to suggest that this would not mean learning to produce [æ] with longer relative durations, though. Rather, it would suggest that Brazilians from the Porto Alegre area would need to learn to produce shorter [ɛ] durations. We could hypothesise that this would entice a system rearrangement to allow for the composite L1-L2 [ɛ] category to be shorter than a BP monolingual's L1 [ɛ] category is expected to be.

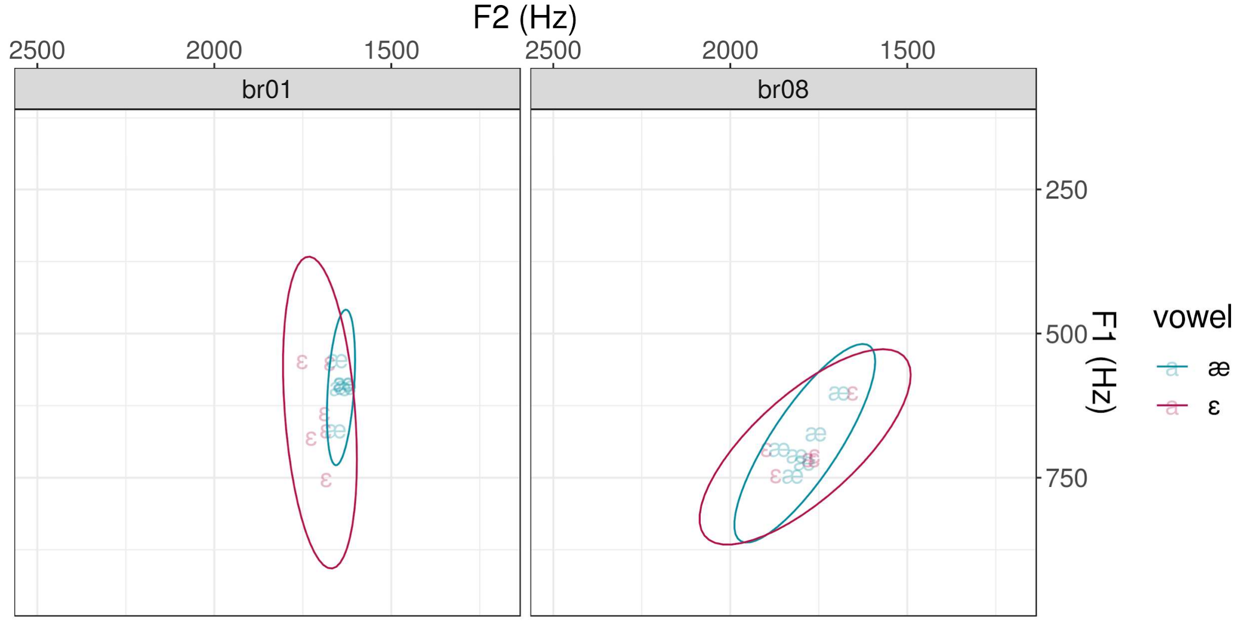
Comparing beginner-level relative durations to those of more proficient learners shown in Figure 4.1, we could further hypothesise that there is ongoing development in regard to duration for L2 [æ] and [ɛ]. Assuming beginners reflect a more L1-like pattern of relative duration for [ɛ] – as there is a native Brazilian Portuguese (BP) [ɛ] category –, we could consider longer relative durations produced by intermediate and advanced learners as a developmental stage. That is, they could have lengthened the relative durations of both vowels, as an attempt to lengthen [æ] productions. We could take this as a first step, which would be followed by a process of 'keeping' [æ] relative durations longer, and 'retracting' [ɛ] durations to the shorter average it has in BP (portrayed by beginners in Figure 4.1). Nonetheless, even if such a trajectory does take place, Figure 4.1 still suggests that the relative duration of the [ɛ] tokens produced by Brazilian learners would still be longer than those of native speakers of English, if we take the beginners' productions as a parameter for native BP [ɛ] length.

Having analysed vowel durations, it is important that we now look at formant frequencies produced by the speakers in our study. As we have mentioned in section 2.2, perception is a process that takes more than one cue into consideration. Even though one or more of the cues present in the speech signal might be weighed as more informative, in a Complex, Dynamic approach we ought to remember that the whole is not just the sum of its parts. In other words, a given cue that is not as relevant as another might still shift perception towards one or another identification, within the interaction of cues.

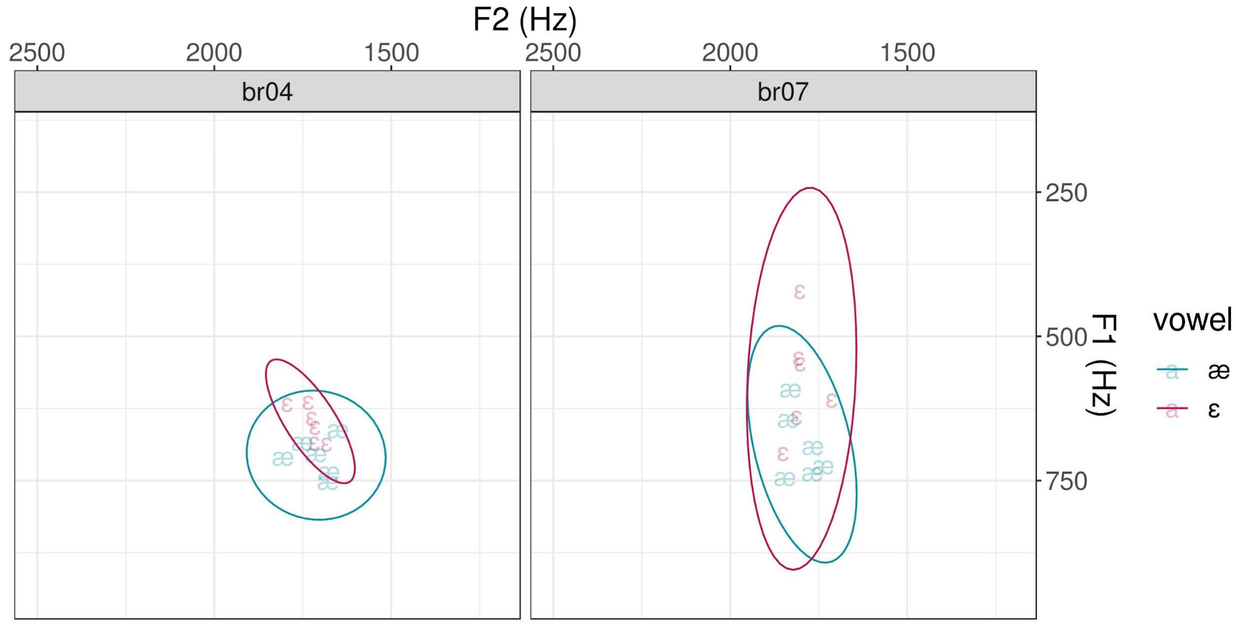
Figure 4.2 compares native speakers of English (baseline) and each Brazilian learners' productions of [æ] (blue) and [ɛ] (salmon), in terms of F1 and F2.

Figure 4.2 – Individual F1 x F2 plots of the vowels [æ] and [ɛ], according to the participants' proficiency level

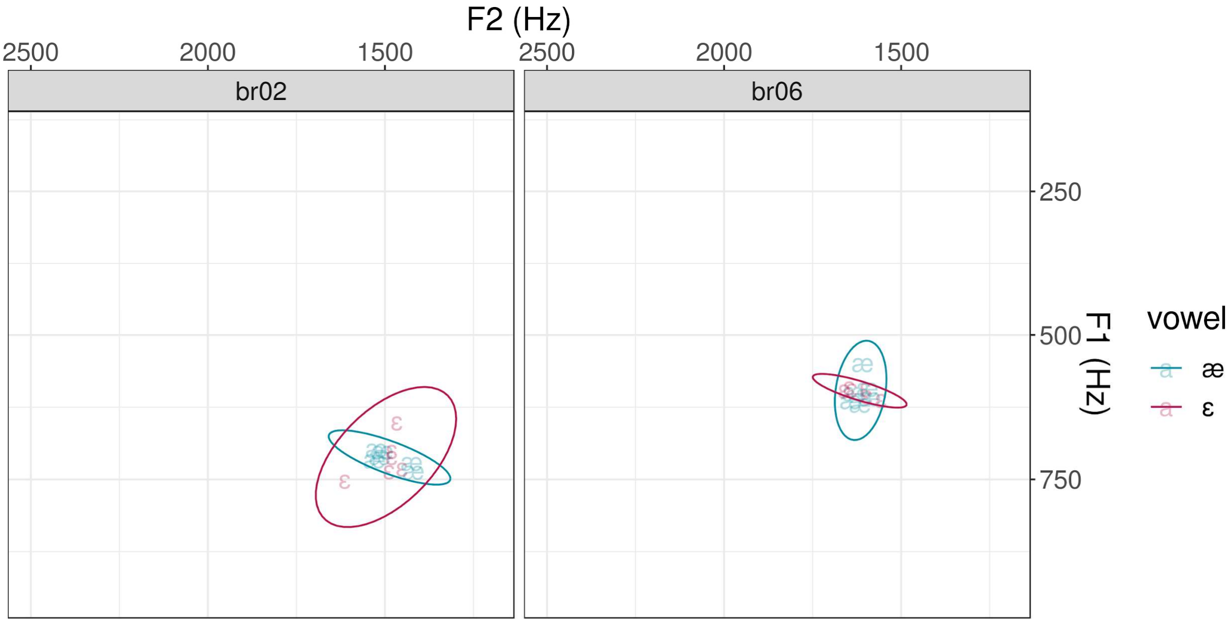
(a) beginner



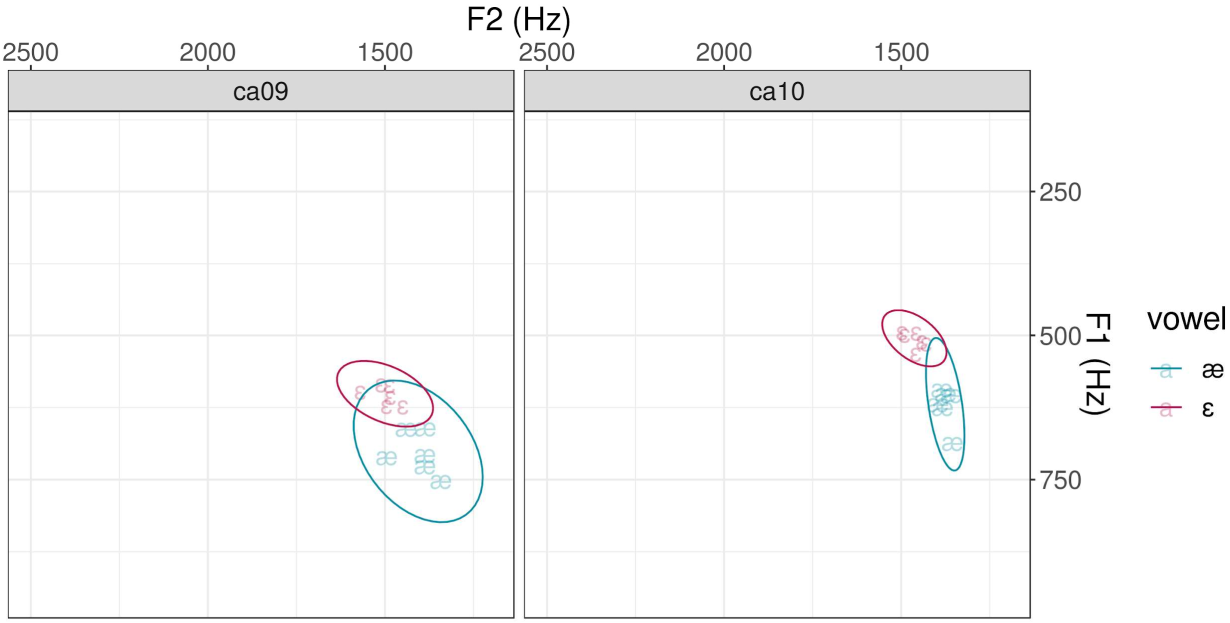
(b) intermediate



(d) advanced



(d) baseline (native speakers)



Source: present study.

As can be seen in Figure 4.2, baseline productions occupy distinct areas in the phonetic space, with just a small portion of partial overlap for [æ] and [ɛ][102]. Brazilian productions, however, show dispersion areas that overlap completely or partially, depending on the learner's proficiency level. It is this degree of partial overlapping that could be taken as an effect of an ongoing process of system rearrangement, as reported by Pereyron (2017). The author describes

[102] Labov, Ash and Boberg (2006) observe that the Canadian dialect is characterised by the Canadian Shift, "a downward and backwards movement of the short vowels /e, æ, o/, which is triggered by the low back merger". (p. 132). The authors also report a "three-way merger, in which *marry* is identical to *Mary* and *merry*" (p. 220).

that shifts in the vowel space could be an effect of a system rearrangement that is attempting to 'make room' for a new category.

> [T]he systems of a multilingual speaker are located within the same acoustic space. [Thus], the insertion of new vowel segments (those of the L2) should contribute to [the systems'] rearrangement that occurs during the process of acquiring a foreign language.[103] (*op. cit.*, p. 115)

Beginner learners have completely overlapping areas, though [æ] seems to occupy a smaller space. Productions also vary between participants BR01 and BR08[104] in terms of how restricted vowel spaces are. BR01 shows more variation in terms of F1 than F2, whilst BR08 varies both to a similar degree.

Compared to the beginners, the intermediate learners seem to have already started producing some distinction between the two vowels. Again, dispersions in the F1 and F2 dimensions vary differently from one speaker to the other. BR04 seems to have lower F1 values for [ɛ], which is consistent with the literature on native productions (LADEFOGED, 2010; YAVAS, 2011). However, her F2 values are only slightly more restricted (less dispersed) for [ɛ] than for [æ]. BR07 shows an almost complete overlap in the F2 dimension. Nonetheless, his F1 areas for [ɛ] and [æ] are more prominently different than they are for his female counterpart. As reported by Pereyron (2017), an enlargement of the dispersion area occupied by a vowel can be part of a category formation process. Within a Complex, Dynamic view, this can be seen as a destabilisation stage of the system that occurs before it reaches a new attractor state. In other words, enlarging the area of [ɛ] towards lower F1 values could be a first step to raising the entire category upwards, leaving the higher F1 area to be occupied, further in the development process, only by the [æ] vowel. Likewise, as we look at BR04's F2 values, enlarging the [æ] F2 values could be seen as part of the process of 'pushing' the category further back, leaving the more anterior space to the [ɛ] vowel.

Our advanced speakers seem to portray this reduction of the area occupied by a vowel, as we have just hypothesised. BR06 has small areas for both vowels, though they still show some partial overlapping. His system appears to have been raised entirely, which could be

[103] "Considerando que os sistemas do falante multilíngue estão localizados no mesmo espaço acústico, a inserção de novos segmentos vocálicos (da L2) deve contribuir para este movimento que ocorre no processo de aquisição de uma língua estrangeira (LE)."

[104] Participants were identified by code in order to maintain their anonymity. Number codes were assigned in the same order participants took part in the study, which is why they do not follow a sequential order across proficiency levels. Participants' codes also contain a letter section, referring to their L1. Hence, BR refers to Brazilian participants and CA to Canadian participants in the production task, whereas AR refers to Argentinian participants and DE to German participants in the perception task.

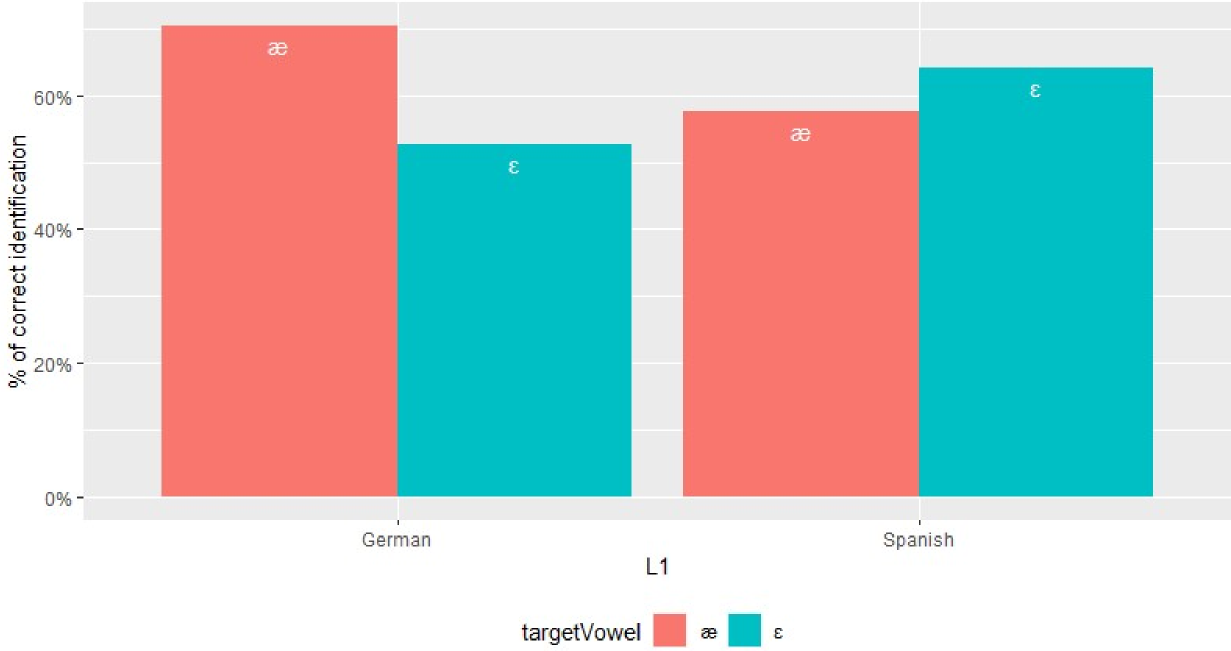
interpreted as a result of raising the [ɛ] vowel, and previous to a possible lowering of the [æ] category. BR02, in her turn, seems to be moving to the opposite direction: though she has not yet reduced the [ɛ] dispersion as much, she seems to have already lowered the acoustic area occupied by [æ].

4.1.1.2 Stimuli perception

The perception of the stimuli is descriptively analysed in this section, as a way of introducing some seminal information to the inferential analysis that attempts to answer the three Research Questions in the following section. We reiterate that the present study analyses both word-level accuracy rates (RQ1) and vowel-level identification, regardless of accuracy (RQ2). We also take a closer look at the role played by acoustic cues in both scenarios (RQ3, see section 3.2).

Firstly, as for the identification of the stimuli, it is worth noting that Argentinian and German participants showed different patterns of perception and in their accuracy rates. For instance, as shown in Figure 4.3, the native speakers of German correctly identified words with [æ] better than the native speakers of Spanish. However, the Argentinian listeners showed the opposite tendency, having higher accuracy rates in the identification of words with [ɛ]. It also seems that Germans have a harder time than Argentinians when trying to correctly identify words with [ɛ], in comparison to those with [æ]. Figure 4.3 shows the proportion of accurate identifications of words with [æ] and [ɛ], according to the listener's L1 groups.

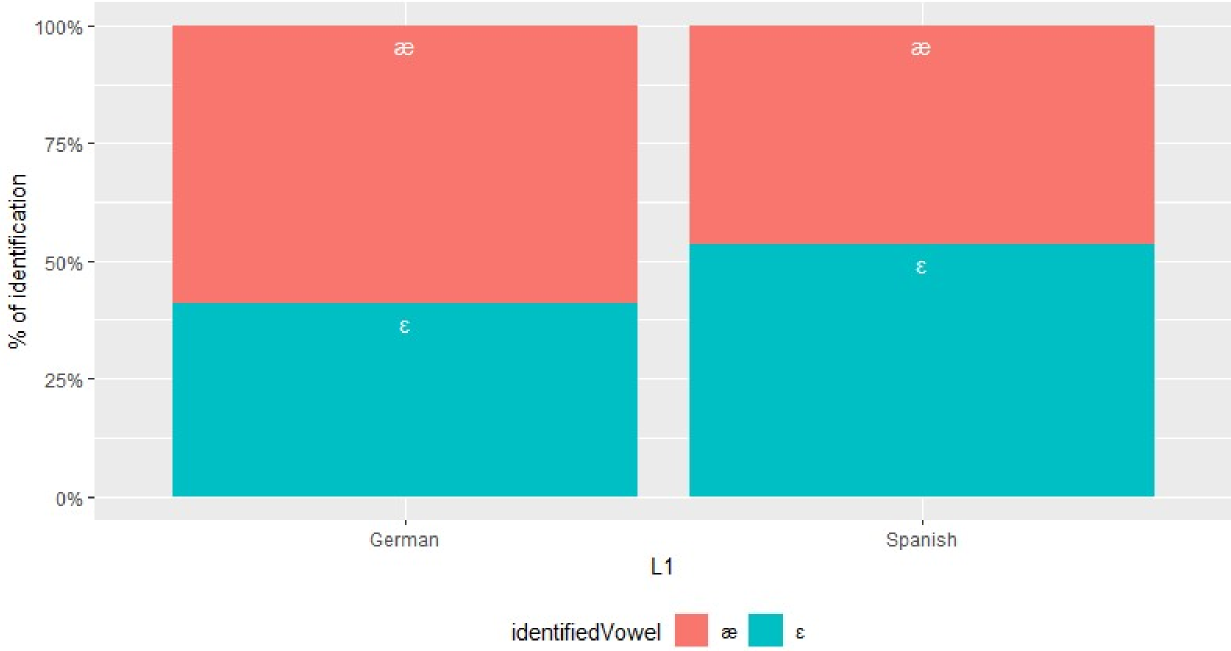
Figure 4.3 – Proportion of accurate identifications of words with [æ] and [ɛ], according to the listener's L1 groups



Source: present study.

In the same fashion, regardless of the target stimuli, Germans also identified more tokens as having the vowel [æ] (salmon) than [ɛ] (blue), whereas Argentinians identified more tokens as having [ɛ] than [æ]. We reiterate that the two members of each minimal pair had the same number of tokens in the stimuli – that is, the same number (32) of words with [æ] and the same number (32) of words with [ɛ]. Figure 4.4 shows the identification patterns for both L1 groups, regardless of their accuracy in relation to the target stimulus.

Figure 4.4 – Proportion of vowel [æ] and [ɛ] identifications, regardless of accuracy, by L1 listener group



Source: present study.

Exploratorily, it could seem that higher accuracy levels relate to a general higher number of identifications of a given vowel. In other words, as German listeners identify a greater number of tokens as [æ] (salmon), they also accurately identify words with [æ] more often. The same goes for Argentinians, who congruently identify more tokens as [ɛ] (blue) and have better accuracy rates for words with the [ɛ] vowel. As we will see in section 4.1.2, our mixed-models showed that the listener's L1 is a significant predictor of word identification accuracy, as well as of vowel identification in general. We move now to the inferential analyses performed with mixed-effects logistic models.

4.1.2 Identification accuracy of words with [æ] and with [ɛ] (RQ1)

The first Research Question the present study aims to answer is: does the L1 have an effect on L2 word intelligibility? As mentioned in section 2.2, the SLM predicts that an 'L1 filter' will play a role in the development of additional languages. That is to say that characteristics of the L1 system will somewhat shape the perception and/or production of L2 speech, even when they would not make sense or would not match characteristics of a native speaker of that language. This 'filter' is what motivated us to work with two groups of listeners with different L1 backgrounds, as explained in section 2.2. Thus, we ask if this prediction would hold true for our dataset and our speaker-listener pairs.

In an attempt to answer this question, a mixed-effects logistic model was fitted to our dataset. The model tested an interaction between L1 and target vowel as predictor variables. L1 was a categorical variable with two levels, 'Spanish' and 'German'. Target vowel was also a categorical variable, with two levels: vowel [æ] grouping ('pat' and 'sat' target words), and vowel [ɛ] grouping ('pet' and 'set' target words). The four lexical items 'pat', 'sat', 'pet' and 'set' are calculated as random intercept effects. The response variable was the accurate identification of the target word. As this question relates to local intelligibility (see section 2.3), we highlight that accuracy rates (response variable) were measured as a complete match between target word and listener's choice. Thus, a 'pat' token was only considered as correctly identified if the listener's identification was 'pat'. If a listener identified a 'pat' token as 'sat', the answer was considered incorrect, despite the fact that the two words share the same nuclei vowel.

We decided to operationalise accuracy this way because, considering the previous example, though the vowel was correctly identified, the onset consonant was not. This was taken as meaning that there was a level of unintelligibility in communication. If we take a sentence like 'He sat me down', the exemplified lack of intelligibility of the onset consonant leads to a misinterpretation of the speaker's intention. That is, 'seating someone down' (making them rest on a seat) is semantically different from 'patting someone down' (frisking them). In other words, our local intelligibility (response variable) referred to in this question relates to word intelligibility, not segment/vowel intelligibility. As we have mentioned in section 2.3, Munro and Derwing (2015) define 'local intelligibility' as encompassing both segment level intelligibility and word level intelligibility. We reinforce, as discussed in section 2.3, that our choice to study word intelligibility is based on two factors: on the one hand, avoiding the use of technical language in labelling '[segments]' as opposed to 'words' (layman's notation); on the other hand, the possibility to allow for a semantic level of analysis to be present, as the Complex, Dynamic view we adopt assumes that all subsystems of a system are in constant interaction.

Taking correct identification as a response variable for the interaction between the predictor variables 'L1' and 'target vowel', the fitted model shows that both the L1 and target vowel are significant in predicting the level of word identification accuracy by non-native listeners of English. Both predictor variables, as well as their interaction, affect the probability of a correct identification of target words with vowels [æ] and [ɛ]. Estimates are provided in

Table 4.2, in log-odds[105] – speaker and target word are calculated as random intercept effects[106]. The intercept shows correct identifications (response variable) by German (DE) listeners identifying target words with [æ].

Table 4.2 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of interacting L1 and target vowel on accurate identification rates of [æ] and [ɛ]

Predictors	Estimates	std. Error	z value	Pr(>|z|)
Intercept ([æ] – DE)	0.978	0.295	3.319	**<0.001**
target vowel [ɛ]	–0.829	0.226	–3.668	**<0.001**
L1 ES	–0.603	0.118	–5.121	**<0.001**
target vowel [ɛ] : L1 ES	1.120	0.164	6.835	**<0.001**
Observations	2,944			AIC = 3,682.5
Marginal R2 / Conditional R2	0.021 / 0.154			

Model: glmer (accuracy ~ targetVowel + L1 + targetVowel * L1
+ (1 | speaker) + (1 | targetWord), family = binomial, data = dados)
Intercept: target words with [æ] by German listeners
Response variable: correct identification
Source: RStudio 1.4.1103 (2021).

All estimates refer to the probability of a correct identification of the target word intended by the speaker, as opposed to incorrect identifications. As shown in Table 4.2, the direction of the effects of L1 and target vowel is different for each group of listener, considering each minimal pair. German listeners of L2 English tend to show better accuracy rates when identifying target words ‘pat’ and ‘sat’ ($\beta_0 = 0.978$, $p = <0.001$) than their respective minimal pairs ‘pet’ and ‘set’. L1 is a significant predictor, as Argentinians show a tendency to incorrectly identify target words with [æ] ($\beta = –0.603$, $p = <0.001$). On the other hand, Argentinian listeners of L2 English tend to show better accuracy rates (in comparison to Germans) identifying target words ‘pet’ and ‘set’ ($\beta = 1.120$, $p = <0.001$) for the predictor L1 in interaction with target vowel. This interaction can be clearly seen in Figure 4.3, in which we see that Germans outrank Argentinians when accurately identifying [æ], but Argentinians outrank Germans when accurately identifying [ɛ].

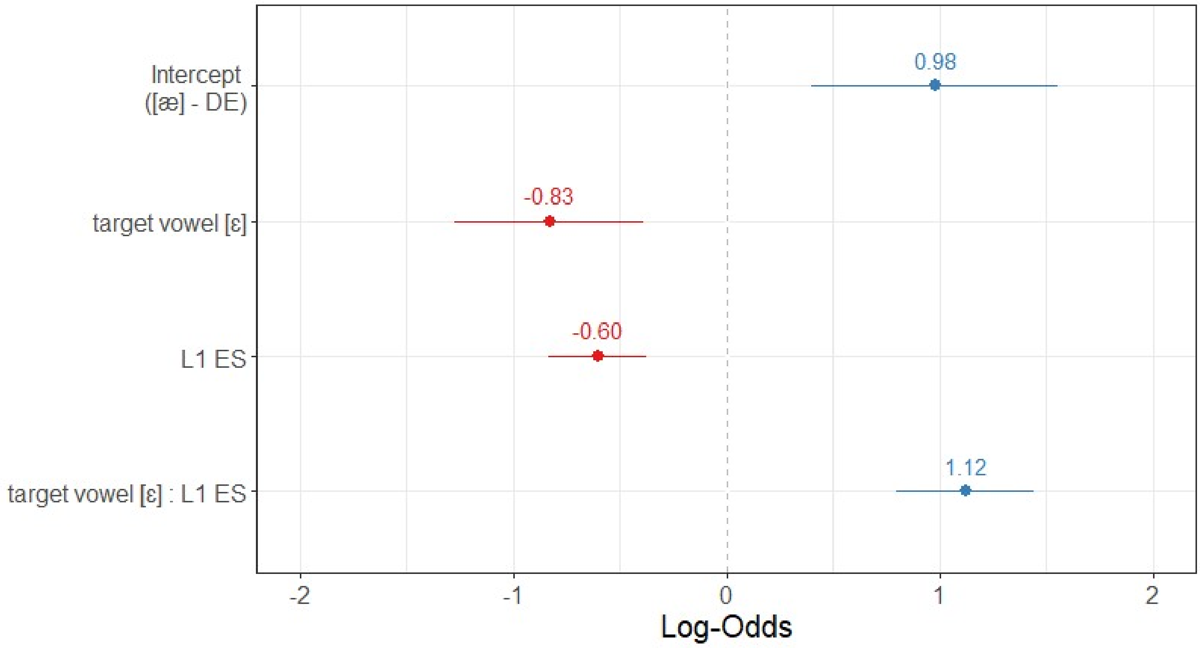
Figure 4.5 shows plotted estimates, and Figure 4.6 shows both groups’ predicted performances based on the mixed-effects logistic model. Blue lines indicate tendencies towards

[105] As Winter (2020) points out, “[a] good thing to remember about log odds [*sic*] is that a log odds value of 0 corresponds to a probability of 0.5, and that positive log odds correspond to p > 0.5 and negative log odds correspond to p < 0.5.” (*op. cit.*, p. 203). Figure 4.5 signals positive log-odds with the blue colour, whereas negative log-odds are printed in red.

[106] We attempted to also use ‘listener’, our other repeated measure in the study, as a random intercept effect. However, models with the three random variables did not converge or were singular fits.

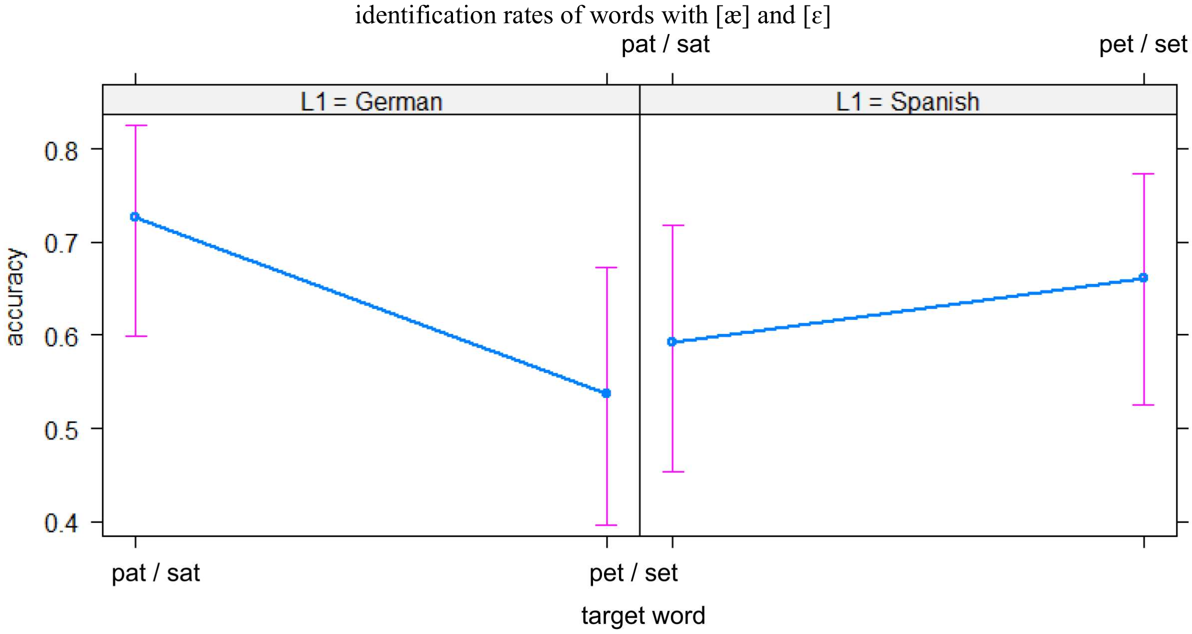
accurate identifications (positive estimates), whereas red lines do not favour correct identifications (negative values).

Figure 4.5 – Log-odd estimates of the mixed-effects logistic model for effects of interacting L1 and target vowel on accurate identification rates of [æ] and [ɛ]



Source: RStudio 1.4.1103 (2021).

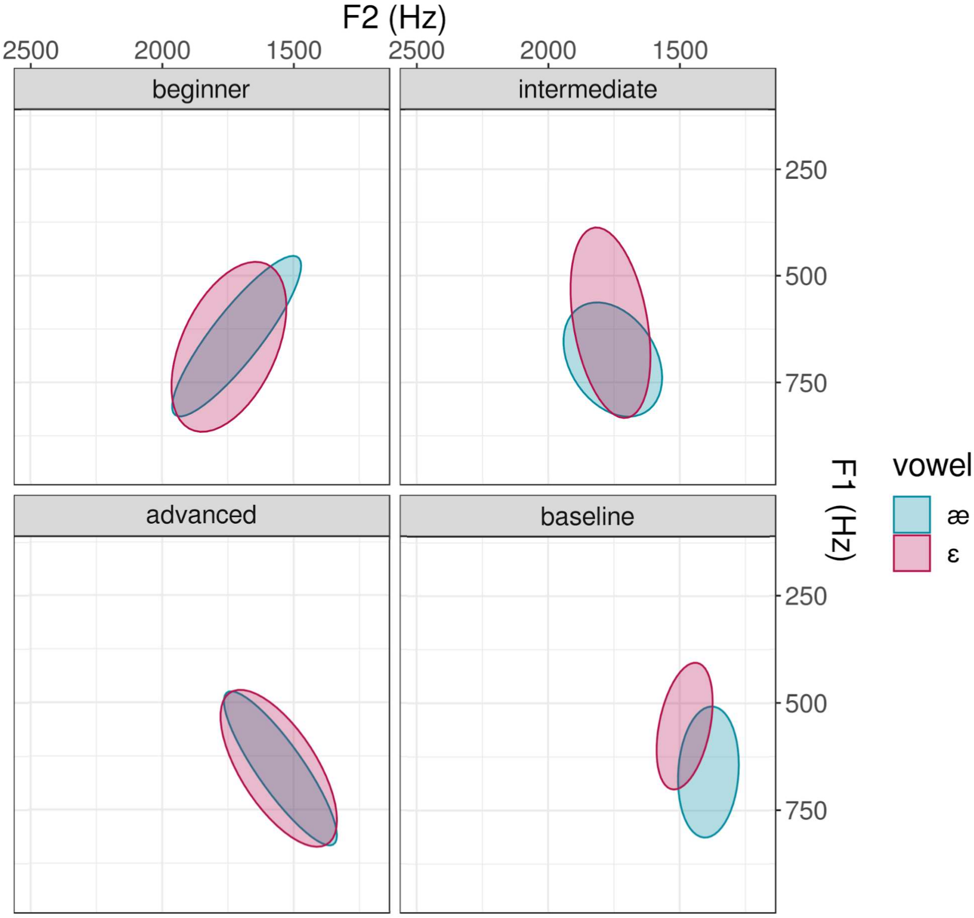
Figure 4.6 – Mixed-effects logistic model's predicted effects of interacting L1 and target vowel on accurate identification rates of words with [æ] and [ɛ]



Source: RStudio 1.4.1103 (2021).

As already hinted by the descriptive analysis, the Germans' accuracy rates are much lower for target words with vowel [ɛ] ('pet' and 'set' tokens). Their accuracy is even lower than that shown by Argentinians in the identification of words with [æ] ('pat' and 'sat' tokens), which are the hardest of the two sets or tokens (with [æ] and with [ɛ]) for the latter group. Figure 4.6 also portrays an interaction between L1 and target vowel, as different target vowels lead to opposite tendencies in accuracy rates for each L1 group. In order to explain the inferential results described here, we once again discuss the vowel production data explained in the previous sections. Figure 4.7 shows vowel dispersion of [æ] and [ɛ] productions in English grouped by proficiency level, in the same fashion. Figure 4.2 (section 4.1.1.1) shows the same dispersions by participant (individually) in each proficiency level.

Figure 4.7 – Vowel dispersion of [æ] and [ɛ] productions in English according to speaker's proficiency level



Source: present study.

As seen in Figure 4.7, the L1 and target vowel effects shown by the model might be tentatively explained by the fact that Brazilian learners are not fully producing a distinction between vowels [ɛ] and [æ] in those tokens. As the native BP vowel system has an [ɛ] category, the SLM predicts that [æ] would be assimilated into that native category of [ɛ]. Unlike the

Canadian native speakers[107] of English (bottom right), the Brazilian learners in all proficiency levels produce [æ] and [ɛ] tokens from totally to partially overlapping areas of their common phonetic space, as seen in section 4.1.1. That is consistent with the literature on the relative difficulty that Brazilian learners have in producing [æ] tokens (RAUBER, 2006; NOBRE-OLIVEIRA, 2003, 2007). Since the distinctions between the two target vowels are not being fully produced, it is hard to imagine that they could be highly intelligible.

When taken together with the fact that an accurate identification rate goes in opposite ways for Argentinian and German listeners, the lack of distinction in production could mean that one or more acoustic cues are being perceived differently by each group of listeners. In other words, as [æ] and [ɛ] tokens are produced in totally/partially overlapping areas, local intelligibility of these tokens is also dependent on listeners' perception, reinforcing the importance of analysing perception and production as a conjoined effort of the speaker-listener pair.

4.1.3 Effects of predictor variables on the identification of [æ] and [ɛ] (RQ2)

As seen in the previous section, an accurate identification of a target word implies accurate identifications of onset and coda consonants and nuclei vowel, considering our dataset of CVC monosyllables. That is to say, in a way, that vowel identification is but a part of that process. Thus, we now turn our analysis to vowel identification, regardless of identification accuracy. Our second Research Question aimed to look at which predictor variables could explain a token being identified as having [æ] or [ɛ], as this segment identification will, in turn, be used in a broader process of word identification.

In order to answer the question "which predictor variables might have an effect on vowel identification?", we built a mixed-effects logistic model. We wanted to check each predictor variable and each possible interaction between those variables. Thus, our first model had, as predictor variables: listener's L1, speaker's proficiency level, stimulus's target vowel, F1, F2 and relative vowel duration. All possible interactions (signaled by the '*' character) amongst those predictors were also included at first: L1 * proficiency level, L1 * target vowel, L1 * F1, L1 * F2, L1 * relative duration; proficiency level * target vowel, proficiency level * F1, proficiency level * F2, proficiency level * relative duration; target vowel * F1, target vowel *F2, target vowel * relative duration; F1 * F2, F1 * relative duration; and F2 * relative duration.

[107] As mentioned in section 4.4.4, native speakers of English produce [æ] with an average of 820Hz in F1 and 1,808Hz in F2, whereas [ɛ] is produced with 704Hz in F1 and 1,910Hz in F2.

Predictors that did not yield significance were excluded from the model, since we only had 2,944 identification tokens for the [æ] – [ɛ] pair. Small datasets do not allow for models with too many predictors, as there is little data to work with when calculating estimates (LEVSHINA, 2015; GRIES, 2013). Thus, the second fit we attempted had fewer predictor variables, from which non-significant ones were also excluded to run a third possible fit. This was done consecutively and resulted in all interactions being non-significant. Even then, though, some predictor variables showed high collinearity rates and were also excluded. As will be seen further on, F1 did not yield significance, but we maintained its presence in the model because of a linguistic point of view. That is, F1 is a heavily weighted cue in native English perception (ESCUDERO, 2009; HOLT; LOTTO, 2006). Also, our plots define the acoustic space in terms of F1 (height) and F2 (frontness).

Thus, the model being reported in the present thesis has L1, target vowel, F1 and F2 as predictor variables. As for random intercept effects, we were able to include all three repeated measures, namely: listeners, speakers and lexical items. The response variable is the identification of a token as having the vowel [ɛ]. This does not take accuracy into account, nor does it distinguish between target words 'pet' and 'set'. The analysis of vowel identification here turns the spotlight to the listener and how they are processing the stimulus. Therefore, the intercept estimates the log-odds of a vowel being identified as [ɛ] when a German (DE) participant listens to a token that was produced from a target vowel [æ].

Our fitted mixed-effects logistic model shows that native language, target vowel and F2[108] are all significant predictors of vowel identification. By vowel we mean a binary, categorical variable [æ] or [ɛ], the first one comprising the identification of tokens as both 'pat' and 'sat', and the latter, tokens identified as 'pet' and 'set' – as mentioned, target words (lexical items) are only used for random intercept effects. Correct matches[109] between target vowel and identified vowel are not accounted for either. Table 4.3 shows the estimates in log-odds.

[108] F1 and F2 values were rescaled to log, as suggested by the RStudio modelling script. This suggestion was made because formant frequency measures had indexes that were too different to compare.

[109] A total of 27 tokens identified as [æ] or [ɛ] from target words with [i] or [ɪ] were excluded from the complete dataset. This was done because of the disproportion of cases (0,92%). A lack of enough variability could have hurt the model's fitting.

Table 4.3 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of L1, target vowels, F1 and F2 on vowel identification (regardless of accuracy) of response variable [ɛ]

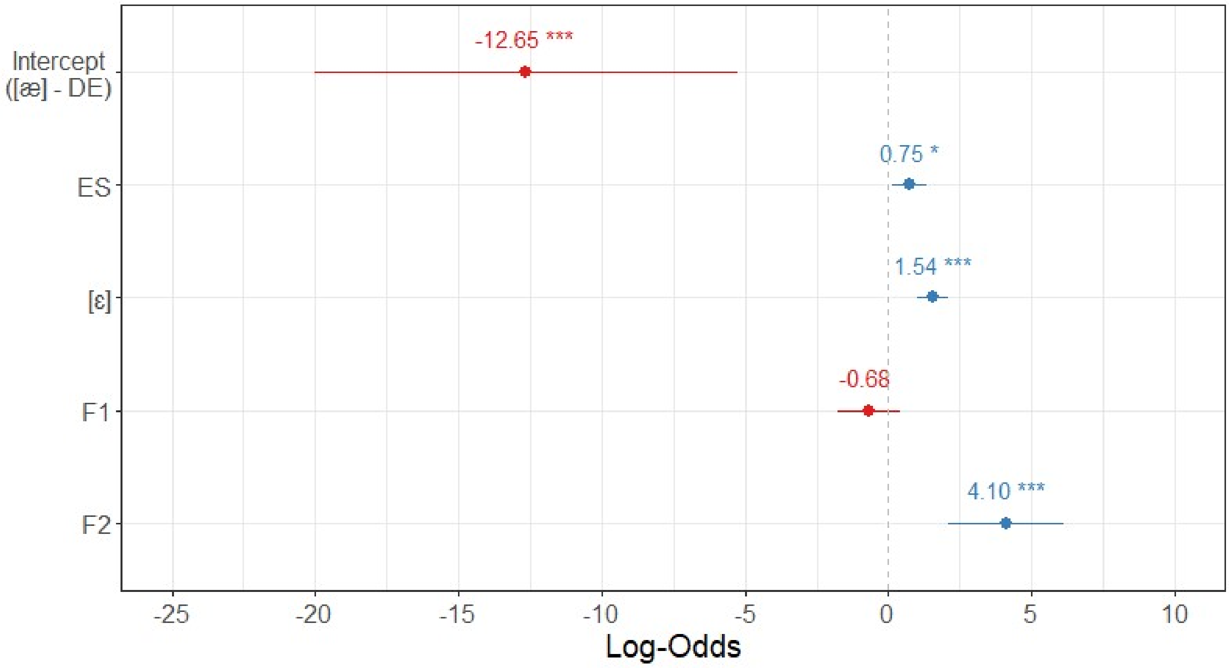
Predictors	Estimates	std. Error	z value	Pr(>|z|)
Intercept ([æ] – DE)	–12.652	3.757	–3.368	**<0.001**
L1 ES	0.750	0.309	2.425	**0.015**
target vowel [ɛ]	1.540	0.275	5.595	**<0.001**
F1	–0.675	0.550	–1.228	0.220
F2	4.102	1.032	3.977	**<0.001**
Observations	2,909			AIC = 2,901.4
Marginal R^2 / Conditional R^2	0.112 / 0.527			

Model: glmer (identifiedVowel ~ targetVowel + L1 + logF1norm + logF2norm + (1 | speaker) + (1 | listener) + (1 | targetWord), family = binomial, data = dados)
Intercept: target vowel [æ] by a German listener
Response variable: vowel identified as [ɛ]
Source: RStudio 1.4.1103 (2021)

All estimates refer to the probability of having the tokem identified as with the vowel [ɛ], as opposed to not being identified that way. As logistic models operate with binary predictions, the model also predicts the log-odds of listeners *not* making this identification. In our data set, *not* identifying a token as [ɛ] means identifying it as [æ], because it was a forced-choice task. Thus, the intercept of the model is the probability of having a German listener (DE) identify a vowel as [ɛ] when the speaker is producing a target word with vowel [æ]. Negative estimates mean there is a higher likelihood of the negative scenario, that is, of a listener *not* identifying a token as [ɛ]. Germans show a low probability of identifying [ɛ] from [æ] stimuli (β_0 = –12.652, p = <0.001). That is to say: when the word 'pat' or 'sat' is produced by a Brazilian learner, there is a higher likelihood that it will be (correctly) identified by a German listener (intercept) as a word with [æ] – in our forced-choice task, tokens 'pat' and 'sat' (see section 3.4.4).

Figure 4.8 plots the log-odd estimates of the mixed-effects logistic model of the effects of L1, target vowel, F1 an F2 on the vowel identifications of [ɛ].

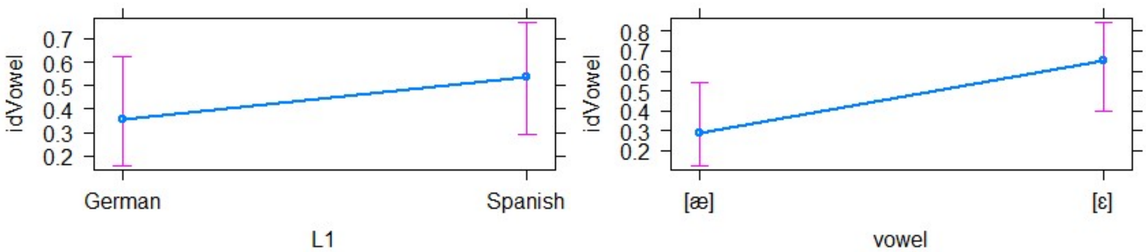
Figure 4.8 – Log-odd estimates of the mixed-effects logistic model for the effects of L1, target vowel, F1 and F2 on vowel identifications of [ɛ]



Source: RStudio 1.4.1103 (2021).

As shown by the blue notation in Figure 4.8, L1 is a significant predictor, as Argentinian participants tend to identify vowels as [ɛ] ($\beta = 0.750$, $p = 0.015$), but Germans do not. When the target word should be produced with [ɛ], there is a higher probability ($\beta = 1.540$, $p = <0.001$) of listeners of either L1 identifying it as [ɛ][110].

Figure 4.9 – Mixed-effects logistic model's predicted effect on [ɛ] vowel identification considering L1 (left) and target vowel (right) as predictor variables



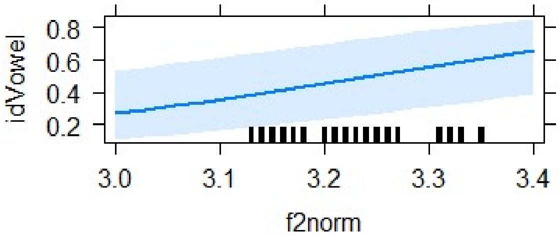
Source: RStudio 1.4.1103 (2021).

Figure 4.9 indicates probabilities and standard errors of the predictions by the fitted model. Predictions refer to a vowel being identified as [ɛ]. As mentioned, Argentinians are more likely to identify a token as having [ɛ] (on the left, L1 predictor variable). This is in line with

[110] Interaction between L1 and target vowel was tested in an earlier fit, but yielded a high collinearity and no significance ($p > 0.05$), allowing us not to include it in our final model.

the descriptive analysis reported in section 4.1.1. Likewise, target words with [ɛ], when produced by Brazilian learners, yield a higher likelihood of being identified as [ɛ] by listeners of both L1 groups (right, target vowel predictor variable). Furthermore, as shown by Figure 4.10, F2 is also a significant predictor variable of [ɛ] vowel identification.

Figure 4.10 – Mixed-effects logistic model's predicted effect on [ɛ] vowel identification considering F2 as a predictor variable



Source: RStudio 1.4.1103 (2021).

Of the acoustic measures under analysis in this study, only F2 is a significant predictor of vowel identification, according to our model. In productions of 'pat' and 'sat', the model indicates that increases in F2 enhance the probability of having a listener identify the tokens as [ɛ]. That is true for both German and Argentinian participants. Thus, that prediction is consistent with the literature on native productions (LADEFOGED, 2010; YAVAS, 2011)[111]. Native speakers produce [ɛ] with higher F2 values (more fronted) than [æ].

Figure 4.7 illustrates the baseline (native) values for both F1 and F2 formants, in contrast with the Brazilian productions in the three proficiency levels assessed in the present study. The baseline plot (bottom right in Figure 4.7) allows us to see that [æ] and [ɛ] are produced with distinctive height (F1) and frontness (F2) values. Brazilian productions do not show this distinction fully developed, though. This will be further discussed in section 4.1.4. For now, we mean only to point out that the model's prediction of higher F2 values leading to higher [ɛ] identification seems to endorse the Complex, Dynamic view that even a partially produced distinction can have an effect on vowel identification. We may also hypothesise that other cues that play a role in the listeners' L1 or composite L1-L2 categories could have led to different identification tendencies by each group of listeners.

Finally, it should also be noted that the response variable 'identified vowel' does not have a balanced number of tokens, unlike the predictor variable 'target vowel'. Target vowels

[111] As mentioned in section 4.4.4, native speakers of English produce [æ] with an average of 820Hz in F1 and 1,808Hz in F2, whereas [ɛ] is produced with 704Hz in F1 and 1,910Hz in F2.

inserted in target words were controlled prior to the identification task (see section 3.4.4). Identified vowels, on the other hand, are the very result of the participants' identification rates. As has already been illustrated in Figure 4.4, there is a different proportion of tokens identified as [æ] and [ɛ] by each group of participants. Germans (18 participants) identified 673 tokens as having [æ] and 472 as having [ɛ], whereas Argentinians (28 participants[112]) identified 824 tokens as having [æ] and 967 as having [ɛ]. This is relevant because it shows that each group favours different identifications. This will be the focus of the analysis in the next section.

4.1.4 Exploratory analysis of listeners' composite L1-L2 categories for [æ] and [ɛ] (RQ3)

As mentioned in section 2.1, a Complex, Dynamic system is in constant change over time. In the present study, this means that the results reported here reflect the state of each participant's system at the time they took part in the experiment. Thus, we assume that the moment in time recorded in the data is somewhat representative of a participant's current system state (at that specific moment in time). Therefore, we now look at how the identified vowels (both accurate and inaccurate) may reflect the participants' categories with regard to the acoustic cues we have investigated. We reiterate that the layout we attempt to outline is a 'still frame' of the participants' systems at the moment of data collection[113].

Within a Complex, Dynamic framework, language is a process, rather than a product that can achieve some perceived end-state. That is true for both native languages and additional ones. Language is also a system that changes through time, as a result of the individual's embodied experiences. As we have detailed before, a person moving to a location where a different variety of the native language is spoken can lead to changes in the way production and perception of that native language are processed. In other words, even the native language is subject to change, no matter how proficient one might be in it (KUPSKE, 2016; PEREYRON, 2017; DE LOS SANTOS, 2017). This adaptive characteristic of the system also brings another characterisation that is relevant to the present analysis: the fact that a process has gradiences, rather than categorical, static states.

Moreover, as we have mentioned in section 2.1, the complex nature of a system implies that changes happen as a result of the interaction of subsystems. Unlike linear development, though, complexity means that the whole cannot be taken as a sum of its parts. This means that

[112] The uneven number of participants is also an unbalancing factor. This will be further discussed in section 5.2.
[113] As mentioned before, from a Complex, Dynamic perspective, there is a possibility that participating in the study might have had an impact on the organisation of the system.

small changes to a subsystem can lead to big changes in the system as a whole, whereas big changes to a subsystem can have little to no effect in the entire system. Interaction between variables, once again, is the key. For the present study, this is of particular relevance, as it is one of the motivations we have for taking a second look at acoustic characteristics, including those that have not yielded significance in our inferential statistics. In other words: our models might have taken into account a factor *per se*, failing to compute its emerging effect within the system interactional dynamics. Furthermore, as each individual is also a subsystem of a speech community, the multiple interactions and their effects might have eluded the statistical approach.

Thus, our third Research Question intends to investigate the different composite L1-L2 categories of our participants. The Complex, Dynamic view allows us to assume that a certain degree of gradience will be present in all systems under analysis (Brazilian speakers' and Argentinian/German listeners' systems). Such gradient characteristics, however, might not be picked up by an inferential analysis, which is not to say that those gradiences could not provide valuable information about these developmental stages. Hence, in RQ3 we asked: "How can the identification patterns in the perception task shed light on the listeners' composite L1-L2 categories?"

The inferential statistics reported and discussed in sections 4.1.1 and 4.1.2 have already estimated the effect of listener's L1 and target vowel in the rates of identification accuracy (RQ1), as well as of listeners' L1, stimulus's F1, F2 and target vowel in vowel identification (RQ2). We reiterate that the first question had 'target word' as the unit of analysis for local intelligibility, which included the correct identification of all segments in the CVC monosyllabic tokens used in the perception task – namely, onset consonant, nuclei vowel and coda consonant. Research Question 2, on the other hand, disregarded accuracy and target stimulus, analysing the identified vowels only. In order to answer Research Question 3, we will go back and look at both levels: accuracy rates (relating to vowels inserted in *words*) and vowel identification (relating to *vowels* as they were perceived). It is our understanding that those are two different levels of analysis and that, together and on their own, they can shed light on developmental aspects of our participants' composite L1-L2 categories. Accuracy rates will be referred to in terms of vowels [æ] and [ɛ], but were measured, as in RQ1, considering a fully accurate identification of each word in the minimal pairs ('pat' identified as 'pat' and 'sat' identified as 'sat' for the accurate [æ] vowel identifications, as well as 'pet' identified as 'pet' and 'set' identified as 'set' for the accurate [ɛ] vowel identifications).

Our exploratory analysis will focus on the acoustic cues that might have played a role in listeners' identifications. As mentioned in section 2.2, acoustic cues present in the speech signal may be weighed differently by each listener or each group of listeners, depending on factors such as cue informativeness, among others. Considering that the process of cue weighting is language-specific, we need to take the listeners' L1s into account. As Escudero (2009) sums up,

> [c]ross-linguistic studies (cf. Strange 1995) have shown, for instance, that experience with the fine-grained acoustics of a specific language environment shapes listeners' perception of the speech signal in a linguistic way. This environmental dependence is observed in two of the basic properties of speech perception, namely the categorization of acoustic continua and the perceptual integration of multiple acoustic dimensions. (p. 2)

As we have mentioned in section 2.2, Escudero's (2009) experiment is one of the inspirations for the present study. She hypothesised that the native parameters of weighting interacting cues could help predict patterns in L2 speech perception. Escudero and Polka (2003) and Escudero (2009) presented stimuli produced by native speakers of Canadian French to native speakers of Canadian English (and vice-versa) and contrasted the boundaries of listeners' native categories to dispersion areas of vowels identified in the L2. The authors highlight that formant frequency is a primary cue for native speakers of English, but when F1 and F2 are not as informative, the temporal aspect has a stronger effect on listeners' perception. For native speakers of French, the second language group investigated in Escudero and Polka (2003) and Escudero (2009), duration is not usually regarded as an informative cue. Moreover, though both [æ] and [ɛ] categories were identified by both English and French speakers, the acoustic characteristics of each sound was perceived differently by native speakers of each language. We report back to Figure 2.1 for a visual inspection. As can be seen in that figure, because duration does not play a relevant role in French, native speakers produced highly variable durations, but in a much stricter F1 dispersion. Native speakers of English, on the other hand, produced variable F1 values combined with variable durations, in a way that a token with a lower F1 had a longer duration, whereas higher F1 values were combined with shorter durations. In both studies, when F1 and/or duration in an L2 stimulus were outside L1 category boundaries, listeners had a lower accuracy rate in vowel identification.

Our research questions used different methods to try and verify if this phenomenon, among others, would also be present in our dataset. We adopted inferential statistics to answer RQ1 and RQ2, and will now discuss cue weighting processes of acoustic characteristics to try and answer RQ3. Thus, our intention with the following analysis is to observe which cues were taken as relevant cues, and what their effect on listeners' perception was. We assume that, as cue weighting is language-specific, listeners' patterns of identification might provide evidence of which (and how) cues were weighed. This is what justifies our choice for an exploratory, stimulus-by-stimulus analysis.

In order to consider identifications of a given token as a 'pattern', we needed to make sure that such identifications were not done at random. Here we will make a slight digression to detail how we arrived at what we will take as a 'pattern' in this section.

If we consider the flip of a coin, heads or tails occur at random. That is to say that there is a 50% chance that the flip will turn heads, as much as there is a 50% chance it will turn tails. In layman's terms, tokens in the present study could be compared to a coin. In the same metaphor, each time an identification is made could be taken as a coin flip. That comparison would yield a 50% chance that a token would be identified as [æ], as much as a 50% chance that it would identified as [ɛ] – regardless of the speaker's intention. That analogy would mean that the token does not have any intrinsic characteristic that leads to one or the other identification – a non-adulterated coin, if you will. Nonetheless, from our descriptive and our inferential analyses, we already know that this is not the case for our tokens: identification does not happen at random. Each of our [æ] and [ɛ] tokens were identified 46 times: 28 times by native speakers of Spanish and 18 other times by native speakers of German. If a given token was identified 23 times as [æ] and 23 times as [ɛ], we cannot infer that any intrinsic characteristic has had an effect on its identification, because 50/50 is the very definition of chance. Even though linguistically we know that it is not precisely 'chance' that led to those identifications, mathematically we cannot consider them as yielding a proper pattern. Therefore, to make sure that it was not 'chance' that led to one or the other identification, we defined our 'pattern' as mathematically 'above chance'.

If we go back to the flip of the coin analogy, this criterion would be equivalent to asking: if we flip the coin 100 times, how many tails would indicate to us that the coin is doctored? That is, are there intrinsic characteristics that are skewing the odds? The syntax for that calculation in RStudio is *qbinom(0.05, 100, 0.5, lower.tail = F)*, where '0.05' stands for a 95% confidence interval; '100' means how many times the coin was flipped; and '0.5' is the standard 50% chance given by the categorical nature of the coin (two sides, one head and one tail). The 'lower tail' function is set to 'false' (F) to return the right wing of the normal distribution. In our dataset, that command line reads *qbinom(0.05, 28, 0.5, lower.tail = F)* for tokens listened to by Argentinians, and *qbinom(0.05, 18, 0.5, lower.tail = F)* for the ones listened to by

Germans. One will notice that the confidence interval is the same (95%), and that a normal distribution (50% chance) is also assumed, as in the coin example – we also have a two-level categorical factor. The numbers ‘28’ and ‘18’ refer to ‘coin flips’, that is, the amount of times the token was identified in each group of listeners. If we look at our Spanish listeners only, a ‘chance’ identification would mean that if a token had 14 to 17 identifications as a given vowel, those identifications do not configure a ‘pattern’. From 18 identifications or more, then we can mathematically assert that something about our token is skewing the distribution, that is, has an ‘above-chance’ effect on listeners’ perceptions.

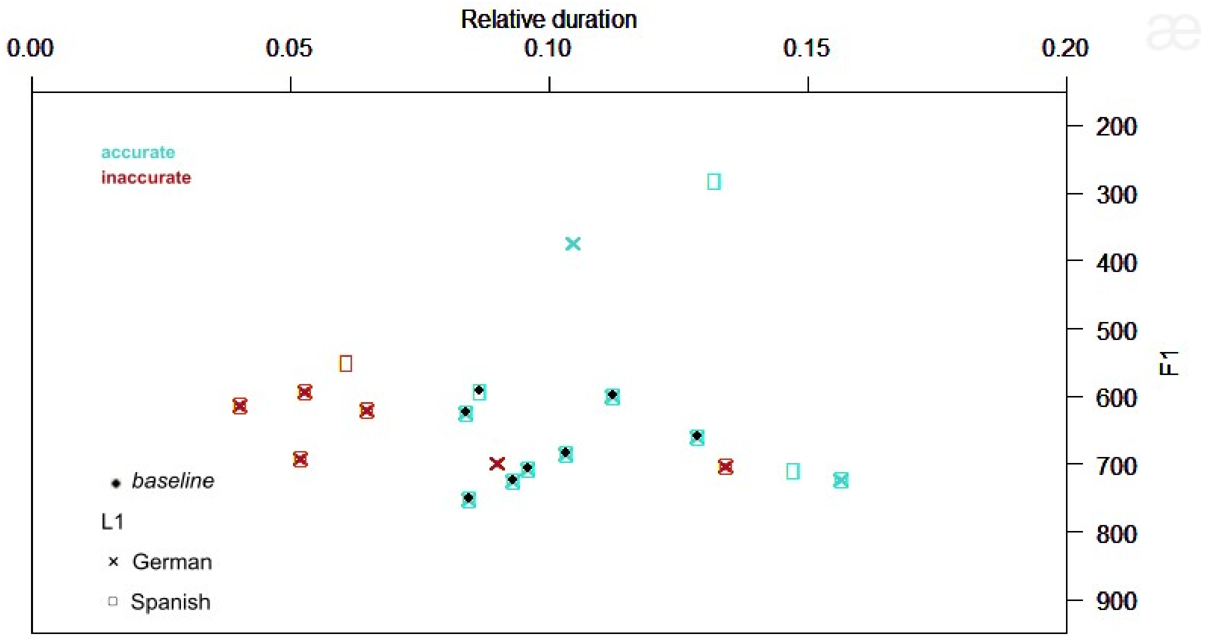
Thus, according to *qbinom* calculations, this is how we selected the subsets of tokens to compose our ‘patterns’: 18 or more identifications mean ‘above-chance’ identification for the Argentinian group, and 12 or more identifications mean ‘above-chance’ identification for the German group. Now that we have explained the criteria by which we selected the subset of identifications that we henceforth call ‘pattern’, let us move on to the analysis *per se*.

As reported in section 4.1.2, results of the fitted model that was built to answer RQ1 estimates that both L1 and target vowel, as well as the interaction of these variables, are significant predictors of identification accuracy. We will now look at how those accurate identifications relate the acoustic data in our stimuli. As in previous sections, we will analyse duration x F1 prior to F1x F2[114] aspects, since Brazilian learners tend to perceive temporal cues better than F1 and F2 distinctions (RAUBER, 2006; NOBRE-OLIVEIRA, 2007). Vowels will also be presented in the same fashion as they have been so far, first [æ] and then [ɛ], for the duration of the lower vowel is longer.

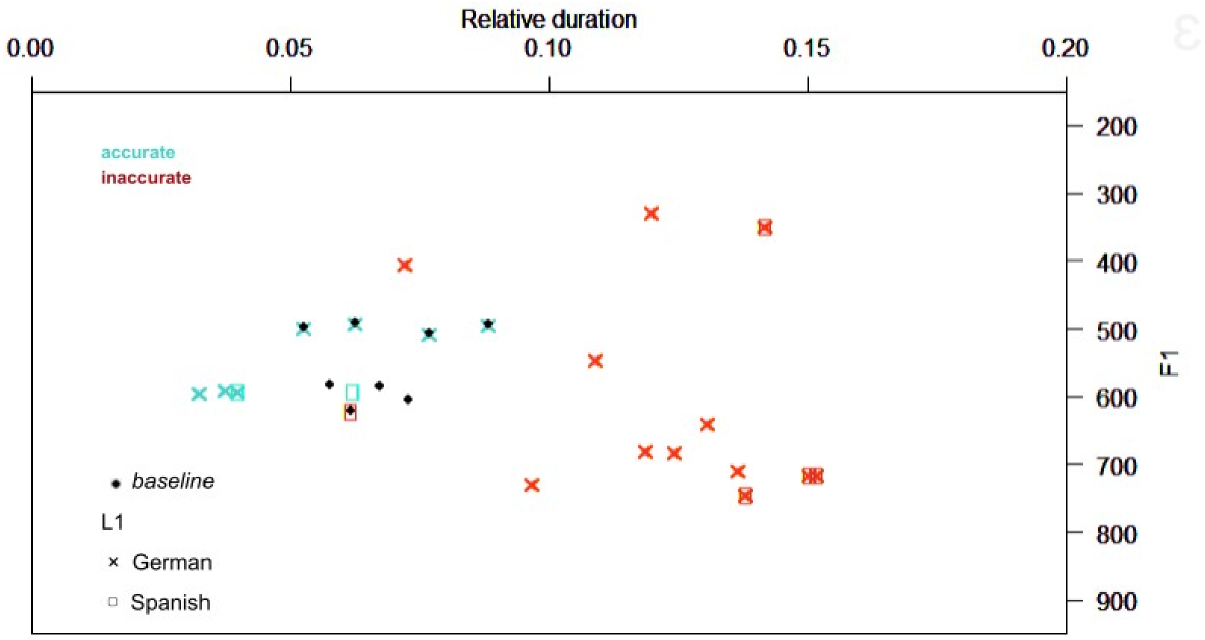
[114] We followed Escudero’s (2009) design and chose to only plot duration x F1, but not duration x F2.

Figure 4.11 – F1 x relative duration plots of [æ] and [ɛ] tokens with above-chance correct and incorrect[115] identification, by target vowel

(a) target vowel [æ]



(b) target vowel [ɛ]



Source: present study.

[115] As mentioned in the beginning of this section, accuracy here is grouped in terms of vowel, but was measured by word. This way, a ‘pat’ token identified as ‘sat’ was deemed an incorrect identification. As vowel quality shows an effect of adjacent segments, we sustained this criteria of accuracy in this section (cf. SILVA *et al.* (2019) for an account on acoustic characteristics of vowels).

The figures in 4.11 plot items[116] with above-chance correct identifications (green) and incorrect identifications (red) in relation to F1 and relative duration. German datapoints are styled as cross signs, whereas Argentinian identifications are shown as squares. Baseline tokens (produced by the Canadian participants) are styled as black dots.

For target vowel [æ], our exploratory analysis shows that duration seems to be a highly relevant dimension in identification accuracy for both groups of listeners. All but one and two tokens shown in the plot (for Argentinians and Germans, respectively) were incorrectly identified when productions presented a longer relative duration. We will notice that F1 does not seem to have been very informative as a cue for these tokens, as both correct and incorrect identifications have a rather large variation in F1 values for the two L1 groups. We will discuss those patterns further later on.

Based on the plots above, German listeners seem to use duration as a main cue in their correct identifications, which leads us to believe that duration was very informative for both vowels under analysis in this section. However, as mentioned before, Brazilian productions did not significantly yield much distinct relative durations across all proficiency levels. This could have led German participants to take as distinctive a cue that was not (as) distinctive.

Accordingly, Germans had lower accuracy rates identifying [ɛ] than they did identifying [æ], as can be seen in Figures 4.3 and 4.1a. For [æ], in which German native speakers had more than 70% identification accuracy, virtually only tokens with longer relative durations were correctly identified above chance. The opposite pattern is indicated by the [ɛ] plot. We could speculate that, despite not being as distinctive, the relative duration produced by Brazilian speakers was filtered by German listeners as a temporal cue for [æ] and [ɛ]. As we have seen in section 4.1.1.1, both [æ] and [ɛ] productions by Brazilian speakers have longer durations than native German [ɛ] and both native North American English and Canadian English [ɛ] and [æ]. This could have meant that the duration cue carried the information of a longer vowel, thus yielding [æ] identifications. Accordingly, for [ɛ], it would seem that German listeners expected lower relative durations than the ones produced by Brazilians. The [ɛ] above-chance correct identification pattern seems to corroborate that hypothesis.

Additionally, Figure 4.11(a) seems to indicate that Argentinians also took relative duration as a highly informative cue to correctly identify target words with [æ]. We notice almost as many above-chance datapoints for the Argentinian participants for [æ] accurate

[116] We reiterate that all plots in this section will have an uneven number of tokens, as they show only a subset of the data we collected, based on the number of above-chance identifications. Each figure will have a different number of datapoints, according to each plot criteria.

identifications as we do for German participants. However, we have observed that, in terms of accuracy rates, native speakers of Spanish had a worse performance for [æ] tokens than did their German counterparts, as discussed in section 4.1.2. A tentative explanation could be that the [æ] temporal cue was somewhat informative for native speakers of Spanish, but was not sufficient when taken into interaction with other cues, such as formant frequency. This contrasts with the Argentinians' performance regarding [ɛ].

Though [ɛ] was correctly identified more often by these participants, it is worth mentioning that there are very few datapoints representing accurate identifications of this vowel *above chance*. This small list of only two stimuli with above-chance accurate identification could be read as meaning that the duration cue had a stronger effect as informative for Argentinians in [æ] identifications, but as not strong for [ɛ]. Considering that Argentinians had higher accuracy rates for [ɛ], a tentative explanation for the patterns we see could be that the temporal cue is usually less informative for those Argentinian listeners, in comparison to other cues, due to its long durations. We could even hypothesise that the relative shorter vowel durations in Spanish in comparison to English are somewhat a reason for this effect. However, when these other cues fail to clearly inform a sound category, vowel duration has a stronger effect on identification. Conversely, we could hypothesise that when the information from the temporal cue is too dissimilar to what an L2 category is expected to present, it 'throws confusion' into the identification process. In other words, we could imagine that Argentinians always expect longer vowel durations in L2 vowels, but they know that [æ] is longer than [ɛ]. Thus, the cue is not as informative for [æ] productions. On the other hand, the expected shorter duration of [ɛ] was not matched by productions, which might have led listeners to disregard the information of that cue and proceed their identification based on more informative ones – thus justifying the higher [ɛ] identification rates, despite the longer relative durations in comparison to native productions.

Additionally, as was the case in Escudero and Polka (2003) and Escudero (2009), the language groups we selected to take part in the present study also have different native characterisations of this acoustic cue. Brazilian Portuguese and Spanish speakers do not use temporal cues distinctively, whereas German speakers do. German words like 'bad' ('bath') and 'bat' (asked) are minimal pairs in German, for the final consonant in 'bad' is produced with devoicing, leaving vowel length as a distinctive source of acoustic information: [ba:t] vs. [bat]. This could be somewhat a reason for German listeners to regard vowel duration as a relevant cue more consistently. Thus, we expected that the temporal cue would have stronger effects for German participants than they would for Spanish participants. Moreover, RQ3 intended to verify whether Escudero's (2009) findings regarding the role played by the 'L1 filter' in the perception of L2 sounds could be replicated for L1 Spanish and L1 German. As seen in Table 4.1 and discussed thus far, Brazilian's vowel productions (in both their L1 and L2) are generally longer than L1 vowel productions by Germans and Argentinians. This fact also highlights how some effects seem clearer for some stimuli, which can be taken as evidence that the speaker-listener pairs might affect the way acoustic cues interact.

Considering the speaker-listener pair, it is additionally relevant to point out that it is not only the language-specific cue weighting parameters from L1 Spanish or L1 German participants that play a role. Our data seem to suggest that while both groups of listeners take the temporal cue as informative for words with [æ], we have also seen that Brazilian speaker participants' longer duration pattern seems to have an effect on the speaker-listener communicational interaction[117]. In other words, the longer duration of native BP [ɛ] seems to be carried to L2 English [ɛ] as well as L2 [æ] (because it is assimilated into [ɛ]). As a longer duration is a characteristic of English [æ], tokens with [ɛ] are not perceived as such, for Brazilian productions of L2 [ɛ] are too long to be interpreted as the target. This is a relevant finding, insofar as previous studies show that Brazilian learners find it difficult to produce the [æ] vowel, given that they need to create a new category for it in their composite L1-L2 system. Our data suggest, however, that though this may be true for formant frequency distinctions, perhaps for temporal distinctions what Brazilian learners need to do is work on reducing vowel length in their productions of [ɛ].

Finally, we must point out that the two different methods of analysis applied in this thesis – inferential in RQ1 and RQ2, and exploratory in this section regarding RQ3 – are complementary to each other, so far as the durational cue is concerned. While our mixed-effects logistic models did not take duration as a relevant predictor of vowel identification, our qualitative analysis suggests that vowel length did have an effect on listeners' perception. We highlight that, within the Complex, Dynamic view of language that we adopt, this is not contradictory, but rather evinces that statistical and descriptive analyses are complementary. On the one hand, we suspect that the data we were able to collect from Brazilian participants did not entice enough variation of duration in itself to cause a statistically significant effect in the inferential statistics. On the other hand, we understand that the combination of acoustic cues

[117] When we are not reporting inferential results, the word 'interaction' does not carry a statistical meaning. Rather, it indicates "mutual or reciprocal action or influence" (MERRIAM-WEBSTER, 2021). More often than not, in this thesis, it points to a communication scenario, that is, a speech produced by a speaker which is perceived by a listener in a meaningful exchange.

leads to emerging patterns that are too varied for the small dataset we worked with in the present study. Moreover, as we have pointed out, because we are working with multiple hybrid systems, due to our groups of non-native learners, it is likely that only a very large-scale experiment, with a larger number of participants, would be able to yield statistical significance for such gradient predictors.

Let us move further to analyse how other cues might have had an effect on accurate vowel identification. Figure 4.11 plotted datapoints in terms of relative duration and F1. We can see that F1 does not seem to have been perceived as such an informative cue for either group of listeners, nor for either vowel, in comparison with duration. We can see that it does seem to play a role, though. We notice that there is a much wider F1 range (300Hz to 750Hz) in which productions are correctly identified as [æ], with a concentration of those tokes in the area below 600Hz. Tokens accurately labelled as words with [ɛ], in their turn, show a more condensed distribution in the 550 to 600 Hz range in the F1 axis, with a wider distribution of inaccurate identifications.

Based on the descriptive analysis provided in section 4.1.1, we could hypothesise that the gradient F1 distinction produced by Brazilian speakers for [æ] and [ɛ] did not yield enough saliency for that cue to have a strong effect on listeners' perception. The fact that all baseline [æ] tokens were correctly identified above chance could be taken as suggesting that listeners have perceived the F1 cue in a native-like fashion, which was not the case for Brazilian productions. Conversely, baseline [ɛ] tokens were only accurately identified above chance when their F1 frequency was lower (as it is expected in native English productions). This could, in turn, indicate that our listeners' L1-L2 [ɛ] composite category allows for cues to play a similar role to what they would for a native speaker of English, but not quite as native-like as the data for [æ] suggest. It could also mean that, in the interaction of all cues being weighed by listeners, F1 had a weaker effect for [ɛ] – than, for instance, did duration.

Finally, it is also worth noticing that there are two[118] tokens with longer durations that we would expect to be correctly identified as [æ], but were incorrectly identified above chance[119]. Besides duration, these tokens also have rather high F1 values, which would typically characterise [æ] tokens. Thus, from our previous hypothesis, they should have been correctly

[118] For Argentinian listeners, that was the case for only one of the two tokens. Nonetheless, it coincides with one of the two 'difficult' tokens for Germans, which is why we chose to look at them both.

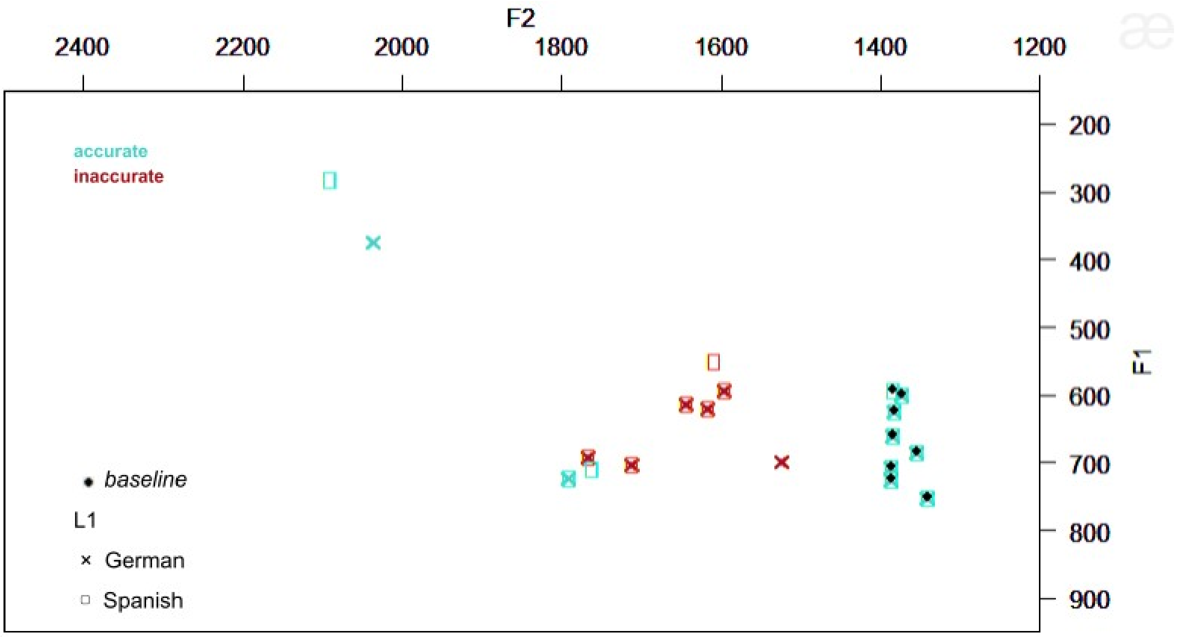
[119] To phrase it in accordance with the criteria we posed for subsetting tokens: it had above-chance identifications as the wrong vowel. That is to say: those two [æ] tokens showed above-chance identifications as [ɛ].

identified. What, then, explains those two outliers? We could assume some other acoustic cue played a role in the identification process.

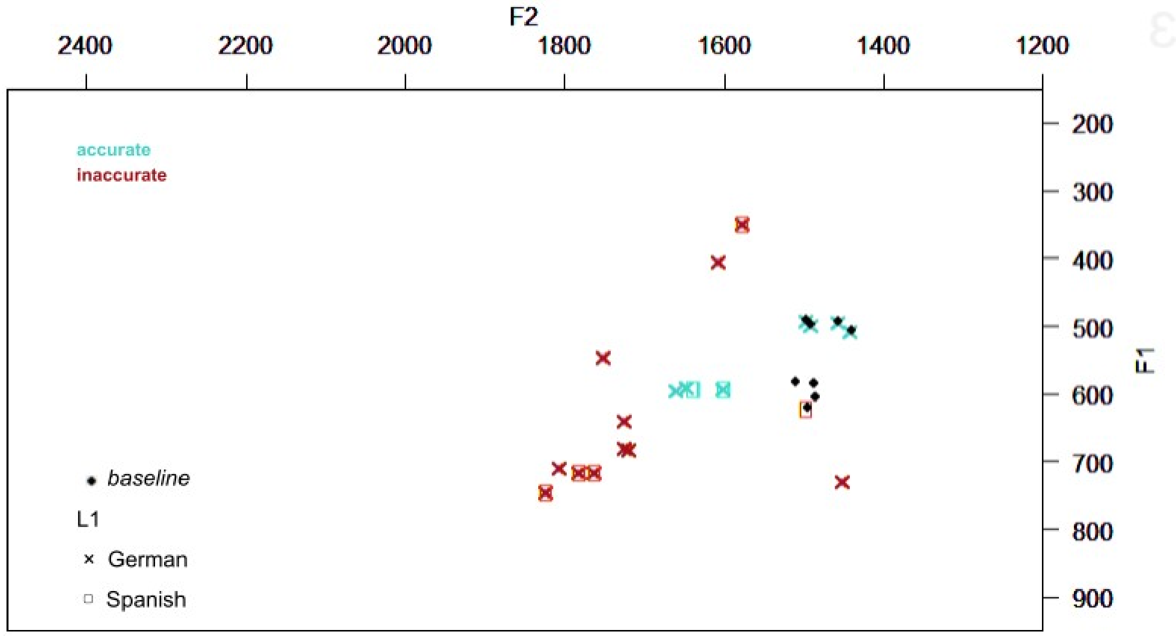
Let us look at the outlier stimulus that caused difficulties to both German and Argentinian listeners. Upon further inspection of that datapoint in our dataset (duration: 13,4%; normalised F1: 712.84Hz), that token is the second repetition of the word 'pat' produced by a female, intermediate-level speaker. We have already established that this token has a longer relative duration and a higher F1 value than some other productions in our dataset, which should lead to a correct identification. However, this token's F2 is around 1,951.34Hz (1,712Hz normalised), which may have been perceived as higher than the F2 values for other tokens or for prototypical [æ] tokens. That could indicate that, in this particular case, F2 had a stronger effect than relative duration in the interaction of decisive cues. This will be made clear as we look at how F2 values of other tokens relate to accurate identification patterns.

Figure 4.12 shows the same above-chance accurately identified tokens we have discussed hitherto, this time in an F1 by F2 plot.

Figure 4.12 – F1 x F2 plots of [æ] and [ɛ] tokens with above-chance correct and incorrect identification, by target vowel
(a) target vowel [æ]



(b) target vowel [ɛ]



Source: present study.

The plots above seem to indicate that F2 works as a cue to accurate identification (signs in green) for both groups of listeners. Consistent with the literature on native English production, most [æ] tokens with above-chance correct identification exhibit lower F2 values (less fronted position), whereas [ɛ] tokens do not show a distinctive pattern. It is also interesting to observe that, once again, both listener groups show higher above-chance accuracy rates in identified baseline tokens of [æ] compared to tokens of [ɛ]. Of the [ɛ] tokens, only the male productions show above-chance accuracy, and only by German listeners. It is also distinctive that baseline [æ] tokens have very high F2 values, in comparison with the Brazilian productions.

Moreover, though Argentinians had better accuracy rates in identifying [ɛ] tokens, we have already pointed out that only two of those tokens were correctly identified above chance by this group. Considering that for the [ɛ] vowel neither relative duration nor F2 yielded strong effects in above-chance results for these native speakers of Spanish, we might hypothesise that F1 was the cue that was taken as the most informative by those participants. We have already seen evidence for that when we analysed Figure 4.11.

Both figures, indeed, show that accurately-identified [æ] tokens seem less 'spread' in terms of F1 than [ɛ] tokens do. Correct [æ] identifications by native speakers of Spanish range from 550Hz to 700Hz, a similar range to that of German participants, who correctly identified [æ] tokens with F1 measures from 600Hz to nearly 750Hz[120]. The Argentinians' two datapoints

[120] Notwithstanding the two outliers, one for each L1 group of listeners.

of correct [ɛ] identifications are both around 600Hz, which is less spread than the tokens that were intelligible to German participants, condensed in a 500Hz to 600Hz range. Therefore, we could speculate that F1 had a stronger effect as a cue for native speakers of Spanish, whereas the F2 cue had a stronger effect for native speakers of German. In other words, it seems that different patterns emerge from the interaction between the different sources of acoustic information in the stimuli, as well the listeners' language-specific cue weighting process. Furthermore, we understand that these patterns are in line with the findings in Escudero (2009).

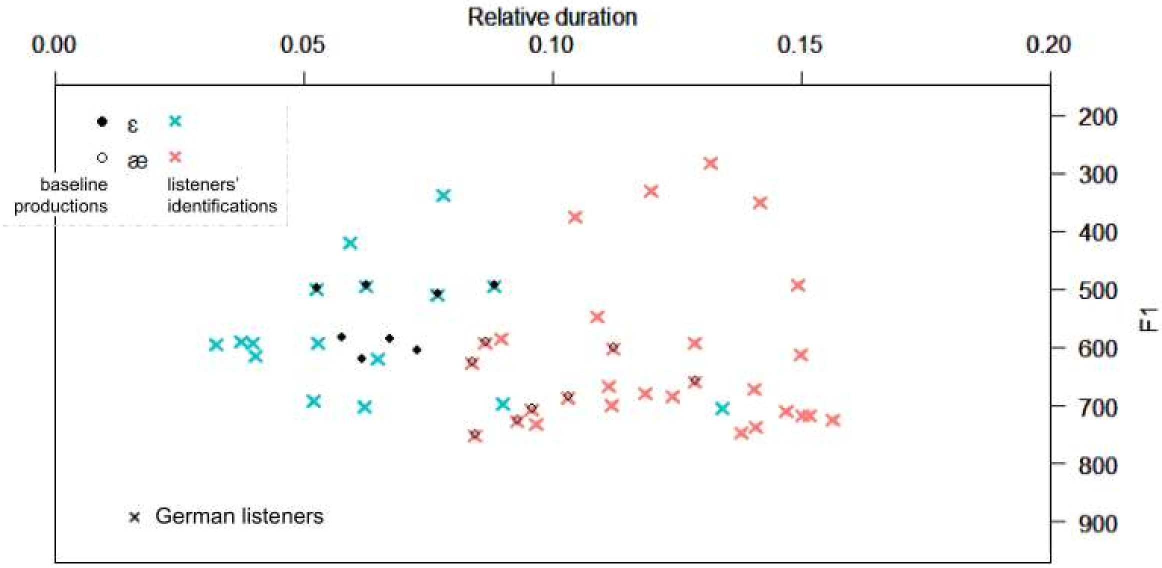
In sum, our data suggest that multiple acoustic cues are playing a role, and, depending on the stimuli, the effects of one or another become more evident. Once again, there seems to be a degree of hybridism in the way composite L1-L2 categories are established for both speaker and listener participants. We highlight that both groups are non-native learners, and that their language systems are in constant change. As we have discussed, speakers seem to carry L1 temporal patterns onto L2 categories. Meanwhile, listeners also seem to weigh acoustic cues based partially on L1 parameters. We take those findings as evidence of a composite L1-L2 category system in their common phonetic space. Moreover, our data seem to provide evidence of the relevance of the speaker-listener pair. Within a Complex, Dynamic framework, it would seem that cues will interact and yield different effects based on the language-specific parameter of the listener, as well as based on characteristics of the stimuli produced by the speaker.

Furthermore, interacting cues are not all primary cues. Adaptive systems can show small effects from large subsystems changes, as well as show large effects as an outcome of small subsystem destabilisations. In our dataset, we could hypothesise that secondary cues might show an emerging effect as a result of the communication situation that involves hybrid systems from non-native speakers and non-native listeners producing and perceiving speech sounds. Inferential statistics is perfectly apt to estimate the main effect of predictor variables, but much larger datasets would be needed to verify the many subsystem interactions – with small to large effects – that we assume when adopting a Complex, Dynamic view of language. Additionally, by default inferential statistics are meant to model patterns, which means that it may not pick up on slighter, more individual processes. An individual's language development is likely to yield a nearly unique learning trajectory, if looked up more closely. Though this does not mean that group or statistics analyses are not relevant, it does suggest that a closer look at gradient phenomena is also beneficial (LOWIE, 2017).

As we have reasoned thus far, accuracy patterns relate in a great deal to the speaker-listener pair. In order to further explore the vowel systems of listeners only, we will now discuss identification patterns, regardless of accuracy. That is, a token 'pat' identified as 'set' will be taken as an [ɛ] identification, for we are looking at how the vowel was perceived, notwithstanding the speaker's intention of producing an [æ] vowel or any given onset consonant[121]. We will use the same criteria as we did for the previous analysis and understand a 'pattern' as an 'above-chance' scenario. Again, we will focus our exploration on temporal and formant frequency cues being weighed by each L1 group when categorizing tokens.

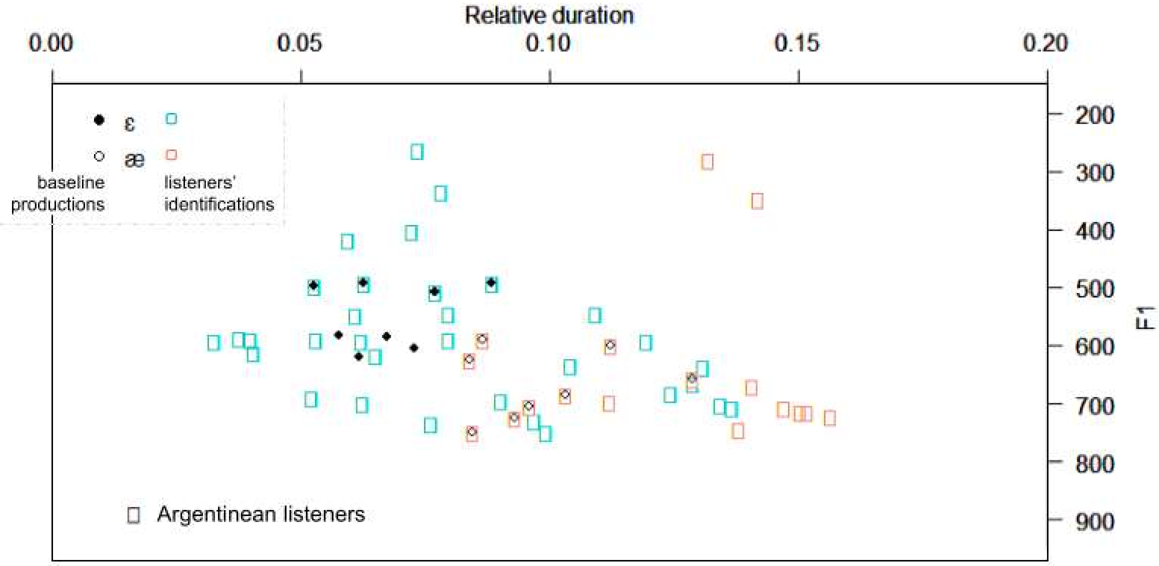
As Figure 4.13 reveals, relative duration seems to be the main cue for both [æ] and [ɛ] identifications, respectively, for both L1 groups, which is consistent with the accuracy patterns just shown.

Figure 4.13 – F1 x relative duration plots of [æ] and [ɛ] tokens with above-chance vowel identification, by L1
(a) L1 = German



[121] All of our tokens had the same [t] coda consonant. As the task was a forced-choice and all options ended with [t], there was no chance of error in view of coda consonant. See details in sections 3.4.3.1 and 3.4.4.2.

(b) L1 = Spanish



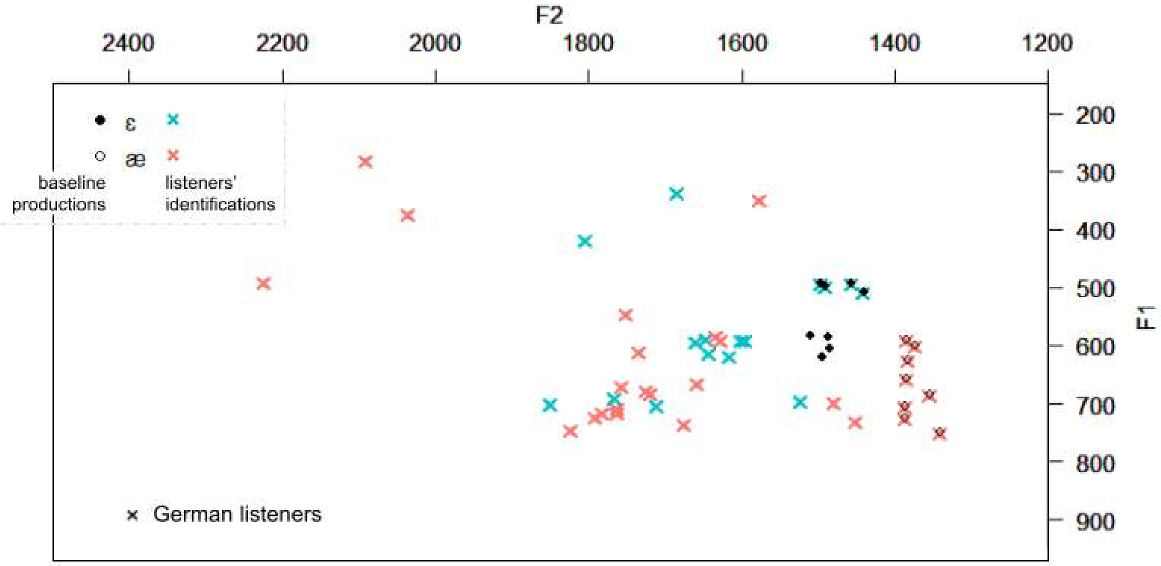
Source: present study.

A vast majority of tokens identified by Germans (datapoints styled as crosses) as [æ] (salmon) has a longer relative duration, with little overlapping between those tokens and [ɛ] identifications (blue). Vowel duration also exhibits consistent results in the identification by Argentinians (datapoints styled as squares). In other words, there is consistency in the identification of [ɛ] with very short relative duration, and of [æ] with very long relative duration. However, when relative duration is between approximately 8% and 13% (not very short nor very long, within the dataset), token identification is somewhat mixed. This could mean that listeners cannot identify this cue as a clear indication source of vowel category, hence the mixed identifications. We could hypothesise that this is partially due to the hybrid nature of the L1-L2 composite categories of both the non-native speakers and the non-native listeners. As we have already established, listeners' composite categories are characterised by L1 and L2 temporal cues being constantly weighed and balanced, through language development, to establish category boundaries. Speakers, in their turn, go through a similar process. In our dataset, the speakers' 'L1 filter' seems to lead learners to produce too long durations for [ɛ], rendering the temporal cue a confusing source of information for listeners. This seems even more relevant as we consider, as shown by Escudero (2009), that the L1 filter might also play a role in how listeners establish their category boundaries and cue weighting processes.

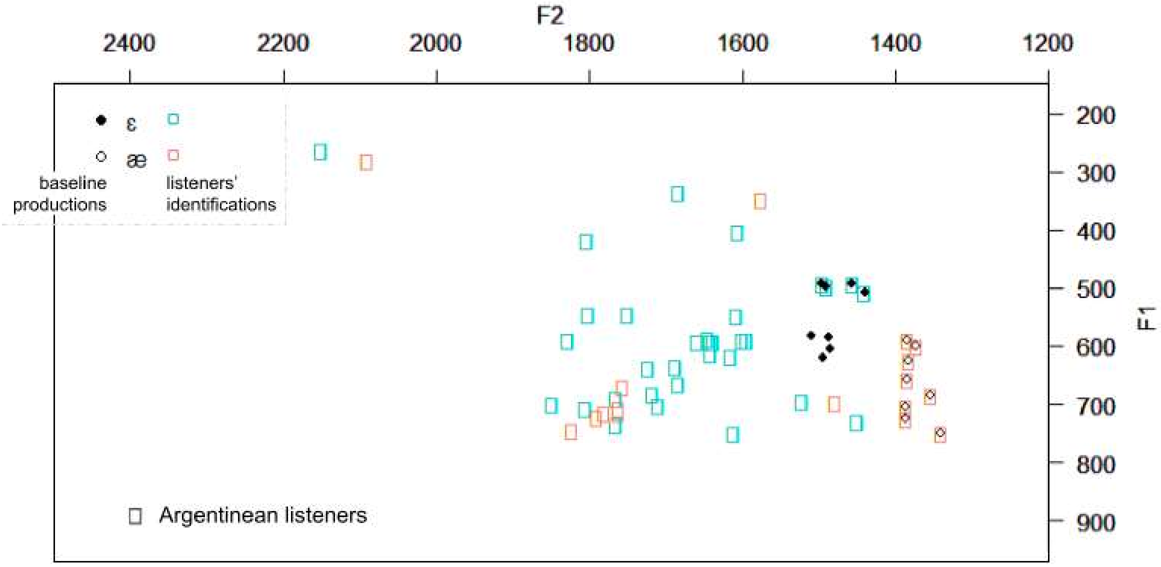
If relative duration seems to portray a clear pattern, the same is not true for vowel quality. The F1 x F2 plottings in 4.14 show identification patterns found in each language group, regardless of accuracy, for above-chance tokens. Salmon styles the [æ] identifications, and blue, the [ɛ] ones.

Figure 4.14 – F1 x F2 plots of [æ] and [ɛ] tokens with above-chance vowel identification, by L1

(a) L1 = German



(b) L1 = Spanish



Source: present study.

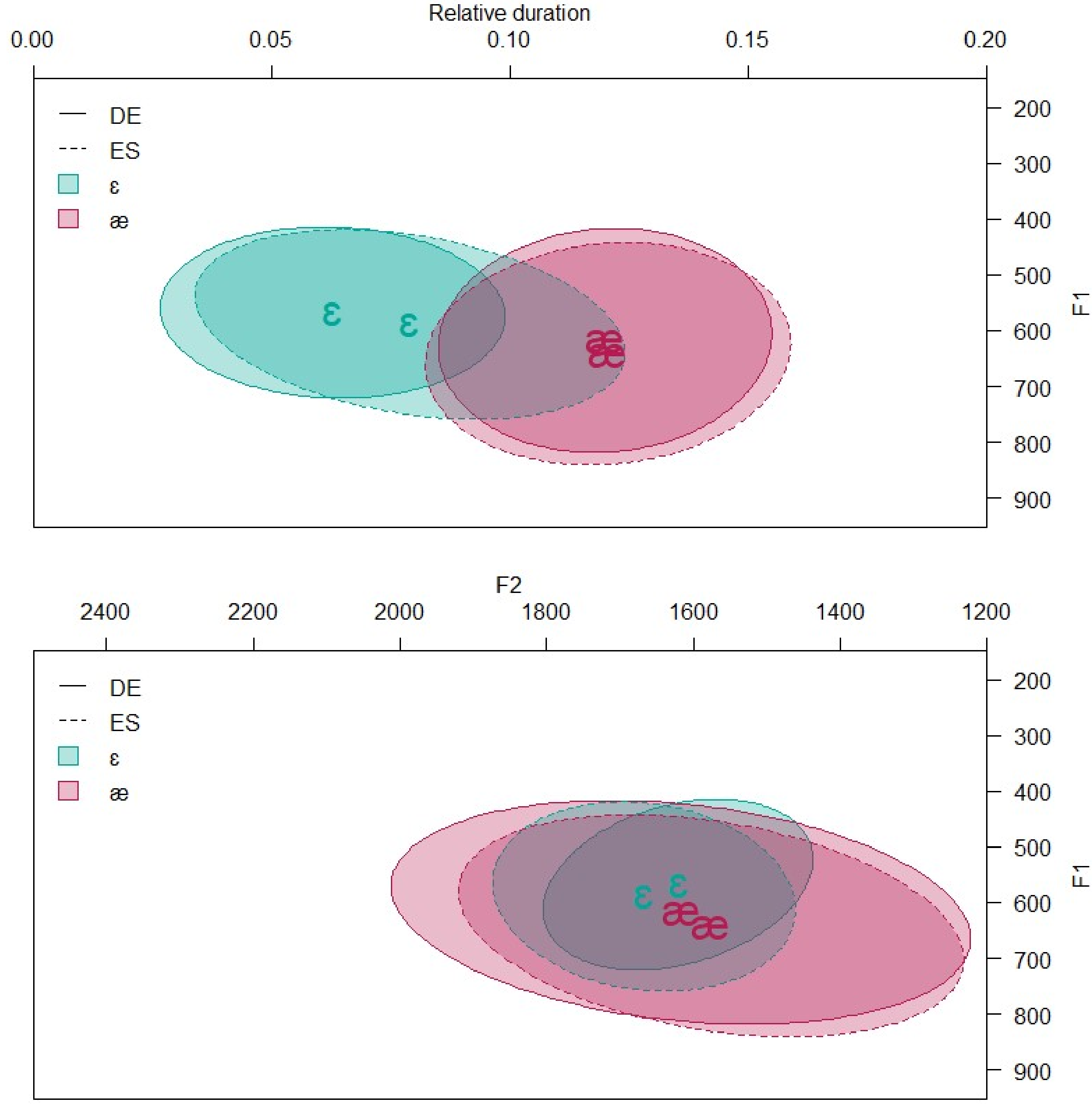
As far as F1 and F2 are concerned, at a first glance, it appears that there are no consistent category boundaries for English [ɛ] and [æ] for either L1 group. A dispersion of identified vowels across the F1 axis, as well as across the F2 axis, does not clearly outline these boundaries. Nonetheless, there are indications of how the categories might be operating. For F1, we observe (as we did in Figure 4.13) that most tokens identified above chance are distributed between 550Hz and 750Hz, but there is a large overlap in that range of both [æ] and [ɛ] above-chance identifications. For the second formant, there seems to be a clearer [æ] pattern for values below 1,400Hz, though all identified tokens in that area are baseline [æ] productions. Above that 1,400Hz limit, even baseline [ɛ] tokens (styled with circles) are not consistently identified.

For German listeners, there seems to be an emerging, albeit slight effect of the interaction between F1 and F2 cues. Some tokens with higher F2 and lower F1 values are identified as [æ], and some others, which show the reverse pattern, are identified as [ɛ]. We highlight that these patterns contradict the literature on native productions. Argentinians seem to portray a stronger effect of F2, but not of F1, as lower F2 values yield [æ] identifications and those between 1,550Hz and 1,750Hz yield [ɛ] identifications. However, above 1,750Hz, there is a perceptible overlap in the identifications of [æ] and [ɛ] by Argentinians. In that 'blurry' area, tokens with higher F1 values appear to be more often identified as [æ]. This seems to reinforce the emerging stronger effect of F1 as a relevant cue for native speakers of Spanish, in comparison to native speakers of German.

A tentative explanation could be that each group's composite L1-L2 categories are made up of acoustic characteristics that present distinct emerging patterns, in view of the combined role played by more than one cue. We could further assume that these differences arise from the development of all languages that the listener knows, and the fact that all categories share the same common phonetic space. The emerging patterns seen in relation to F1 and F2 seem to concur with the reasoning we presented in relation to vowel length. Both discussions would be consistent with the SLM hypothesis that similar new (L2) speech sounds would be assimilated into L1 categories at first, being dissimilated later in the learning trajectory. Thus, the composite category could provide evidence of this developing dissimilation. Our data seem to endorse that view of these processes. Though our logistic model did not estimate statistical significance for some cues (such as speaker's proficiency or vowel length), Figures 4.13 to 4.14 seem to indicate that these cues are being weighed, and that different patterns are emerging from their interaction. For native speakers of German, F1 seems to be a more informative cue for [æ] identification, though a more mixed pattern in [ɛ] identification also suggests that F2 can have stronger effects, depending on how informative other cues are. A similar pattern is found in the identification by native speakers of Spanish – though, as for accuracy patterns, F1 values seem more concentrated in a more restricted space for [æ] than for [ɛ].

The effects of relative duration, as well as of F1 and F2 as weighted cues can be more clearly seen in dispersion patterns. Partial and total overlaps in listeners' identifications seem to support our exploratory analysis hitherto. Figure 4.15 shows dispersion areas by F1 x relative duration and F1 x F2.

Figure 4.15 – F1 x relative duration and F1 x F2 dispersion plots of [æ] and [ɛ] tokens with above-chance vowel identification



Source: present study.

It seems that Germans mostly identify each vowel of the pair based on relative durations, as the dispersion area of each vowel is almost completely separated (solid exterior line). On the other hand, Argentinians (dashed exterior line) show strong effects of the durational cue, but those effects are consistently entangled with effects of formant cues as well. The areas of dispersion appear to endorse this analysis. A case could be made in that direction if we look at baseline identifications shown in Figures 4.13 and 4.14. All [æ] tokens are accurately identified by Argentinians in an above-chance fashion, though [æ] is the hardest of the two tokens for these listeners, according to the fitted model. As mentioned, a tentative explanation could be that vowel length is distinctive in German, whereas Riverplate Spanish does not have such a distinction, and has overall very short vowels. As a consequence, Germans could be seen as more attuned to vowel length, due to L1 previous experience and acoustic category boundaries, whereas Argentinians perceive all vowels as 'long' in comparison to L1 standards, thus having a harder time setting short/long boundaries.

Finally, as we look at L1-L2 composite categories, it is worth noticing that [ɛ] is not a distinctive vowel in L1 Riverplate Spanish, but it is a native vowel of both L1 Brazilian Portuguese and L1 Central German. As discussed earlier, within the SLM framework it could be argued that Brazilian learners have assimilated [æ] into the native [ɛ] category. It could also be that Germans have done so. Additionally, it could mean that, for both Germans and/or Brazilians, the spectral and temporal specifications of each L1 [ɛ] are operating as a 'filter' for the L2 category. This, in turn, might lead to a mismatch between BP-like [ɛ] and German-like [ɛ] boundaries, as well as in relation to native English [ɛ] boundaries.

We reiterate that not all of the effects portrayed by Figures 4.11 to 4.15 are present in the statistical estimates reported in section 4.1.3 (RQ2). Rather than interpreting one result as better or more accurate than the other, we see them as complementary. As Lowie (2017) points out, there is a limit to the contribution of inferential analysis to the study of language development, as any statistical method, by definition, attempts to group factors and effects. That is not to say, he continues, that statistics has no value. To the contrary, there is a valuable contribution that can be made by using inferential methods, as long as it is combined with other methods of analysis. As we detailed in section 2.1, whereas a Complex, Dynamic system changes through time, different fractal time windows may be taken as (static) portraits of the state of a language system at a given moment. This fractality also means that we can carry out group analyses, so long as we do not forget that each individual trajectory, if looked up close, will reveal its particularities.

The present study attempts to combine the potential benefits of inferential statistics with a qualitative approach, thus valuing a process-based approach, albeit in a cross-sectional study. As we discussed throughout this section, inferential statistics did not estimate proficiency level nor duration as significant predictors of vowel identification. However, upon looking at our dataset from a Complex, Dynamic perspective, we have observed that different developmental stages are accompanied by distinct degrees of temporal and formant frequency acoustic boundaries for both speakers and listeners. It is from that perspective that we understand it as relevant to look at acoustic cues that did not yield statistical significance, yet seem to show emerging effects in listeners' perception patterns.

Our descriptive analysis has already pointed out that little distinction in vowel length is produced between [æ] and [ɛ] by the two speakers at each of the three proficiency levels. However, differences are apparent accross proficiency levels, and seem to point to an ongoing dissimilation process. We have also noticed that there are different acoustic category boundaries (as far as formant frequencies are concerned) across proficiency levels, as well as between participants at the same level. Though those gradient distinctions in regard with proficiency level might not have been statistically significant, it is our understanding that they did have a lot to contribute to our analyses. In other words, upon disregarding the speakers' proficiency level variable in the inferential analyses, many dimensions of the listeners' perception processes present in our experiment would have eluded us.

We will now report and discuss results pertaining to the second vowel minimal pair we set out to analyse in the present study, that of [i] and [ɪ], before we draw our final considerations.

4.2 MINIMAL PAIRS 'FEET' – 'FIT' AND 'SEAT' – 'SIT'

We will now look at data referring to the minimal pairs with the vowels [ɪ] and [i]. This section is structured in the same way as the previous one. We provide descriptive analyses of stimuli production and perception, then move on to report inferential and qualitative results, discussing how our findings can provide answers to our Research Questions.

4.2.1 Descriptive analyses

In this section, we will begin by providing a descriptive analysis of stimuli production, and following that, of stimuli perception. These descriptions aim to provide some context for the description and discussion of the inferential statistics and exploratory analyses in the following sections.

4.2.1.1 Stimuli production

As discussed in section 2.4, the Brazilian Portuguese vowel system has seven vowels in stressed position, one of which is the high front vowel [i]. In BP, there is no lax counterpart to this tense vowel. However, in English, vowel tenseness has a contrastive value, which can be seen in minimal pairs such as 'feet' – 'fit' and 'seat' – 'sit'. Therefore, the SLM predicts that

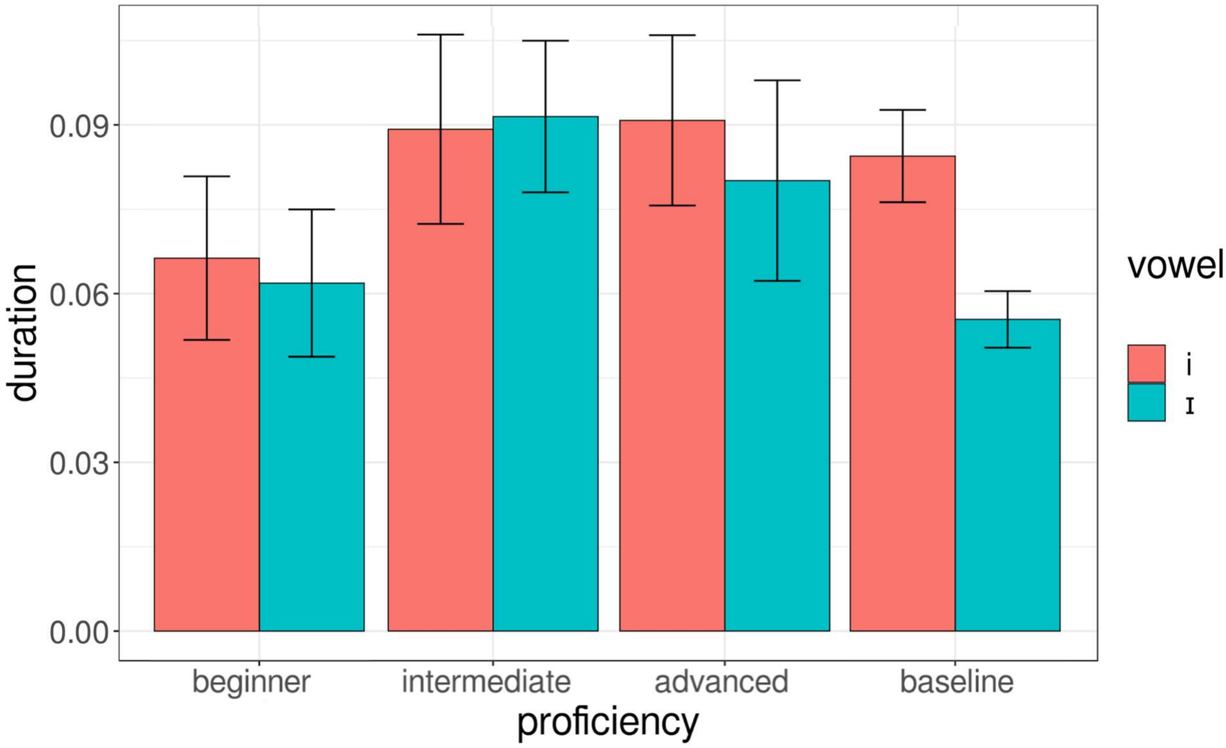
Brazilian learners of English would need to create a new category for the lax [ɪ] in order to perceive and produce distinct lexical items, as the ones used as stimuli in the present study. It is also worth reiterating that the existence of the high front vowel in both BP and English native inventories does not mean that they are produced with similar acoustic characteristics[122].

Previous studies have shown that Brazilian learners initially assimilate the high front lax vowel into the native [i] category (PEREYRON, 2017; ZIMMER; SILVEIRA; ALVES, 2009; NOBRE-OLIVEIRA, 2003). A dissimilation process is thus necessary for the new [ɪ] category to be created and, consequently, in order for the learner to distinctively produce and perceive pairs like 'feet' – 'fit' and 'seat' – 'sit'. This dissimilation process[123] tends to happen more easily than the dissimilation process of [æ], discussed in section 2.4.1 (RAUBER, 2006; NOBRE-OLIVEIRA, 2003, 2007). Based on the literature, we expected that different proficiency levels would show different stages of the [ɪ] dissimilation process.

The literature also reports that Brazilian learners tend to use extrinsic duration earlier and more often than formant frequencies when distinctively producing [ɪ] and [i] (ZIMMER; SILVEIRA; ALVES, 2009). Therefore, as we did in the previous section of this chapter, we will analyse duration before F1 or F2 values. Figure 4.16 shows relative duration patterns for productions in all proficiency levels.

[122] In this sense, we ought to highlight that indeed the native North American English [i] is produced in a higher and more fronted area than the native Brazilian Portuguese [i].
[123] There is also a similar dissimilation process concerning the native BP [i] and the L2 English [i]. As we have mentioned, the high front vowels are not produced with the same acoustic characteristics in both languages.

Figure 4.16 – Relative durations of the productions of [i] and [ɪ], according to proficiency level



Source: present study.

Figure 4.16 compares the relative durations produced by each proficiency level. A visual analysis shows that the relative durations of [i] and [ɪ] are really similar for most proficiency levels of non-native speakers. An exception can be found in the productions by the advanced learners, in which a larger distinction in duration seems to be present. We could take this small difference as portraying a developmental stage of those speakers, like a 'halfway' on the way to a more prominent and systematic distinction – as seen for the native/baseline speakers' productions.

Moreover, we can see, in general, that the tense vowel is produced by Brazilian learners with similar relative durations to those of native English speakers'. The one exception is the beginner group, in which we see shorter relative durations, closer to the native lax [ɪ]. The lax vowel produced by non-native speakers, however, is longer than baseline average durations, and this includes productions by the beginner group. Beginners have the closest [ɪ] productions to baseline tokens, with relative durations averaging a little over 6%, whereas intermediate learners show 9%, and advanced learners produce it with about 7,5%. The native [ɪ] produced by the Canadian participants has a relative duration of under 6%.

As shown in Table 4.4, longer absolute durations are actually characteristic of the Porto-Alegrense variety of Brazilian Portuguese our participants speak as their native language.

Pereyron (2017) reports an average of 145.60ms for monolingual native BP [i] duration, which is longer than what she reports for monolingual speakers of the Riverplate variety of Spanish (70.48ms). It is also longer than what Maack (1949) reports for monolingual native speakers of German (111.67ms). Comparing the Argentinian bilinguals in Pereyron's study (2017) and the Brazilian speakers in our study, averages of both [i] and [ɪ] show a longer duration in Brazilian productions – 117.85ms for [i] and 112.81ms for [ɪ] in the case of native BP speakers, in comparison to 84.87ms for [i] and 76.55ms [ɪ] in the case of native Spanish speakers. Again, we see that the Brazilian's productions of L2 [ɪ] are even longer than the Argentinians'.

Table 4.4 – Average measures of absolute duration (ms) of native and L2 vowels produced by native speakers of Brazilian Portuguese, Riverplate Spanish, German, Canadian English and North-American English

Speaker and their languages	**Vowel duration (in ms)**				
	L2 [i]	**L2 [ɪ]**	**L1 [i]**	**L1 [ɪ]**	**L1 [e]**[124]
Monolingual Brazilian native speakers of the Porto-Alegrense variety of Portuguese (PEREYRON, 2017)			145.60	-	175.13
Monolingual Argentinian native speakers of the Riverplate variety of Spanish (PEREYRON, 2017)			70.48	-	78.03
Monolingual German native speakers of L1 Silesian/Bavarian variety of German[125] **(MAACK, 1949)**			-[126] / 111.67**		-* / 149.17**
Plurilingual[127] **Canadian native speakers of the L1 Ontario variety of English (present study)**	154.41	102.53			
Monolingual North-American native speakers of L1 English (compiled by PEREYRON, 2017)	87-130***	68-103***			
Bi/Plurilingual Brazilian speakers of L1 Porto-Alegrense BP and L2 English (present study)	117.85[128]	112.81	-	-	-
Bi/Plurilingual Argentinian speakers of L1 Riverplate ES and L2 English (PEREYRON, 2017)	84.87	76.55			

*Short vowel[129]
**Long vowel
***Measures compiled from Rauber (2006) and Lima Junior (2013), apud Pereyron (2017).

Source: elaborated by the author (2021) based on the data from the present study; Pereyron (2017); Maack (1949).

Table 4.4 also allows us to see that Brazilian productions of the lax [ɪ] are too long when compared to any other L1 or L2 values for the lax vowel from language groups in our dataset.

124 As neither BP nor Spanish have the lax vowel [ɪ] sound, we list values for their closest L1 vowel in terms of F1 and F2, namely [e] in both languages.
125 See section 2.4.3 for details.
126 As mentioned in section 2.4.3, German does not have a short [i] vowel, as it does for other vowels we have analysed thus far. (Cf. KÖNIG, 2004).
127 We did not set monolingualism as an inclusion criteria for our native speakers of English, due to our difficulty in finding native speakers of English as participants. We understand that additional languages our participants speak may have had an effect on their vowels systems. See section 5.2 for details.
128 It is worth mentioning that, on average, Brazilian productions of all three levels of proficiency are shorter than what Figure 4.16 indicates for each separate level for [i]. The average is also shorter than the native [i] productions measured by Pereyron (2017).
129 As mentioned in section 2.4.3, German has distinctive long/short vowel durations (Cf. KÖNIG, 2004).

The tense [i] is produced by Brazilian learners with an average absolute duration that is within the native parameters shown in the table. The pattern for [i] and [ɪ] is, thus, somewhat similar to that of [æ], insofar as: (a) both sounds are longer than those of Argentinian speakers; and (b) the Brazilian productions of both [i] and [ɪ] are closer to German [i] and to Canadian and North-American native [i] than to [ɪ]. Those are average results, though.

Our data show that beginner learners are closer to baseline speakers in regards to the relative duration of the lax [ɪ] – allegedly, the one that will only appear in a later developmental stage. Contrary to what we would expect from the literature on native productions, we also see that intermediate learners produced longer durations for the lax than for the tense high front vowel. Finally, we see advanced speakers produce the largest distinction between the two vowels in the pair, though not (yet) in the same way native Canadian speakers do.

The description we have provided thus far could allow us to interpret these patterns as different L2 developmental stages for our learners. We could imagine that in the course of their language development, learners firstly produced an intrinsic temporal distinction only, likely using L1 temporal parameters to produce both L2 vowels. In a second stage, as vowel duration is perceived as a distinctive cue for the tense-lax pair, learners produce larger overall durations, affecting both vowels of the pair. Having found a 'long vowel' standard, so to speak, they 'leave' the long, tense vowel at that pattern and shorten the lax vowel back to the durational length they used to produce at first. We could further assume that a next step would be to enlarge the distinction we already see in the advanced learners' productions. That is, learners would need to shorten [ɪ] a bit more, as well as reduce the dispersion rates in vowel durations – though we do not have data to showcase this latter stage.

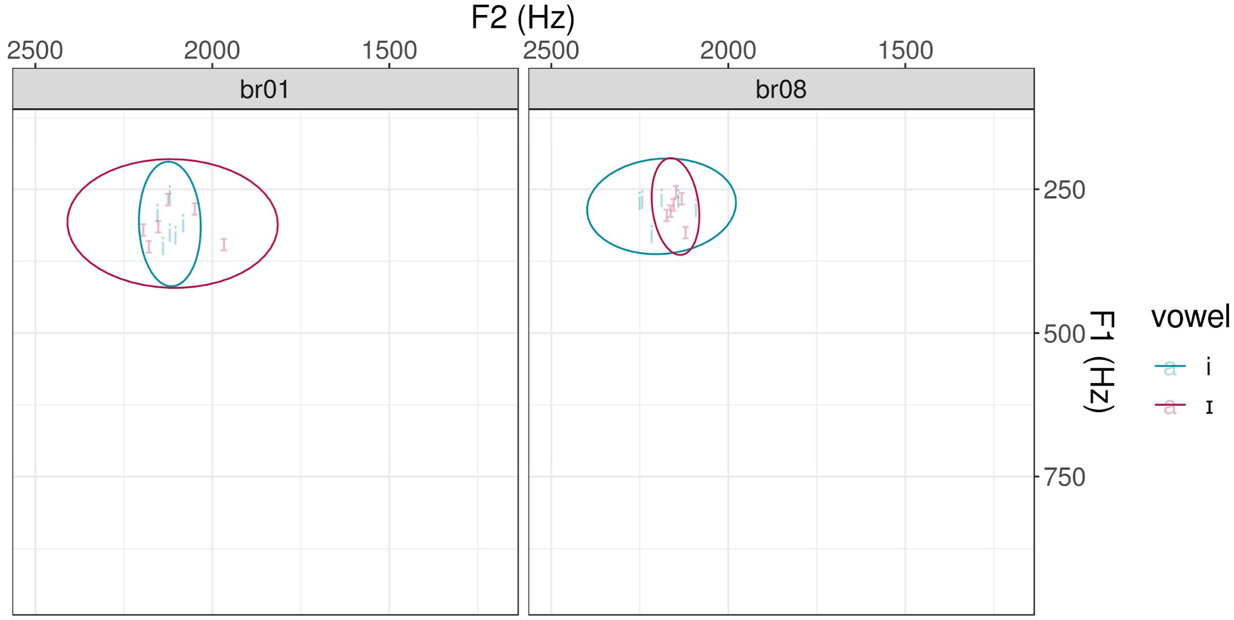
As we have mentioned, the literature points to duration as the acoustic cue that Brazilian learners rely on to distinguish the tense and the lax high front vowels. We have also pointed out that previous studies predict that the distinction between [i] – [ɪ] minimal pairs is easier to develop than the distinction between the [æ] – [ɛ] pair. Thus, the first observation we ought to make is that the patterns our study found in our participants' productions are different for both pairs, as far as relative duration is concerned. In section 4.1.1.1, we described that all [æ] productions were longer than [ɛ] productions, though not in a native-like fashion. We also described that [ɛ] productions were way too long for the native [ɛ] average duration, approximating Brazilian [ɛ] productions to native [æ] category parameters. We call these results back because our data show a similar situation for the tense and lax high front vowels, as we have previously exposed.

Though there is evidence that vowel duration is a highly informative cue for Brazilian learners, we must also look at formant frequency characteristics of those learners' productions. Spectral information is regarded as heavily weighted by native speakers of English, meaning that we cannot only look at vowel length to try and understand how acoustic characteristics of the stimuli might have effects on listeners' perception (as we will attempt to do in section 4.2.4).

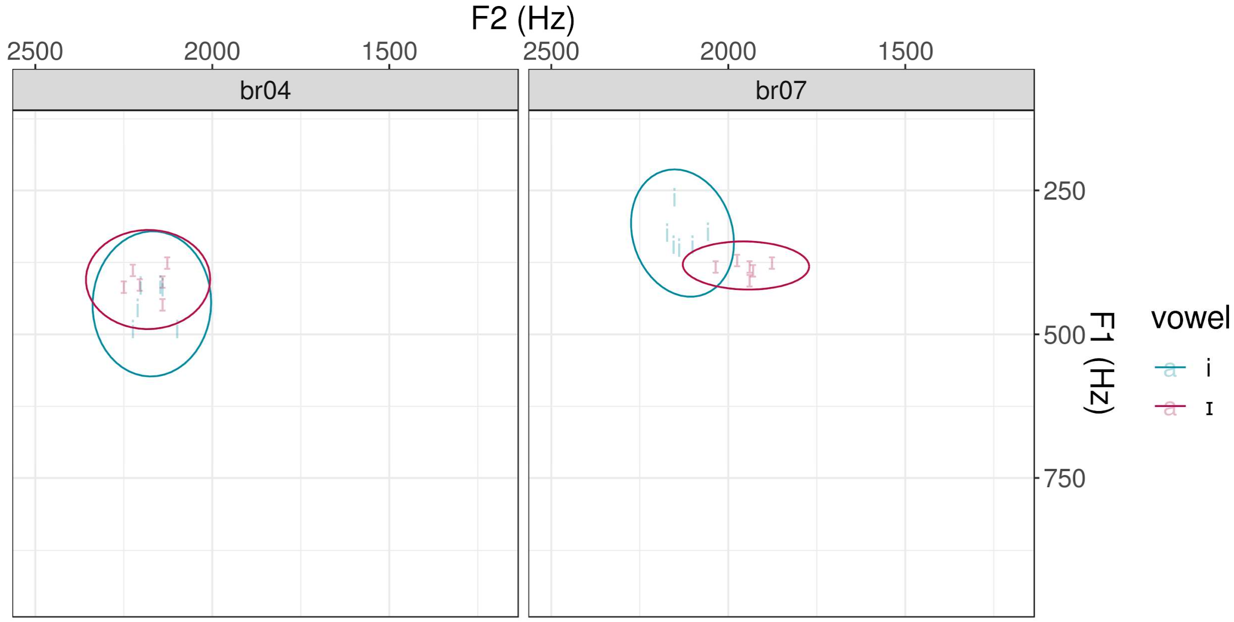
Figure 4.17 compares native speakers of English (baseline) and Brazilian learners' productions of [i] (blue) and [ɪ] (salmon), in terms of F1 and F2.

Figure 4.17 – Individual F1 x F2 plots of the vowels [i] and [ɪ], according to the participants' proficiency level

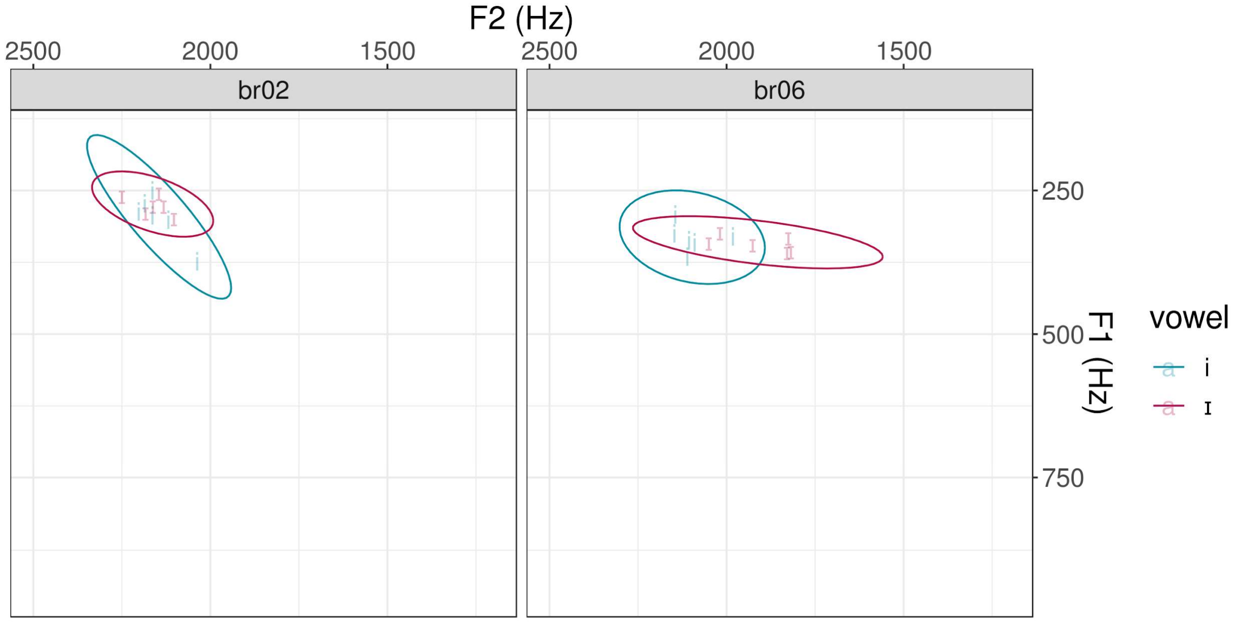
(a) beginner



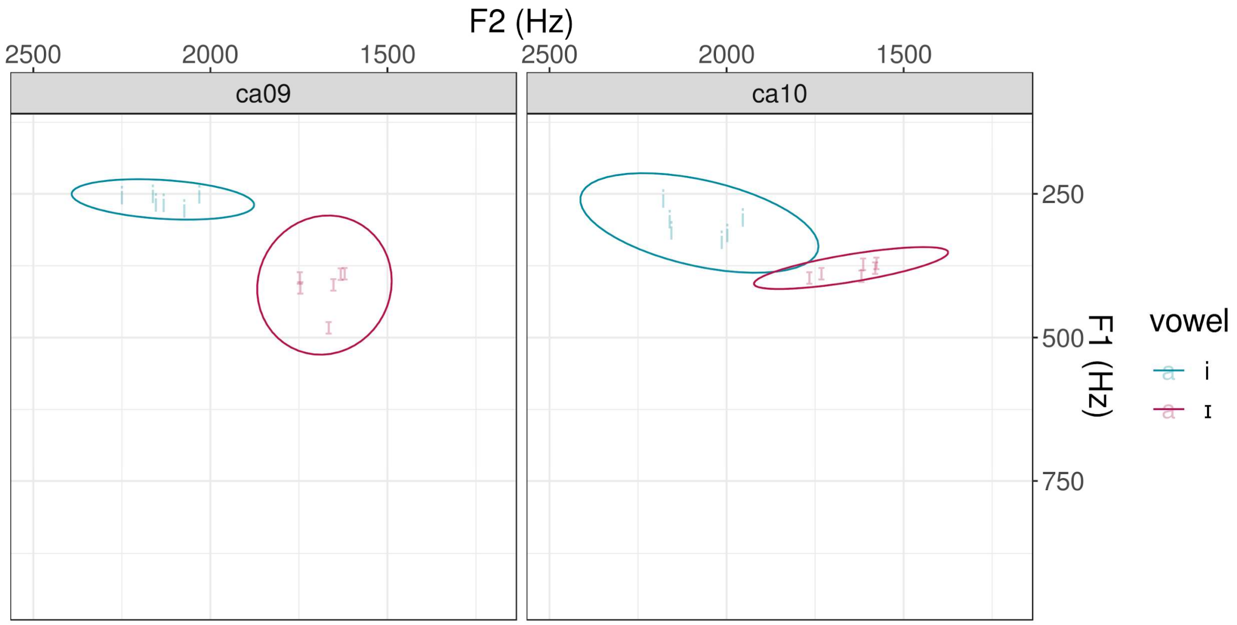
(b) intermediate



(d) advanced



(d) baseline (native speakers)



Source: present study.

As we have previously observed with [æ] and [ɛ] productions, the tense and the lax high front vowels seem to totally or partially overlap in the speakers' common phonetic space. Intermediate and advanced proficiency level speakers appear to be developing the new [ɪ] category, as the F1 values for [i] are more varied and the F2 values seem to show distinct patterns for each vowel of the pair.

When we observe the learners' productions by their proficiency level, both beginner speakers show a complete overlap of [i] and [ɪ] areas, which is consistent with the literature on the assimilation process of the lax vowel into an L1 [i] category (GONÇALVES, 2014; NOBRE-OLIVEIRA, 2003). It is also noticeable that BR01 has a larger F2 dispersion for the lax vowel. Conversely, BR08 has a larger F2 area for the tense vowel.

Compared to beginners, the vowels produced by the intermediate learners are distributed in a different pattern. Both speakers already have distinct ranges of F1 values for each segment, with [ɪ] having a more restricted area than [i]. Moreover, compared to beginners, their [ɪ] dispersion shows a higher F1 area. Additionally, BR07 shows a distinction in relation to F2 as well, in a way that [ɪ] and [i] dispersions overlap in just a small area.

The different patterns found in the F1 and F2 values of each vowel can also be observed for advanced learners. BR02 shows a larger dispersion for [i] on the F2 axis, with the lax [ɪ] already occupying a higher F1 area. BR06, on the other hand, shows a larger F2 dispersion for [ɪ], extending to a less fronted space, whereas its F1 dispersion is more condensed. Compared to baseline productions, we could expect that language development would lead BR02's L2 [i] category to be raised (and have a less variable F1 value), whereas BR06 seems to tend to leave his [i] category in the space it already occupies in the common phonetic space and lower the [ɪ] space (again, reducing the variability in his F2 values for both vowels).

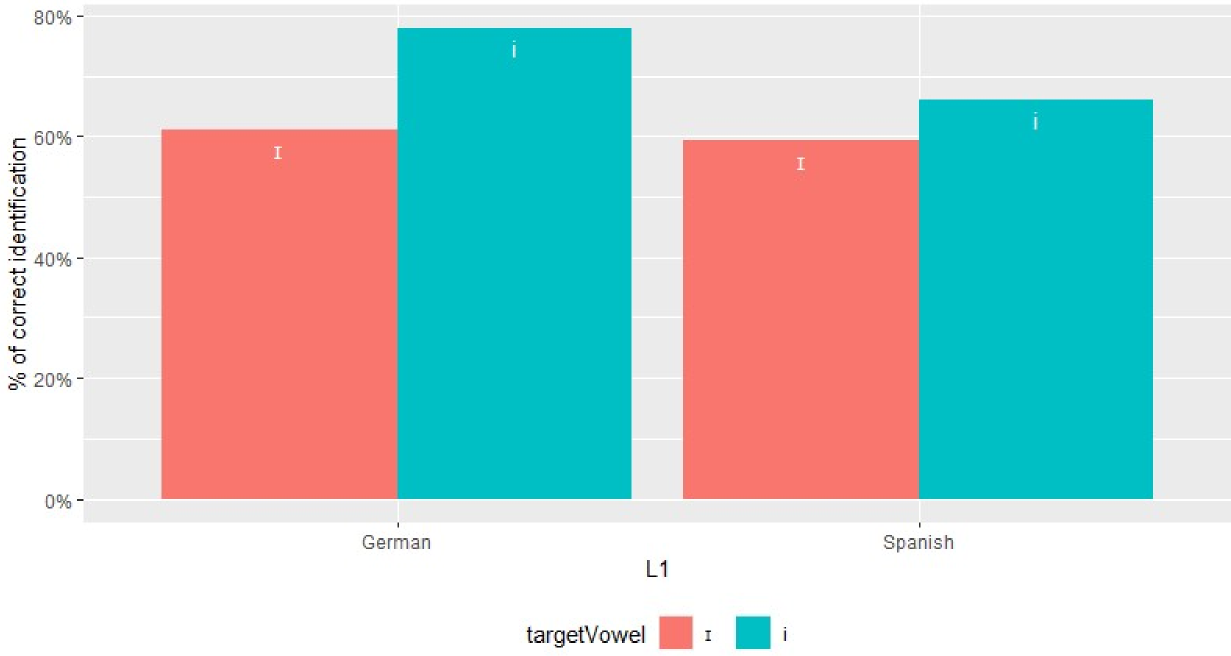
Finally, it is worth mentioning that our male intermediate speaker (BR07) appears to have the most dissimilated [ɪ] category of all six Brazilian participants. We would expect that our participants at the advanced level would also show this dissimilation, which was not the case. We take the fact that one of the intermediate speakers, but neither of the advanced speakers, is this far along in the dissimilation process of [ɪ] as a limitation to the study, because we grouped speakers by proficiency levels expecting this grouping to showcase different developmental stages (see section 3.3.1).

4.2.1.2 Stimuli perception

We will now provide a description of the perception patterns of our stimuli. This brief section aims to provide seminal information to the inferential analysis that will follow. Our description will focus on the two levels of analysis that will appear in our results, that is: word-level accuracy rates and vowel-level identification patterns, regardless of accuracy.

As mentioned, previous studies have shown that Brazilian learners of English find it easier to dissimilate the lax vowel [ɪ] than the low central [æ] from the vowel categories in their L1 (RAUBER, 2006; NOBRE-OLIVEIRA, 2003, 2007). Accordingly, Argentinian and German listeners have shown better accuracy rates when identifying stimuli of the [i] – [ɪ] pair.

Figure 4.18 – Proportion of accurate identifications of words with [i] and [ɪ], according to the listener's L1 groups

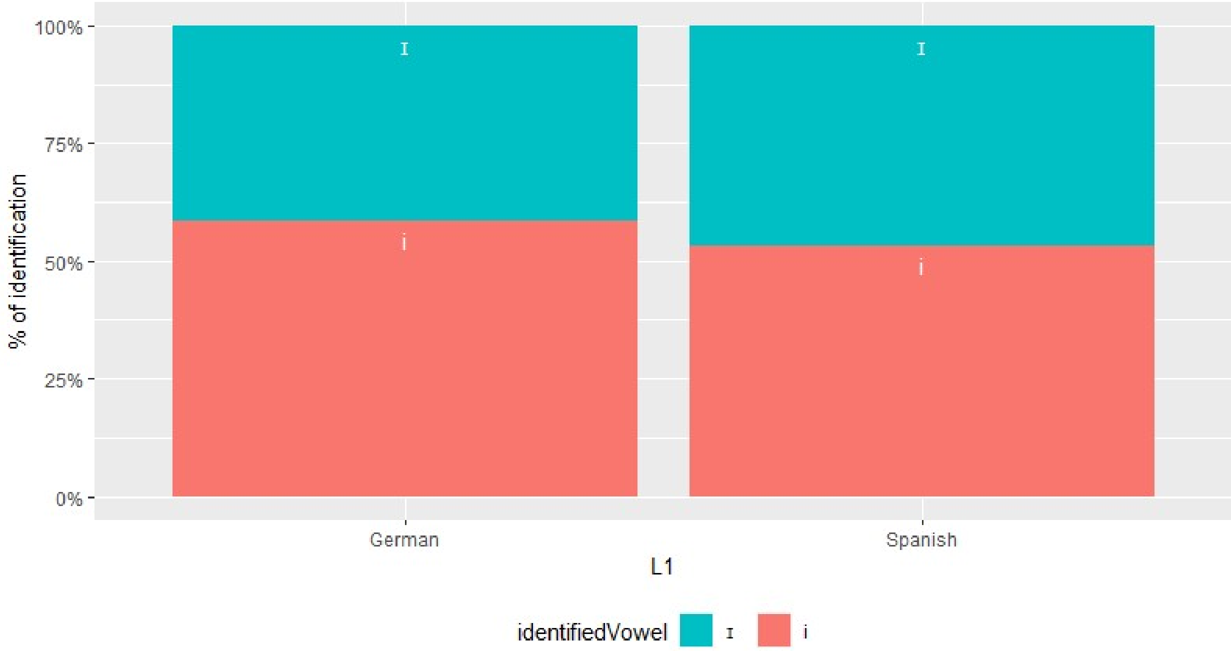


Source: present study.

Unlike what was seen in [æ] vs. [ɛ], Figure 4.18 shows that both groups of listeners have similar (and lower) accuracy rates when identifying words with the lax [ɪ]. Moreover, vowel identifications of [i] and [ɪ] also have more similar accuracy rates across the groups of listeners, especially in regard to the lax vowel, as shown in Figure 4.18. This will be discussed in the coming sections.

Figure 4.19 shows the proportion of vowel [i] and [ɪ] identifications, regardless of accuracy, by L1 listener group. As can be seen, the proportion is more balanced between the two vowels than they were between [æ] and [ɛ] for both groups of listeners.

Figure 4.19 – Proportion of vowel [i] and [ɪ] identifications, regardless of accuracy, by L1 listener group



Source: present study.

As a final remark, it is worth noticing that [i] is the only vowel present in all four vowel systems analysed in this study, namely L2 English and the three L1s (Portuguese, Spanish and German). However, we highlight yet again that the existence of [i] in more than one native inventory does not mean that the high front vowel is produced with the same acoustic characteristics across those languages. Nonetheless, we can expect that, at first, the learner will not dissimilate their L1 [i] and the L2 [i], as they are still similar, when compared to other categories. Within our theoretical framework, the SLM assumes a common phonetic space, where categories of all languages a person speaks coexist. This fact can also have an effect on the results of the present study.

We move now to reporting and discussing the inferential statistics.

4.2.2 Identification accuracy of words with [i] and with [ɪ] (RQ1)

In our first Research Question, we inquired whether the listeners' L1 has an effect on the intelligibility of a speaker's production of an L2 word. As mentioned earlier, this question arises from the SLM prediction that an 'L1 filter' will play a role in the development of additional languages. This 'filter' will act in the speaker as well as in the listener systems. As we have stated with regard to the [æ] – [ɛ] pair, this leads us to expect that cues used by the L1 system will be somewhat weighed in the perception and/or production of L2 speech, though they may not match native-like processes. Also, Munro and Derwing's (2015) intelligibility construct highlights the speaker-listener pair, which is another motivation for our choice of having two L1 groups of listeners.

As we are interested in observing intelligibility effects, our unit of analysis to determine an ‘accurate’/‘inaccurate’ identification was the target word. Our mixed-effects logistic model tested an interaction between L1 and target vowel as predictor variables. The response variable was the accurate identification of the target word. We reiterate that accuracy rates were measured as a match between target word and listener's choice, which means correct identification of onset consonant, vowel and coda consonant of the CVC monosyllables used as stimuli in the perception task. For instance, a ‘fit’ token was only considered as correctly identified if the listener chose ‘fit’ as the word they heard. Identifications such as ‘sit’ (correct vowel) or 'feet’ (correct onset and coda) were considered incorrect. Correct identifications were grouped, for the statistical analyses, by target vowel, i.e. as a categorical variable with two levels, [i] and [ɪ]. Again, we highlight that local intelligibility in the word level means understanding the speaker’s intended message, which would not be the case if a sentence like ‘this is a good fit’ (match) was heard as ‘this is a good feat[130]’ (collaborative work).

The fitted model shows that the interaction of L1 and target vowel is significant in predicting the level of identification accuracy by non-native listeners of English for words with [i]. However, the L1 in itself is not a significant predictor when the target word has the vowel [ɪ]. Estimates are provided in Table 4.5, in log-odds – lexical item is also taken as a random intercept effect[131]. The intercept shows correct identifications (response variable) by German (DE) listeners identifying target words with [ɪ].

[130] Our dataset includes the word ‘feet’ because we took word frequency into account. We use ‘feat’ in the example because ‘feat’ and ‘feet’ are homophones and ‘feat’ is more likely to cause ambiguity. As discussed in section 3.4.3.1, ‘fit’ as a noun is singular, whereas ‘feet’ is plural, which could mean that contextual information (such as verb conjugation) could counterbalance the lack of intelligibility at the segment or word level and allow for a higher level of global intelligibility – albeit with lower comprehensibility (cf. MUNRO; DERWING, 2015 for a discussion on the relationship among intelligibility, comprehensibility and accentedness).

[131] We have attempted to also use ‘listener’ and ‘speaker’, the two other repeated measures in the study, as random intercept effects. However, the models with the three random variables did not converge or were singular fits, that is “the parameters are on the boundary of the feasible parameter space: variances of one or more linear combinations of effects are (close to) zero” (BATES et al, 2015, R package, command line ‘?isSingular’). The authors also explain that "[s]ingular fits are common in practical data-analysis situations, especially with small- to medium-sized data sets and complex variance-covariance models." (*op. cit.*, p. 25)

Table 4.5 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of interacting L1 and target vowel on accurate identification rates of [i] and [ɪ]

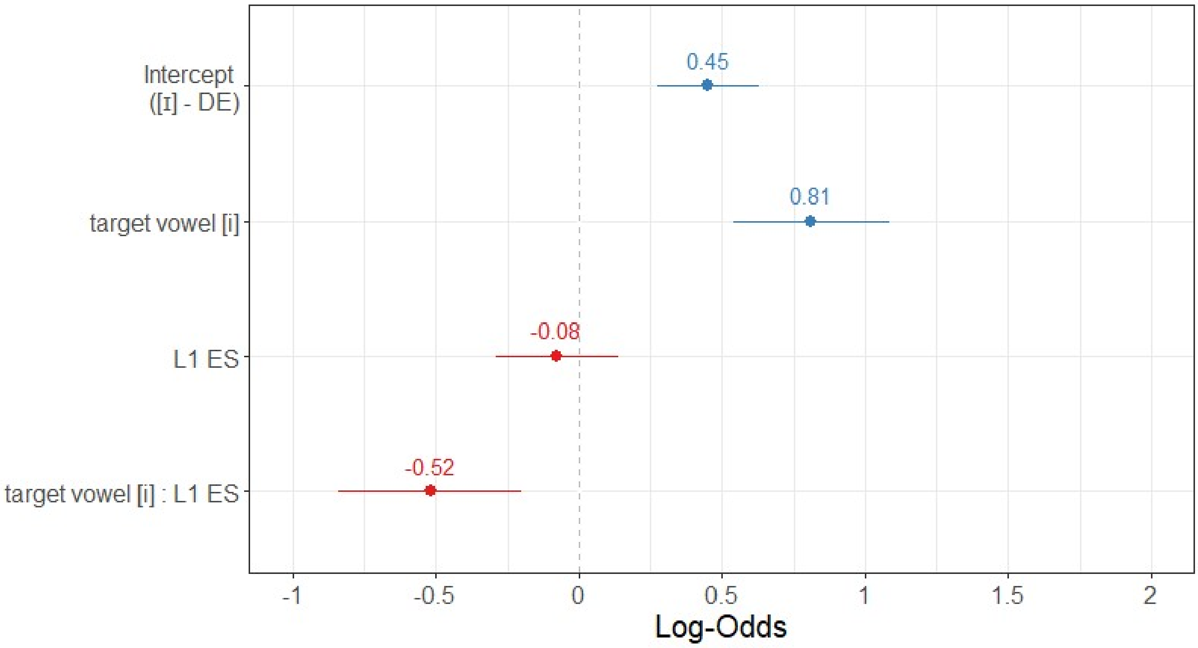
Predictors	Estimates	std. Error	z value	Pr(>|z|)
Intercept ([ɪ] – DE)	0.452	0.908	4.983	**<0.001**
target vowel [i]	0.811	0.139	5.844	**<0.001**
ES	–0.077	0.109	–0.706	0.480
target vowel [i] : ES	–0.519	0.164	–3.160	**0.002**
Observations	2,944			AIC = 3,746.4
Marginal R^2 / Conditional R^2	0.031 / 0.031			

Model: accuracy ~ targetVowel + L1 + targetVowel * L1
+ (1 | targetWord), family = binomial, data = dados)
Intercept: target words with [ɪ] by German listeners
Response variable: correct identification
Source: RStudio 1.4.1103 (2021)

The model predicts that Germans tend to correctly identify target words with vowel [ɪ] (β_0 = 0.452, p = <0.001). However, L1 is only a significant predictor in interaction with target vowel, but not on its own (p > 0.05), as it does not have an individual effect. According to the model, German listeners are also likely to accurately identify target words with the vowel [i] (β = 0.811, p = <0.001). The relationship between target vowel and L1 is not constant in the data. The difference between Argentinian learners and German learners is not the same, when we compare target [ɪ] and [i] conditions. As seen in Figure 4.18, on the one hand, the identification rates of the lax vowel by the two groups of listeners is very close – which might explain why L1 was not a significant predictor on its own. On the other hand, Germans present a higher accuracy rate for [i] identifications than Argentinians do (negative sign in the interaction, β = –0.519, p = 0.002).

Figure 4.20 presents the model's estimates. Blue lines indicate that the model estimates a tendency towards correct identifications, whereas red lines portray a tendency towards inaccurate identifications, considering the conditions expressed in the intercept. Estimates are also provided in log-odds.

Figure 4.20 – Log-odd estimates of the mixed-effects logistic model for effects of interacting L1 and target vowel on accurate identification rates of [i] and [ɪ]

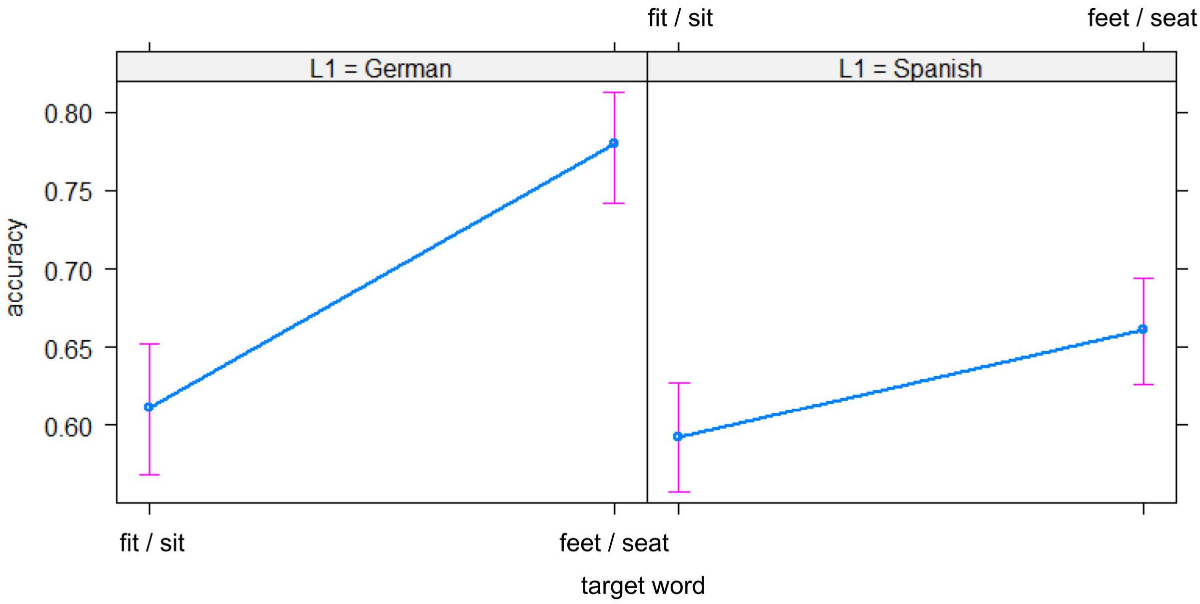


Source: RStudio 1.4.1103 (2021)

Figure 4.20 shows predicted performances for both L1 groups based on the mixed-effects logistic model. As mentioned before, the native language is not a significant predictor of accurate identification *per se* (red estimate line crossing the dashed line), only in interaction with the target vowel. That is, as we have mentioned, the listener's L1 predicts different estimates for accurate identifications only in interaction with the target vowel. Once again, as we see in Figure 4.18, target words with [ɪ] yield similar accuracy rates for both Germans and Argentinians, but for tokens with [i] the German listeners' rates are much higher than that of the Argentinian listeners'.

Below, Figure 4.21 plots the mixed-effects logistic model's predicted effects of interacting L1 and target vowel on accurate identification rates of words with [i] and [ɪ]. The plots group 'feet' and 'seat' together as tokens of words with [i], as well as 'fit' and 'sit' as tokens with [ɪ]. Nevertheless, we reiterate that accuracy indexes were calculated as a perfect match between target word and identified word. That is to say, 'feet' was considered intelligible when it was identified as 'feet', but not when it was identified as 'seat' (correct vowel identification, but incorrect word identification).

Figure 4.21 – Mixed-effects logistic model's predicted effects of interacting L1 and target vowel on accurate identification rates of words with [i] and [ɪ]



Source: RStudio 1.4.1103 (2021)

In Figure 4.21, it is possible to see that the standard error bars (pink) of both Germans and Argentinians predicted accurate identifications are very similar for lax vowel [ɪ]. In other words, the performance of both groups is pretty similar, as the overlapping standard error bars allow us to see. That is consistent with the lack of significance of the L1 predictor. Performance predictions for the tense vowel, however, once again show that German listeners tend to have better accuracy rates in target word identification. In Figure 4.18, provided in section 4.2.1.2, we have already seen the same patterns just described.

The fitted model and the descriptive analysis (such as in Figure 4.18, already shown in section 4.2.1.2) seem to show that productions with the tense vowel [i] are generally more intelligible than their minimal pair counterparts with the lax vowel [ɪ]. As [i] corresponds to a vowel category present in Brazilian Portuguese, it seems to cause less difficulty for learners to produce it in English. However, it should be noted that vowel quality and temporal characteristics of BP [i] might not coincide with those of English [i]. For instance, as we have seen in Table 4.4, Pereyron (2017) measured average monolingual Brazilian [i] duration in 145.60ms, and the same author reported native North-American [i] duration as an average of 87ms to 130ms. In other words, insofar as duration, Porto-Alegrense BP [i] productions tend to be longer than native North-American [i] productions. Moreover, as the SLM assumes a common phonetic space, we can also observe that formant frequencies of the tense vowel have distinct boundaries for native BP speakers and native English speakers. Table 4.6[132] portrays such a difference, along with values for native German [i] production and native Spanish [i] productions.

Table 4.6 – Average measures of F1 (Hz), F2 (Hz) and absolute duration (ms) of the native [i] vowel produced by native speakers of Porto-Alegrense Brazilian Portuguese, Riverplate Spanish, German, Ontario Canadian English and North-American English

Speaker and their languages	L1 [i] cate		
	F1 (Hz)	F2 (Hz)	duration (ms)
Monolingual Brazilian native speakers of the Porto-Alegrense variety Portuguese (PEREYRON, 2017)	361.3	1,936.5	145.60
Monolingual Argentinian native speakers of the Riverplate variety of Spanish (PEREYRON, 2017)	343.6	2,105.7	70.48
Monolingual German native speakers of the standard variety of German[133] (SENDLMEIER; SEEBODE, 2010)	275.9	2,313.6	111.67*
Plurilingual[134] Canadian native speakers of the L1 Ontario variety of English (present study)	325.1	2,706.1	154.41
Monolingual North-American native speakers of L1 English (compiled by PEREYRON, 2017)	270.0-249.0**	2,290.0-2,790.0**	87-130**

*Long vowel[135]
**Measures compiled from Rauber (2006) and Lima Junior (2013), apud Pereyron (2017).
Source: elaborated by the author (2021) based on the data from the present study; Pereyron (2017); Sendlmeier and Seebode (2010).

The existence of a previous, albeit L1-based category could also explain why tokens with that vowel were more intelligible to our listeners. However we cannot assume this beforehand. We called forth the measures presented in Table 4.6 to highlight that the existence of a native category does not presuppose that this category's acoustic specifications will be either a match to a given native (L2) standard, or necessarily intelligible. Besides, as we have argued, 'intelligibility' is highly dependent on the speaker-listener pair, which means that many

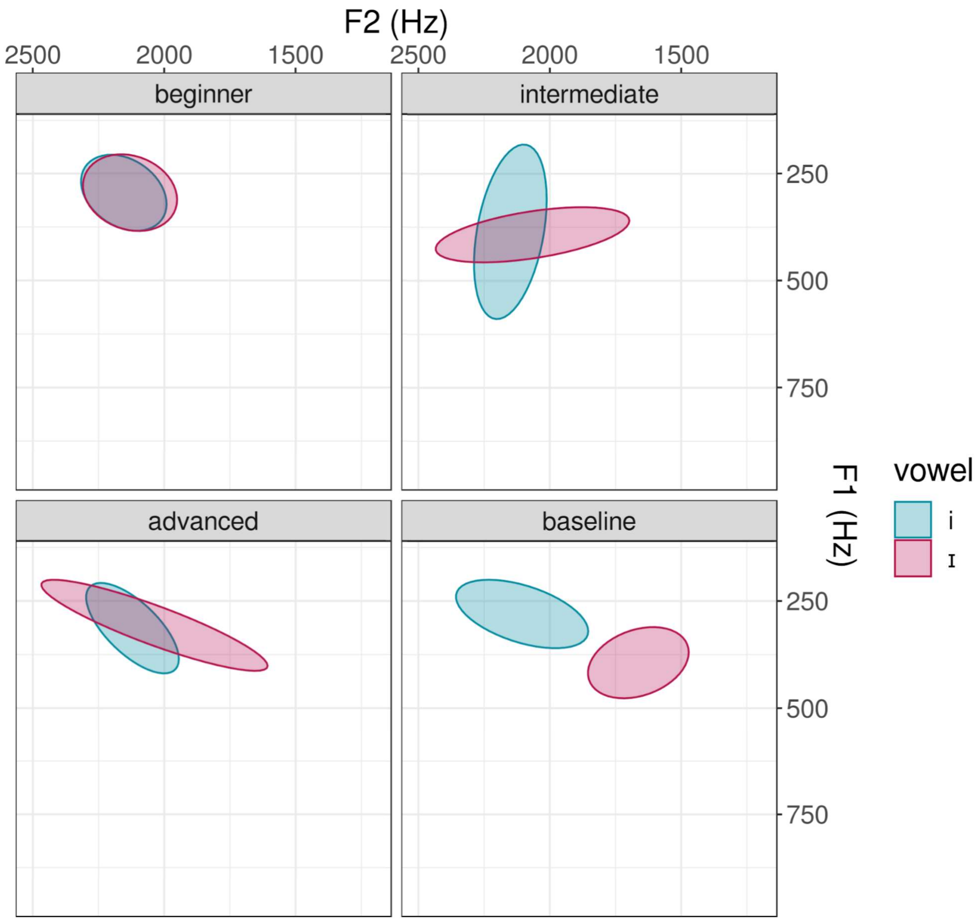
[132] Again, we provide this table because we understand it relevant, given that [i] is a native category in all four languages in interaction in this study – namely L2 English, L1 BP, L1 German and L1 Spanish.
[133] See section 2.4.3 for a detailed account.
[134] We reiterate that we did not set monolingualism as an inclusion criteria for our native speakers of English, due to our difficulty in finding native speakers of English as participants. We understand that additional languages our participants speak may have had an effect on their vowels systems. See section 5.2 for details.
[135] As mentioned in section 2.4.3, German does not have a short [i] vowel, as it does for other vowels we have analysed thus far (Cf. KÖNIG, 2004).

systems might be interacting simultaneously. That is precisely the case of this study, in which we have speaker's L1, listener's L1 and their shared L2 English, which might entice two distinct L1-L2 composite categories – the listener's and the speaker's. We will discuss this further in section 4.2.4, but for now we want to highlight that the model estimates higher accuracy rates of identification of target words with [i], a category present in all four language systems.

Nevertheless, those tokens seem to present a certain degree of unintelligibility to native speakers of Riverplate Spanish, who have lower accuracy rates for the [i] vowel (66,07%) than their German counterparts (77,95%). A tentative explanation to explain this result could be that the L1 Spanish 'filter' leads perception to a non-categorisation scenario. Pereyron (2017) reports that BP native [i] is produced at a similar height to Riverplate Spanish [i], but Brazilian monolinguals produce the vowel in a much more posterior position than Argentinian monolinguals do. Details are provided in Table 4.6. Thus, we could hypothesise that the relative ease with which Brazilian learners produce [i] tokens in English comes from producing it in a similar fashion to what it is produced in Portuguese, which in turn lowers Brazilian English intelligibility for Argentinian listeners, as far as [i] is concerned. We will look at that in more detail in section 4.2.4.

As for the lack of intelligibility of tokens with the lax vowel [ɪ], as predicted by our mixed-effects logistic model, a speculative explanation could be that Brazilian learners did not produce a (sufficient) distinction between tense and lax vowels. As Brazilian learners tend to assimilate both [i] and [ɪ] as tokens of the [i] category, dissimilating the [ɪ] category should be part of their learning trajectory, in order to enhance learners' local intelligibility. In our dataset, proficiency seems to only slightly reflect that ongoing dissimilation. Figure 4.22 shows the vowel dispersions of [i] and [ɪ] productions in English grouped by proficiency level, in the same fashion that Figure 4.17 has previously shown the same dispersions by participant in each proficiency level.

Figure 4.22 – Vowel dispersion of [i] and [ɪ] productions in English according to speaker's proficiency level



Source: present study.

As seen in Figure 4.15, intermediate and advanced learners show signs of ongoing development of the new category, though one that is not yet fully dissimilated. This will be further discussed in section 4.2.4, but for now we want to point out that there is still a large, albeit partial overlap between the [i] and [ɪ] areas in the common phonetic space of those intermediate and advanced learners. This could indicate that the two categories are not yet fully dissimilated. Therefore, this could be a tentative explanation for why some productions are not (yet) as distinctive as it would be needed for higher rates of local intelligibility.

When taken together with the fact that an accurate identification rate differs between Argentinian and German listeners, the lack of distinction in production could mean that one or more acoustic cues are being perceived differently by each group of listener. In other words, as [i] and [ɪ] tokens are produced in totally/partially overlapping areas, local intelligibility of these tokens is also dependent on listeners' perception, reinforcing the importance of analysing perception and production as a factor of the speaker-listener pair.

4.2.3 Effects of predictor variables on the identification of [i] and [ɪ] (RQ2)

Thus far we have analysed accurate word identification, and we have repeatedly pointed out that it was operationalised as correct identification of onset consonant, nuclei vowel and coda consonant of the CVC monosyllables that our participants recorded in the production task. As vowel identification is part of the word (correct) identification process, we will now turn our attention to how vowels were identified by listeners, regardless of accuracy. Our second Research Question asked which predictor variables can explain the identification of a token as having [i] or [ɪ].

In order to answer this question, we built a mixed-effects logistic model. Initially, we intended to check each possible interaction between predictor variables, as well as their effect on their own. Therefore, we started off with the listener's L1, speaker's proficiency level, target vowel, F1, F2 and relative vowel duration as predictor variables. All possible interactions (signaled by the '*' character) amongst those predictors were also included at first: L1 * proficiency level, L1 * target vowel, L1 * F1, L1 * F2, L1 * relative duration, proficiency level * target vowel, proficiency level * F1, proficiency level * F2, proficiency level * relative duration; target vowel * F1, target vowel *F2, target vowel * relative duration; F1 * F2, F1 * relative duration, and F2 * relative duration. Predictors that did not yield significance were excluded from the model. This was done because we had only 2,944 data points, and models with too many predictors do not work properly with small datasets (LEVSHINA, 2015; GRIES, 2013). After removing non-significant predictors, a second fit was run, from which once again non-significant predictors were excluded. This was done consecutively and resulted in all interactions being non-significant. We then used L1, proficiency level, target vowel, F1, F2 and relative duration as predictors, only this time without interactions. Some predictor variables showed high collinearity rates and were also excluded. As will be seen further on, L1 did not yield significance, but we kept this variable in this model because it was considered to be an essential variable, as the 'L1 filter' is the underlying motivator of all three of our research questions.

The following results are thus estimated by a model in which L1, target vowel, F1 and F2 are the predictor variables. As for random intercept effects, we tried including all three repeated measures (listeners, speakers and lexical items), but models did not converge or were singular fits. Therefore, we only succeeded in keeping 'listener' as a random intercept effect. The response variable is the identification of a token as having the vowel [i]. Once again, it should be said that this does not take accuracy into account, nor does it distinguish target words ‘feet’, ‘fit’ ‘seat’ or ‘sit’. The analysis of vowel identification turns the spotlight to the listener and how they are processing the stimulus. Therefore, the intercept estimates the log-odds of a vowel being identified as [i] when a German participant (DE) identifies a token that was produced from a word with the target vowel [ɪ].

Our mixed-effects logistic model shows that target vowel, F1 and F2[136] are all significant predictors of a listener’s vowel identification. By vowel we mean a binary variable [i] or [ɪ], the first one comprising identifications as both tokens ‘feet’ and ‘seat’, and the latter, tokens identified as ‘fit’ and ‘sit’. Correct matches between target vowel and identified vowel are not accounted for either. Table 4.7 shows the estimates in log-odds, and Figure 4.23 shows a plot of estimates for effects of L1, target vowels, F1 and F2 on vowel identification (regardless of accuracy) of response variable [i].

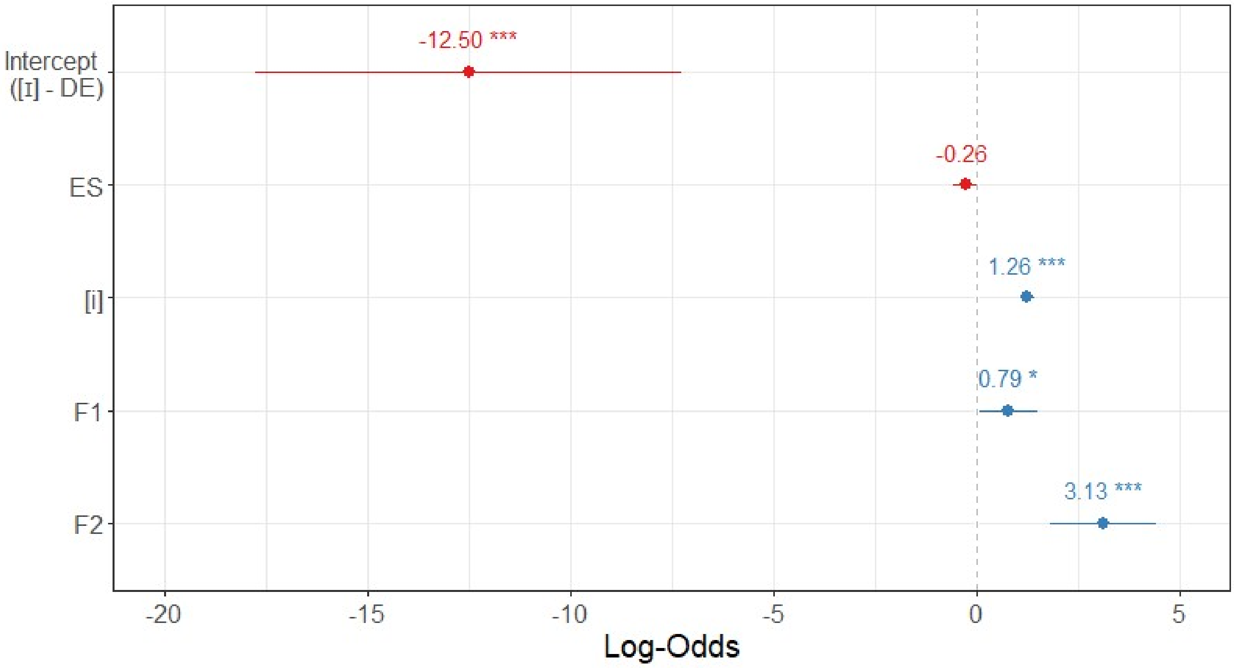
Table 4.7 – Mixed-effects logistic model estimates and associated standard errors, z-values, and p-values for effects of L1, target vowels, F1 and F2 on vowel identification (regardless of accuracy) of response variable [i]

Predictors	Estimates	std. Error	z value	Pr(>|z|)
Intercept ([ɪ] – DE)	–12.505	2.682	–4.663	**<0.001**
ES	–0.255	0.151	–1.688	0.091
target vowel [i]	1.256	0.091	13.774	**<0.001**
F1	0.795	0.356	2.232	**0.026**
F2	3.128	0.673	4.646	**<0.001**
Observations	2,917			AIC = 3,620.1
Marginal R^2 / Conditional R^2	0.142 / 0.186			

Model: accuracy ~ targetVowel + L1 + logF1norm + logF2norm + (1 | listener)
Intercept: target vowel [ɪ] by a German listener
Response variable: vowel identified as [i]
Source: RStudio 1.4.1103 (2021)

[136] F1 and F2 values were rescaled to log, as suggested by the RStudio modelling script. This suggestion was made because frequency measures had indexes that were too different to compare.

Figure 4.23 – Log-odd estimates of the mixed-effects logistic model for the effects of L1, target vowel, F1 and F2 on vowel identifications of [i]



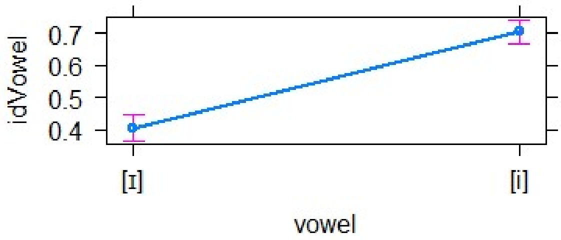
Source: RStudio 1.4.1103 (2021)

The model estimates that tokens with the target vowel [ɪ] will be likely identified as [ɪ] tokens by German listeners ($\beta_0 = -12.505$, $p = <0.001$). We reiterate that logistic models predict the probability of the occurrence of a phenomenon versus its *non*-occurrence. Our task design was a forced-choice, so 'not' identifying a token as [i] necessarily means (in our design) identifying it as [ɪ][137]. Thus, the negative estimate (lines in red) is interpreted as a tendency towards [ɪ] identifications, as [i] is the response variable. The model also estimates that tokens with the target vowel [i] will likely be identified as [i] tokens ($\beta = 1.256$, $p = <0.001$). Those estimates refer to both L1 groups of participants, as L1 was not a significant predictor of vowel identification (red estimate line crossing the dashed line).

Figure 4.24 shows the mixed-effects logistic model's predicted effect on [i] vowel identification considering target vowel as a predictor variable.

[137] A total of 35 tokens identified as [i] or [ɪ] from target words with [æ] or [ɛ] were excluded from the complete dataset. This was done because of the disproportion of cases (1,19%). A lack of enough variability could have hurt the model's fitting.

Figure 4.24 – Mixed-effects logistic model's predicted effect on [i] vowel identification considering target vowel as a predictor variable

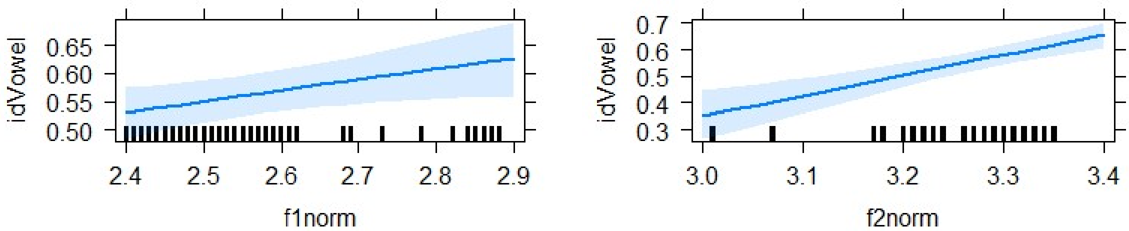


Source: RStudio 1.4.1103 (2021)

The model also estimates that the higher the F1, the more likely a token is to be identified as [i] (β = 0.795, p = 0.026). This is not consistent with the literature on native productions, in which [ɪ] is produced with higher F1 values (lower height) than [i]. A tentative explanation for this phenomenon could be that our data shows a high variability in F1 values for [i] productions by the Brazilian learners that took part in the present study. As we have seen in section 4.2.1.1, one of our participants (BR04) had lowered both high front vowels, whereas another had expanded de F1 boundaries of her [i] to encompass [i] and [ɪ] prototypical F1 values. We will look closely at acoustic data in section 4.2.4, but for now we want to point out that this result is not entirely surprising. Along with the high variability in F1 shown by productions of [i], we understand that listeners are taking more than one acoustic cue into account, as we will explore further in our next Research Question. Thus, our Complex, Dynamic view of language would lead us to assume that this 'reverse pattern' of identification is an emerging effect of the interaction of the hybrid systems at play. This will also be further discussed in section 4.2.4.

Figure 4.25 plots the mixed-effects logistic model's predicted effect on [i] vowel identification considering F1 (left) and F2 (right) as predictor variables, as we have been discussing.

Figure 4.25 – Mixed-effects logistic model's predicted effect on [i] vowel identification considering F1 (left) and F2 (right) as predictor variables



Source: RStudio 1.4.1103 (2021)

For native productions, Figure 4.25 should yield a negative slope coefficient for F1, that is, an inclination downwards. This would indicate that the higher the F1, the less likely a token would be identified as [i], as F1 is the spectral inverse correlate of vowel height. We can also see that the F2 slope is more acute. This is a visual way of portraying the fact that, by our model's estimates, F2 is having a stronger effect on vowel identifications on our study than F1. Unlike F1, the model predicts that higher F2 values will enhance the log-odds of [i] identification ($\beta = 3.218$, $p = <0.001$). The model estimates for the F2 effects on identification are consistent with the literature. In native speech, [i] is produced in a more fronted position (higher F2) than [ɪ]. In relation to F1, we have already noticed that Brazilian productions show slightly less varied F2 values for [i] and [ɪ] (see section 4.2.1.1).

Once again, this prediction may be a result of the lack of distinction in Brazilian productions of [i] and [ɪ] tokens. As already mentioned in the previous section, Brazilian learners produce the tense and the lax vowels in a totally or partially overlapping area of the common phonetic space. This could be seen as an indication that both vowels of English are being assimilated into the same [i] category of Brazilian Portuguese. Moreover, as the [i] category is present in BP, it could be the case that English tokens are being produced with characteristics of the L1 rather than L2 category, or a composite L1-L2 category (FLEGE; BOHN, 2021). Additionally, if we recall Figure 4.17, our intermediate participants show a distinct pattern from other non-native proficiency groups we collected data from. Both speakers had lowered their [i] productions, in comparison to other Brazilian speakers that participated in the present study. BR04 [i] productions were showed F1 values that were even higher than her [ɪ] productions. Meanwhile, BR07 had the smallest overlap between [i] and [ɪ] dispersions amongst all non-native speakers in the present study.

Finally, it is worth noticing that F1 is a significant predictor of [ɪ] – [i] identification, but not of [æ] – [ɛ]. If we consider each minimal pair as a subsystem of the vowel system, we can hypothesise that they are in different developmental stages. Indeed, previous research has shown evidence that [æ] is harder for Brazilian learners to perceive and produce than [ɪ]. It is also possible that F1 is having an emerging effect on [i] and [ɪ] identifications, which it did not for [æ] and [ɛ]. That, however, would bear the question if this effect is speaker/stimulus-related, listener-related, or a hybrid scenario with all of them. Indeed, as we will detail in section 4.2.4, the Complex, Dynamic view of language we adopt leads us to believe that these effects are emerging precisely from the interaction of the hybrid systems of the speaker-listener pair. The characteristics we recalled of the intermediate speakers' productions seem to point to that.

Lastly, we should also note that the response variable ‘identified vowel’ does not have a balanced number of tokens. Target vowels were controlled prior to the identification task (see section 3.4.4), whereas Identified vowels are the result of the perception task. We have already illustrated in Figure 4.19 that there is a different proportion of tokens identified as [i] and [ɪ] by each group of participants. Germans (18 participants) identified 676 tokens as having [i] and 483 as having [ɪ], whereas Argentinians (28 participants[138]) identified 948 tokens as having [i] and 845 as having [ɪ]. This is relevant because, though both groups favour [i] identifications, they do not do so with the same results. This will be the focus of the analysis in the next section.

In the next section, we will further explore the developmental stages of our participants’ systems and their possible effects on perception.

4.2.4 Exploratory analysis of listeners’ composite L1-L2 categories for [i] and [ɪ] (RQ3)

This section will be organised in the same fashion as the one regarding [æ] and [ɛ] tokens. We turn our attention now to our third Research Question, which inquires how the identification patterns in the perception task can shed light on the listeners' composite L1-L2 categories.

Hitherto we have taken a closer look at the intelligibility of Brazilian productions, that is, the accuracy rates of target word identification. This answered RQ1. We have also analysed how vowels were identified by listeners in the two L1 groups, regardless of accuracy. That was the aim of RQ2. The answer to those previous questions was obtained by means of inferential statistics, which yielded significant predictor variables of word identification accuracy and general vowel identification tendencies. In RQ3, we want to observe gradient phenomena that might have been overseen by group statistics, as well as provide a stimuli-by-stimuli analysis of the perceptual patterns followed by our two groups of listeners.

As mentioned in section 4.1.4, we understand that language is a Complex, Dynamic system, and that phonology, morphology, semantics, etc. are subsystems of this system. Likewise, a speaker is taken as a subsystem of the speech community system. All these systems and subsystems are constantly adapting, as a result of an individual's embodied experiences. We take the data collected in the present study as a portrait of the listeners' systems at the moment they participated in the experiment.

---

[138] The uneven number of participants is also an unbalancing factor. This will be further discussed in section 5.2.

It is also worth reiterating that a complex system changes as a result of the interaction of its subsystems. We will highlight again that those changes cannot be linearly presumed from the sum of each subsystem's changes, as complexity entails that interaction can render local, small shifts as having a large impact on general alterations. This emerging nature of change is the ground for our two different approaches in our research questions. RQ1 and RQ2 employed inferential statistics to find predictor variables in the identification process. RQ3, on the other hand, will use exploratory and individual analyses to turn the spotlight to emerging phenomena that might not have been picked up by the model fittings. Moreover, as we understand that each different learning trajectory should yield a distinct systems status, we expect that an exploratory analysis can shed light on gradient processes that would only show effects on inferential statistics if we had a huge dataset – and possibly not even then. In sum, this section attempts a more process-based-like analysis than a product-based one (LOWIE, 2017), combining the benefits of both methods to try and understand how listeners' L1-L2 composite categories are established and which effects can emerge when they interact with the speakers' productions.

Our previous sections have already analysed the speaker-listener binomial as one aspect of the communication between non-native learners of a foreign language (RQ1). We have noticed that the interaction between a Brazilian learner and an Argentinian learner will have different success rates than that between a Brazilian and a German non-native English speaker. We have also looked at which stimulus-related variables can have an effect on how listeners identify the vowels under analysis, as well as how those listeners' L1 may also be of relevance. Thus, we already know that our data suggest that the L1 is not a significant predictor for the [i] – [ɪ] pair (as both groups seemed to exhibit the same identification patterns)[139], though it was for [æ] – [ɛ]. This difference suggests that different aspects may play a role in the success of communication, as well as in speakers and listeners' production and perception – considering that they co-evolve (FLEGE; BOHN, 2021).

On the one hand, previous studies have already shown that the distinction between the tense and the lax high front vowels is easier to develop than the one between [æ] and [ɛ] (RAUBER, 2006; NOBRE-OLIVEIRA, 2007). Additionally, the tense [i] is a vowel that is present in all the three native languages of our participants as well as in their shared L2, whereas the lax [ɪ] is only native to the German system. In turn, with [æ] and [ɛ], we had [ɛ] in both Portuguese and German, but not in Spanish, and [æ] in none of the three L1s, only in L2 English.

[139] L1 is only a significant predictor in interaction with target vowel. See section 4.2.3.

The fact that a vowel is part of a native system is of particular relevance because it may lead to a stronger presence of L1 characteristics being passed on to L2 perception and production. It can also make a stronger case for an assimilation process, as there could be a higher perceived similarity[140] between an L1 category and an L2 sound, according to the SLM (FLEGE, 1995; FLEGE; BOHN, 2021). This, in turn, could have effects on both perception and production. Thus, we are talking about three possible dimensions, namely: assimilation in perception and/or production by Brazilian learners (speaker group in this study), as well as perception and/or production by Argentinian/German learners (listener groups). Moreover, as Flege and Bohn (2021) point out, perception and production co-evolve, so we could be looking at a fourth scenario: that of a simultaneous adaptation of both processes (on the speaker side and on the listener side). This hybrid nature of the systems, as well as their interaction (in a communicational setting) is what might have eluded the statistical approach we used in previous sections, as we have mentioned before, justifying our methodological choice for the present analysis.

Therefore, as explained in section 4.1.4, our exploratory analysis will look closely at the subset of above-chance accurate identifications and above-chance identification patterns. This was done for [i] and [ɪ] in the same fashion as it had been for [æ] and [ɛ]. We refer to the details provided in that section, and allow ourselves to reiterate only that this choice allowed us to look at more consistent behaviours, to which we refer as 'patterns'. Additionally, we highlight that the discussion in the present section takes the vowels as the unit of analysis to determine an ‘accurate’/‘inaccurate’ identification in relation to the target. As when looking at accurate identifications, we grouped the vowels binarily. That is, [i] vowels as the ones in tokens identified as 'feet' and 'seat', and [ɪ] vowels from tokens identified as 'fit' and 'sit'. Accuracy ratings were also obtained the same way as in section 4.2.2, when not only the vowel, but also the onset and the coda consonants were correctly identified ('fit' identified as 'fit' is an accurate [ɪ] identification, but 'fit' identified as 'sit' or as 'feet' is an inaccurate [ɪ] identification).

As with previous sections, we will analyse duration x F1 prior to F1 x F2[141], since the Brazilian learners’ dissimilation process of [ɪ] from [i] tends to emphasise relative duration, rather than spectral dimensions. Vowels will also be presented in the same order as they have been so far, [i] and then [ɪ], for the duration of the higher vowel is longer. Figure 4.26 shows

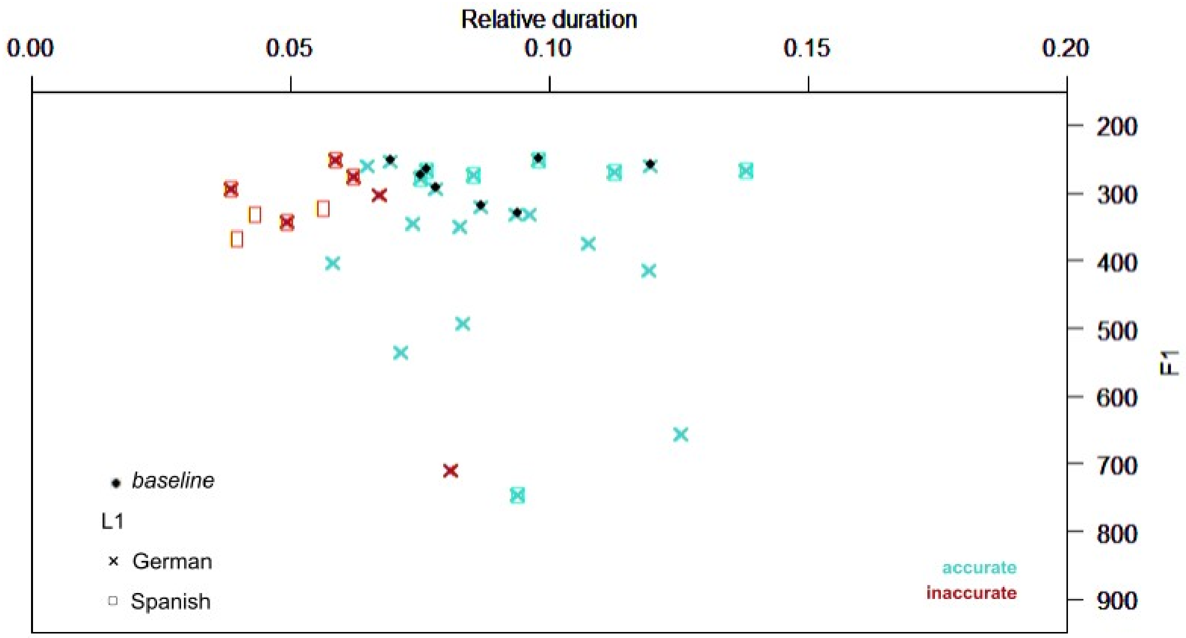
---

[140] Flege and Bohn highlight that methods for attesting (dis)similarity still need to be polished. “It remains to be determined how best to measure cross language [*sic*] phonetic dissimilarity. The importance of doing so is widely accepted but a standard measurement procedure has not yet emerged (for discussions see Bohn, 2002; Strange, 2007).” (*op. cit.*, p.30)

[141] As we have mentioned, we follow Escudero (2009) in this concern.

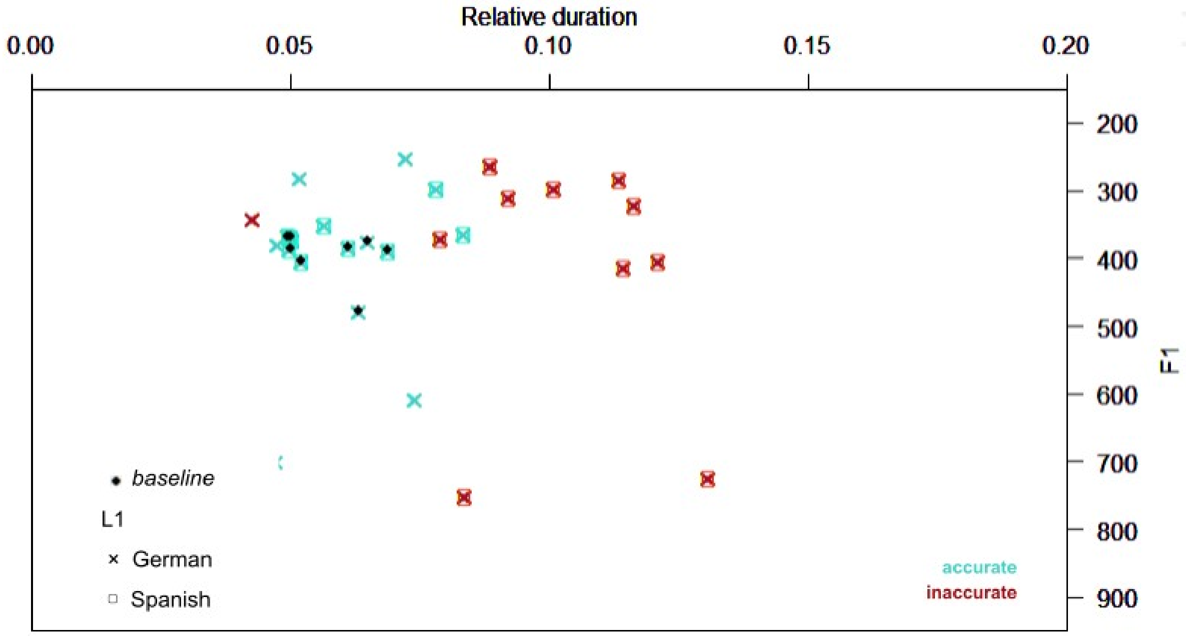
F1 x relative duration plots of [i] and [ɪ] tokens with above-chance correct and incorrect identification, by target vowel.

Figure 4.26 – F1 x relative duration plots of [i] and [ɪ] tokens with above-chance correct and incorrect[142] identification, by target vowel
(a) tense [i]



[142] As mentioned in the beginning of this section, accuracy here is grouped in terms of vowel, but was measured by word. This way, a 'pat' token identified as 'sat' was deemed an incorrect identification. As vowel quality shows effects of adjacent segments, we sustained this accuracy criterium in this section (cf. SILVA *et al.* (2019) for an account on acoustic characteristics of vowels).

(b) lax [ɪ]



Source: present study

The plots in Figure 4.26 style accurately identified tokens in green, and inaccurately identified words in red. Crosses signal German listeners' datapoints, whereas squares signal Argentinian listeners' datapoints. Baseline tokens produced by native speakers of Canadian English are styled as black dots. The same pattern will be used in the following figures relating to accuracy rates.

As can be seen in Figure 4.26, our exploratory analysis indicates that relative duration appears to be, indeed, a relevant cue for accurate identification by both listeners' groups in our dataset, when it comes to above-chance identifications. The images also appear to highlight a relatively stronger effect of duration for German listeners (many more cross signs than square datapoints). This could be a result of an 'L1 filter' (Flege, 1995; FLEGE; BOHN, 2021), given that the relative duration of L1 German vowels is distinctive, as it is in English, which is not the case for Brazilian Portuguese or Riverplate Spanish. Thus, a tentative explanation for this difference in above-chance accuracy rates for the Germans could be that it is reflecting the weighing of an L1 decisive cue[143], even when these individuals are perceiving L2 speech. Conversely, we could hypothesise that the Argentinian listeners' perception does not show a

[143] Though [i] in German does not have a shorter (lenis) counterpart (MAACK, 1949), we stand by this hypothesis, as we understand that this cue is distinctive for many other vowels, which would render it somewhat systematic in the German language, and therefore in the 'L1 filter'.

strong effect of duration in the interaction of the acoustic cues being weighed. This hypothesis seems fit for both minimal pairs, as duration was also less informative for this L1 group when accurately identifying words with [æ] and [ɛ].

We also ought to point out that Brazilian productions did not show (much) distinction in terms of [i] and [ɪ] relative durations, as we have described in section 4.2.1.1. Moreover, we have noticed, in the same section, that Brazilian [ɪ] productions had similar relative durations to those of native [i] productions. Those two factors could have rendered relative duration as not so informative a cue for Argentinian listeners. Additionally, as Riverplate Spanish native vowels are generally shorter than all L1 and L2 Brazilian productions, we could hypothesise that Argentinian learners perceive all Brazilian productions as 'long', which may be another reason why it does not seem to provide distinctive data between minimal pairs such as [i] and [ɪ] – or [æ] and [ɛ], as seen in section 4.1.4. Despite all that, duration appears to have an emerging informative relevance for native speakers of Spanish when it comes to Brazilian [ɪ] productions in L2 English. Tentatively, we could imagine that those listeners were expecting much smaller durations, even within an 'all too long' parameter for L2 sounds. Hence, the 'too long' relative durations of [ɪ] productions could have had a stronger effect as a confusion factor. As we can see in Figure 4.26, there are many more datapoints for Argentinians (squares) in the [ɪ] plot than in the [i] one, indicating that duration had a stronger effect for those identifications.

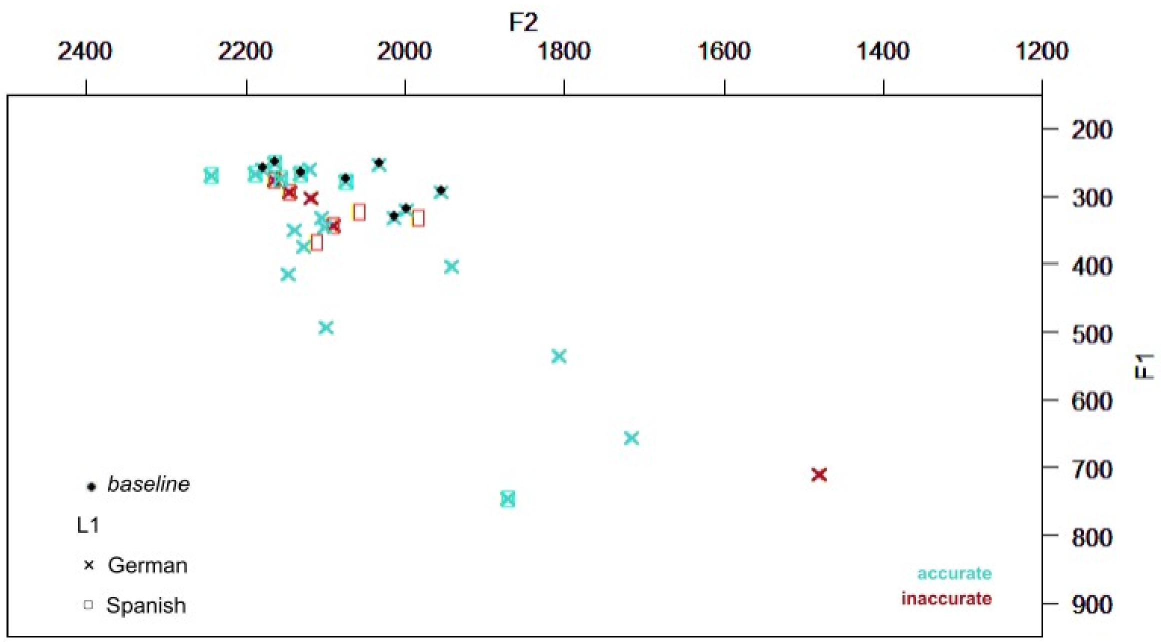
Finally, our stimulus-by-stimulus analyses suggest that the hybrid systems of both listeners and speakers are interacting and leading to the emergence of different effects. As cue weighting is a language-specific process, as we have reiterated, we cannot assume that cues are weighted differently by participants in the same L1 group in view of the target vowel, but that this weighting process is subject to complex effects arising from the interaction of the systems. In other words, assuming that duration is not a decisive cue for native speakers of Riverplate Spanish does not mean that this acoustic characteristic might not play a role when multiple cues interact. As we have mentioned, complex systems can undergo big changes due to small phenomena in its subsystems. Therefore, we ought to look at other cues that are informing listeners' perception and that are interacting with relative durations.

Moreover, we can also see that above-chance correct [i] identifications are spread across the F1 axis, whereas incorrect identifications are more concentrated in one area of the acoustic space. For [ɪ] identifications in relation to accuracy, we see less dispersion for tokens both accurately and inaccurately tokens identified above chance. If we take only the more densely populated areas, we see there is a slight difference in height, as [i] has F1 values of under 300Hz, and [ɪ] tokens have values between 250Hz and 350Hz, in its majority. We ought to consider that there may be other cues that play a role in those cases.

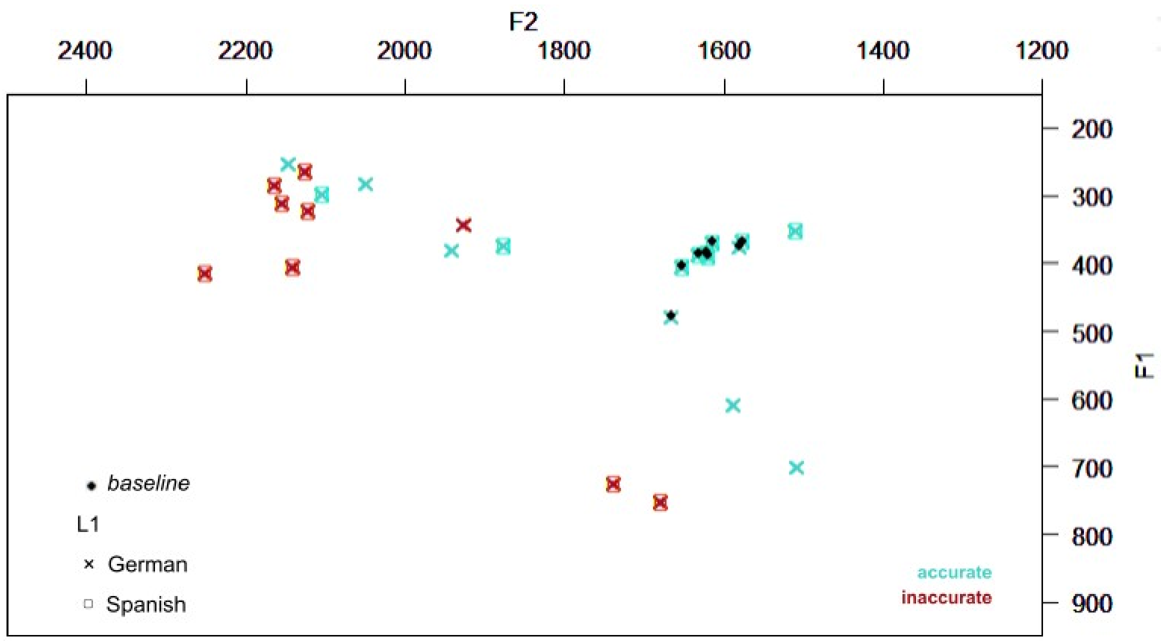
Figure 4.27 shows the same [i] and [ɪ] tokens with above-chance correct and incorrect identification, by target vowel, only this time portraying the F1 by F2 dispersion. Again, green datapoints indicate above-chance accurate identifications, whereas red datapoints indicate above-chance inaccurate identifications. German listeners' results are noted with crosses, and Argentinian results are noted with squares.

Figure 4.27 – F1 x F2 plots of [i] and [ɪ] tokens with above-chance correct and incorrect identification, by target vowel

(a) tokens with [i]



(b) tokens with [ɪ]



Source: present study.

As can be seen in Figure 4.27, accurate and inaccurate above-chance identifications of [i] partially overlap in the 1,950Hz to 2,150Hz F2 area for low F1 values. This could mean that F2 is not being highly informative for the identification of those tokens – if it was, we would expect a majority of one or of the other pattern of (in)accurate identification. It would seem that F1 appears to be a more decisive cue than F2 for those tokens, considering that strict error dispersion area. However, that is not consistent with the literature on native productions (which states that [i] has lower F1 values), nor does it explain why the same F1 area also has so many above-chance correctly identified tokens. This 'mixed' scenario could be taken as an indication that F1 and F2 effects are emerging from the interaction with some other acoustic cue. Given what we have seen in Figure 4.26, we could even say this cue is duration, as it seems to set more clear boundaries between [i] and [ɪ] tokens identified in an above-chance fashion.

As for [ɪ], Figure 4.27 suggests that higher F2 values lead to incorrect [ɪ] identifications, which is consistent with the literature on native speakers. Likewise, lower F2 values seem preponderant in correct [ɪ] identifications, which holds true for a wide variety of corresponding F1 values. We could hypothesise that, as Brazilian productions are too long for the lax vowel, a stronger F2 effect emerges when duration is not as informative as listeners expected. This idea seems to fit our tentative explanation as to why neither F1 or F2 yielded a clear effect on [i] identifications. In other words, our data suggest that a long duration has a strong effect in [i] identifications, but that long [ɪ] productions render this cue less informative, which leads to other cues having a more decisive effect.

In sum, the exploratory analysis seems to point to the need for Brazilian productions to have a more salient duration distinction between vowels [i] and [ɪ], shortening [ɪ] vowel length, as that appears to be the decisive cue for the lax vowel for both German and Argentinian listeners. As for formant frequencies, our data seems to suggest that producing the [ɪ] vowel in a less anterior space (lower F2 values) would be more beneficial to enhance local intelligibility, when compared to focusing teaching and training on producing the lax vowel in a less high position (higher F1 values). We highlight, however, that F1 might be a decisive cue for other non-native speakers of English, besides those native from Central Germany and the Buenos Aires state in Argentina. Further studies would need to be conducted to verify how native speakers of other varieties and/or languages perceive productions in L2 English by Brazilian learners (see section 5.3). What our dataset does seem to strongly endorse is the relative importance of both the speaker and the listener in the success of communication, instead of just 'blaming' one of them for a possible lack of success.

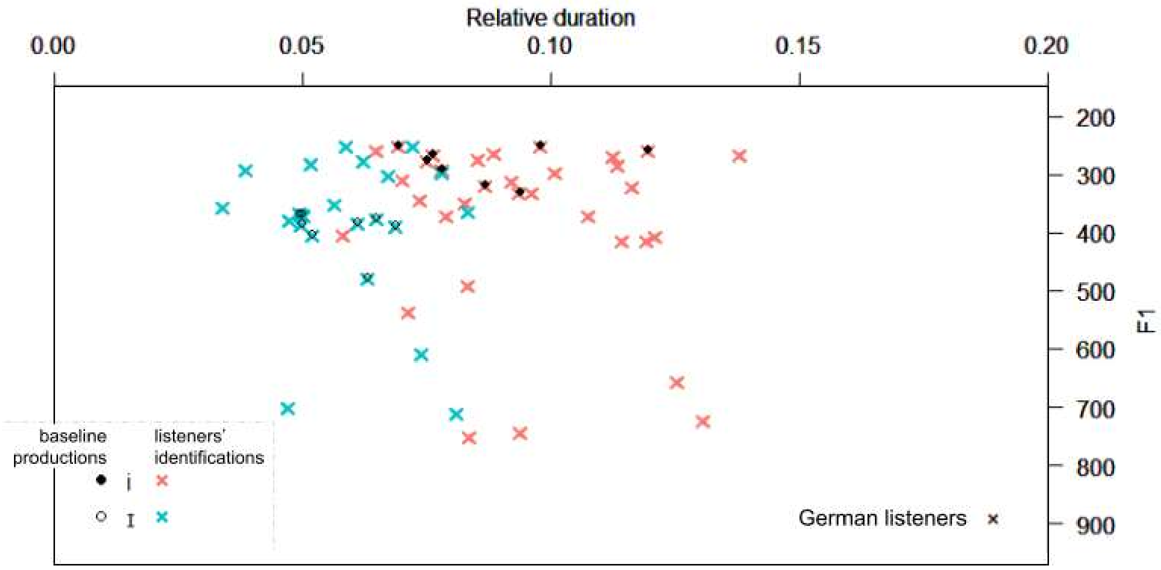
Having asserted the importance of the speaker-listener pair in regards to word intelligibility, we turn our attention to a more listener-related analysis. As we have mentioned, accurate word identification entices correctly identifying the onset consonant, the nuclei vowel and the coda consonant in our CVC monosyllables. That is, vowel identification is but a part of word identification. Therefore, we will now discuss the patterns of vowel identification shown by our groups of listeners, regardless of accuracy in relation to target vowel or target word. In this scenario, a token identified as 'fit' will be taken as an [ɪ] identification, regardless of whether the target word was 'fit', 'feet', 'sit' or 'seat'. We do so in order to try and understand how the acoustic cues present in the stimulus have had an effect on how listeners categorised it.

Again, we start by observing temporal effects, and following that we look specifically at spectral cue effects. We will once more present the tense vowel first and the lax vowel later in our comparisons. The above-chance subset was also selected by the same means we have previously described.

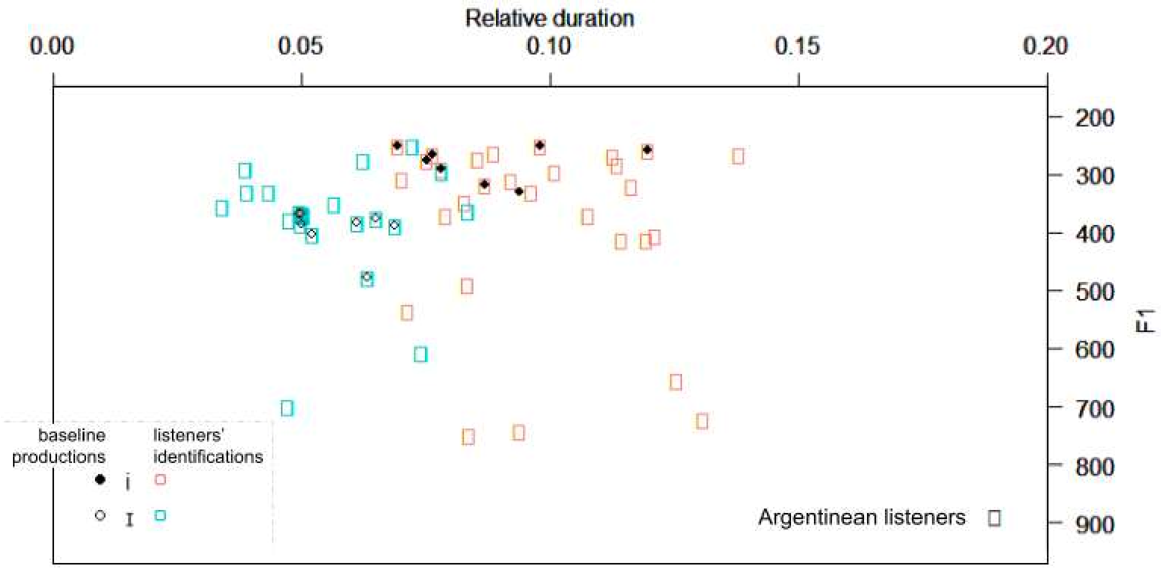
The plots in Figure 4.28 show identification patterns, regardless of their accuracy in relation to target vowels. Tokens are, once again, plotted as a function of F1 by relative duration. As we move on to look at identifications, regardless of accuracy, our plots will compare minimal pairs as they were identified by each group of listeners – we present the German participants' results before presenting the Argentinian's results. Cross signs still consistently indicate identifications by Germans, and squares, by Argentinians. Furthermore, as we are no longer talking about correct/incorrect identifications, we will use colours to signal the vowel of the minimal pair, with [i] being noted as salmon datapoints, and [ɪ] as blue ones.

Figure 4.28 – F1 x relative duration plots of [i] and [ɪ] tokens with above-chance vowel identification, by L1

(a) L1 = German



(b) L1 = Spanish



Source: present study.

As with accuracy patterns, above-chance consistent identifications seem to show that relative duration is a relevant dimension for German listeners. Nonetheless, contrary to what was observed with [æ] and [ɛ] tokens, for the high front vowels there is a larger area in which duration does not seem to be a determinant cue for this group of listeners.

As for Argentinian listeners, it is important to first highlight that, though duration does not seem such a relevant cue for accurate identifications, it does seem to be weighed as a decisive cue when we look at consistent (above-chance) identifications more broadly. We say that based on the amount of tokens with above-chance identification – 50 in contrast to the 17 above-chance *accurate* identifications. The general identification pattern (regardless of accuracy) seems to point to an important role played by duration as far as identification goes. However, Argentinian category boundaries appear not to match the distinction being produced in our stimuli. That is, though duration has a strong effect on identification in general, it may lead to *inaccurate* identifications in relation to Brazilian speakers' intended vowels.

Moreover, Figure 4.28 shows that Argentinian listeners do not have the same identification pattern across both minimal pairs. Here, we see that relative durations between the area of about 7% to 8% seem to cause confusion, whereas for [æ] and [ɛ] the 'unclear' region was between about 8% and 13% of relative duration (see also Figure 4.13), the same for German participants (see section 4.1.4). Additionally, it is worth mentioning that, for Germans, the 'blurry' relative duration area is wider, between about 6% and 8% of relative duration. That difference could be a sign that Germans are using a native boundary as a filter to the L2 relative duration, whereas Argentinians, who do not have native boundaries to filter tokens through, may be likely using category boundaries developed over their L2 learning trajectory. Here, contrary to what was discussed in section 4.1.4, we hypothesise that native speakers of Spanish have created a new category for [ɪ], hence the presence of a boundary that does not come from an L1 category. It could also be the case that the native [i] category is more similar to the L2 [ɪ] category. Of course, we ought to hypothesise, as well, that the way Brazilians produce duration distinctions might be leading to those different perception patterns, rather than this being a listener-only effect. That is, because Argentinian vowels are short, Argentinians may have a hard time distinguishing what we could call 'long' (attempts to produce the short, lax vowel) and 'too long' (attempts to produce the tense vowel), taking their L1 parameters of short vowels. The discussion we have made thus far seems to support emerging effects of cues across the hybrid systems of both speakers and listeners.

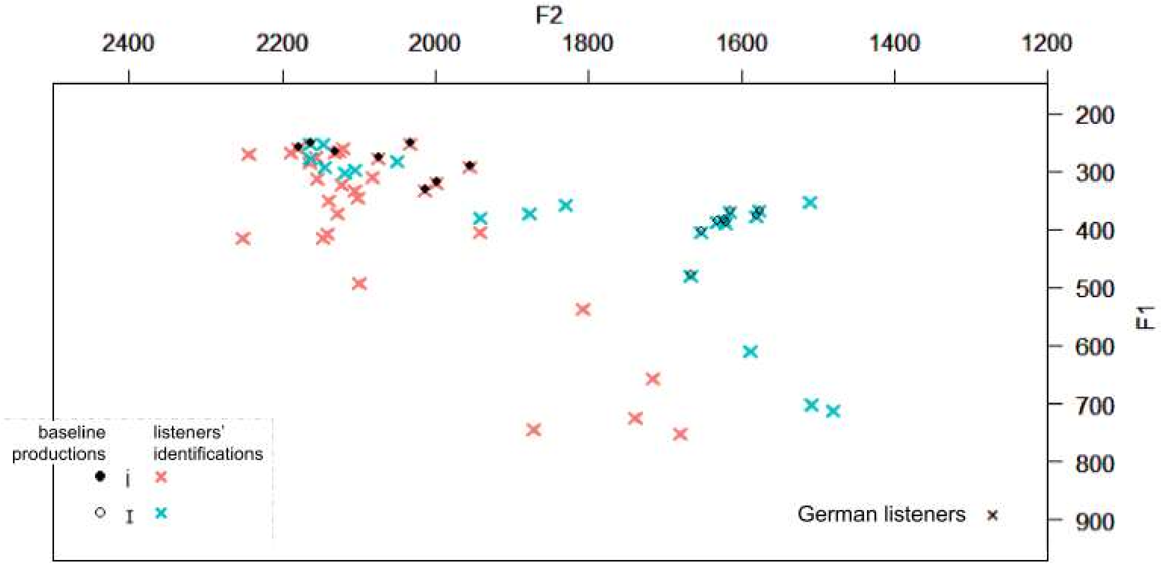
As illustrated by Figure 4.16, it is important to remember that speakers from different proficiency levels produce different relative durations. Beginner and intermediate level learners barely make any distinction between [ɪ] and [i] in that dimension. Besides, beginners produce relative durations closer to (albeit still longer than) baseline [ɪ] tokens, whereas intermediate level learners produce both vowels with longer relative durations that are closer to (and even longer than) baseline [i] tokens. In other words, though there is hardly any distinction in the pair, the durations approximate productions of each proficiency level to different categories. As for advanced learners, it seems that they are developing a temporal distinction between the tense and the lax vowels – though not yet at the same level that baseline/native speakers produce them. These characteristics might explain why the fitted model did not estimate relative duration as a significant predictor: too much variation within a small dataset. As we have seen, though, there are relevant processes that relate to vowel length, reiterating the importance of an exploratory, individual analysis.

Finally, we can observe that, as seen in [æ] and [ɛ], F1 does not seem to be such a decisive cue, as tokens identified above chance are spread across the F1 axis, with values ranging from about 250Hz to 750Hz for both [i] and [ɪ]. This seems to contradict the model presented in RQ2. That inferential fit (see section 4.2.3) estimates that both F1 and F2 are significant predictors of vowel identification, regardless of whether this identification is a match to what the speaker intended, and considering the whole of the dataset – instead of just the above-chance subset we analyse in the present section. We argue, however, that this is not a contradiction, but rather a gradient phenomenon in which F1 might have a more or less clear effect depending on how cues are interacting and informing each stimulus identification by the listeners. This is a phenomenon, we reiterate, that might have eluded the inferential statistics, though it is noticeable by applying a complimentary, exploratory method.

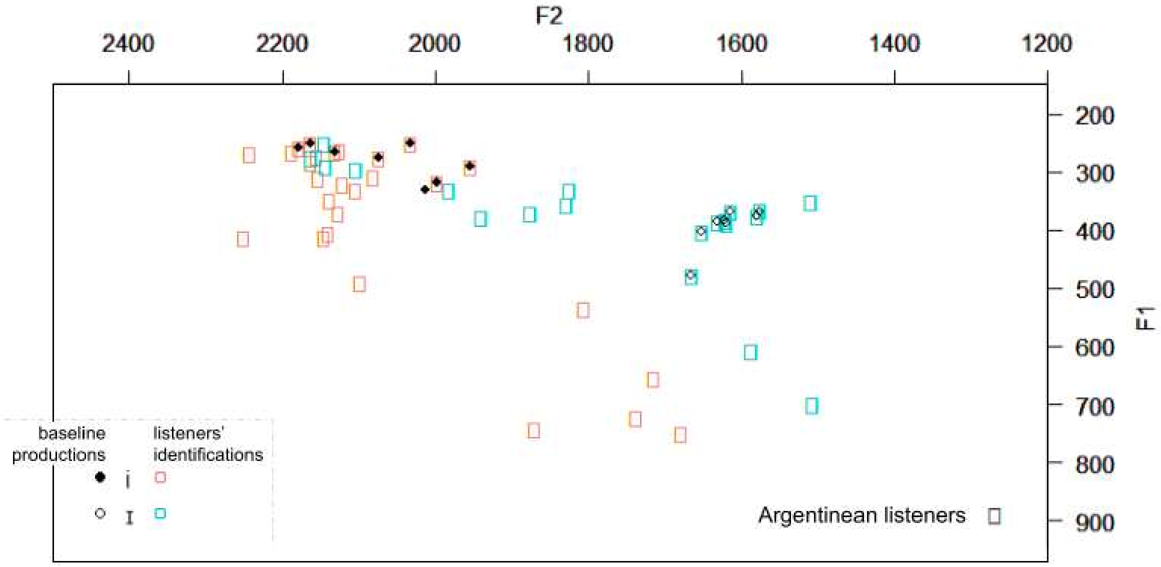
Figure 4.29 shows F1 x F2 plots of [i] and [ɪ] tokens with above-chance vowel identifications, by L1. Once more, German listeners' identifications are styled with crosses, and Argentinians' with squares. Salmon denotes [i] tokens in both plots, and blue, [ɪ] tokens.

Figure 4.29 – F1 x F2 plots of [i] and [ɪ] tokens with above-chance vowel identification, by L1

(a) L1 = German



(b) L1 = Spanish



Source: present study.

For both L1 groups, F2 seems to play a larger role, as most tokens of [i] are identified in the most fronted area of the figure. Likewise, most tokens of [ɪ] are in the less fronted part of the plot. Our inferential analysis in section 4.2.3 had already shown high estimates for that effect. As for F1, it seemed less informative compared to relative duration only. However, when in interaction with F2, F1 seems to play a more decisive role in the identification of [i] and [ɪ] for both groups of listeners.

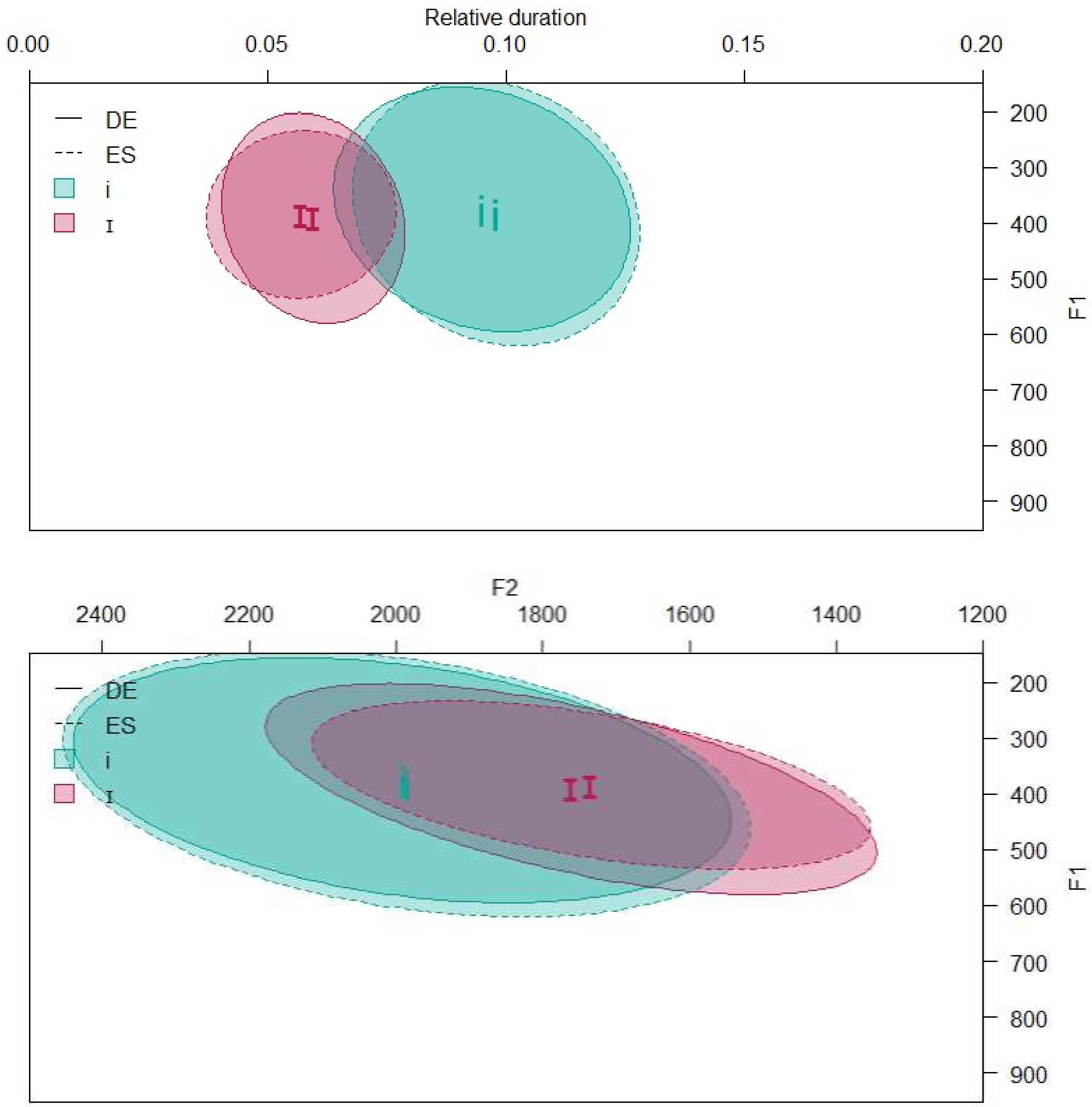
In acoustic areas in which F2 values are lower and F1 is high, identifications tend towards [ɪ], which is consistent with the literature on native productions – as we had already noticed in section 4.2.3. Likewise, when F2 is higher and F1 is lower, identifications tend towards the tense vowel, as they would in native productions. As a matter of fact, we see that almost all baseline productions were consistently identified in an above-average fashion. Furthermore, we should note that though our inferential analyses indicated that the higher the F1, the more likely a token would be identified as [i], our exploratory analysis shows that is not a uniform pattern. In other words, when all productions are grouped, F1 had an effect that goes the opposite way of what would be expected; however, a stimulus-by-stimulus observation of tokens identified above chance shows that this cue interacts with others and yields different identification patterns.

Additionally, an individual look reveals that despite the fact that the combination of F1 and F2 shows a somewhat regular pattern, we still find some confusing tokens in the low F1 by high F2 area on the top left part of the figure. It would seem that something about those tokens threw both German and Argentinian listeners into the opposite direction from the one they more consistently followed in most above-average identifications of [ɪ]. We could hypothesise, again, that another cue is emerging as more informative in the identification of those confusing tokens.

Finally, it is also worth mentioning that the Spanish vowel system, with five vowels, has a less spread distribution for each vowel in the acoustic space (PEREYRON, 2017). Thus, the fact that identifications by L1 Spanish participants are more broadly distributed along the F1 axis could be evidence of an ongoing system adaptation that aims to accommodate the new L2 vowel category – namely, here, that of the lax high front vowel. Nonetheless, this tentative explanation does not seem to apply to L1 German participants, who would have an L1 [ɪ] category already in place and with distinctive F1 values. Indeed, that category tends to have lower F1 values than the native [i] category (SENDLMEIER; SEEBODE, 2010), as is the case with English. The 'L1 filter' would lead us to assume that Germans should identify tokens in areas of high F1 and low F2 values as [ɪ], as would native English speakers. Our data does not show that pattern in L2 perception, though. For German listeners, therefore, perhaps relative duration and F2 carry a greater weight in vowel identification than does F1 – at least when perceiving L2 sounds. Moreover, it could be the case that the F1 values that make up the boundaries for their native German [ɪ] category do not match Brazilian [ɪ] productions. However, this hypothesis seems less likely, given that native German [ɪ] has an average F1 of 369Hz for males and 433Hz for females, which should – had there been an F1 filter in place – have yielded less varied identification patterns across the F1 axis. Additionally, given that German listeners consistently identified baseline tokens in an above-average fashion, we would again argue that those listeners' perception patterns are more likely due to the cues produced by Brazilian speakers, or at the very least due to emerging interactional patterns.

The role played by the three acoustic cues we have discussed thus far can be more clearly seen in the dispersion plots in Figure 4.30. The ellipses refer to above-chance identifications, regardless of accuracy, in terms of relative duration by F1, and F1 by F2. The tense [i] is noted in blue, and the lax [ɪ], in salmon. The solid line indicates perception patterns by German listeners, and the dashed line signals Argentinians' perceptions.

Figure 4.30 – F1 x relative duration and F1 x F2 dispersion plots of [i] and [ɪ] tokens with above-chance vowel identification



Source: present study.

As we have argued, relative duration appears to have a major role in vowel identifications for both groups. Figure 4.30 shows but a little overlap of the ellipses for both language groups. That is interpreted as a somewhat clear category boundary with regard to relative duration. We ought to remember, though, that categories are not outlined from a single cue only. To the contrary, our data suggest that the role played by acoustic cues may show a clearer effect depending on how these cues interact in each stimulus provided by the speaker. In that sense, we could interpret the overlapping area between the vowel ellipses in two ways. On the one hand, maybe the temporal cue was not informative enough. On the other hand, that other cue(s) has(ve) emerged as more decisive for a given stimulus.

The dispersion of above-chance identification of [i] and [ɪ] when F1 and F2 are considered seems to support that analysis. We see that, as we would expect for native productions, the dispersion of the lax vowel may be found in a less fronted area, whereas the tense vowel is in a more fronted portion of the figure. This is in line with what we have inferentially and exploratorily reported thus far, that F2 plays a major role in vowel identification by the Argentinian and German listeners who took part in our study. The first formant seems to play a smaller part in the process, if we consider the large range of values it takes in the identifications. However, we can still see that the [ɪ] area is less spread than [i], and that the ellipses present a slight angle, attesting to the fact that an F1 effect appears when it operates together with F2. As we have mentioned, this inclination of the ellipses suggests that the higher the F2 and lower the F1, the more likely an identification will tend towards [i], which is in line with the literature on native productions. Finally, we see that there is a larger overlap between [i] and [ɪ] ellipses in Figure 4.30. We could hypothesise that tokens in that intersection area are presenting emerging effects of other cues, such as duration, as we have previously argued.

In sum, we understand that our analysis of how cues operate together and affect the identification of [i] and [ɪ] by our listeners has shown, time and time again, that we cannot take acoustic characteristics of speech sounds at face value. As we have argued thus far, the hybrid system of the non-native Brazilian speakers produced cues that are interpreted by the hybrid systems of the non-native German and Argentinian listeners. From that communicational interaction, on a stimulus-by-stimulus basis, different cues emerge as having a major role.

We will now move on to answering our Research Questions, based on the results and analyses we have presented in this and in the previous sections.

4.3 FINAL CONSIDERATIONS

Once again, we highlight that the present thesis adopts a Complex, Dynamic view of language, which poses that language learning is a process, as opposed to a static product. The Complex, Dynamic framework assumes change through time as an inherent characteristic of systems, and those changes will emerge from the non-linear, complex interaction of multiple subsystems. Thus, the way we conceive it, a Complex, Dynamic account is able to perceive gradiences and individualities in the production and perception of speech sounds, as well as of language development as a whole.

As detailed in section 2.1, our study assumes the Complex Dynamic Systems Theory as a metatheory. That is, it informs our view of language, though our methodology is cross-sectional. We follow Lowie's (2017) recommendation and couple inferential, group statistics with individual, exploratory analyses of the data we have collected. As we draw near to our conclusion, this claim is of particular importance, as this methodological choice has a great impact on the way we look at our data, as well as on the results we report from those analyses.

We will argue once again that our inferential statistics were able to highlight which predictor variables were significant in estimating our response variables. Despite our relatively small dataset, those model fittings could predict general, group behaviours, indicating tendencies. That analysis alone, however, was not able to pick up on the gradient nature of the language development process. As we have seen, though variables such as listener's L1 and stimuli F1 and F2 were estimated as significant, other variables such as the relative duration of the vowel stimuli and speaker's proficiency level were disregarded. Upon performing an exploratory analysis of our dataset, however, we have been able to see that those variables do play a decisive role in vowel identification in general, and accurate vowel identification in particular. That role emerges from the interaction among different cues, as well as from the interaction between the speaker's and the listener's hybrid systems, as in our study not only speakers, but also listeners are both non-native learners of L2 English. Moreover, our analysis has suggested that this interaction took place on a stimulus-by-stimulus basis.

We understand, thus, that the methods adopted in the present thesis are complementary, in that one highlights variables that have a more general effect, whereas the other allows us to see gradient phenomena that are also playing a decisive role in the processes under analysis. Having recapitulated that, we will now move to answer our three Research Questions, based on the results and discussion presented in the previous sections. Answers will be provided by each question in a separate subsection, within which we will consider the [æ] – [ɛ] minimal pair first, followed by the [i] – [ɪ] minimal pair.

4.3.1 Research Question 1: does the L1 have an effect on L2 word intelligibility?

As we have operationalised it, word intelligibility entails a correct identification of onset consonant, nuclei vowel and coda consonant in our CVC monosyllables. Once again, we recapitulate that what we conceive as an 'accurate identification' encompasses all three segments of the stimulus. This way, a 'pat' token identified as 'sat' is considered an inaccurate identification, for the onset consonant was mistakenly identified, though the vowel was correctly identified.

Our inferential analysis in section 4.1.2 supports that a listener's native language (L1) will indeed have an effect on L2 word intelligibility. Our native speakers of the Central variety of German tend to correctly identify productions of words with [æ] by our Brazilian non-native speakers of English and of our native speakers of Canadian English better than our native speakers of the Riverplate variety of Spanish do. Argentinians, on the other hand, perform better in the correct identification of words with [ɛ]. That is, accuracy, in those cases, is affected by L1. Conversely, for productions of words with [i] and [ɪ], L1 is not a significant predictor *per se* (section 4.2.2).

We have also calculated the effect of target vowel as a predictor variable. Target vowel is a grouping factor that gathers accurate identifications of target words with [æ] (tokens of 'pat' and 'sat'), [ɛ] (tokens of 'pet' and 'set'), [i] (tokens of 'feet' and 'seat') or [i] (tokens of 'fit' and 'sit'). The target vowel was a significant predictor of accurate word identifications for tokens of both minimal pairs. Target words with [æ] are more likely to be accurately identified than those with [ɛ]. Conversely, the latter are likely to be misidentified. Target words with [ɪ] tend to be correctly identified, as do the ones with [i]. Estimates for tokens with the tense vowel show that the likelihood of accurate identification is higher than for the lax vowel.

Finally, our mixed-effects logistic models have also estimated the effect of L1 in interaction with target vowel. Our inferential analysis shows that our German listeners are more likely to correctly identify words with target [æ] than our Argentinian listeners. The same is true for words with target vowel [i]. On the other hand, words with target vowel [ɛ] have a higher likelihood of being unintelligible to German listeners, whereas they tend to be intelligible to Argentinian listeners. Our model fitting estimates that L1 is not a significant predictor of accurate identification for words with target vowel [ɪ].

4.3.2 Research Question 2: Which predictor variables can explain the identification of a token as having [æ], [ɛ], [i] or [ɪ]?

As we have mentioned, word intelligibility entailed the correct identification of the three segments in our CVC monosyllables. Therefore, vowel identification is but a part of the process of word identification. That is why our second research question enquired which variables might have an effect on how listeners perform such vowel identifications. In order to do so, we disregarded accuracy in relation to the speaker's intention (target vowel), as we wanted to observe this process on the listener's side. Additionally, as hypothesised by the SLM (FLEGE, 1995; FLEGE; BOHN, 2021), and as evidenced by Escudero and Polka (2003) and Escudero (2009), we expected to be able to evaluate the presence of a possible 'L1 filter' in the non-native listeners' perception.

Our inferential statistics have estimated that L1 is a significant predictor variable for the identification of a vowel as [ɛ]. However, the model fittings did not yield L1 as a significant predictor for the identification of a vowel as [i]. These results for general vowel identification are in line with what we have found with our model fitting for accurate word identification (including the accurate identification of the nuclei vowel) and reported in section 4.3.1. Although in our first attempt to fit the model to our data we tried to estimate the effect of L1 in interaction with target vowel as a predictor variable for vowel identification, it yielded no significance. As we have mentioned, this interaction was removed from the model which we report in the present thesis. As the interaction of those variables was not estimated, we cannot compare these results with the ones in section 4.3.1 in that respect.

We have also found that target vowel is a significant predictor of vowel identification, regardless of accuracy. Our mixed-effects logistic models estimate that when speakers try to produce a word with [ɛ], listeners are more likely to identify it as a word with [ɛ] – that is, correctly. The same is true when speakers produce a word with target vowel [i]: in this scenario, listeners are more likely to (accurately) identify those stimuli as words with the [i] vowel. As we have discussed, previous studies have shown that [æ] tends to be assimilated into [ɛ] and that [ɪ] tends to be assimilated into [i] by Brazilian learners of L2 English. Thus, we could hypothesise that [ɛ] and [i] target vowels are being produced in an L1 fashion, as they are native BP categories. Likewise, as the speaker has more experience with their native language, we could assume that [ɛ] and [i] vowels do not present difficulty for production (if they are indeed being produced in an L1 fashion). However, we refer to the discussion in section 4.1.3, about the long native BP durations of [ɛ] being carried into L2 [ɛ] productions, and how those L1-like durations seem to hinder the informativeness of the L2 temporal cue.

Our inferential analyses also estimated that F1 is not a significant predictor variable of [ɛ] identifications. Conversely, for vowels identified as [i], F1 was a significant predictor variable. The model estimates that the higher the F1, the more likely listeners are to identify a vowel as [i]. However, this is not consistent with the literature on native productions, as [i] is produced in a lower F1 area of the acoustic space than [ɪ].

Finally, unlike the first formant, F2 values were estimated to significantly predict [ɛ] and [i] identifications. In both cases, the higher the F2, the more likely the identification as one of these vowels was. That is consistent with the literature on native productions. Native speakers produce [ɛ] in a more fronted area than [æ], and the same is true for [i] productions in comparison to [ɪ] tokens.

4.3.3 Research Question 3: how can the identification patterns in the perception task shed light on the listeners' composite L1-L2 categories?

As the Complex, Dynamic view of language assumes change over time as an inherent characteristic of a system, a group-based, inferential approach may not be sufficient to provide information about individual, gradient processes. Therefore, in order to answer Research Question 3, we adopted an exploratory analysis methodology, since we understand the need for a stimulus-by-stimulus look at our dataset is complementary to that provided by inferential statistics.

The individual analyses we carried out proved able to suggest that there are, indeed, gradient phenomena that our statistical approach did not pick up – maybe due to the small number of datapoints we have collected. As in sections 4.1.4 and 4.2.4, we will look at vowel length first, moving forward to F1 and F2 acoustic cues later. We recall that we worked with subsets of data in order to look closely at more consistent patterns of behaviour (see section 4.1.4).

As we have mentioned, the temporal cue was not significant in the inferential initial attempts – which is why it was removed in further fittings and is not present in the final models reported here (see sections 4.1.2 and 4.2.2, as well as 4.1.3 and 4.2.3). We assume that this lack of significance was due to the small dataset we collected, in which there was not enough variation. We have also described, in sections 4.1.1.1 and 4.2.1.1, that the Brazilian learners who partook in our study did not produce (much) distinctive relative durations between [æ] and [ɛ], nor between [i] and [ɪ], which might have contributed to why the model ignored the effects of this variable.

Our exploratory analyses, however, suggest that relative duration does play a role in the listeners' identification patterns, both in terms of accuracy and regardless of target vowel, and for the two pairs of vowels. For the German listeners, duration seemed to have a clearer effect than it did for the Argentinian listeners. Also, for the [æ] – [ɛ] pair, it showed stronger effects than it did for the [i] – [ɪ] pair.

As we have mentioned, Brazilian productions did not show (highly) distinctive durations between tokens of each vowel in the pairs. Given these descriptions of the participants' productions, one might be led to believe that the answer to RQ3 is that relative duration is a dimension defined exclusively in our listeners' L2 category. In other words, if duration is not being produced (speaker side), but is being perceived, it ought to be a listener-related process. That, however, would be a misleading answer.

We have assessed that duration plays a strong role, but we have also seen that this dimension alone is not decisive for vowel identification. As we have extensively discussed, duration is taken as being more or less informative depending on how it is interacting with other acoustic cues. In turn, we ought to point out that our listeners' category boundaries cannot be entirely inferred from their perception, as their identifications present emerging effects that are due to stimulus characteristics, which brings the speaker-listener communicational interaction into light. In other words, though we see some indications of how listeners' L1-L2 categories are established, we also see evidence that in a communication setting those boundaries will not linearly predict identification, as different combinations of acoustic cues in the stimulus will lead to different cues playing an emerging major role in vowel identification.

In our dataset, upon looking at each participant's data, as well as at each proficiency level, it does appear that there is an ongoing language development process in regard to vowel length. That is, speakers also have a hybrid system in their L1-L2 composite categories. Moreover, when compared to baseline productions by our native speakers of Canadian English, we were able to observe two main facts that are relevant to answering RQ3. The first one is that native speakers of the Porto-Alegrense variety of Brazilian Portuguese tend to produce large vowel lengths, both in L1 and in L2 – we refer to Tables 4.1 and 4.4. The second is that these long durations seem to approximate Brazilian productions of both [æ] and [ɛ] to the native productions of [æ]. Accordingly, Brazilian productions of [i] and [ɪ] also appear to have relative durations that are closer to native productions of [i]. Not surprisingly, our inferential statistics have evidenced that [æ] and [i] are more likely to be intelligible (ie. yield correct identification) to our listeners, as shown by target vowel effects reported in sections 4.1.2 and 4.2.2.

We have also pointed out that [ɛ] and [i] are native categories in BP, whereas [æ] and [ɪ] are not. Thus, we expected that the 'L1 filter' would be present, especially considering the perceived (dis)similarity of L2 sounds in the formation of L1-L2 composite categories of Brazilian learners. The same can be said about our listeners' hybrid systems. We have highlighted the fact that duration has a distinctive role in native German vowel categories, unlike in native BP or Spanish categories. Furthermore, we have more than once recalled that Argentinians produce much shorter vowel lengths in Spanish than Germans and Brazilians do in their native languages. We recapitulate those facts in order to reinforce that the communicational interaction of the hybrid, non-native systems is a key factor in the analysis of both intelligibility and perception processes of L2 vowels.

With that in mind, our answer to RQ3 in relation to formant frequencies is, accordingly, that those acoustic dimensions play a strong role in word and vowel identification. Nonetheless, as we have seen thus far, that role is more or less decisive depending on how F1 and F2 are interacting with each other, as well as with other acoustic cues – like duration –, on a stimulus-by-stimulus base.

We have observed that, in comparison with duration, F1 seems to play a smaller role for both our L1 groups of listeners. Accurate identifications of [æ] and [ɛ] show a more condensed dispersion – each dispersion in a different F1 area of the acoustic space –, according to what we would expect based on the literature on native productions. A similar pattern is found for the tense and lax high front vowels, except that this effect is clearer for Argentinian listeners than it is for German listeners. For [i], Germans present a much larger F1 range in which tokens are correctly identified. Conversely, as we have mentioned, Germans seem to show a clearer role played by duration in comparison to Argentinians. We highlight that the native speakers of Spanish had fewer above-chance correct identifications for all vowels, in comparison to the number of datapoints of the native speakers of German. As for identification in general (regardless of accuracy), the role of F1 is not as distinctive for the [æ] – [ɛ] pair as it is for the [i] – [ɪ] pair. We refer to Figure 4.13, in which we see areas with mixed identifications of both [æ] and [ɛ]. Conversely, Figure 4.28 presents a less mixed pattern.

Additionally, our analyses suggest that the role played by the F2 cue would be stronger than F1 for the four vowels we are analysing and for both groups of listeners. We see a clearer role played by the second formant cue for [i] and [ɪ] vowels than we do for [æ] and [ɛ]. Again, we say that based on areas of the F2 axis in which there are mixed identification patterns. It is worth highlighting that for *accurate* identifications of [æ] and [ɛ], the role played by F2 is clearer than it is for general identifications of this pair. Conversely, for the high front vowels, the role played by F2 is much clearer in the identifications in general than is in the *accurate* identifications – particularly for the tense vowel. This is surprising, considering that the tense vowel yielded the higher accurate rates of all four. Moreover, we would like to highlight that Germans seem to use duration as the most decisive cue of the three under analysis in the present thesis. Nonetheless, the combined role of F1 and F2 appears very clear in their accurate and general identification patterns, which we take as yet another sign of the interaction of these cues in the perception of those listeners.

In sum, two processes appear to be salient in the analyses we have presented in order to answer RQ3. On the one hand, we seem to have found evidence for language-specific cue weighting processes, that is, some cues appear to be primary and others, secondary. On the other hand, they do not act as a static hierarchy, for we have also noticed, time and time again, that the role played by a cue also depends on the interaction with other cues. That is to say that the informativeness of a cue (or lack thereof) might lead to emerging effects of other cues. Furthermore, when taken together with major effects estimated by inferential statistics, we have also observed that a cue can be or become decisive depending on each stimulus.

All in all, in having attempted to answer our three research questions, we understand, thus, that it was highly beneficial that we applied complementary methods in answering the research questions we posed. The inferential statistics yielded major predictor variables for group effects (RQ1 and RQ2), whereas the exploratory analysis picked up on the gradience of those statistically significant effects, as well as of the role played by other cues that were not statistically significant (RQ3). We will discuss this further in the next chapter.

5 CONCLUSION

This chapter will summarise the results obtained from the inferential and exploratory analyses of the data collected in the present study. There is also a subsection detailing the limitations of the study. Lastly, we will point to the contributions we believe our study has made to the fields of Laboratory Phonology and Applied Linguistics, as well was the aspects we believe might generate further investigation.

5.1 SUMMARY OF RESULTS

The present study collected speech samples from six Brazilian learners of English as an additional language (L2). Those participants self-assessed their proficiency level as beginner, intermediate or advanced. All learners (1 male and 1 female in each group) were born and raised until the age of 7 years old in Porto Alegre and its metropolitan area, thus speaking the same Porto-Alegrense variety of Brazilian Portuguese (BP) as their native language (L1). Those learners participated in the production task by recording samples in L2 English. Recordings were elicited by a sentence-reading task. Target words were inserted in the carrier sentence “The word is ___ too”. To set a baseline, two native speakers of Canadian English (1 male and 1 female) also recorded the same tokens, in the same manner. A total of 128 sentences were selected as stimuli in the identification task.

The identification task was done online. A total of 46 listeners participated in the perception task. Argentinian participants (28) were born and raised in the Buenos Aires province, and hence spoke the same Riverplate variety of Spanish as their L1. German participants (18) were born and raised in states where the Central variety of German (L1) is spoken.

Listeners were instructed to listen to each stimulus only once and then select the word that filled the gap in the carrier sentence ‘The word is ___ too’, based on what they had just heard. The forced-choice task exhibited, for all stimuli, the same 8 options, namely all the eight target words included in the task: ‘feet’, ‘fit’, ‘pat’, ‘pet’, ‘sat’, ‘seat’, ‘set’ and ‘sit’.

We had three research questions that we intended to answer with our experiment. RQ1 enquired whether the listeners’ L1 would have an effect on the intelligibility of the L2 speech produced by the speakers. Intelligibility was operationalised as ‘accurate word identification’, when all three segments of our CVC monosyllables were correctly identified. This way, a ‘sat’ token was considered to be correctly identified if it was identified as ‘sat’, but not if it was identified as ‘pat’ (correct vowel) or ‘set’ (correct onset/coda consonants). RQ2 enquired which predictor variables would allow us to estimate how (nuclei) vowels in our stimuli were identified by listeners, regardless of accuracy. Both RQ1 and RQ2 were answered by fitting mixed-effects logistic models. To answer RQ3, on the other hand, we moved away from inferential statistics and the main effects they yield, and performed an exploratory, stimulus-by-stimulus analysis. We did that because RQ3 enquired about the listeners’ L1-L2 composite categories, and what could be inferred about them based on vowel identification patterns (accurate and general).

In chapter 4, we analysed each of our minimal pairs, [æ] – [ɛ] and [i] – [ɪ] separately. In this section, we will follow the same organisation.

5.1.1 Minimal pairs with [æ] and [ɛ]

Our mixed-effects logistic model estimated that both listener’s L1 and target vowel, as well as an interaction of those variables, are significant in predicting accuracy identification rates of target words with [æ] and [ɛ]. The German participants were likely to accurately identify target words with [æ], whereas Argentinian participants were not likely to do so. However, Argentinian listeners performed better than German listeners in accurately identifying target words with [ɛ]. Accordingly, Germans identified words with [æ] correctly much more often than they identified words with [ɛ]. Argentinians, on the other hand, correctly identified words with [ɛ] more often than they identified words with [æ]. It must be said, though, that the accuracy rates of each set of tokens – though still significantly different – was closer for Argentinians than it was for Germans. That is, though Argentinians accurately identified [ɛ] more often, it was not much more often than they correctly identified [æ]. We thus concluded, in answering our RQ1, that Brazilian productions of words with [æ] are more intelligible for Germans than for Argentinian listeners. Conversely, tokens with [ɛ] have higher intelligibility rates for Argentinian listeners.

Also using a mixed-effects logistic model, in answering the second research question, the present study obtained significant results for target vowel, L1 and F2 as predictor variables of vowel identification (regardless of accuracy). When disregarding accuracy and looking only at which vowel was identified, our inferential statistics showed that the target vowel in the stimulus is a significant predictor of vowel identification, as German speakers are unlikely to identify target words with [æ] as tokens of words with [ɛ]. L1 is also a significant predictor, as native speakers of Spanish have a tendency to identify target words with [æ] as tokens of words with [ɛ]. Conversely, if the target word a Brazilian learner is trying to produce has an [ɛ] vowel, both Argentinians and Germans tend to identify it correctly as an [ɛ] token. Acoustic measures of F2 were also a significant predictor variable. However, results are not in consonance with the literature on native English productions: for our participants, the higher the F2, the more likely participants were to identify a token as [æ]. For native perception, because [æ] is less fronted than [ɛ], higher F2 frequencies tend to yield [ɛ] identifications. The F1 values were included in the model fitted to answer RQ2, but did not yield significance.

Our exploratory analyses (RQ3) aimed to complement our statistical models, in a way that allowed us to investigate gradient processes that might not have been picked up by the inferential statistics, given our small dataset. Therefore, besides the statistically significant predictor variables, these analyses also took speakers' proficiency levels and vowel durations into account. Additionally, the stimulus-by-stimulus analyses focused on above-chance accuracy and identification patterns. Above-chance accuracy and identification were operationalised as more than 18 occurrences of a given identification by Argentinians, a group with 28 participants, and as more than 14 occurrences of a given identification by Germans, a group with 18 participants (see 3.3.3). Throughout sections 4.1 and 4.2, we analysed accuracy and identification patterns separately. In the present section, we will summarise those findings all together, pointing to eventual differences in the patterns if necessary.

Our exploratory analyses seem to indicate that vowel duration was a highly informative cue for both German and Argentinian listeners. For German listeners, duration appears to play a stronger role than other cues for both [æ] and [ɛ] identifications. For Argentinians, however, duration seems to have an emerging, decisive role in the [æ] identifications, but not in the [ɛ] identifications. We speculate that the native distinctiveness of vowel length in German might lead listeners of that L1 to perceive the temporal cue as more informative, in comparison to the other group of listeners. Moreover, as Spanish native vowels are all short, our tentative explanation is that duration tends to be more or less decisive in conjunction with other cues. That is, since L2 vowels are all 'long', considering the L1 Spanish filter, that cue needs to be combined with others to be informative enough. Finally, we ought to reiterate that our Brazilian participants (a) did not produce much distinctiveness between [æ] and [ɛ] in terms of vowel length; (b) showed longer durations in L2 [æ] and [ɛ] tokens than our baseline/native Canadian English; and (c) are native speakers of the Porto-Alegrense variety of BP, which tends to produce longer L1 vowel lengths than native L1 Spanish or L1 German productions. In our understanding, those speaker-related characteristics also played a role in how their productions were perceived.

Besides the temporal cue, our exploratory analyses also investigated the role played by the frequency (spectral) cues. Though F1 was not a significant predictor in the inferential analysis, our stimulus-by-stimulus investigation shows that F1 appears to have a slight effect on [æ] identifications for both groups, as tokens with F1 above 600Hz seem more consistently identified as [æ] – which is also consistent with a native production in English. For Argentinian listeners, the role played by the F1 cue seems stronger than it does for German listeners, especially in what concerns [ɛ] identifications. The German participants' identification pattern appears to combine the F1 and F2 cues, though the role played by this combination seems still less strong than the role of vowel duration. Moreover, the way German listeners' identifications seem to combine F1 and F2 is not consistent with the literature on native productions of English. That is, German listeners show a pattern in which higher F1 and lower F2 yielded [ɛ] identifications, whereas native English listeners would take those values as informative for [æ] identifications. Our inferential analysis also shows this pattern in relation to the F2 estimates, in which the higher the F2, the more likely the participant will make an [æ] identification. Once again, we highlight that our speakers' productions did not show much distinctiveness in terms of F1 or F2, and besides showed high variation with regard to the frequency cues in [æ] and [ɛ] productions.

Considering both our statistical and our exploratory results, we conclude that the L1 is an important variable when it comes to the intelligibility of words with the vowels [æ] and [ɛ]. The way that variable will impact results is not uniform, though. The role of both speakers' and listeners' L1 filters in their production and perception of a common L2 (English) seems to depend on other interacting factors. We have suggested, for instance, that the role played by vowel length in the native language of our participants has somehow shaped how they produce/perceive L2 sounds (for example, with Brazilian speakers producing longer vowels as they do in their L1, and German speakers relying strongly on vowel duration, as they do in their L1). The participant´s L1 is not uniform. Neither is it the sole predictor of vowel identifications. To the contrary, our data seem to show that intelligibility is susceptible to effects that emerge from the communication within the speaker-listener pair. In other words, as the hybrid, non-native systems of speaker and listener 'encounter' in a communication setting, the informativeness of acoustic cues present in the speech signal might be perceived as stronger or weaker depending on other cues, as well as on participant-related variables (such as their L1, L2 proficiency level and individual learning trajectories, to name but a few).

5.1.2 Minimal pairs with [i] and [ɪ]

We will now summarise the results of the words with vowels [i] and [ɪ]. The high front vowels differ from those in the other minimal pair analysed in this study. On the one hand, the literature in the field (RAUBER, 2006; NOBRE-OLIVEIRA, 2003, 2007) posits that [ɪ] is easier to develop for native speakers of BP than [æ] – [i] and [ɛ] are already established as L1 categories[144]. Secondly, our descriptive analyses of the Brazilian learners' productions by relative duration, F1 and F2 are consistent with that expectation, showing different developmental stages of [i] and [ɪ] for our participants at all proficiency levels. Additionally, [i] is the only one of the four segments investigated in this thesis that is present in all three L1s (Portuguese, German and Spanish), as well as in L2 English – though we highlight once again that this does not mean they are produced with the same acoustic characteristics across languages.

Our mixed-effects logistic model for RQ1 estimated that the relationship between target vowel and L1 is not constant in the data. The German participants were more likely to accurately identify target words with [i] vowels than the Argentinian participants were. However, Argentinian and German listeners performed similarly (and more poorly) when accurately identifying target words with the lax [ɪ]. We thus concluded, in answering our RQ1, that Brazilian productions of words with [i] are more intelligible than those with [ɪ]. Moreover, productions with [i] are more intelligible for German listeners than they are for Argentinian listeners.

The second mixed-effects logistic model was fitted to analyse the identified vowels, regardless of accuracy, in RQ2. The model estimates that target vowel, F1 and F2 are significant predictor variables. Words with target vowel [ɪ] are likely to be identified as tokens with [ɪ] by German listeners, and the same is true for the identification of words with the target vowel [i] being identified as a token with [i]. For F1, the literature on native productions (LADEFOGED, 2010) predicts lower F1 values for the tense [i]. However, the fitted model showed that the higher the frequency of the first formant, the higher the likelihood of a token being identified as [i] by both groups of listeners (L1 was not significant). Conversely, higher F2 values yielded a higher likelihood of having a target [i] being identified as [i]. That is consistent with the literature, as native productions of the lax [ɪ] are more posterior than [i], hence having lower F2 values than [i].

[144] Once again, we highlight that the presence of those vowels in native inventories does not mean that they are produced with the same acoustic characteristics that they would show in other varieties/languages.

Once again, we performed exploratory analyses in order to answer RQ3. These analyses aimed to complement our statistical models, in a way that allowed us to investigate gradient processes that might not have been picked up by the inferential statistics, due to our small dataset. Therefore, besides the statistically significant predictor variables, the exploratory analyses also took speakers' proficiency levels and vowel durations into account. Additionally, the stimulus-by-stimulus analyses focused on above-chance accuracy and identification patterns. Above-chance accuracy and identification were operationalised for the [i] – [ɪ] pair in the same way as it had been for the [æ] – [ɛ] pair. (that is, as more than 18 occurrences of a given identification by Argentinians, a group with 28 participants, and as more than 14 occurrences of a given identification by Germans, a group with 18 participants). Throughout sections 4.1 and 4.2, we analysed accuracy and identification patterns separately. In the present section, we will summarise those findings all together, as we did in the previous section.

As we have already seen in analysing the mid/low vowel pair, vowel duration seems to play a strong role in the identification of the high front vowels. Again, German participants appear to show a clearer effect of the temporal cue than do Argentinian participants. Conversely, duration seems to play a stronger role for Argentinians when identifying the lax vowel, that is, the shorter of the two vowels. In a similar fashion, when we analysed the other minimal pair, it showed a stronger role for the [æ] identifications, the longer vowel of that pair.

Unlike what we observe for [æ] and [ɛ], the Brazilian productions of [i] and [ɪ] show distinct duration patterns across proficiency levels. Advanced speakers are already producing distinctive vowel lengths for [i] and [ɪ], albeit not in a native fashion. This not only means that there is a distinction to be perceived, but it also suggests that there is more variation amongst productions – which may be why duration was not a significant predictor in the inferential analyses.

Also clearer in this pair than in the previous one, the role played by a combination of F1 and F2 seems quite relevant. These acoustic measures are also more varied in the Brazilian productions of the [i] – [ɪ] pair than they were in the [æ] – [ɛ] pair. This difference in development is consistent with previous studies (RAUBER, 2006; NOBRE-OLIVEIRA, 2003) that show that Brazilian learners establish an [ɪ] category faster than an [æ] category. In other words, as the distinctiveness between the tense and the lax vowel is being produced more prominently, we hypothesise that it might play a stronger role, that is, it might be perceived by listeners as more informative. We also highlight that F2 seems to be taken as slightly more informative than F1, because F2 presents less dispersion than F1 in our stimuli.

5.2 LIMITATIONS OF THE STUDY

As in any Laboratory Phonology experiment, some design choices, while allowing for more control of some variables, also imply limitations on other aspects. Firstly, our experiment adopted self-assessment of proficiency levels (SCHOLL; FINGER; FONTES, 2017; MARIAN, BLUMENFELD; KAUSHANSKAYA, 2007), which might not yield the other types of assessment[145]. Additionally, our descriptive analysis indicates that the development patterns predicted by previous studies were not entirely reflected in the productions by the speaker participants in the present study. For example, in terms of vowel length, our advanced learners did not show a distinction in the [æ] and [ɛ] productions in comparison to our intermediate level learners. Conversely, in terms of F1 and F2, one of the intermediate level learners showed a larger distinction in the [i] and [ɪ] productions than both advanced speakers.

Moreover, the small number of participants in the production task (six Brazilian learners and two Canadian native speakers) might have yielded little gradience to the data. Additionally, both native speakers of English who participated in our study were also highly proficient in other languages, with Portuguese among those foreign languages spoken by both Canadian participants – one of them participants was a simultaneous bilingual, born in Canada from Portuguese parents. Though her native Portuguese variety is the European – thus different from that of our Brazilian participants –, a Complex, Dynamic perspective leads us to believe, therefore, that those participants' native systems may be different from that of monolinguals. That is, all foreign languages they have learned are likely to have had an effect on their linguistic system, with phenomena such as attrition being assumed to have happened – and, as language is a process, be still in course (SCHERESCHEWSKY; ALVES, 2019; DE LOS SANTOS, 2017; PEREYRON, 2017; KUPSKE, 2016).

The selection of target words was based on word frequency and syllabic and segmental characteristics. However, both controls hindered our ability to control functional load on a lexical level. For instance, in a pair like 'pat' and 'pet', verb and noun functions might be somewhat even. The same cannot be said about 'fit' and 'feet', because even if we are to consider both noun functions, singular and plural nouns might affect the perception process. The case of 'sit' and 'seat' might be more 'uneven' yet, for verbs and nouns are never in equivalent syntagmatic positions.

[145] Regardless of which assessment is used, though, we expect that no test will ever be perfect in describing a learner's proficiency. We say that because, within a Complex, Dynamic framework, each listener will be unique in that respect.

It is also worth mentioning that the present study collected data during the Covid-19 pandemic. Hence, a limitation of audio quality was inevitable, as participants in the production task recorded their samples using mobile phones and under less than ideal background noise conditions. The issue was addressed by selecting the best of the three repetitions to be used as stimulus in the perception task (see section 3.4.4).

Additionally, our target words were inserted in the carrier sentence “The word is ___ too”, in order to allow for a prosodic context, without giving listeners a semantic context. However, in the same fashion as word class, this is a laboratory condition that would hardly be replicated in a real-life communication context (SALVES; WANGLON; ALVES, 2020; CRUZ, 2017). That is, sentence context might prevail over segment identification, rendering Brazilian productions more intelligible than they were found to be in the results we present in this thesis.

It should also be mentioned that, for the perception tasks, we have chosen to use a forced-choice design, in order to avoid data loss from transcriptions that did not present a coda consonant, as explained in section 3.4.4. This helped us avoid a ‘see’ transcription of a ‘seat’ token, for example. Nonetheless, this might have influenced listeners' perception, for the written forms of the target words were shown as alternatives to be chosen from. Moreover, mixing oral input with written input might have helped or hindered our listener participant’s choice process, though we did not test for that.

Also, as in any online participation design, listeners’ data were collected in unknown circumstances. We attempted to control this variable by instructing participants, prior to the beginning of data collection, to wear headphones and to play each audio file only once.

The uneven number of listener participants in each L1 group is also a limitation of the study. We intended to have as many German participants as we did of Argentinians. Invitations were sent though unofficial channels, such as email groups and social media groups, but seemed to be more enticing to the Argentinian learners than to the German ones. We hypothesise that cultural differences, as well as lack of payment for partaking in the study, might have motivated that difference. The first invitations were sent only to speakers of each L1 variety that we initially intended to analyse, that is: Riverplate Spanish and Central German. For native speakers of Spanish, we reached the target number of participants with those invitations alone. For native speakers of German, however, participation was very low when invitations were sent only to people of the Central variety. Thus, we expanded the invitation to any native speaker of German with advanced English proficiency. Our goal was to choose the German variety with the higher number of participants. By the time the data collection phase was over, Central

German ended up being that variety, but that still meant disregarding information from 11 participants from other areas/language varieties of the country. We do, however, intend to use those data in future investigations, gathering data from all varieties and observing if the observed patterns are the same or different.

In regards to listeners, it is also worth mentioning that participants had a varied pool of additional languages they spoke. Their number of languages, the order in which they were learned, and also the amount of input and active use, among other factors, might have had an effect on the way each learner's common phonetic space was organised. Moreover, six Argentinian and four German participants reported speaking Portuguese as a foreign language, a fact that might also have had an effect on the perception patterns they presented in the experiment. None of these scenarios were statistically tested, and an exploratory analysis is left for a further study (see section 5.3).

Another limitation of the present study was the technology used. Due to anonymity requirements, the programming used on the website where data were collected did not allow for participants to close the task and reopen it without losing previous responses. This might have led participants to quit participation half way through. We tried to go around the issue by displaying, at all times, a warning that the tab/browser should not be closed or the data would be lost. Additionally, the randomisation programming did not allow for a countdown. Because the experiment had 128 tokens and was relatively long, many participants might have quit the experiment for not knowing how much longer it would take them to conclude their participation[146].

We have made a choice to operationalise 'word intelligibility' as a complete match between target word and the word that listeners identified. That included onset consonant, nucleus vowel and coda consonant. However, given the small dataset, we chose to group the target words by their nuclei vowels, when it came to defining the predictor variables in the mixed models we ran. This way, correct identifications of 'sat' and 'pat' tokens, for instance, were grouped as correct [æ] identifications. We understand, though, that the different lexical items might have yielded different estimates and effects, had they been analysed individually instead of grouped by vowel. This is a limitation of the present study, and we would suggest that future studies might benefit from larger data collections, as well as from analyses that investigate each target word separately in regard to accuracy.

[146] From 120 people that signed the TCLE, only 63 concluded their participation (see details in section 3.3.3). Prior to signing the TCLE, those participants were informed that their participation was estimated to last about 35 minutes.

Finally, considering our theoretical framework, having a cross-sectional study might be seen as a limitation, as our participants' learning trajectories are only estimated and somewhat self-reported. We understand that future studies might choose to further investigate our findings in a longitudinal fashion. Nonetheless, we highlight that we attempted to provide a more individual analysis by investigating perception patterns in a stimulus-by-stimulus fashion. We also included participants and lexical items as random intercept effects on our models, whenever possible.

5.3 CONTRIBUTIONS TO THE FIELD AND FURTHER STUDIES

We hope that our study has contributed to the field of Psycholinguistics, as our data seem to suggest that intelligibility is dependent on both the speaker and the listener in a communication setting (ALBUQUERQUE, 2019). Of our findings, we highlight especially the high accuracy rates of the words with [æ]. Our results point to a somewhat 'easier' learning path in distinguishing between [æ] and [ɛ]. As Porto-Alegrense Brazilian Portuguese categories already entail a longer vowel length, it seems that shortening the [ɛ] productions might be a more effective path to dissimilate [æ] and [ɛ] than focusing on vowel height or frontness. We reiterate, however, that our findings suggest that there is an emerging effect of acoustic cues, depending on how informative they are perceived on a stimulus-by-stimulus basis.

Moreover, we believe that our findings regarding these emerging effects may also contribute to research in the field of Laboratory Phonology. As we consider the learner's L1 in combination with the speaker's L1, we may be able to investigate L2 production and perception processes that may lead us to a better understanding of how phonetic-phonological categories are established. Our analyses regarding acoustic cues might also be helpful in further investigations of the speech continuum and its perception.

Results of the present research study might have pedagogical implications in pronunciation teaching and training. Given the significant predictor variables yielded in our mixed-effects logistic models and our exploratory analyses, teachers of English should invest time in highlighting the distinctiveness of minimal pairs such as [i] – [ɪ] and [æ] – [ɛ]. More specifically, as those pairs yielded different tendencies in accurate and general vowel identifications, it seems that learners could benefit from some sort of phonetic training showcasing this distinctiveness in word contexts, as Nobre-Oliveira (2007)[147] did. Though local

[147] Nobre-Oliveira's groundbreaking study trained Brazilian learners with both natural and synthesised speech samples in order to help speakers to perceive the distinction between the high front tense/lax and the mid/central

intelligibility is only a part of global intelligibility, it can be assumed that enhancing this aspect would potentially enhance the general intelligibility of L2 speech.

Additionally, results suggest that teachers could choose to focus their instruction on vowel length, as much as on tongue height and frontness, when approaching the distinctiveness of the minimal pairs that we analysed. Previous studies (RAUBER, 2006; NOBRE-OLIVEIRA, 2003) have already shown that producing some distinction in vowel duration is a common process for Brazilian learners, so instruction that enhances spectral dimensions, along with the development of temporal cues, should be helpful in leading productions towards more intelligible patterns.

It is worth highlighting that pronunciation coursebook materials could also benefit from the present results, as well as from the literature on the topic, by suggesting activities that help speakers hit their communicative goals when it comes to the production of words with high front tense/lax vowels and mid/low minimal vowel pairs.

Finally, the significance of L1 as a predictor of both identification and accuracy seems to indicate that learners could benefit from being exposed to different patterns of L2 speech. Our study investigated Brazilian productions of English, which might be more common to Argentinian participants, due to geographical proximity – and, conversely, less common to German listeners. Having students listen to L2 speech produced by different L1 learners, as well as from different native speakers with varying native accents, might prove helpful to allow for system adaptations.

In line with Flege and Bohn's (2021) recommendation, we suggest that further studies on L2 perception may include the quantity and quality of input received by participants as part of the analysis. Controlling participant's other foreign languages (besides their L1 and L2) might also prove helpful in evaluating a possible effect of those languages in the participants' system organisations. Moreover, we understand that an investigation comparing more than one L1 variety[148] – be it of speaker's and/or listener's L1 – might also shed light on how the non-native hybrid systems interact and lead to emerging effects on speech perception.

Finally, we hope that our study can contribute to future investigations that further our understanding of the construct of 'intelligibility', as well of the processes involved in it. We

---

vowels we have also studied in the present thesis, as well as the high back [u] – [ʊ] pair. Her study found that the improvements from the group that was trained with synthesised samples transferred to natural listening. Nobre-Oliveira (*op. cit.*) also partially confirmed her hypothesis that training would improve learners' ability to distinguish the pairs of vowels, as well as lead to improvements in the students' productions.

[148] Lucena and Alves (2009, 2010) have studied the influence of native L1 varieties on L2 acquisition, though their investigation focuses on acoustic characteristics in speech production, and not on intelligibility.

hope the conclusions reached in this study may contribute to the debate surrounding L2 learning, echoing the growing concern over communication success, rather than over any ‘nativeness’ goal. Lastly, we also hope that the language processes we investigated in our study can shed some additional light on the growing evidence towards the Complex, Dynamic view of language, and that further studies can continue to analyse language development as a constantly-changing process over time.

REFERENCES

ALBUQUERQUE, J. I. A. de. **Caminhos dinâmicos em inteligibilidade e compreensibilidade de línguas adicionais:** um estudo longitudinal com dados de fala de haitianos aprendizes de português brasileiro. 2019. Tese (Doutorado em Letras) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2019. Available online at http://hdl.handle.net/10183/207592. Retrieved in: 27 mar. 2021.

ALBUQUERQUE, J. I. A. de; ALVES, U. K. Uma visão dinâmica sobre a inteligibilidade de fala: um estudo longitudinal com haitianos e brasileiros. **Revista Entrepalavras**, v. 10, p. 232-255, 2020.

ALVES, U. K. Teoria dos Sistemas Dinâmicos e desenvolvimento fonético-fonológico em uma nova língua. *In*: ORTIZ-PREUSS, E.; FINGER, I. (Org.). **A dinâmica do processamento bilíngue**. 1ed. Campinas-SP: Pontes Editores, 2018, v. , p. 117-161.

ALVES, U. K; BRISOLARA, L. B.; ROSA, L. C. da; BUSKE, A. C. S. Efeitos da duração do vozeamento da fricativa [z] na identificação, por brasileiros, de pares mínimos produzidos por hispânicos. **DIACRITICA**, v. 32, p. 437-465, 2019.

ALVES, U. K; LUCHINI, P. L. Entrenamiento perceptivo o instrucción explícita? Percepción y producción de los patrones de Voice Onset Time iniciales del inglés (LE) por estudiantes brasileños. **FORMA Y FUNCION**, v. 33, p. 135-165, 2020.

ALVES, U. K; MOTTA, C. S.. Focusing on the right cue: Perception of voiceless and voiced stops in English by Brazilian learners. **Phrasis** - Studies in Language and Literature, v. 50, p. 31-50, 2014.

ALVES, U. K; ZIMMER, M. Percepção e produção dos padrões de VOT do inglês por aprendizes brasileiros: O papel de múltiplas pistas acústicas sob uma perspectiva dinâmica. **Alfa**, v. 59, n. 1, p. 157-180, 2015.

ASSMANN, P. F.; KATZ, W. F. Time-varying spectral change in the vowels of children and adults. **Journal of the Acoustical Society of America**, v. 108, p. 1856-1866, 2000.

BAAYEN, R. H.; SHAFAEI-BAJESTAN, E. **languageR**: Analyzing Linguistic Data: A Practical Introduction to Statistics. [Source code] R package version 1.5.0. 2019. https://CRAN.R-project.org/package=languageR

BARBOSA, P.; MADUREIRA, S. **Manual de Fonética Acústica Experimental** - com aplicação aos dados do Português. São Paulo: Cortez, 2015.

BATES, D.; MAECHLER, M.; BOLKER, B.; WALKER, S. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v. 67, n.1), p.1-48, 2015.

BECKNER, C.; BLYTHE, R.; BYBEE, J.; CHRISTIANSEN, M.; CROFT, W.; ELLIS, N.; HOLLAND, J.; KE, J.; LARSEN-FREEMAN, D.; SCHOENEMANN, T. Language is a complex adaptive system: Position paper. **Language Learning**, v. 59, n. 1, p. 1-26, 2009.

BEST, C. A direct realist view of cross-language speech perception. *In*: STRANGE, W. (Ed.). **Speech perception and linguistic experience**: Theoretical and methodological issues in cross-language speech research. Timonium: York Press, 1995. p. 167-200.

BEST, C.; TYLER, M. D. Nonnative and second-language speech perception: Commonalities and complementarities. **Language experience in second language speech learning**: In honor of James Emil Flege, v. 1334, p. 1-47, 2007.

BLANK, C. A. **A influência grafo-fônico-fonológica na produção oral e no processamento de priming em multilíngues**: uma perspectiva dinâmica. 2013. Tese (Doutorado em Letras) - Centro de Educação e Comunicação, Universidade Católica de Pelotas, Pelotas, 2013. Available online at https://wp.ufpel.edu.br/ppgl/files/2018/11/A-Influ%C3%AAncia-Grafo-F%C3%B4nico-Fonol%C3%B3gica-na-Produ%C3%A7%C3%A3o-Oral-e-no-Processamento-de-Priming-em-Multil%C3%ADngues-Uma-Perspectiva-Din%C3%A2mica-Cintia-Avila-Blank-1.pdf. Retrieved in: 27 mar. 2021.

BOERSMA, P.; WEENINK, D. **Praat**: doing phonetics by computer [Computer program]. Versão 6.1.14. Amsterdam: 2015. http://www.praat.org

BOHN, O.S.; FLEGE, J. E.. The production of new and similar vowels by adult German learners of English. **SSLA**, 14 (2), p. 131-158, 1992.

BUNGE, S.; HEINZ, V. **Code Snippets** [Source code]. Version 2.14.1. https://codesnippets.pro

CALLOU, D.; MORAES, J. A; LEITE, Y. O vocalismo do português do Brasil. **Letras de Hoje**, v. 31, n. 2, p. 27-40, 1996.

CHOMSKY, N.; HALLE, M. **The sound pattern of English**. Nova Iorque: Harper & Row, 1968.

CRYSTAL, D. **English as a global language.** New York: Cambridge University Press, 2003.

CRUZ, N. C. Terminologies and definitions in the use of intelligibility: state-of-the-art. **Revista Brasileira de Linguística Aplicada**, n. 7, v. 1, p. 149-159, 2007.

CRUZ, N. C. Pronunciation intelligibility in Brazilian learners' English and listeners' threshold. **Horizontes de Linguística Aplicada**, ano 16, n. 2, 2017.

CRUZ, N. C.; PEREIRA, M. de A. Pronúncia de aprendizes brasileiros de inglês e inteligibilidade: um estudo com dois grupos de ouvintes. **Revista Virtual de Estudos da Linguagem** – ReVEL, v. 4, n. 7, p. 1-26, 2006.

DAVIES, Mark. **The Corpus of Contemporary American English (COCA)**. 2008-. https://www.english-corpora.org/coca/

DE BOT, K.; LOWIE, W.; THORNE, S. L.; VERSPOOR, M. Dynamic System Theory as a comprehensive theory of second language development. *In* MAYO, P. G; MANGADO, J. G.; ADRIÁN, M. M. (Eds.). **Contemporary perspectives on second language acquisition**. John Benjamins Publishers, 2013, p. 167-189.

DE BOT, K.; LOWIE, W.; VERSPOOR, M. A dynamic systems theory approach to second language acquisition. **Bilingualism**: Language and Cognition, v. 10, n. 1, p. 7–21, 2007.

DERWING, T.; MUNRO, M. **Pronunciation Fundamentals**: Evidence-based perspective for L2 Teaching and Research. Amsterdam, the Netherlands: John Benjamins, 2015.

ELEMENTOR. **Elementor Pro** [Source code]. Version 3.2.1. https://elementor.com

ESCUDERO, P. The linguistic perception of similar L2 sounds. *In:* BOERSMA, P.; HAMANN, S. (eds.). **Phonology in Perception**. Berlin: Mouton de Gruyter, 2009, p. 151-190.

ESCUDERO, P; POLKA, L. A cross-language study of vowel categorization and vowel acoustics: Canadian English versus Canadian French. *In:* SOLÉ, M.; RECASENS, D.; ROMERO, J. (eds.). **Proceedings of the 15th International Congress of Phonetic Sciences**. Barcelona: Causal Productions, 2003, p. 861-864.

FLEGE, J. Second language speech learning: Theory, findings, and problems. *In:* STRANGE, W. (Ed.). **Speech perception and linguistic experience**: Theoretical and methodological issues in cross-language speech research. Timonium: York Press, 1995. p. 233-272.

FLEGE, J. Assessing constraints on second-language segmental production and perception. *In* MEYERM, A.; SCHILLER, N. **Phonetics and Phonology in Language Comprehension and Production, Differences and Similarities**. Berlin: Mouton de Gruyter, 2003.

FLEGE, J; BOHN, O. The Revised Speech Learning Model (SLM-r). *In:* WAYLAND, R. (Ed.). **Second Language Speech Learning**: Theoretical and Empirical Progress. Cambridge: Cambridge University Press, 2021. pp. 3-83.

FOX, J.; WEISBERG, S. An R Companion to Applied Regression, 3rd Edition. Thousand Oaks: 2019.

GELMAN, A.; SU, Y.-S. **arm**: Data Analysis Using Regression and Multilevel/Hierarchical Models. [Source code] R package version 1.11-2. 2020. https://CRAN.R-project.org/package=arm

GONÇALVES, A. R. **In search of speech intelligibility**: the case of English high front vowels. Dissertação (Mestrado em Letras) - Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2014. Available online at https://repositorio.ufsc.br/xmlui/handle/123456789/123415. Retrieved in: 27 mar. 2021.

GONÇALVES, A. R; SILVEIRA, R. Intelligibility research in Brazil: Empirical findings and methodological issues. **Horizontes de linguística aplicada**, v. 14, p. 51-81, 2015

GRIES, S. T. Statistics for Linguistics with R: A Practical Introduction. 2 ed. Berlim: De Gruyer Mouton, 2013

HOLT, L. L.; LOTTO, A. J. Cue weighting in auditory categorization: implications for first and second language acquisition. **Journal of the Acoustical Society of America**, v. 119, n. 5, p. 3059-3071, 2006.

KÖNIG, W. **dvt-Atlas Deutsche Sprache**. 14 ed. München: Deutscher Taschenbuch-Verlag, 2004. *E-book*.

KUPSKE, F. F. **Imigração, atrito e complexidade**: a produção das oclusivas surdas iniciais do inglês e do português por sul-brasileiros residentes em Londres. Tese (Doutorado em Letras) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2016.

KUPSKE, F. F; ALVES, U. K. Orquestrando o Caos: O Ensino de Pronúncia de Língua Estrangeira à Luz do Paradigma da Complexidade. **Fórum Linguístico**, v. 14, n. 4, p. 2771-2784, 2017.

LABOV, W.; ASH, S.; BOBERG, C. **The Atlas of North American English**: Phonetics, Phonology and Sound Change. Berlin: Mouton de Gruyter, 2006.

LADEFOGED, P. **Vowels and Consonants**: An Introduction to the Sounds of Languages. Oxford, UK: Blackwell, 2010.

LADEFOGED, P.; JOHNSON, K. **A course in phonetics**. 7a ed. Stamford: Cengage, 2015.

LARSEN-FREEMAN, D. Complexity Theory: the lessons continue. *In:* ORTEGA, L.; HAN, Z. H. (eds.). **Complexity Theory and Language Development**: in celebration of Diane Larsen-Freeman. Amsterdam: John Benjamins Publishing Company, 2017, p. 11-50.

LEHET, M.; HOLT, L. I. Dimension-Based Statistical Learning Affects Both Speech Perception and Production. **Cognitive Science**, n. 41, v. 4, p. 885-912, 2017.

LEPAGE; A.; LACHARITÉ, D.. Perceptions of Intelligibility and Reaction Times: Accent-tolerant vs. Non Accent-tolerant Listeners. *In:* MOMPEAN, J. A.; FOUZ, J. (eds.). **Investigating English Pronunciation**: Current Trends and Directions. Murcia, Spain: Palgrave Macmillan, 2015.

LEVIS, J. M. Changing contexts and shifting paradigms in pronunciation teaching. **Tesol Quarterly**, v. 39, n. 3, p. 369-377, 2005.

LEVIS, J. M. **Intelligibility, Oral Communication and the Teaching of Pronunciation**. Cambridge University Press, 2018.

LEVSHINA, N. **How to do linguistics with R**: Data exploration and statistical analysis. John Benjamins Publishing Company, 2015.

LIMA JUNIOR, R. M. **A influência da idade na aquisição da fonologia do inglês como língua estrangeira por brasileiros**. 2012. Tese (Doutorado em Linguística) - Instituto de Letras, Universidade de Brasília, Brasília, 2012. Available online at https://repositorio.unb.br/handle/10482/12361. Retrieved in: 27 mar. 2021.

LIMA JUNIOR, R. M. The influence of metalinguistic knowledge of segmental phonology on the production of English vowels by Brazilian undergraduate students. **Ilha do Desterro**, v. 70, n. 3, p. 117-130, 2017.

LOWIE, W. Lost in state space? Methodological considerations in Complex Dynamic Theory approaches to second language development research. *In:* ORTEGA, L.; HAN, Z. H. (eds.). **Complexity Theory and Language Development**: in celebration of Diane Larsen-Freeman. Amsterdam: John Benjamins Publishing Company, 2017, p. 123-141.

LOWIE, W; VERSPOOR, M. Variability and Variation in Second Language Acquisition Orders: A Dynamic Reevaluation. **Language Learning**, vol. 65, n. 1, p. 63-88, 2015.

LUCENA, R. M.; ALVES, U. K. Influência do dialeto materno na aquisição de inglês (L2): o caso das obstruintes em posição de coda. **Letra Viva**, v. 9, p. 19-33, 2009.

LUCENA, R. M.; ALVES, U. K. Implicações dialetais (dialeto gaúcho vs. paraibano) na aquisição de obstruintes em coda por aprendizes de inglês: uma análise variacionista. **Letras de Hoje**, v. 45, p. 35-42, 2010.

LÜDECKE, D. **sjPlot**: Data Visualization for Statistics in Social Science. [Source code] R package version 2.8.7. https://CRAN.R-project.org/package=sjPlot

LUK, G.; BIALYSTOK, E. Bilingualism is not a categorical variable: Interaction between language proficiency and usage. **Journal of Cognitive Psychology**, v. 25, n. 5, p. 605-621, 2013.

MAACK, A. Die spezifische Lautdauer deutscher Sonanten. **STUF** - Language Typology and Universals, v. 3, n. 1-6, 1949.

MARIAN, V.; BLUMENFELD, H. K.; KAUSHANSKAYA, M. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. **Journal of Speech, Language, and Hearing Research**, v. 50, n. 4, p. 940-967, 2007.

MCCLOY, D. **phonR**: tools for phoneticians and phonologists. [Source code] R package version 1.0-7. 2016.

INTERACTION. *In:* DICTIONARY by Merriam-Webster, America's most-trusted online dictionary. Springfield: Merriam-Webster, Incorporated, 2021. Available online at https://www.merriam-webster.com/dictionary/interaction. Retrieved in: 27 mar. 2021.

MUNRO, M.; DERWING, T. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. **Language learning**, v. 45, n. 1, p. 73-97, 1995.

MUNRO, M.; DERWING, T. Intelligibility in Research and Practice: Teaching Priorities. In: REED, M.; LEVIS, J. **The Handbook of English Pronunciation**. Wiley-Blackwell: Chichester, 2015.

NOBRE-OLIVEIRA, D. **Sheep ou ship? Men ou man? O papel da hierarquia de restrições na aquisição das vogais coronais do inglês como língua estrangeira**. Dissertação (Mestrado em Letras). Universidade Católica de Pelotas, Pelotas, 2003. Available online at https://wp.ufpel.edu.br/ppgl/files/2018/11/Sheep_ou_Ship-Denise_Nobre.pdf. Retrieved in: 27 mar. 2021.

NOBRE-OLIVEIRA, D**. The effects of training on the learning of American English vowels by native Brazilian Portuguese speakers**. Tese (Doutorado em Letras). Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2007. Available online at http://repositorio.ufsc.br/xmlui/handle/123456789/90286. Retrieved in: 27 mar. 2021.

PEREYRON, L. **A produção vocálica por falantes de espanhol (L1), inglês (L2) e português (L3)**: uma perspectiva dinâmica na (multi)direcionalidade da transferência linguística. Tese (Doutorado em Letras). Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Available online at http://hdl.handle.net/10183/172391. Retrieved in: 27 mar. 2021.

PINTO, M. O. **Índice de variabilidade** - um critério de avaliação de parâmetros acústicos de vogais orais e consoantes nasais como elemento para a perícia de comparação de locutores. Tese (Doutorado em Linguística). Escola de Humanidades, Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2017. Available online at http://tede2.pucrs.br/tede2/handle/tede/7798. Retrieved in: 27 mar. 2021.

RAUBER, A. S. **Perception and production of English vowels by Brazilian EFL speakers**. Tese (Doutorado em Letras) - Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2006. Available online at http://repositorio.ufsc.br/xmlui/handle/123456789/88701. Retrieved in: 27 mar. 2021.

RSTUDIO TEAM. **RStudio**: Integrated Development for R [Computer program]. Version 4.0.3. Boston: RStudio, 2021. Available online at http://www.rstudio.com. Retrieved in: 27 mar. 2021.

SALVES, D.; WANGLON, P.; ALVES, U. K. The role of L1 English speakers' familiarity with Brazilian-accented English (L2) in the intelligibility of Brazilian learners of English (L2): a discussion on intelligibility from a Complex Dynamic Systems perspective. **Ilha do Desterro**, v. 73, p. 339-362, 2020.

SANTOS, B. da R. de L. **A produção da vogal átona final /e/ por porto-alegrenses aprendizes de espanhol como segunda língua (L2)**: uma investigação sobre atrito linguístico em ambiente de L2 não - dominante. 2017. Dissertação (Mestrado em Letras) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Available online at http://hdl.handle.net/10183/172845. Retrieved in: 27 mar. 2021.

SANTOS, G. **Percepção e produção das vogais médias do espanhol por falantes do português brasileiro**. 2014. Tese (Doutorado em Letras) - Centro de Educação e Comunicação, Universidade Católica de Pelotas, Pelotas, 2014. Available online at http://tede.ucpel.edu.br:8080/jspui/handle/tede/350. Retrieved in: 27 mar. 2021.

SANTOS, G.; RAUBER, A. Percepção e produção das vogais médias do espanhol/LE. *In:* ALVES, U. (Org.). **Aquisição fonético-fonológica de língua estrangeira**: Investigações rio-grandenses e argentinas em discussão. Campinas: Pontes, 2016, p. 151-178.

SCHERESCHEWSKY, L. C.; ALVES, U. K. A variabilidade no desenvolvimento e no atrito linguístico: o papel da experiência individual no estudo sobre os padrões de VOT em sistemas multilíngues. **PROLÍNGUA**, v. 14, p. 146-158, 2019.

SCHOLL, A. P.; FINGER, I. Elaboração de um questionário de histórico da linguagem para pesquisas com bilíngues. **Nonada**: Letras em Revista, v. 2, n. 21, 2013.

SCHOLL, A; FINGER, I.; FONTES, A.B.A.L. Fatores de experiência linguística associados à proficiência autoavaliada por usuários de inglês como língua adicional. **Letrônica**, v. 10, n. 2, p. 689-699.

SCHOORMANN, H; HEERINGA, W.; PETERS, J. Standard German vowel productions by monolingual and trilingual speakers. **International Journal of Bilingualism**, v. 23, n. 1, 2019. p.138-156.

SCHWARTZHAUPT, B. M. **Testing intelligibility in English**: the effects of Positive VOT and contextual information in a sentence transcription task. 2015. Dissertação (Mestrado em Letras) - Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015. Available online at http://hdl.handle.net/10183/131639. Retrieved in: 27 mar. 2021.

SCHWARTZHAUPT, B. M; ALVES, U. K; FONTES, A. B. A. da L. The role of L1 knowledge on L2 speech perception: investigating how native speakers and Brazilian learners categorize different VOT patterns in English. **Revista de Estudos da Linguagem**, v. 23, p. 311-334, 2015.

SENDLMEIER, W. F.; SEEBODE, J. **Formantkarten des deutschen Vokalsystems**. 2010. Available online at https://www.kw.tu-berlin.de/fileadmin/a01311100/Formantkarten_des_deutschen_Vokalsystems_01.pdf. Retrieved in 27 mar. 2021.

SILVA, T.C. ; SEARA, I. C. ; SILVA, A. H. P. ; RAUBER, A.S. ; CANTONI, M. **Fonética Acústica** - os sons do português brasileiro. 1. ed. São Paulo: Contexto, 2019.

SILVEIRA, R. **The influence of pronunciation instruction on the perception and production of English word-final consonants**. 2004. Tese (Doutorado em Letras) - Centro de Comunicação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2004. Available online at http://repositorio.ufsc.br/xmlui/handle/123456789/87379. Retrieved in: 27 mar. 2021.

SILVEIRA, R. PL2 Production of English Word-Final Consonants: The Role of Orthography and Learner Profile Variables. **Trab. Ling. Aplic.**, Campinas, v. 51, n. 1, p.13-34, 2012.

SMITH, L.; NELSON, C. International intelligibility of English: directions and resources. **World Englishes**, v. 4, n. 3, p. 333-342, 1985.

STOECKLE, P. **Country Report Germany**. 2009. Available online at http://dx.doi.org/10.13140/RG.2.1.2797.0165. Retrieved in 27 mar. 2021.

STRANGE, W. Cross-language study of speech perception: a historical review. *In:* W. Strange (ed.), **Speech Perception and Linguistic Experience**: Issues in Cross-Language Research, 3-45. Timonium, MD: York Press, 1995.

STRANGE, W.; BOHN, O.S. Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. **Acoustical Society of America**, v. 104, n. 1, 1998. p.488-504.

STRANGE, W.; BOHN, O. S.; TRENT, S. A.; NISHI, K. Acoustic and perceptual similarity of North German and American English vowels. **Acoustical Society of America**, vol. 115, n. 4, 2004. pp.1791-1807

THE BRITISH NATIONAL CORPUS. Version 3 (BNC XML Edition). 2007. Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium. Available online at http://www.natcorp.ox.ac.uk/. Retrieved in: 27 mar. 2021.

VAN LEUSSEN, J.; ESCUDERO, P. Learning to perceive and recognize a second language: the L2LP model revised. **Frontiers in Psychology**, v. 6., 2015.

WICKHAM, H. **ggplot2**: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016. https://ggplot2.tidyverse.org

WICKHAM, H. *et al*. Welcome to the tidyverse. **Journal of Open Source Software**, v. 4, n.43, 1686, 2019.

WICKHAM, H.; SEIDEL, D. **scales**: Scale Functions for Visualization. [Source code] R package version 1.1.1. 2020. https://CRAN.R-project.org/package=scales

WINTER, B. **Statistics for Linguists**: An Introduction Using R. Routledge: New York, 2020.

WOOD E. R.; MENZEL, I. **Show yourself**. Walt Disney Records, 2019. Available at: https://www.youtube.com/watch?v=md7dK5-qvHc . Accessed in: Mar 25, 2008

WPVIBES. **Form Vibes** - Database Manager for Forms. [Source code] Version 3.0.9. 2021. https://wpvibes.com/products/form-vibes/

YAVAS, M. **Applied English Phonology**. 2 ed. Wiley-Blackwell: Chichester, 2011.

YU, H.; LOWIE, W. Dynamic Paths of Complexity and Accuracy in Second Language Speech: A Longitudinal Case Study of Chinese Learners. **Applied Linguistics**, 2019.

ZIMMER, M.; SILVEIRA, R.; ALVES, U. K. **Pronunciation instruction for Brazilians**: Bringing theory and practice together. New Castle: Cambridge Scholars Publishing, 2009.

APPENDIX A – Informed Consent Form in Portuguese for speaker participants (TCLE-BR)

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Você está sendo convidada(o) a participar de uma pesquisa sobre a pronúncia de brasileiros falando inglês. Para professores poderem ensinar pronúncia, é preciso saber o que ensinar e o que corrigir. Nesse sentido, é necessário identificar o que é sotaque e o que pode ser um empecilho para a comunicação. A fala em língua estrangeira tende a ter características que vêm da língua nativa. Assim, o inglês falado por brasileiras(os) vai soar diferente do inglês de argentinas(os), por exemplo, e ambos serão diferentes do inglês de falantes nativas(os). Algumas dessas características peculiares são apenas parte do sotaque da pessoa, mas outras podem significar alterações que impedem quem ouve de entender o que está sendo dito.

Esta pesquisa busca identificar na pronúncia de brasileiras e brasileiros quais características podem fazer parte de uma categoria de um som ou de outro som. Para realizar a pesquisa, você responderá a um questionário sobre sua experiência linguística (média de 15 minutos para preenchimento), adicionalmente às respostas que você já deu na etapa de triagem sobre sua proficiência na língua.

Você e demais participantes brasileiras(os) também vão gravar frases em inglês. Os áudios não terão identificação de quem está falando. Cada pessoa recebe um número antes da gravação e toda a pesquisa é feita apenas com esse número, sem uso de nome ou outro dado identificador. As frases gravadas por você e pelas(os) outras(os) brasileiras(os) falaram serão usadas em outra etapa da pesquisa, quando aprendizes de inglês de outros países (Argentina, Alemanha) vão transcrever o que foi dito nas gravações. Essa segunda etapa com falantes não-nativas(os) será feita online, através de site dedicado à pesquisa e após convite para pessoas desses países. Sua identidade também será mantida em sigilo nessa fase posterior do estudo, quando outras(os) participantes vão ouvir as gravações realizadas na etapa em que você está sendo convidada(o) a participar.

A gravação será realizada na sua casa, preferencialmente em um cômodo em que esteja só você, com porta e janelas preferencialmente fechadas, em um espaço silencioso. Você vai ler uma lista de frases em inglês enquanto as grava usando seu celular. As frases serão apresentadas em um PowerPoint que você deverá colocar para rodar em seu computador. Haverá uma sessão de prática, com uma lista curta de 4 itens, para você verificar se entendeu a tarefa e tirar dúvidas. Essa prática tem duração estimada em 2 minutos. Depois, haverá a gravação da lista para a pesquisa (24 itens, repetidos 3 três vezes cada, num total de 72 frases), com duração aproximada de 18 minutos. Essa segunda etapa será dividida em três partes, de cerca de 5 minutos cada, para você descansar a voz entre uma e outra. Ao final, você responderá um breve questionário acerca da gravação que realizou (tempo estimado em 2 minutos). O tempo total de participação na pesquisa é de cerca de 37 minutos, contando as gravações, os intervalos e os questionários.

Você pode sentir-se cansada(o) ou entediada(o) ao longo das gravações. Por isso há intervalos programados, mas se você quiser também pode parar a gravação a qualquer momento e retomá-la depois, recomeçando a apresentação e a gravação do zero. Também é possível desistir de participar a qualquer momento, sem precisar justificar a decisão. Há risco de você se sentir constrangida(o) em frente ao microfone. Lembre-se, nesse sentido, que as pessoas que vão ouvir você também são falantes de outras línguas nativas e que tem o inglês como língua adicional. Finalmente, há risco remoto de as(os) participantes estrangeiras(os) que ouvirem os áudios reconhecerem sua voz. Cabe dizer que você receberá um número, que é usado também na gravação, para garantir sua privacidade e preservar sua identidade.

Os resultados referentes à percepção das suas produções e das(os) suas(seus) demais colegas brasileiras(os) serão disponibilizados publicamente, sempre mantendo o sigilo das informações de todas(os) as(os) participantes. Concluída a etapa perceptual com estrangeiras(os), na qual serão apresentados os áudios de suas produções, o site dedicado será imediatamente retirado do ar com consequente exclusão de quaisquer arquivos de áudio e de respostas do servidor da página, arquivos esses que serão gravados em mídia física (DVD-ROM) e arquivados, em móvel chaveado, no gabinete do orientador desta pesquisa. As gravações e demais dados serão armazenados por até 5 (cinco) anos. Após período de 5 anos do encerramento da pesquisa aqui proposta, as mídias físicas serão destruídas.

Não há benefício direto para você na participação neste estudo. Entretanto, as descobertas deste estudo podem ajudar professores e pesquisadores em atividades de sala de aula e de avanço científico, e por isso sua participação é tão importante. Ainda assim, você pode se recusar a participar, e também pode deixar o estudo em qualquer fase posterior sem nenhum prejuízo a você.

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 Secretaria Municipal de Saúde de Porto Alegre (SMSPA):

Comitê de Ética em Pesquisa/SMSPA
Rua Capitão Montanha, 27 – 7º andar – 90010-040 – Porto Alegre, RS
Telefone: (51) 3289-5517
E-mail: cep_sms@hotmail.com.br e cep-sms@sms.prefpoa.com.br
Horário de atendimento externo: 8h até às 14h, sem intervalo.

____
DATA

____
Participante

____
Déborah Salves (Mestranda)

____
Prof. Dr. Ubiratã Kickhöfel Alves (Orientador)

APPENDIX B – Informed Consent Form in English for speaker participants (TCLE-CA)

INFORMED CONSENT FORM

You are being invited to take part in a research experiment about non-native pronunciations of English. In order for teachers to work on pronunciation, it is important to know what to teach and what to correct. To do that, it is essential to identify which characteristics of foreign speech can be considered to be accented, and which can lead to an unsuccessful communication. Non-native speech tends to carry characteristics related to an individual's native language. That is to say that the English spoken by an Argentinian student will sound different from the one spoken by a German learner, and both will also differ from the English spoken by a native English speaker. Some of these distinct characteristics are just part of an individual's accent, but some can have alterations that will prevent the listener from understanding what is being said.

This research study aims to identify which characteristics of non-native speech may help include a sound in one category or in another. The speakers in this study have distinct proficiency levels (from beginner to advanced), are all from a specific country, one that is different from yours and from all other participants in this stage of the study that you are being invited to participate. The listeners, like yourself, are all highly proficient (advanced) in English and from two different countries: Argentina and Germany.

As you take part in this research, you and all other participants in this stage will listen to a series of recordings containing sentences in English (288 total) and will be asked to transcribe the target word (that is, the word that precedes 'too', which is now being shown as a gap: "The word is ___ too"). The sentences are all the same, the only different word is the second last, which is why you only need to type that word. After the transcription, you will be asked to choose, on a scale of 1 to 9, how hard it was for you to understand the recording. We ask you to use headphones or earphones to participate in this stage, which will take about 35 minutes of your time.

Initially, you will read the instructions in an introductory screen, and then you will have a practice session, with a short list of two recordings, in order to give you the chance to see if you have any doubts. After you are comfortable with your understanding of the method of the study, you can click the "next" button and start the data collection. Each recording will be played only once, after which you will type down the target word in the corresponding space, and drag the scale to annotate the difficulty you had in the provided scale. When you are done you can click the "next" button and the next sentence will be played. This part of the research will be divided into five blocks, of about 6 and half minutes each, so you can get some rest between the blocks. These pauses are scheduled every 60 sentences. When you feel that you have rested enough, you can click the "next" button to start the next block.

After all recordings have been transcribed, you will be asked to fill in a questionnaire about your experience with English and other foreign languages, so we can have an insight on how you have learned the language and how you use it (estimated time to fill the form is about 15 minutes). The total time of your participation in the present research is about 50 minutes, which include listening and transcribing the words on the recordings, taking the breaks and filling in the questionnaire.

You may feel tired or bored throughout your participation. That is why there are breaks scheduled between the blocks of recordings, but you can also pause the task at any given time and restart it after whatever time you deem needed. You can also quit the participation at any point, without needing to justify your decision. Because the task is entirely in English, you might feel embarrassed or insecure for not being able to understand some of the recordings. In order to avoid this feeling, we remind you that the sentences were recorded by learners with different proficiency levels, which reinforces that they are all, like you, people who are learning/have learned English as a foreign language.

The stage of the research to which you are being invited to participate is entirely done online, on the same web address where you are reading the present document, and no personal information about you will be collected. Your identity will not be requested at any point of this study.

The results concerning how you and all Argentinian and German participants listen and transcribe the sentences in English that you have listened to will be made publically available, always preserving the confidentiality of any data collected from participants. After this stage of the study is concluded, this website will be immediately taken down, with the consequent exclusion of any audio or answer file from the online server. The files with your answers will be recorded in hard copies (DVD-ROM) and filed, in a locked cabinet, inside the office of the professor who advises this study. After five years of the conclusion of this study, all hard copies will also be destroyed.

There are no direct benefits from you in participating in this study. Any discovery made with its results, however, can help teachers and researchers during class activities and allow for scientific advance, which is why it is so important that you accept the invitation to participate.

Still, you may refuse to do so, and you can also, at any moment, change your mind and leave the study, without any harm to you.

If you have any questions about this study, please contact the professor who advises this thesis project:

Prof. Ubiratã Kickhöfel Alves
Prédio Administrativo do Instituto de Letras – Sala 220 – Campus do Vale
Av. Bento Gonçalves, 9500 – 91501-000 – Porto Alegre, RS
Phone: +55 (51) 3308-7081
Email: ukalves@pq.cnpq.br

If you have any questions regarding your rights or your participation in this study, please write to the Research Ethics Committee from City Health Department of Porto Alegre (SMSPA):

Comitê de Ética em Pesquisa/SMSPA
Rua Capitão Montanha, 27 – 7º andar – 90010-040 – Porto Alegre, RS
Phone: +55 51 3289-5517
E-mail: cep_sms@hotmail.com.br e cep-sms@sms.prefpoa.com.br
Office hours: 8am to 2pm.

If you wish, you can inform your email to receive a digital copy of the present document.

____
DATE

____
Participant

____
Déborah Salves (Research advisee)

____
Prof. Ubiratã Kickhöfel Alves, PhD (Research adviser)

APPENDIX C – Informed Consent Form in English for listener participants (TCLE-EN)

INFORMED CONSENT FORM

You are being invited to take part in a research experiment about non-native pronunciations of English. In order for teachers to work on pronunciation, it is important to know what to teach and what to correct. To do that, it is essential to identify which characteristics of foreign speech can be considered to be accented, and which can lead to an unsuccessful communication. Non-native speech tends to carry characteristics related to an individual's native language. That is to say that the English spoken by an Argentinian student will sound different from the one spoken by a German learner, and both will also differ from the English spoken by a native English speaker. Some of these distinct characteristics are just part of an individual's accent, but some can have alterations that will prevent the listener from understanding what is being said.

This research study aims to identify which characteristics of non-native speech may help include a sound in one category or in another. The speakers in this study have distinct proficiency levels (from beginner to advanced), are all from a specific country, one that is different from yours and from all other participants in this stage of the study that you are being invited to participate. The listeners, like yourself, are all highly proficient (advanced) in English and from two different countries: Argentina and Germany.

As you take part in this research, you and all other participants in this stage will listen to a series of recordings containing sentences in English (240 total) and will be asked to transcribe the target word (that is, the word that precedes 'too', which is now being shown as a gap: "The word is ___ too"). The sentences are all the same, the only different word is the second last, which is why you only need to type that one word. We ask you to use headphones or earphones to participate in this stage, which will take about 35 minutes of your time.

Initially, you will read the instructions in an introductory screen, and then you will have a practice session, with a short list of two recordings, in order to give you the chance to see if you have any doubts. After you are comfortable with your understanding of the method of the study, you can click the "start" button and start the data collection. Each recording should be played only once, after which you will type down the target word in the corresponding space. When you are done you can click the "next" button and the next sentence will be displayed.

You can click the “play” icon whenever you are ready, and you can take breaks to rest if you feel like it. When you feel that you have rested enough, you can click the “play” icon to resume.

After all recordings have been transcribed, you will be asked to fill in a questionnaire about your experience with English and other foreign languages, so we can have an insight on how you have learned the language and how you use it (estimated time to fill the form is about 15 minutes). The total time of your participation in the present research is about 50 minutes, which include listening and transcribing the words on the recordings, taking the breaks and filling in the questionnaire.

You may feel tired or bored throughout your participation. That is why you can take breaks during your participation. You can also quit the participation at any point, without needing to justify your decision. Because the task is entirely in English, you might feel embarrassed or insecure for not being able to understand some of the recordings. In order to avoid this feeling, we remind you that the sentences were recorded by learners with different proficiency levels, which reinforces that they are all, like you, people who are learning/have learned English as a foreign language.

The stage of the research to which you are being invited to participate is entirely done online, on the same web address where you are reading the present document, and no personal information about you will be collected. Your identity will not be requested at any point of this study.

The results concerning how you and all Argentinian and German participants listen and transcribe the sentences in English that you have listened to will be made publically available, always preserving the confidentiality of any data collected from participants. After this stage of the study is concluded, this website will be immediately taken down, with the consequent exclusion of any audio or answer file from the online server. The files with your answers will be recorded in hard copies (DVD-ROM) and filed, in a locked cabinet, inside the office of the professor who advises this study. After five years of the conclusion of this study, all hard copies will also be destroyed.

There are no direct benefits from you in participating in this study. Any discovery made with its results, however, can help teachers and researchers during class activities and allow for scientific advance, which is why it is so important that you accept the invitation to participate. Still, you may refuse to do so, and you can also, at any moment, change your mind and leave the study, without any harm to you.

If you have any questions about this study, please contact the professor who advises this thesis project:

Prof. Ubiratã Kickhöfel Alves
Prédio Administrativo do Instituto de Letras – Sala 220 – Campus do Vale
Av. Bento Gonçalves, 9500 – 91501-000 – Porto Alegre, RS
Phone: +55 (51) 3308-7081
Email: ukalves@pq.cnpq.br

If you have any questions regarding your rights or your participation in this study, please write to the Research Ethics Committee from City Health Department of Porto Alegre (SMSPA):

Comitê de Ética em Pesquisa/SMSPA
Rua Capitão Montanha, 27 – 7º andar – 90010-040 – Porto Alegre, RS
Phone: +55 51 3289-5517
E-mail: cep_sms@hotmail.com.br e cep-sms@sms.prefpoa.com.br
Office hours: 8am to 2pm.

If you wish, you can inform your email to receive a digital copy of the present document.

____

DATE

____

Participant

____

Déborah Salves (Research advisee)

____

Prof. Ubiratã Kickhöfel Alves, PhD (Research adviser)

APPENDIX D – Self-Assessment Proficiency section for Brazilian speaker participants (screening stage)

SELF-ASSESSMENT OF PROFICIENCY LEVEL

Pronome:

[ ] Masculino
[ ] Feminino
[ ] Outro

1. Liste as línguas que você sabe na ordem em que as aprendeu (a Língua 1 deve ser a língua nativa). Depois, circule, na escala de 1 a 6, seu nível de proficiência em cada língua que sabe (1 = muito baixo, 2 = baixo, 3 = razoável, 4 = alto; 5 = muito alto e 6 = fluente):

Língua 1: _______						
Leitura	1	2	3	4	5	6
Escrita	1	2	3	4	5	6
Compreensão auditiva1	2	3	4	5	6	
Fala	1	2	3	4	5	6
Língua 2: _______						
Leitura	1	2	3	4	5	6
Escrita	1	2	3	4	5	6
Compreensão auditiva1	2	3	4	5	6	
Fala	1	2	3	4	5	6
Língua 3						
Leitura	1	2	3	4	5	6
Escrita	1	2	3	4	5	6
Compreensão auditiva1	2	3	4	5	6	
Fala	1	2	3	4	5	6
Língua 4						
Leitura	1	2	3	4	5	6
Escrita	1	2	3	4	5	6
Compreensão auditiva1	2	3	4	5	6	
Fala	1	2	3	4	5	6

2. Você teve alguma experiência de imersão em país anglofalante por período superior a seis meses?

[ ] Sim.
[ ] Não.

3. Você já teve algum treinamento fonético/fonológico/de pronúncia na língua inglesa?

[ ] Sim.
[ ] Não.

APPENDIX E – Language History Questionnaire in Portuguese for Brazilian speaker participants

QUESTIONÁRIO DE HISTÓRICO DA LINGUAGEM
(adaptado de Scholl e Finger, 2013 e Scholl, Finger e Luz Fontes, 2017)

Olá, agradecemos sua participação neste estudo. Este questionário tem a função de permitir entender melhor como você aprendeu inglês e como utiliza o idioma. Esses dados são importantes para que pesquisadores e professores possam compreender a maneira como o aprendizado se desenvolve e como pode ser auxiliado de maneira pedagógica e eficiente. Seus dados servem para propósitos descritivos estatísticos e não serão identificados, conforme o Termo de Consentimento Livre e Esclarecido que você assinou. Obrigada.

Participante nº: _______

Parte 1 – Aprendizado de idiomas

1. Liste todas as línguas que você sabe em ordem de aquisição, sendo a língua 1 a sua língua nativa (se você sabe mais do que 4 línguas, liste apenas as primeiras 4 que você aprendeu):

Língua 1	Língua 2	Língua 3	Língua 4

2. Indique onde você aprendeu essas línguas (você pode marcar mais de uma opção em cada língua):

	Língua 1	Língua 2	Língua 3	Língua 4
Casa				
Escola				
Curso de idiomas				
Sozinho				
Outro				

3. Informe a idade (em anos) em que você:

	Língua 1	Língua 2	Língua 3	Língua 4
Começou a aprender				
Começou a utilizar ativamente				
Tornou-se fluente				

4. Indique, em uma escala de 0 a 6 (0 = nada, 1 = muito pouco, 2 = pouco, 3 = mais ou menos, 4 = razoavelmente, 5 = consideravelmente, e 6 = muito), o quanto cada um destes fatores contribuiu para a aprendizagem das suas línguas:

	Língua 1	Língua 2	Língua 3	Língua 4
Interação com a família				
Interação com os amigos				
Leitura de revistas e jornais				
Leitura de livros				
Leitura de textos acadêmicos				
Televisão (filmes, séries, YouTube, etc)				
Música / rádio / podcast				
Videogame				
Redes sociais				
Curso de idiomas				

5. Você já realizou algum treinamento de pronúncia ou cursou alguma disciplina de Fonética e/ou Fonologia? Escreva SIM ou NÃO no quadrado correspondente:

	Língua 1	Língua 2	Língua 3	Língua 4
Treinamento de pronúncia				
Disciplina de Fonética / Fonologia				

Parte 2 – Uso dos idiomas

6. Estime em número de horas o quanto você usa cada língua para as seguintes atividades diariamente:

	Língua 1	Língua 2	Língua 3	Língua 4
Interação com a família				
Interação com os amigos				
Leitura (livros, revistas, jornais)				
Leitura de textos acadêmicos				
Televisão (filmes, séries, YouTube, etc)				
Música / rádio / podcast				
Videogame				
Redes sociais				
Curso de idiomas				

7. Informe o número de meses que você passou em cada um destes ambientes (use 0 se não teve passou nenhum tempo):

	Língua 1	Língua 2	Língua 3	Língua 4
País em que a língua é falada ativamente				
Família em que a língua é falada				
Escola em que a língua é falada				
Trabalho em que a língua é falada				

8. Caso haja alguma outra informação que você ache importante sobre o aprendizado ou o uso das suas línguas, por favor, escreva abaixo:

____

____

Parte 3 – Demográfico

Sexo:
( ) F
( ) M
( ) Outro

Idade:

Cidade de nascimento:

____

Cidade em que você morou até os 7 anos de idade (liste em sequência a partir do nascimento, se mais de uma:

____

Cidade de nascimento da mãe ou da figura materna em sua vida:

____

APPENDIX F – Language History Questionnaire in English for Canadian speaker participants and Argentinian and German listener participants

LANGUAGE EXPERIENCE QUESTIONNAIRE
(adapted from Scholl and Finger, 2013 and Scholl, Finger and Luz Fontes, 2017)

Hi, we thank you for participating in this study. This questionnaire aims to help us have a better understanding of how you have learned English and how you use the language. The data are important for researchers and teachers to comprehend the ways the learning process develops and how it can be aided in a pedagogical and efficient way. Your data will be used for descriptive statistics and you will not be identified, in accordance with the Informed Consent Form you signed electronically. Thank you.

Part 1 – Language learning

1. List the languages you speak, in a chronological order according to the one you have learned first, using your native language as Language 1 (if you speak more than 4 languages, list just the first 4 you have learned):

Language 1	Language 2	Language 3	Language 4

2. Indicate where you have learned those languages (you can check more than one option for each language):

	Language 1	Language 2	Language 3	Language 4
Home				
School				
Language school				
By yourself (self taught)				
Other				

3. Inform at which age (in years) you:

	Language 1	Language 2	Language 3	Language 4

started to learn				
started to actively use				
became fluent				

4. Inform, on a scale from 0 to 6 (0 = nothing, 1 = very little, 2 = little, 3 = somewhat, 4 = reasonably, 5 = considerably, e 6 = a lot), how much each of these factors has contributed to you learning of the languages:

	Language 1	Language 2	Language 3	Language 4
Interaction with your family				
Interaction with your friends				
Reading magazines and newspapers				
Reading books				
Reading academic literature				
Television (movies, TV shows, YouTube, etc)				
Music / radio / podcast				
Videogame				
Social networks				
Language school				

5. Have you ever had any pronunciation training or have you ever taken a course on Phonetics/Phonology at college/university? Mark YES or NO in the corresponding field:

	Language 1	Language 2	Language 3	Language 4
Pronunciation training				
Course on Phonetics/Phonology				

Part 2 – Use of languages

6. Estimate, in number of hours (on a daily basis), how much you use each language you speak in the following activities:

	Language 1	Language 2	Language 3	Language 4
Interaction with your family				
Interaction with your friends				
Reading (books, magazines, newspapers)				
Reading academic literature				
Television (movies, TV shows, YouTube, etc)				
Music / radio / podcast				
Videogame				
Social networks				
Language school				

7. Inform the number of months that you have spent in each of these environments (use 0 if you have not spent any time in one or more of them):

	Language 1	Language 2	Language 3	Language 4
Country in which the language is actively spoken				
Family in which the language is actively spoken				
School in which the language is spoken				
Workplace in which the language is spoken				

8. In case there is any other information about your language learning or about your language use that you find relevant, please mention it below::

____

____

Part 3 – Self-assessment of proficiency level[149]

1. Mark, on a scale from 1 to 6, your proficiency level in each language you speak (1 = very low, 2 = low, 3 = reasonable, 4 = high; 5 = very high and 6 = fluent):

Language 1: _______

Reading	1	2	3	4	5	6
Writing	1	2	3	4	5	6
Listening	1	2	3	4	5	6
Speaking[150]	1	2	3	4	5	6

Language 2: _______

Reading	1	2	3	4	5	6

[149] Ainda que já tenham informado nível avançado de proficiência, interessa saber se sua autoavaliação será de 5 ou de 6 na escala likert, bem como se essa autoavaliação evidencia níveis diferentes de percepção para habilidades diferentes (ainda que a presente pesquisa não utilize um delas, speaking).

[150] "Marque, em uma escala de 1 a 6, seu nível de proficiência em cada uma das línguas que você fala (1 = muito baixo, 2 = baixo, 3 = razoável, 4 = alto; 5 = muito alto e 6 = fluente):"

Writing	1	2	3	4	5	6
Listening	1	2	3	4	5	6
Speaking	1	2	3	4	5	6
Language 3: _______						
Reading	1	2	3	4	5	6
Writing	1	2	3	4	5	6
Listening	1	2	3	4	5	6
Speaking	1	2	3	4	5	6
Language 4: _______						
Reading	1	2	3	4	5	6
Writing	1	2	3	4	5	6
Listening	1	2	3	4	5	6
Speaking	1	2	3	4	5	6

Part 4 – Demographic information

Gender:
( ) F
( ) M
( ) Other

Age:

City of birth:

____

City where you lived until you were 7 years old (if there was more than one, list them chronologically from your birth):

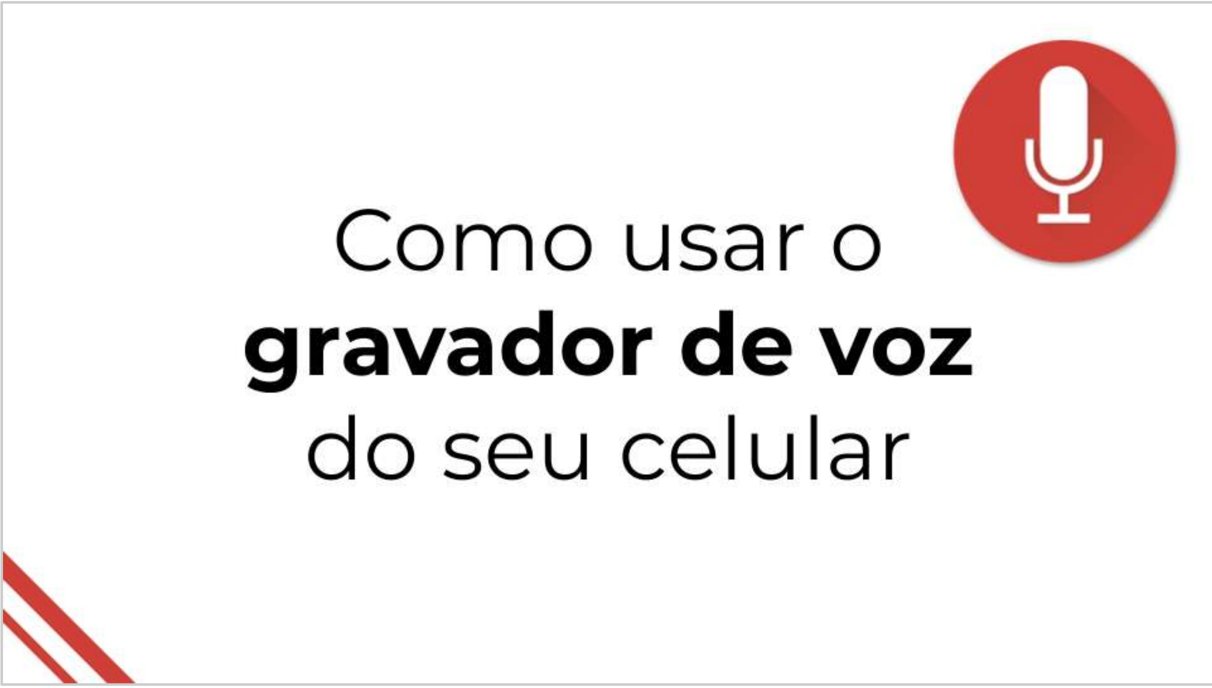
____

City where your mother (or the motherly figure in your life) was born:

____

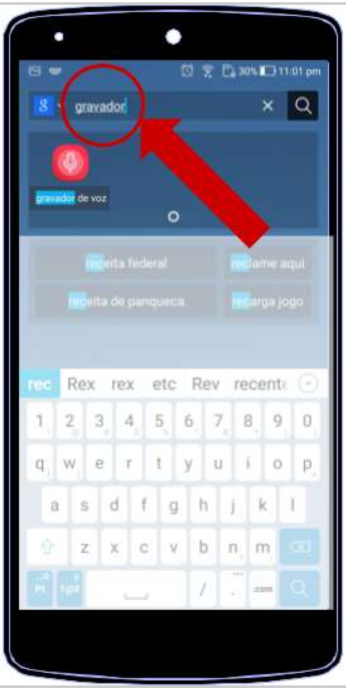
APPENDIX G – Supporting instructions (in Portuguese) on how to use the mobile phone's native voice recording app

COMO USAR O GRAVADOR DE VOZ DO SEU CELULAR

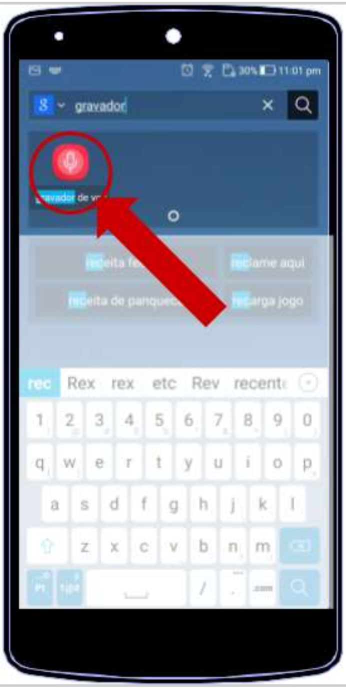


2 Procure pelo ícone de Busca (lupa)

3 Digite: "gravador"

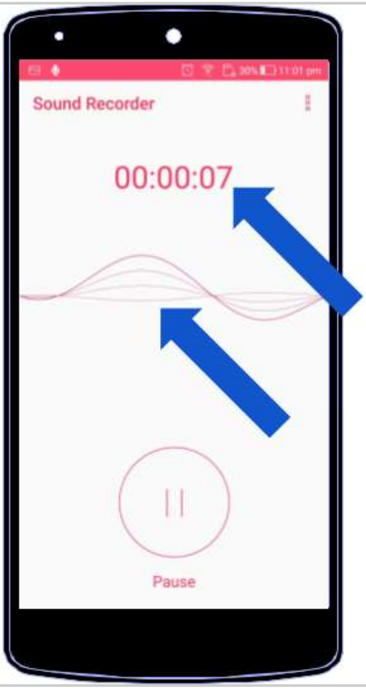


4 Toque sobre o ícone do aplicativo/app de "Gravador de voz"



5 Toque sobre o botão do microfone para começar a gravar

Você vai ver que o tempo começa a contar e que as ondas no meio da tela se movem quando você fala.

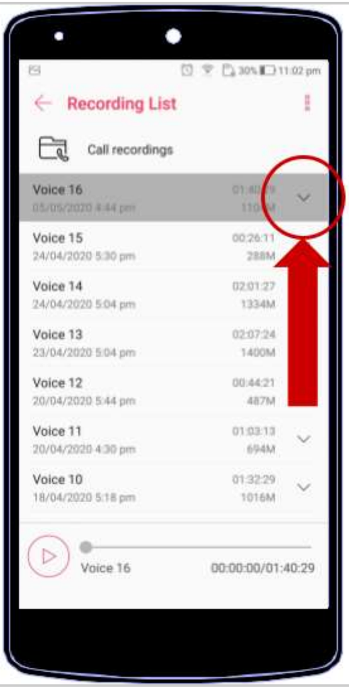


6 Quando terminar a gravação, aperte sobre o botão com o QUADRADO.

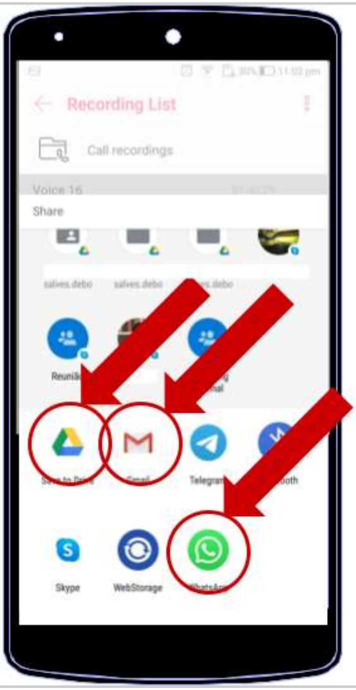
(Dependendo do seu celular, você pode precisar ter que apertar Pause antes de aparecer o quadrado.)

7 Na lista de gravações, a que você acabou de fazer será a primeira no topo. Clique na seta na parte direita dessa gravação e, depois, selecione a opção "Compartilhar".

(Se a lista não aparecer após você apertar o quadrado, clique no ícone do canto superior direito.)

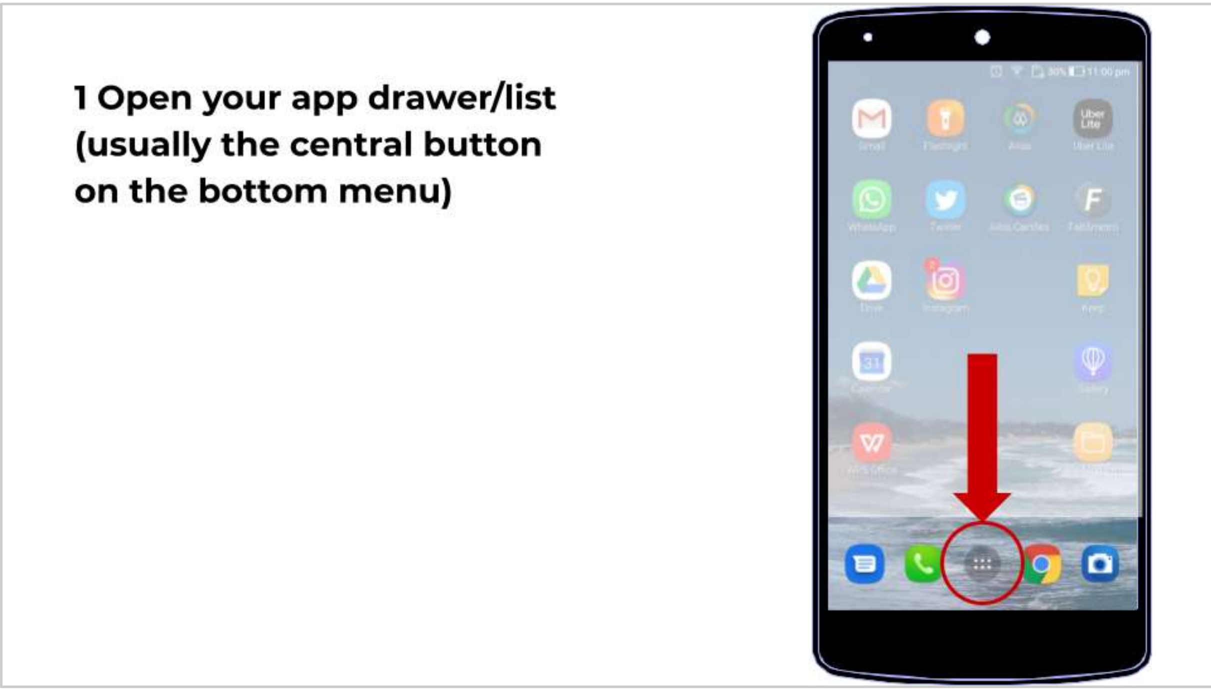


8 Você pode fazer o envio pelo WhatsApp, por email ou pelo Google Drive, como ficar mais fácil pra você :)



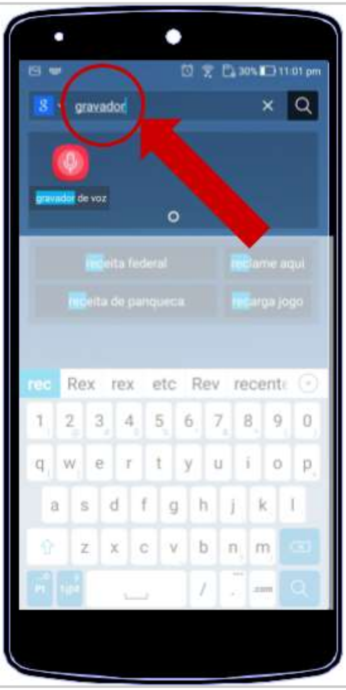
APPENDIX H – Supporting instructions (in English) on how to use the mobile phone's native voice recording app

HOW TO USE YOUR MOBILE PHONE'S VOICE RECORDER APP

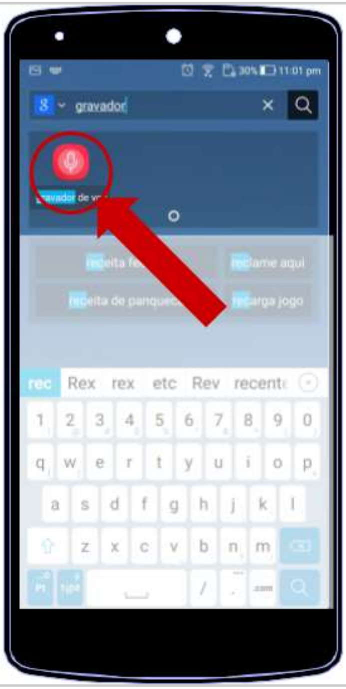


2 Look for the Search icon (magnifying glass)

3 Type: "recorder"

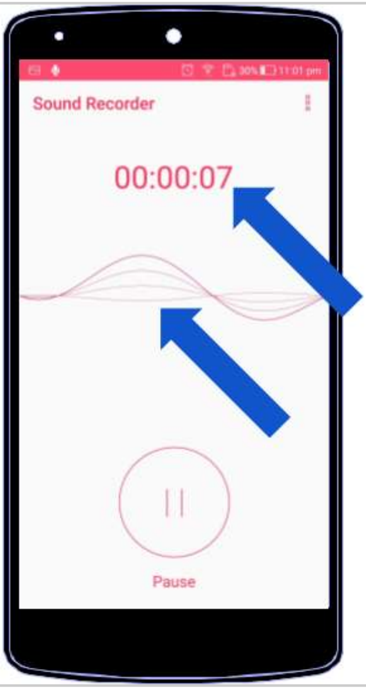


4 Touch on the icon of the "Voice recorder"



5 Touch on the microphone button to start recording

You will see that the timer starts to run and that the waves on the middle of the screen move when you talk.

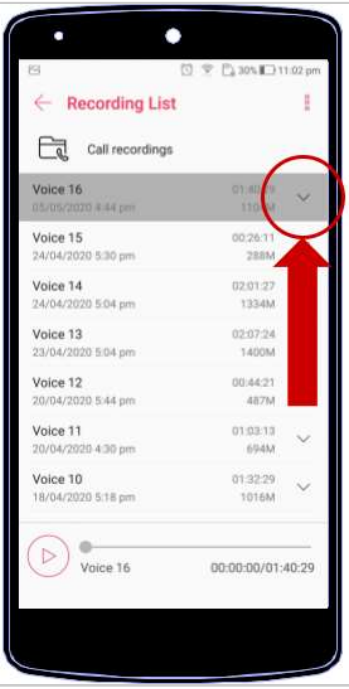


6 When you are finished recording, touch the button with the SQUARE symbol.

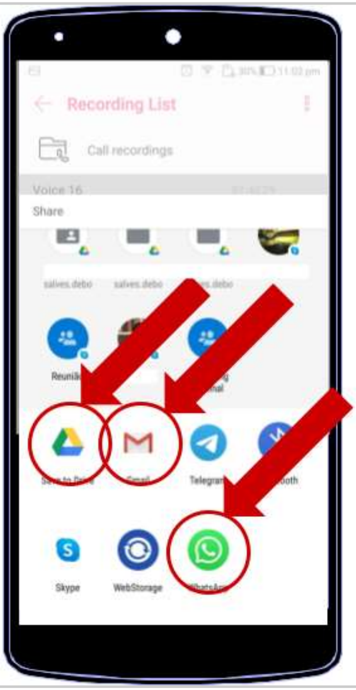
(Depending on your phone's make and model, you might need to push the Pause button before the Stop/square button appears on the screen.)

7 On your recordings' list, the recording you have just finished should be the first on top. Touch the arrow pointing down (on the right side of the your recording) and then select the "Share" option

(If your list does not appear automatically after you stop recording, touch the icon on the top right corner to see it.)



8 You can share your recording via WhatsApp, email ou Google Drive, whatever is easiest for you :)



APPENDIX I – MS PowerPoint presentation file for the recording of stimuli by Brazilian speaker participants

Gravação de estímulos

Pesquisa de mestrado

Déborah Salves

@deborahsalves | salves.deborah@gmail.com | http://lattes.cnpq.br/9996910978436484

PRÓXIMA →

Boas vindas

Olá, participante! Agradeço por seu tempo em participar da pesquisa.

Por favor, abra esta apresentação em um computador ou notebook e deixe seu celular parado durante a gravação (em vez de segurá-lo na mão, por exemplo).

Para que a qualidade do áudio seja a melhor possível, por favor, de preferência utilize fones de ouvido (como os que vêm com o celular), feche portas e janelas e esteja sozinha(o) no cômodo, sem animais de estimação. Por favor, deixe seu celular em modo avião durante a gravação, que leva cerca de 20 minutos.

PRÓXIMA →

Por favor, **inicie o gravador do seu celular** antes de prosseguir para a próxima tela.

JÁ INICIEI A GRAVAÇÃO →

Instruções

Ao clicar no botão a seguir, você iniciará sua participação. Você verá uma tela branca com uma frase em inglês.

Diga a frase em voz alta.

O slide passa sozinho, o que demora 5 segundos, durante os quais nada acontece. Você não precisa clicar em nada, nem no mouse, nem no teclado.

As frases são todas iguais, alterando-se apenas uma das palavras, a penúltima. As mesmas frases aparecem mais de uma vez.

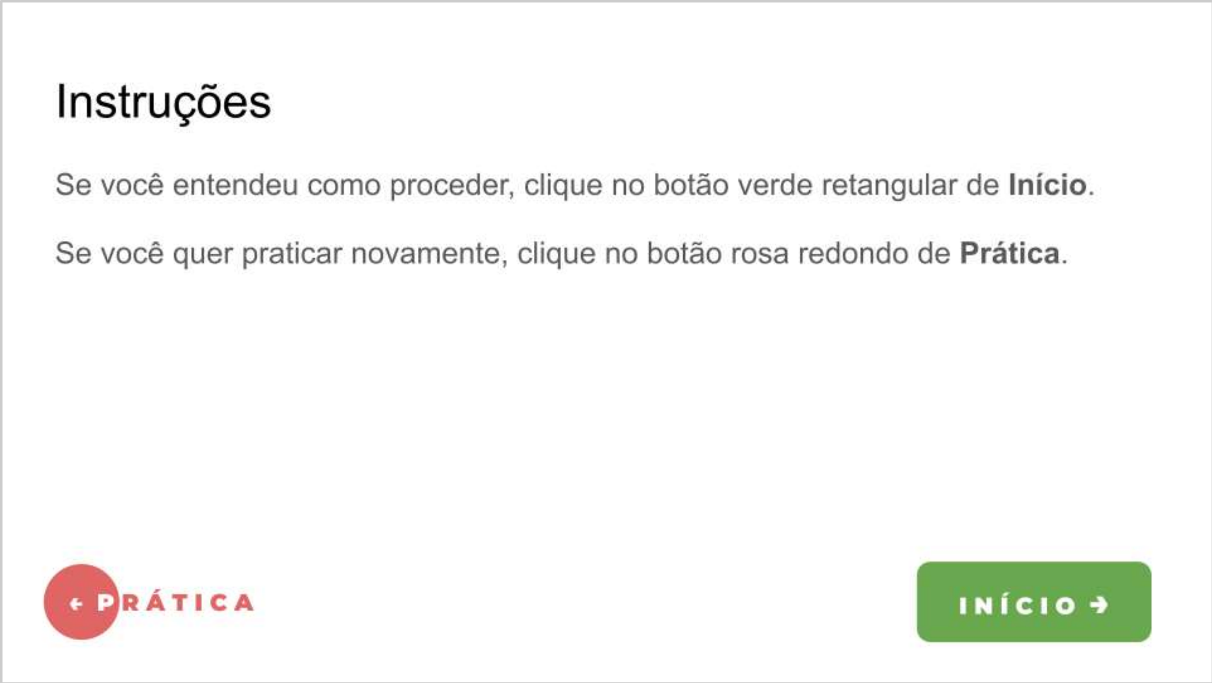
PRÁTICA →

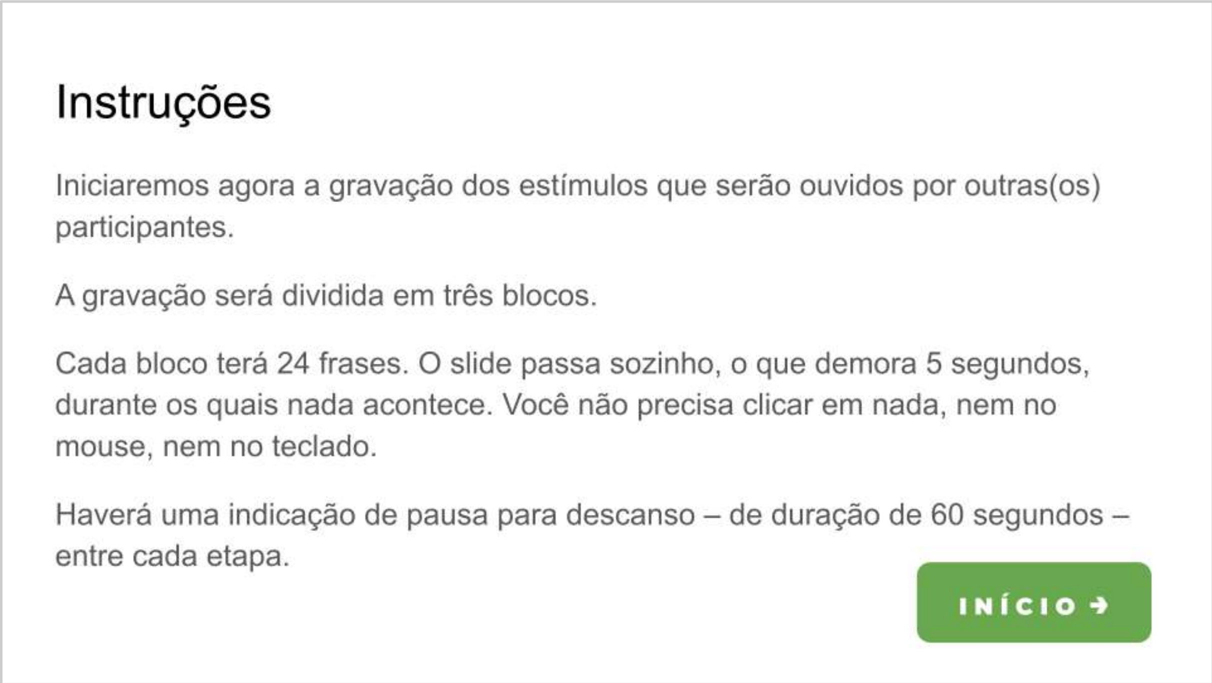
The word is **home** too.

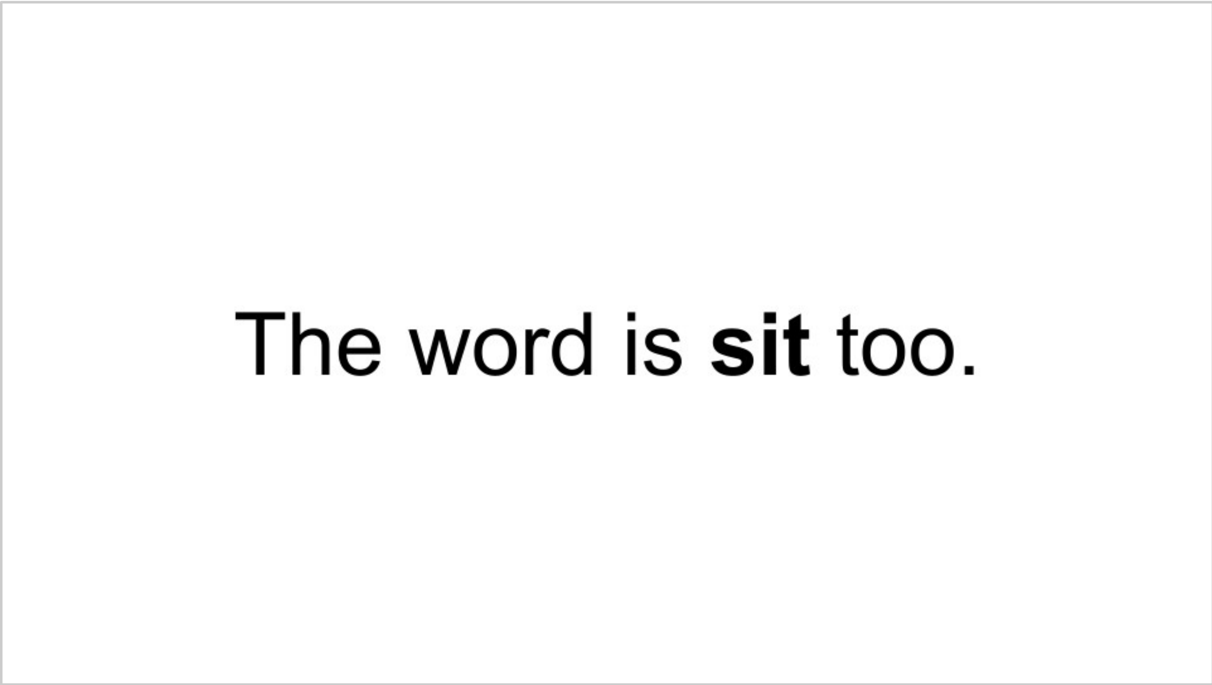
The word is **moon** too.

The word is **home** too.

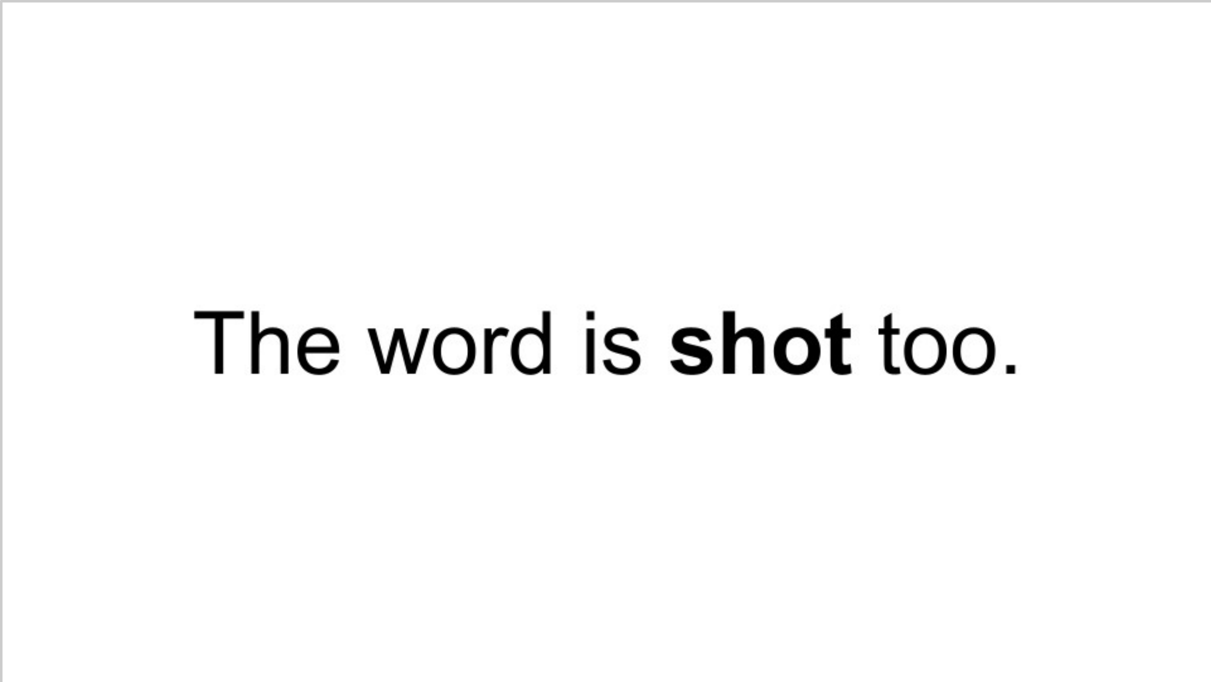
The word is **moon** too.


Instruções
Se você entendeu como proceder, clique no botão verde retangular de **Início**.
Se você quer praticar novamente, clique no botão rosa redondo de **Prática**.
← PRÁTICA
INÍCIO →


Instruções
Iniciaremos agora a gravação dos estímulos que serão ouvidos por outras(os) participantes.
A gravação será dividida em três blocos.
Cada bloco terá 24 frases. O slide passa sozinho, o que demora 5 segundos, durante os quais nada acontece. Você não precisa clicar em nada, nem no mouse, nem no teclado.
Haverá uma indicação de pausa para descanso – de duração de 60 segundos – entre cada etapa.
INÍCIO →



(the next 23 carrier sentences of the first repetition followed)



(the next 23 carrier sentences of the second repetition followed)

The word is **cut** too.

(the next 23 carrier sentences of the third repetition followed)

MUITO OBRIGADA!

Você já pode encerrar a gravação.

Por favor, envie sua gravação da forma que for melhor para você:

Email: @gmail.com

WhatsApp: (

Google Drive/Nuvem (clique aqui)

APPENDIX J – Vocabulary Questionnaire

QUESTIONÁRIO SOBRE VOCABULÁRIO

Olá, agradecemos sua participação neste estudo. Este questionário tem a função de permitir entender o nível de familiaridade que você tem, no momento desta coleta de dados, sobre o vocabulário usado.

Participante nº: _______

Circule o número que corresponde ao nível de familiaridade que você tem com cada palavra. Utilize a seguinte classificação:

0 = Eu nunca vi/ouvi essa palavra antes.

1 = Eu já vi/ouvi essa palavra antes, mas não sei o que significa.

2 = Eu já vi/ouvi essa palavra antes e sei o que ela significa em um determinado contexto, mas não sei dar uma definição para ela.

3 = Eu já vi/ouvi essa palavra antes, sei o que ela significa e consigo dar uma definição.

boot	0	1	2	3
bought	0	1	2	3
but	0	1	2	3
caught	0	1	2	3
cut	0	1	2	3
feet	0	1	2	3
fit	0	1	2	3
food	0	1	2	3
foot	0	1	2	3
heat	0	1	2	3
hit	0	1	2	3
pat	0	1	2	3
pet	0	1	2	3
sat	0	1	2	3

seat	0	1	2	3
set	0	1	2	3
shoot	0	1	2	3
shot	0	1	2	3
should	0	1	2	3
shut	0	1	2	3
sit	0	1	2	3

APPENDIX K – Brazilian speaker participants' responses in the Language History Questionnaire

Table 7.1 – Brazilian speaker participants' responses to QuExPli's Linguistic section

	Participante	BR01				BR02				BR04				BR06				BR07				BR08			
Línguas que você aprendeu		L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4
		PB	EN	ES	—	PB	EN	ES	—	PB	EN	FR	—	PB	EN	ES	—	PB	EN	—	—	PB	EN	ES	IT
Onde você aprendeu essas línguas	Casa	sim			—	sim			—	sim			—	sim			—	sim		—	—	sim			
	Escola regular	sim	sim	sim	—	sim			—		sim		—	sim	sim	sim	—	sim	sim	—	—			sim	
	Escola de idiomas				—		sim	sim	—		sim	sim	—		sim		—		sim	—	—		sim		sim
	Sozinha(o)	sim	sim	sim	—		sim		—		sim		—		sim		—	sim	sim	—	—	sim			
	Outro	sim	sim	sim	—				—			sim	—				—		sim	—	—	sim	sim	sim	sim
Idade (em anos) em que você	começou a aprender		11	11	—		13	23	—		12	18	—		9	9	—		17	—	—		12	8	28
	começou a utilizar ativamente				—		15		—		16	21	—		14		—			—	—				
	tornou-se fluente				—		21		—				—				—			—	—				
O quanto os seguintes fatores contribuíram para sua aprendizagem	Interação com a família	4	0	0	—	6	0	0	—	6	0	0	—	6	0	0	—	6	0	—	—	6	0	0	0
	Interação com os amigos	3	3	3	—	6	4	3	—	6	4	4	—	6	3	0	—	4	2	—	—	6	5	2	0
	Leitura de revistas e jornais	6	4	3	—	6	6	6	—	3	4	4	—	6	4	3	—	5	5	—	—	6	6	5	4
	Leitura de livros	6	0	3	—	5	6	6	—	4	4	5	—	6	5	1	—	5	5	—	—	6	6	5	5
	Leitura de textos acadêmicos	6	4	3	—	4	4	6	—	1	0	3	—	4	1	0	—	5	5	—	—	6	4	6	6
	Televisão (filmes, séries, YouTube, etc)	3	6	4	—	1	6	6	—	2	6	6	—	3	5	2	—	2	6	—	—	1	6	3	0
	Música / rádio / podcast	4	6	3	—	1	6	5	—	1	6	5	—	1	6	2	—	5	6	—	—	1	6	6	4
	Videogame	1	3	3	—	1	6	6	—	1	4	4	—	1	6	0	—	1	6	—	—	0	0	0	0
	Redes sociais	1	5	3	—	0	4	4	—	5	5	6	—	1	1	0	—	1	4	—	—	0	4	0	0
	Curso de idiomas	0	0	0	—	0	6	6	—	0	6	6	—	1	6	0	—	0	6	—	—	0	6	0	6
Treinamento de pronúncia ou participação em disciplina de Fonética / Fonologia		não	não	não	—	não	sim	não	—	não	não	sim	—	não	não	não	—	não	não	—	—	não	sim	não	sim
Uso diário (em horas) da língua com	Interação com a família		0	0	—		0	0	—		0	0	—		0	0	—		0	—	—		0	0	0
	Interação com os amigos		0.5	0.5	—		1	0	—		0	0	—		0	0	—		0	—	—		0.5	0	0
	Livros, revistas e jornais		1	0.5	—		2	1	—		0.5	1	—		1	0	—		0	—	—		0.5	0	0
	Leitura de textos acadêmicos		0.5	0.5	—		0	0	—		0	0	—		0	0	—		0	—	—		1	0.5	0.5
	Televisão (filmes, séries, YouTube, etc)		2	1	—		2	1	—		2	0.5	—		2	0	—		1	—	—		1	0	0
	Música / rádio / podcast		2	0.5	—		1	0	—		2	0.5	—		8	0	—		4	—	—		0.5	0	0
	Videogame		0.5	0	—		0	0	—		0	0	—		6	0	—		2	—	—		0	0	0
	Redes sociais		2	1	—		1	1	—		0	0.5	—		0	0	—		0	—	—		0	0	0
	Curso de idiomas		0	0	—		8	0	—		0	0	—		0	0	—		0	—	—		0	0	0
Experiência de imersão (em meses)	País	2	1		—		0	0	—		0	0	—	0	0		—		0	—	—		0	0	0
	Família	0	1		—		0	0	—		0	0	—	0	0		—		0	—	—		0	0	0
	Escola	0	0		—		0	0	—		18	0	—	84	0		—		25	—	—		0	0	0
	Trabalho	0	0		—		48	0	—		0	0	—	28	28		—		3	—	—		0	0	0

APPENDIX L – Brazilian speaker participants' responses in the Language History Questionnaire

Table 7.2 – Brazilian speaker participants' responses to QuExPli's Demographic section

Participante	**BR01**	**BR02**	**BR04**	**BR06**	**BR07**	**BR08**
Idade (em anos)	23	24	22	29	28	30
Gênero	Masculino	Feminino	Feminino	Masculino	Masculino	Feminino
Cidade de nascimento	Porto Alegre	Porto Alegre	Porto Alegre	Porto Alegre	Porto Alegre	Porto Alegre
Cidade(s) em que você morou até os 7 anos de idade	Viamão	Canoas	Porto Alegre	Porto Alegre, Esteio	Porto Alegre	Porto Alegre
Cidade de nascimento de sua mãe (ou da figura materna em sua vida)	Porto Alegre	Porto Alegre	Porto Alegre	Santa Maria	Porto Alegre	Porto Alegre
Caso haja alguma outra informação que você ache importante sobre o aprendizado ou o uso das suas línguas, por favor, escreva abaixo:	Acredito que o auxílio na formação escolar desde os 11 anos de idade foi importante para conhecer a língua. Entretanto, a plataforma em que mais aprendi foi assistindo os filmes legendados. Igualmente aprendi ouvindo músicas. Como toco violão e procuro as cifras, por vezes uma música em inglês ou espanhol me agrada aos ouvidos e, no exercício de tocá-la no violão e cantar, me desperta a curiosidade de saber seu significado. Aí começo a entender a letra e, consequentemente, memorizar um certo vocabulário. Também acredito que em função da música, minha habilidade na pronuncia se desenvolveu, ao "imitar" o som dos cantores estrangeiros que ouvia. Inclusive existem palavras que não sei o significado, mas sei pronunciar e escrever.		O número de meses passados em escola que fala a língua 2 eu considerei como sendo a escola de idiomas em que estudei. Só falávamos inglês e eu passava cerca de 4h semanalmente lá.	2 Anos e 4 meses trabalhando em uma multinacional fizeram eu dar um grande salto qualitativo na aprendizagem das línguas 2 e 3 - principalmente da 2.		

APPENDIX M – Canadian speaker participants' responses in the Language History Questionnaire

Table 7.3 – Canadian speaker participants' responses to QuExPli's Linguistic section

	Participant	CA09				CA10			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		EN	FR	IT	ES	EN	PT	FR	ES
Indicate where you have learned those languages (you can check more than one option for each language)	Home	yes		yes		yes	yes		
	School	yes	yes	yes	yes	yes	yes	yes	yes
	Language school					yes	yes		
	By yourself (self taught)			yes	yes	yes	yes		yes
	Other					yes	yes	yes	yes
Inform at which age (in years) you	started to learn		6	13	16		0	8	19
	started to actively use						2	8	19
	became fluent						6	25	24
How much has each of these factors contributed to you learning	Interaction with family		2	4	2		6	1	2
	Interaction with friends		2	4	2		1	1	2
	Reading magazines and newspapers		3	3	3		2	2	1
	Reading books		3	2	4		1	1	1
	Reading academic literature		2	3	4		1	1	4
	Television (movies, TV shows, YouTube, etc)		2	4	3		3	4	4
	Music / radio / podcast		2	4	1		5	4	6
	Videogame		2	2	1		0	0	0
	Social networks		2	2	1		5	1	1
	Language school		5	4	4		6	5	5
Pronunciation training or course on Phonetics / Phonology at college / university		No	No	No	No	Yes	No	Yes	Yes
Daily usage (in hours)	Interaction with family		0	0	0		2	0	0
	Interaction with friends		0	0	0		1	0	0
	Reading books, magazines and newspapers		0	0	0		1	0	0
	Reading academic literature		0	0	0		0	0	0
	Television (movies, TV shows, YouTube, etc)		0	1	0		1	1	1
	Music / radio / podcast		0	0	0		2	1	1
	Videogame		0	0	0		0	0	0
	Social networks		0	0	0		1	0	0
	Language school		0	0	0		0	0	0
Immersion experience (in months)	Country		3	3	1		12	0	1
	Family		1	1	0		12	0	1
	School		0	0	0		24	100	3
	Workplace		0	1	0		1	0	0
Proficiency level	Reading		4	3	3		5	4	5
	Writing		2	2	2		5	4	5
	Listening		3	3	3		6	4	5
	Speaking		2	3	3		6	3	5

APPENDIX N – Canadian speaker participants' responses in the Language History Questionnaire

Table 7.4 – Canadian speaker participants' responses to QuExPli's Demographic section

Participant	CA09	CA10
Age (in years)	37	38
Gender	Male	Female
City of birth	Canada	Cambridge, ON, Canada
Cities lived in until you were 7 years old	Canada	Cambridge, ON, Canada; London, ON, Canada; Yeosu, Jellonam-do, South Korea; Brantford, ON, Canada; Paris, ON, Canada
City of birth of your mother (or the motherly figure in your life)	Hamilton	Aveiro, Portugal
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:		Being born in Canada to Portuguese immigrants, naturally my first language was primarily Portuguese which was used in the household consistently. Later, I learned English through television programs, small neighbourhood friendships and school (Age 5). I became more fluent in speaking, reading and writing in Portuguese starting at the age of 8. Also, at that age, we begin to learn French in school (official 2nd language of Canada) which I took up until the 10th grade. After many travels, to Cuba and other Latin-speaking places, I took Spanish due to interest at the University level. I also obtained my Bachelor of Arts in English Literature via the University of Waterloo, Canada.

APPENDIX O – Argentinian and German listener participants' responses in the Language History Questionnaire (QuExPli)

Table 7.5 – Argentinian and German listener participants' responses to QuExPli's Linguistics sections

	Participant	AR01				AR02				AR03				AR04			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		ES	EN	DE	—	ES	EN	RU	—	ES	EN	PT	—	ES	EN	DE	—
Where have you learned those languages	Home				—				—				—				—
	School				—		L2		—				—		yes		—
	Language school				—		yes	yes	—		yes		—		yes		—
	By yourself (self taught)		yes		—		yes		—			yes	—		yes	yes	—
	Other				—				—				—				—
Age (in years) at which you	started to learn		5	22	—		7	24	—		3	30	—		8	35	—
	started to actively use		6		—		7	24	—		3	35	—		27		—
	became fluent		12		—		11		—		12		—		30		—
How much has each of these factors contributed to you learning	Interaction with family		0	0	—		1	0	—		0	0	—		0	0	—
	Interaction with friends		0	0	—		5	0	—		3	6	—		0	1	—
	Reading magazines and newspapers		5	2	—		6	1	—		5	6	—		1	2	—
	Reading books		6	0	—		6	0	—		6	0	—		3	4	—
	Reading academic literature		5	1	—		6	0	—		6	0	—		3	0	—
	Television (movies, TV shows, YouTube, etc)		4	5	—		6	4	—		4	6	—		6	6	—
	Music / radio / podcast		6	1	—		6	4	—		5	6	—		4	2	—
	Videogame		0	0	—		4	0	—		0	0	—		0	0	—
	Social networks		0	1	—		5	4	—		2	6	—		4	3	—
	Language school		3	3	—		6	6	—		6	0	—		1	0	—
Pronunciation training or course on Phonetics / Phonology at college / university		Yes	Yes	No	—	No	Yes	No	—	Yes	Yes	No	—	No	Yes	No	—
Daily usage (in hours)	Interaction with family		0	0	—		2	0	—		0	0	—		0	0	—
	Interaction with friends		10	0	—		1	0	—		5	0	—		0	0	—
	Reading books, magazines and newspapers		12	0	—		6	1	—		5	2	—		0	0.5	—
	Reading academic literature		4	1	—		3	0	—		8	0	—		1	0	—
	Television (movies, TV shows, YouTube, etc)		4	2	—		8	1	—		3	1	—		1	0.5	—
	Music / radio / podcast		4	0	—		3	1	—		3	0	—		0	0	—
	Videogame		0	0	—		0	0	—		0	0	—		0	0	—
	Social networks		5	0	—		3	0	—		3	2	—		0.5	0	—
	Language school		12	0	—		0	2	—		8	0	—		0	0	—
Immersion experience (in months)	Country		9	0	—		0	0	—				—		1	0	—
	Family		1	0	—		0	0	—				—		0	0	—
	School		0	0	—			12	—				—		22	0	—
	Workplace		3	0	—		24	0	—				—		34	0	—
Proficiency level	Reading		6	2	—		6	1	—		6	4	—		6	3	—
	Writing		6	0	—		6	2	—		6	1	—		5	1	—
	Listening		6	2	—		6	1	—		6	3	—		6	3	—
	Speaking		6	0	—		5	0	—		6	3	—		5	2	—

	Participant	AR05				AR06				AR07				AR08			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		ES	EN	FR	DE	ES	EN	—	—	ES	EN	FR	PT	ES	EN	PT	—
Where have you learned those languages	Home							—	—								—
	School			yes				—	—						yes		—
	Language school		yes	yes	yes		yes	—	—		yes				yes	yes	—
	By yourself (self taught)							—	—								—
	Other							—	—			yes	yes				—
Age (in years) at which you	started to learn		10	13	26		8	—	—		11	22	32		7	22	—
	started to actively use		15	21	29		21	—	—		22				17	23	—
	became fluent		19	24			21	—	—		24				18	24	—
How much has each of these factors contributed to you learning	Interaction with family		0	0	0		0	—	—		0	0	3		0	0	—
	Interaction with friends		0	0	5		6	—	—		0	0	0		3	0	—
	Reading magazines and newspapers		3	2	0		5	—	—		0	0	0		3	3	—
	Reading books		4	5	0		5	—	—		4	0	0		4	3	—
	Reading academic literature		4	3	0		6	—	—		4	0	0		4	3	—
	Television (movies, TV shows, YouTube, etc)		5	5	4		5	—	—		6	0	2		6	3	—
	Music / radio / podcast		6	6	0		4	—	—		6	0	1		6	3	—
	Videogame		0	0	0		0	—	—		0	0	0		0	0	—
	Social networks		0	0	0		4	—	—		2	0	0		3	3	—
	Language school		6	6	6		5	—	—		4	6	6		1	6	—
Pronunciation training or course on Phonetics / Phonology at college / university		No	Yes	Yes	No	No	Yes	—	—	—	Yes	No	No	No	Yes	No	—
Daily usage (in hours)	Interaction with family		0	0	0		0	—	—		0	0	0		0		—
	Interaction with friends		0	0	0		1	—	—		0	0	0		6		—
	Reading books, magazines and newspapers		3	0	0		1	—	—		1	0	0		5		—
	Reading academic literature		5	0	0		1	—	—		1	0	0		6		—
	Television (movies, TV shows, YouTube, etc)		1	1	1		2	—	—		4	0	0		6		—
	Music / radio / podcast		4	0	0		1	—	—		2	0	0		3		—
	Videogame		0	0	0		0	—	—		0	0	0		0		—
	Social networks		0	0	0		0	—	—		1	0	0		6		—
	Language school		3	0	0		0	—	—		4	0	0		4		—
Immersion experience (in months)	Country		1	1	1		45	—	—		8	0	1		1	0	—
	Family		0	0	0		30	—	—		8	0	0		0	0	—
	School		1	0	0		0	—	—		0	0	0		10	0	—
	Workplace		0	0	0		45	—	—		8	0	0		100	0	—
Proficiency level	Reading		6	6	0		5	—	—		6	1	1		6	4	—
	Writing		5	2	0		5	—	—		4	0	1		6	3	—
	Listening		6	4	0		5	—	—		6	1	2		6	4	—
	Speaking		6	5	0		5	—	—		6	0	2		6	4	—

	Participant	AR10				AR11				AR12				AR13			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		ES	EN	PT	—	ES	EN	—	—	ES	EN	IT	—	ES	EN	DE	PT
Where have you learned those languages	Home				—			—	—				—				
	School		yes		—		yes	—	—		yes		—		yes		yes
	Language school		yes		—		yes	—	—		yes	yes	—		yes		
	By yourself (self taught)				—		yes	—	—			yes	—				
	Other		yes	yes	—			—	—				—			yes	
Age (in years) at which you	started to learn		4	22	—		13	—	—		7	20	—		8	23	52
	started to actively use		14		—		14	—	—		10	22	—				
	became fluent		18		—		16	—	—		15		—				
How much has each of these factors contributed to you learning	Interaction with family		0	0	—		0	—	—		0	0	—		1	0	0
	Interaction with friends		1	4	—		0	—	—		0	0	—		2	2	5
	Reading magazines and newspapers		4	3	—		2	—	—		5	6	—		3	0	4
	Reading books		6		—		5	—	—		6	6	—		4	0	4
	Reading academic literature		6		—		6	—	—		6	3	—		4	4	3
	Television (movies, TV shows, YouTube, etc)		6	4	—		4	—	—		6	6	—		4	0	2
	Music / radio / podcast		6	4	—		4	—	—		6	6	—		5	0	3
	Videogame		3		—		0	—	—		1	0	—		0	0	0
	Social networks		5	4	—		6	—	—		6	6	—		1	0	2
	Language school		6		—		6	—	—		6	6	—		4	4	3
Pronunciation training or course on Phonetics / Phonology at college / university		No	Yes	No	—	Yes	Yes	—	—	No	Yes	No	—	No	Yes	No	No
Daily usage (in hours)	Interaction with family		0		—		0	—	—		0	0	—		0	0	1
	Interaction with friends		1		—		0	—	—		0	0	—		1	0	1
	Reading books, magazines and newspapers		1		—		1	—	—		2	2	—		1	0	1
	Reading academic literature		2		—		3	—	—		1	0	—		2	0	
	Television (movies, TV shows, YouTube, etc)		5		—		1	—	—		5	3	—		1	0	1
	Music / radio / podcast		5		—		1	—	—		5	3	—		2	0	1
	Videogame		0		—		0	—	—		0	0	—		0	0	0
	Social networks		2		—		1	—	—		3	1	—		0	0	0
	Language school		4		—		3	—	—		8	2	—		1	0	0
Immersion experience (in months)	Country				—		0	—	—		0	0	—		0	0	5
	Family				—		0	—	—		0	0	—		1	0	50
	School				—		+12	—	—		0	0	—		50	0	0
	Workplace				—		0	—	—		5	0	—		0	0	0
Proficiency level	Reading		6	3	—		6	—	—		4	3	—		4	1	4
	Writing		6	2	—		4	—	—		4	3	—		3	0	1
	Listening		6	3	—		4	—	—		4	3	—		3	0	4
	Speaking		6	2	—		4	—	—		4	3	—		3	0	1

	Participant	AR14				AR15				AR16				AR17			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		ES	EN	—	—	ES	EN	IT	—	ES	EN	—	—	ES	EN	FR	—
Where have you learned those languages	Home			—	—			yes	—			—	—				—
	School			—	—		yes	yes	—		yes	—	—		yes	yes	—
	Language school		yes	—	—		yes		—		yes	—	—				—
	By yourself (self taught)			—	—				—			—	—				—
	Other			—	—				—			—	—				—
Age (in years) at which you	started to learn		21	—	—		4	3	—		6	—	—		11	33	—
	started to actively use		26	—	—		15		—		20	—	—		16		—
	became fluent			—	—		22		—		20	—	—		21		—
How much has each of these factors contributed to you learning	Interaction with family		0	—	—		0	3	—		0	—	—		0	0	—
	Interaction with friends		3	—	—		3	0	—		2	—	—		2	0	—
	Reading magazines and newspapers		2	—	—		5	0	—		3	—	—		3	3	—
	Reading books		2	—	—		5	0	—		6	—	—		3	3	—
	Reading academic literature		6	—	—		6	0	—		6	—	—		3	0	—
	Television (movies, TV shows, YouTube, etc)		6	—	—		6	2	—		6	—	—		3	3	—
	Music / radio / podcast		6	—	—		6	2	—		6	—	—		3	3	—
	Videogame		1	—	—		0	0	—		6	—	—		0	0	—
	Social networks		6	—	—		4	2	—		6	—	—		3	3	—
	Language school		6	—	—		6	0	—		6	—	—		3	3	—
Pronunciation training or course on Phonetics / Phonology at college / university		Yes	Yes	—	—	No	Yes	No	—	No	Yes	—	—	Yes	Yes	No	—
Daily usage (in hours)	Interaction with family		0	—	—		1		—		0	—	—		0	0	—
	Interaction with friends		0	—	—		1		—		0	—	—		1	0	—
	Reading books, magazines and newspapers		2	—	—		2		—		1	—	—		2	0	—
	Reading academic literature		2	—	—		3		—		2	—	—		2	0	—
	Television (movies, TV shows, YouTube, etc)		2	—	—		2		—		3	—	—		4	0	—
	Music / radio / podcast		3	—	—		2		—		3	—	—		2	0	—
	Videogame		0	—	—		0		—		1	—	—		0	0	—
	Social networks		1	—	—		1		—		4	—	—		1	0	—
	Language school		1	—	—		5		—		4	—	—		4	0	—
Immersion experience (in months)	Country		6	—	—		4	0	—		0	—	—		0	0	—
	Family		0	—	—		0	0	—		0	—	—		0	0	—
	School		0	—	—		100	0	—		0	—	—		90	0	—
	Workplace		6	—	—		160	0	—		0	—	—		10	0	—
Proficiency level	Reading		4	—	—		6	1	—		6	—	—		5	1	—
	Writing		4	—	—		6	0	—		6	—	—		4	1	—
	Listening		4	—	—		6	2	—		6	—	—		3	1	—
	Speaking		4	—	—		6	1	—		6	—	—		3	1	—

	Participant	AR18				AR19				AR20				AR21			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		ES	EN	FR	—	ES	EN	—	—	ES	EN	IT	—	ES	EN	—	—
Where have you learned those languages	Home				—		yes	—	—				—			—	—
	School		yes		—		yes	—	—		yes		—			—	—
	Language school		yes	yes	—		yes	—	—		yes	yes	—		yes	—	—
	By yourself (self taught)				—		yes	—	—				—			—	—
	Other				—			—	—				—			—	—
Age (in years) at which you	started to learn		5	19	—		6	—	—		7	14	—		5	—	—
	started to actively use		6	20	—		10	—	—		16		—		15	—	—
	became fluent		9	20	—		16	—	—		25		—		20	—	—
How much has each of these factors contributed to you learning	Interaction with family		0	0	—		1	—	—		0	0	—		0	—	—
	Interaction with friends		6	0	—		0	—	—		1	0	—		3	—	—
	Reading magazines and newspapers		6	6	—		4	—	—		3	0	—		1	—	—
	Reading books		6	6	—		6	—	—		5	1	—		5	—	—
	Reading academic literature		2	2	—		3	—	—		5	0	—		5	—	—
	Television (movies, TV shows, YouTube, etc)		6	6	—		6	—	—		5	1	—		6	—	—
	Music / radio / podcast		6	6	—		6	—	—		5	1	—		6	—	—
	Videogame		0	0	—		3	—	—		4	0	—		2	—	—
	Social networks		0	1	—		6	—	—		5	1	—		2	—	—
	Language school		6	0	—		3	—	—		5	0	—		5	—	—
Pronunciation training or course on Phonetics / Phonology at college / university		No	Yes	No	—	No	Yes	—	—	—	Yes	No	—	No	Yes	—	—
Daily usage (in hours)	Interaction with family		0	0	—		2	—	—		0		—		0	—	—
	Interaction with friends		10	0	—		0	—	—		1		—		1	—	—
	Reading books, magazines and newspapers		6	2	—		10	—	—		1		—		2	—	—
	Reading academic literature		3	0	—		15	—	—		2		—		4	—	—
	Television (movies, TV shows, YouTube, etc)		12	2	—		10	—	—		2		—		2	—	—
	Music / radio / podcast		6	1	—		8	—	—		2		—		2	—	—
	Videogame		0	0	—		2	—	—		0		—		0	—	—
	Social networks		1	0	—		10	—	—		2		—		0	—	—
	Language school		0		—		2	—	—		3		—		1	—	—
Immersion experience (in months)	Country				—		0	—	—				—		6	—	—
	Family				—		0	—	—				—		1	—	—
	School				—		0	—	—				—		100	—	—
	Workplace				—		0	—	—				—		6	—	—
Proficiency level	Reading		6	3	—		6	—	—		5		—		6	—	—
	Writing		5	2	—		5	—	—		5		—		5	—	—
	Listening		5	3	—		6	—	—		5		—		6	—	—
	Speaking		6	2	—		4	—	—		5		—		6	—	—

	Participant	AR23				AR25				AR26				AR27			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		ES	EN	—	—	ES	EN	—	—	ES	EN	IT	—	ES	EN	—	—
Where have you learned those languages	Home			—	—			—	—				—		yes	—	—
	School		yes	—	—		yes	—	—		yes	yes	—		yes	—	—
	Language school			—	—		yes	—	—		yes	yes	—		yes	—	—
	By yourself (self taught)		yes	—	—		yes	—	—				—		yes	—	—
	Other			—	—			—	—				—		yes	—	—
Age (in years) at which you	started to learn		6	—	—		11	—	—		9	12	—		12	—	—
	started to actively use		22	—	—		17	—	—		20		—		17	—	—
	became fluent		27	—	—		29	—	—		16	15	—		18	—	—
How much has each of these factors contributed to you learning	Interaction with family		0	—	—		0	—	—		0	0	—		0	—	—
	Interaction with friends		0	—	—		2	—	—		1	0	—		1	—	—
	Reading magazines and newspapers		1	—	—		3	—	—		1	1	—		1	—	—
	Reading books		6	—	—		6	—	—		2	1	—		3	—	—
	Reading academic literature		6	—	—		5	—	—		2	0	—		6	—	—
	Television (movies, TV shows, YouTube, etc)		6	—	—		5	—	—		6	2	—		6	—	—
	Music / radio / podcast		6	—	—		5	—	—		6	2	—		4	—	—
	Videogame		0	—	—		0	—	—		6	0	—		3	—	—
	Social networks		1	—	—		0	—	—		6	0	—		6	—	—
	Language school		3	—	—		6	—	—		6	4	—		6	—	—
Pronunciation training or course on Phonetics / Phonology at college / university		No	Yes	—	—	—	No	—	—	No	Yes	No	—	—	Yes	—	—
Daily usage (in hours)	Interaction with family		0	—	—		0	—	—		0	0	—		1	—	—
	Interaction with friends		1	—	—		1	—	—		0	0	—		2	0	0
	Reading books, magazines and newspapers		2	—	—		1	—	—		1	0	—		3	0	0
	Reading academic literature		8	—	—		1	—	—		2	0	—		1	0	0
	Television (movies, TV shows, YouTube, etc)		1	—	—		1	—	—		4	1	—		1	0	0
	Music / radio / podcast		1	—	—		2	—	—		2	1	—		3	0	0
	Videogame		0	—	—		0	—	—		2	0	—		0	0	0
	Social networks		1	—	—		0	—	—		3	1	—		1	0	0
	Language school		2	—	—		2	—	—		4	0	—		0	0	0
Immersion experience (in months)	Country		0	—	—			—	—		0	0	—		1	0	0
	Family		0	—	—			—	—		0	0	—		0	0	0
	School		80	—	—			—	—			0	—		18	0	0
	Workplace		0	—	—			—	—		0	0	—		0	0	0
Proficiency level	Reading		6	—	—		6	—	—		5	5	—		6	0	
	Writing		6	—	—		5	—	—		5	4	—		5	0	
	Listening		5	—	—		3	—	—		5	5	—		6	0	
	Speaking		3	—	—		3	—	—		5	3	—		5	0	

	Participant	AR28				AR29				AR30				AR31			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		ES	EN	—	—	ES	EN	DE	—	ES	EN	—	—	ES	EN	PT	—
Where have you learned those languages	Home			—	—		yes		—			—	—				—
	School		yes	—	—		yes		—		yes	—	—			yes	—
	Language school		yes	—	—		yes		—			—	—		yes		—
	By yourself (self taught)			—	—		yes	yes	—			—	—				—
	Other			—	—				—		yes	—	—				—
Age (in years) at which you	started to learn		4	—	—		3	20	—		7	—	—		15	40	—
	started to actively use		16	—	—		6		—		19	—	—		40		—
	became fluent		20	—	—		9		—		19	—	—		40		—
How much has each of these factors contributed to you learning	Interaction with family		0	—	—		0	0	—		0	—	—		2	0	—
	Interaction with friends		1	—	—		5	0	—		0	—	—		0	0	—
	Reading magazines and newspapers		1	—	—		6	0	—		0	—	—		1	0	—
	Reading books		3	—	—		6	0	—		4	—	—		5	0	—
	Reading academic literature		4	—	—		6	0	—		4	—	—		6	5	—
	Television (movies, TV shows, YouTube, etc)		3	—	—		6	3	—		5	—	—		5	0	—
	Music / radio / podcast		3	—	—		6	3	—		4	—	—		6	0	—
	Videogame		0	—	—		0	0	—		0	—	—		0	0	—
	Social networks		2	—	—		0	0	—		4	—	—		2	0	—
	Language school		6	—	—		6	0	—		6	—	—		2	0	—
Pronunciation training or course on Phonetics / Phonology at college / university		Yes		—	—	Yes	Yes	No	—	No	Yes	—	—	No	Yes	Yes	—
Daily usage (in hours)	Interaction with family		0	—	—		0	0	—		0	—	—		1	0	—
	Interaction with friends		0	—	—		3	0	—		0	—	—		0	0	—
	Reading books, magazines and newspapers		2	—	—		8	0	—		0	—	—		2	0	—
	Reading academic literature		4	—	—		8	0	—		0	—	—		2	0	—
	Television (movies, TV shows, YouTube, etc)		2	—	—		2	1	—		3	—	—		0	0	—
	Music / radio / podcast		1	—	—		2	1	—		1	—	—		1	0	—
	Videogame		0	—	—		0	0	—		0	—	—		0	0	—
	Social networks		1	—	—		0	0	—		0	—	—		0		—
	Language school		5	—	—		3	0	—		2	—	—		0	0	—
Immersion experience (in months)	Country			—	—				—		0	—	—		1	0	—
	Family	1		—	—				—		0	—	—		0	0	—
	School	10		—	—				—		10	—	—		0	0	—
	Workplace	10		—	—				—		0	—	—		0	0	—
Proficiency level	Reading		4	—	—		6	3	—		5	—	—		5	2	—
	Writing		4	—	—		5	3	—		5	—	—		4	2	—
	Listening		5	—	—		6	2	—		5	—	—		4	2	—
	Speaking		5	—	—		6	3	—		4	—	—		5	2	—

	Participant	DE03				DE04				DE05				DE06			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		DE	FR	EN	—	DE	EN	FR	SE	DE	RU	EN	IT	DE	RU	EN	FR
Where have you learned those languages	Home				—		yes					yes					
	School		yes	yes	—		yes	yes			yes	yes			yes	yes	yes
	Language school			yes	—				yes				yes		yes		
	By yourself (self taught)				—				yes			yes					
	Other		yes		—								yes		yes	yes	yes
Age (in years) at which you	started to learn		8	12	—		2	12	19		12	14	26		6	11	12
	started to actively use		12	18	—		5		21		23	14	26		6	11	12
	became fluent		19	20	—		12		21			18				15	15
How much has each of these factors contributed to you learning	Interaction with family	0	0	0	—		6	0	0		1	4	0		2	2	1
	Interaction with friends	0	0	1	—		4	1	4		1	4	1		4	3	1
	Reading magazines and newspapers	0	0	2	—		4	1	4		0	3	0		1	3	1
	Reading books	0	1	2	—		6	2	5		0	4	0		1	3	2
	Reading academic literature	0	1	5	—		6	3	6		0	1	0		0	6	1
	Television (movies, TV shows, YouTube, etc)	0	0	4	—		6	1	4		0	3	1		1	5	1
	Music / radio / podcast	0	1	5	—		4	1	0		0	5	1		1	6	1
	Videogame	0	0	0	—		6	3	2		0	0	0		1	6	0
	Social networks	0	0	4	—		3	1	0		0	1	0		0	2	0
	Language school	0	2	3	—		0	3	6		0	0	3		6	0	0
Pronunciation training or course on Phonetics / Phonology at college / university		No	No	No	—	Yes	Yes	No	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Daily usage (in hours)	Interaction with family		0	0	—		1	0	0		0	0	0		0	0	0
	Interaction with friends		0.5	2	—		1	0	2		0	1	0		0	0.1	0
	Reading books, magazines and newspapers		0	0	—		2	0	2		0	1	0		0	0.5	0
	Reading academic literature		0	3	—		2	0.25	3		0	0	0		0	3	0.1
	Television (movies, TV shows, YouTube, etc)		0	0.5	—		3	0	1		0	1	0		0	0.5	0
	Music / radio / podcast		0	0.5	—		0	0	0		0	1	0		0	6	0
	Videogame		0	0	—		4	0	0		0	0	0		0	2	0
	Social networks		0	1	—		0.5	0	0		0	1	0		0	0.1	0
	Language school		0	0	—		0	0	1		0	0	0		0	0	0
Immersion experience (in months)	Country		14	15	—		0	0	4		1	4	5		0	0	1
	Family		3	0	—		0	0	0		8	0	0		0	0	0
	School		3	6	—		0	0	0		72	96	0		5	0	0
	Workplace		3	9	—		0	0	0		0	0	0		0	0	0
Proficiency level	Reading		4	5	—		6	4	6		3	4	2		2	6	5
	Writing		3	4	—		5	1	5		2	3	1		0	6	3
	Listening		5	5	—		5	2	4		3	4	3		0	6	3
	Speaking		4	6	—		5	1	4		2	3	1		0	6	3

	Participant	DE09				DE12				DE13				DE14			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		DE	EN	IT	—	DE	EN	ES	—	DE	EN	IT	PT	DE	EN	FR	—
Where have you learned those languages	Home			yes	—				—								—
	School		yes		—		yes	yes	—		yes				yes	yes	—
	Language school				—				—				yes		yes		—
	By yourself (self taught)		yes	yes	—				—						yes	yes	—
	Other				—				—			yes			yes	yes	—
Age (in years) at which you	started to learn		12	22	—		8	19	—		8	27	34		10	12	—
	started to actively use		23		—		18	20	—		17	27	35		15	19	—
	became fluent		23		—		20		—		24	30	36		19		—
How much has each of these factors contributed to you learning	Interaction with family		3	4	—		0	0	—		0	0	0		3	0	—
	Interaction with friends		4	0	—		0	0	—		6	6	6		5	2	—
	Reading magazines and newspapers		5	0	—		2	4	—		6	4	5		3	2	—
	Reading books		5	0	—		4	2	—		6	4	5		5	2	—
	Reading academic literature		4	0	—		4	3	—		6	1	5		2	0	—
	Television (movies, TV shows, YouTube, etc)		6	3	—		6	2	—		5	2	5		1	0	—
	Music / radio / podcast		6	3	—		6	2	—		4	1	4		4	1	—
	Videogame		5	0	—		2	2	—		1	0	0		0	0	—
	Social networks		6	3	—		5	3	—		4	0	2		0	0	—
	Language school		3	0	—		4	6	—		0	0	3		1	0	—
Pronunciation training or course on Phonetics / Phonology at college / university		Yes	No	No	—	Yes	Yes	Yes	—	No	Yes	No	No	No	No	No	—
Daily usage (in hours)	Interaction with family		0	0	—		0	0	—		0	0	16		2	0	—
	Interaction with friends		0.5	0	—		0	0	—		0	0	2		1	0	—
	Reading books, magazines and newspapers		1	0	—		2	0.5	—		1	0	1		1	0	—
	Reading academic literature		0.5	0	—		2	0	—		0	0	0		0	0	—
	Television (movies, TV shows, YouTube, etc)		10	0	—		1	0.5	—		0	0	1		1	0	—
	Music / radio / podcast		10	0.5	—		1	0.5	—		0	0	1		2	0	—
	Videogame		5	0	—		0		—		0	0	0		0	0	—
	Social networks		10	0.5	—		2	0.5	—		0	0	1		1	0	—
	Language school		0	0	—		0	0.5	—		0	0	0		0	0	—
Immersion experience (in months)	Country		0	3	—		4	5	—		0	2	42		91	2	—
	Family		0	30	—		4	0	—		0	2	37		1	0	—
	School		0	0	—		4	0	—		0	0	0		0	0	—
	Workplace		0	0	—		0	0	—		0	0	42		4	0	—
Proficiency level	Reading		6	2	—		6	4	—		5	4	5		5	2	—
	Writing		4	2	—		6	3	—		4	2	4		5	2	—
	Listening		6	3	—		6	2	—		5	3	5		5	2	—
	Speaking		4	1	—		6	2	—		5	3	5		5	2	—

	Participant	DE15				DE17				DE19				DE20			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		DE	EN	FR	PL	DE	EN	PT	—	DE	EN	FR	—	DE	EN	—	—
Where have you learned those languages	Home							yes	—				—			—	—
	School		yes	yes			yes		—		yes	yes	—		yes	—	—
	Language school				yes				—				—			—	—
	By yourself (self taught)				yes		yes		—		yes		—			—	—
	Other						yes		—				—		yes	—	—
Age (in years) at which you	started to learn		10	12	24		8	10	—		8	9	—		5	—	—
	started to actively use		16				13	10	—		14	14	—		13	—	—
	became fluent		16				16		—		14	14	—		17	—	—
How much has each of these factors contributed to you learning	Interaction with family		0	0	0		1	6	—		2	0	—		0	—	—
	Interaction with friends		0	0	5		4	0	—		0	2	—		2	—	—
	Reading magazines and newspapers		3	0	0		3	0	—		0	0	—		4	—	—
	Reading books		6	4	0		5	4	—		6	1	—		0	—	—
	Reading academic literature		4	0	0		5	0	—		0	0	—		0	—	—
	Television (movies, TV shows, YouTube, etc)		6	0	0		6	4	—		3	0	—		4	—	—
	Music / radio / podcast		6	0	0		6	5	—		1	0	—		4	—	—
	Videogame		1	0	0		0	0	—		0	0	—		4	—	—
	Social networks		0	0	0		5	3	—		5	0	—		0	—	—
	Language school		0	0	6		4	0	—		1	4	—		0	—	—
Pronunciation training or course on Phonetics / Phonology at college / university		Yes	No	No	No	No	Yes	No	—	No	No	No	—	No	No	—	—
Daily usage (in hours)	Interaction with family		0	0	0		0	1			2	0	—		0	—	—
	Interaction with friends		0	0	0		1	0			0	0	—		0	—	—
	Reading books, magazines and newspapers		0	0	0		2	1			6	0	—		1	—	—
	Reading academic literature		1	0	0		2	0			0	0	—		0	—	—
	Television (movies, TV shows, YouTube, etc)		1	0	0		5	1			0	0	—		0	—	—
	Music / radio / podcast		2	0	0		3	1			1	0	—		2	—	—
	Videogame		0	0	0		0	0			0	0	—		0	—	—
	Social networks		0	0	0		2	1			1	0	—		0	—	—
	Language school		0	0	0		3	0			0	0	—		0	—	—
Immersion experience (in months)	Country		0	0	0		0	9			0	5	—		1	—	—
	Family		0	0			0	216			0	5	—		0	—	—
	School		0	0	0		36	0			0	5	—		0	—	—
	Workplace		0	0	0		0	1			0	0	—		0	—	—
Proficiency level	Reading		6	2	2		6	5			6	3	—		5	—	—
	Writing		6	1	1		4	3			6	3	—		4	—	—
	Listening		4	1	1		6	6			5	4	—		5	—	—
	Speaking		4	1	2		5	4			5	2	—		3	—	—

	Participant	DE21				DE23				DE24				DE25			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		DE	EN	—	—	DE	EN	FR	—	DE	EN	FR	KO	DE	EN	FR	IT
Where have you learned those languages	Home			—	—				—								
	School			—	—		yes	yes	—		yes	yes			yes		
	Language school			—	—				—				yes		yes	yes	yes
	By yourself (self taught)		yes	—	—				—				yes			yes	yes
	Other			—	—				—						yes		yes
Age (in years) at which you	started to learn		18	—	—		9	12	—		5	12	17		10	18	32
	started to actively use		20	—	—				—		13		20		12	19	32
	became fluent		25	—	—				—		15		21		23		
How much has each of these factors contributed to you learning	Interaction with family		0	—	—		0	0	—		2	0	0		0	0	0
	Interaction with friends		5	—	—		1	0	—		4	2	3		6	2	2
	Reading magazines and newspapers		6	—	—		2	2	—		1	2	0		5		1
	Reading books		6	—	—		2	1	—		6	3	2		6	1	0
	Reading academic literature		3	—	—		1	1	—		4	0	0		5	1	0
	Television (movies, TV shows, YouTube, etc)		6	—	—		5	3	—		6	4	5		5	1	1
	Music / radio / podcast		2	—	—		6	6	—		3	3	6		5	1	2
	Videogame		0	—	—		1	0	—		0	0	0		0	0	0
	Social networks		0	—	—		6	6	—		4	2	4		4	0	0
	Language school		0	—	—		6	6	—		0	0	5		2	6	6
Pronunciation training or course on Phonetics / Phonology at college / university		No	No	—	—	No	No	No	—	—	Yes	No	No	—	No	No	No
Daily usage (in hours)	Interaction with family		0	—	—		0	0	—		0	0	0		1	0	0
	Interaction with friends	12	0	—	—		1	0	—		4	0	1		0.5	0	0
	Reading books, magazines and newspapers	12	0	—	—		0	0	—		2	1	0		1	0	0
	Reading academic literature	3	0	—	—		0	0	—		1	0	0		0.2	0	0
	Television (movies, TV shows, YouTube, etc)	5	0	—	—		1	1	—		2	1	2		0.2	0.01	0.01
	Music / radio / podcast	0	0	—	—		2	2	—		1	1	1		1	0.01	0.01
	Videogame	0	0	—	—		0	0	—		0	0	0		0	0	0
	Social networks	4	0	—	—		4	2	—		2	0	1		0.1	0	0
	Language school	0	0	—	—		2	2	—		1	0	1		0	0	0
Immersion experience (in months)	Country		0	—	—		1	0	—		2	1	0		60	4	3
	Family		0	—	—		1	0	—		0	0	0		24	0	1
	School		0	—	—		1	0	—		0	0	0		12	1	1
	Workplace		0	—	—		0	0	—		2	0	0		40	0	1
Proficiency level	Reading		6	—	—		5	3	—		6	4	3		6	3	3
	Writing		6	—	—		4	2	—		6	4	3		6	2	2
	Listening		6	—	—		5	2	—		6	4	5		6	3	3
	Speaking		6	—	—		4	2	—		5	2	2		6	2	2

	Participant	DE27				DE28			
Languages you have learned		*L1*	*L2*	*L3*	*L4*	*L1*	*L2*	*L3*	*L4*
		DE	EN	—	—	DE	EN	SE	ES
Where have you learned those languages	Home			—	—			yes	
	School		yes	—	—		yes		yes
	Language school			—	—				
	By yourself (self taught)			—	—				
	Other			—	—				
Age (in years) at which you	started to learn		6	—	—		3	6	11
	started to actively use		15	—	—		5	6	
	became fluent		17	—	—		12		
How much has each of these factors contributed to you learning	Interaction with family		1	—	—		2	6	0
	Interaction with friends		0	—	—		5	1	0
	Reading magazines and newspapers		1	—	—		0	2	0
	Reading books		3	—	—		6	0	2
	Reading academic literature		2	—	—		6	0	2
	Television (movies, TV shows, YouTube, etc)		6	—	—		6	1	0
	Music / radio / podcast		6	—	—		6	1	0
	Videogame		3	—	—		6	1	0
	Social networks		4	—	—		6	0	0
	Language school		0	—	—		6	0	2
Pronunciation training or course on Phonetics / Phonology at college / university		No	No	—	—	No	No	No	No
Daily usage (in hours)	Interaction with family		0	—	—		0	1	0
	Interaction with friends		0	—	—		3	0	0
	Reading books, magazines and newspapers		0	—	—		4	0	0
	Reading academic literature		0	—	—		4	0	0
	Television (movies, TV shows, YouTube, etc)		2	—	—		2	0	0
	Music / radio / podcast		1	—	—		4	1	0
	Videogame		0	—	—		5	2	0
	Social networks		2	—	—		2	0	0
	Language school		0	—	—		2	0	0
Immersion experience (in months)	Country			—	—		2	8	2
	Family			—	—		1	1	0
	School			—	—		0	0	0
	Workplace			—	—		0	0	0
Proficiency level	Reading		6	—	—		6	3	1
	Writing		6	—	—		6	1	1
	Listening		6	—	—		6	3	1
	Speaking		4	—	—		6	2	1

APPENDIX P – Argentinian and German listener participants' responses in the Language History Questionnaire (QuExPli)

Table 7.6 – Argentinian and German listener participants' responses to QuExPli's Demographic sections

Participant	AR01	AR02	AR03	AR04
Age (in years)	55	25	45	35
Gender	Female	Female	Female	Female
City of birth	Mar del Plata, Argentina	Buenos Aires City	Buenos Aires	Mar del plata
Cities lived in until you were 7 years old	Mar del Plata	Buenos Aires City	Buenos Aires	Mar del plata
City of birth of your mother (or the motherly figure in your life)	Santa Fe, Arg.	Buenos Aires City	Buenos Aires	Carhue
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				Apps should be included in the study. I spend at least 30 min a day using them. The one I use the most is Anki (an app with flashcard that is mostly used to learn things such as vocabulary by heart)

Participant	AR05	AR06	AR07	AR08
Age (in years)	43	41	34	37
Gender	Female	Female	Female	Female
City of birth	Buenos Aires	Mar del Plata	Mar del Plata	mar del plata
Cities lived in until you were 7 years old	Buenos Aires, Río Gallegos	Mar del Plata	Mar del Plats	mar del plata
City of birth of your mother (or the motherly figure in your life)	Buenos Aires	Necochea	Rufino, Santa Fe, Arg	mar del plata
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				

Participant	AR10	AR11	AR12	AR13
Age (in years)	25	26	22	63
Gender	Female	Female	Female	Female
City of birth	Mar del Plata	Balcarce	Coronel Vidal	Capital Federal (Bs. As.)
Cities lived in until you were 7 years old	Mar del Plata	Balcarce	Coronel Vidal	Capital Federal / Gral. Pirán / Mar del Plata
City of birth of your mother (or the motherly figure in your life)	Pehuajó	Esquina, Corrientes	Coronel Vidal	Asunción - Paraguay
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				I have learnt and remember a lot of vocabulary by songs

Participant	AR14	AR15	AR16	AR17
Age (in years)	30	42	26	39
Gender	Female	Female	Female	Female
City of birth	Mar del Plata Argentina	Mar del Plata	mar del plata	Mar del Plata
Cities lived in until you were 7 years old	Mar del Plata - Rosario	Mar del Plata	mar del plata	Mar del Plata
City of birth of your mother (or the motherly figure in your life)	Córdoba Argentina	Córdoba	buenos aires	Mar del Plata
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				

Participant	AR18	AR19	AR20	AR21
Age (in years)	28	26	35	59
Gender	Female	Female	Female	Female
City of birth	AMBA	Mar Del Plata	Olavarría	Comodoro Rivadavia
Cities lived in until you were 7 years old	AMBA	Mar Del Plata	Olavarría	Comodoro Rivadavia
City of birth of your mother (or the motherly figure in your life)	AMBA	Mar Del Plata	Azul	Comodoro Rivadavia
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				

Participant	AR23	AR25	AR26	AR27
Age (in years)	27	55	27	22
Gender	Female	Female	Female	Female
City of birth	Mar del Plata	Villa Elisa (Entre Ríos)	Bahía Blanca, Argentina	Mar del Plata
Cities lived in until you were 7 years old	Mar del PLata	Lobería (Buenos Aires)	Bahía Blanca, Mar del Plata (Argentina)	Mar del Plata
City of birth of your mother (or the motherly figure in your life)	Buenos Aires capital	Villa Elisa (Entre Ríos)	Coronel Suárez, Argentina	Termas de Río Hondo
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				At present time I use my second language a lot in university

Participant	AR28	AR29	AR30	AR31
Age (in years)	26	21	24	41
Gender	Female	Female	Female	Female
City of birth	Mar del Plata	Mar del Plata	Villa gesell	mar del plata
Cities lived in until you were 7 years old	Mar del plata	Mar del Plata	Villa gesell	mar del plata
City of birth of your mother (or the motherly figure in your life)	Mar del plata	Mar del Plata	Rojas	balcarce
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				

Participant	DE03	DE04	DE05	DE06
Age (in years)	26	21	54	30
Gender	Female	Male	Male	Male
City of birth	Berlin	Berlin	Kayna	Berlin (Germany)
Cities lived in until you were 7 years old	Berlin	Brieselang, Germany	Berlin, Potsdam	Berlin (Germany)
City of birth of your mother (or the motherly figure in your life)	Berlin	Hamburg	Goldap (former East Prussia)	Berlin (Germany)
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:				primary school was bilingual (German/Russian)

Participant	DE09	DE12	DE13	DE14
Age (in years)	25	24	38	53
Gender	Female	Female	Male	Female
City of birth	Cologne, Germany	Köthen	Berlin	Bonn Germany
Cities lived in until you were 7 years old	Cologne	Köthen	Berlin	Suburbs of Bonn
City of birth of your mother (or the motherly figure in your life)	Cologne	Köthen	Rostock	Berlin
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:	I use my 2nd language mostly for entertainment, my 3rd language only while learning it			I love languages

Participant	DE15	DE17	DE19	DE20
Age (in years)	33	18	22	22
Gender	Female	Female	Female	Non-binary
City of birth	Taipei	Berlin	Bonn (Germany)	Berlin
Cities lived in until you were 7 years old	Berlin	Berlin	Bonn (Germany)	Berlin, Falkensee
City of birth of your mother (or the motherly figure in your life)	Berlin	Berlin	Belo Horizonte (Brazil)	Staaken
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:	I have been 1 week in Ireland and sometimes for 2 days in a Polish family.			

Participant	DE21	DE23	DE24	DE25
Age (in years)	58	19	22	55
Gender	Female	Female	Female	Female
City of birth	FRANKFURT/M	Berlin	Berlin	Winnweiler
Cities lived in until you were 7 years old	Frankfurt/M	Berlin	Berlin	Winnweiler
City of birth of your mother (or the motherly figure in your life)	Germany	Berlin	Eisenach	Rockenhausen
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:			I spend my day thinking and speaking to myself in English and Korean more than in German. This is how I became fluent in these two languages.	

Participant	DE27	DE28
Age (in years)	19	18
Gender	Female	Female
City of birth	Berlin, Germany	Berlin
Cities lived in until you were 7 years old	Berlin, Germany; Darmsdorf, Germany	Berlin
City of birth of your mother (or the motherly figure in your life)	Potsdam, Germany	Berlin
In case there is any other information about your language learning or about your language use that you find relevant, please mention it below:		

APPENDIX Q – Brazilian participants' responses in the Vocabulary Questionnaire

Table 7.7 – Brazilian speaker participants' responses to the Vocabulary Questionnaire

Participant	BR01	BR02	BR04	BR06	BR08	BR07
bat	3	3	3	3	0	3
bet	0	3	3	3	1	3
book	3	3	3	3	3	3
boot	1	3	1	3	2	3
bought	1	3	2	3	3	3
but	3	3	3	3	3	3
caught	1	3	2	3	2	3
cut	3	3	3	3	1	3
feet	3	3	3	3	3	3
fit	3	3	3	3	2	3
food	3	3	3	3	3	3
foot	3	3	3	3	2	3
heat	3	3	3	3	3	3
hit	3	3	3	3	2	3
pat	1	0	2	3	0	3
pet	3	3	3	3	2	3
sat	1	3	1	3	0	3
seat	0	3	2	3	0	3
set	3	3	3	3	0	3
shoot	3	3	2	3	1	3
shot	3	3	3	3	2	3
should	3	3	3	3	3	3
shut	3	3	3	3	0	3
sit	3	3	3	3	2	3