

UNIVERSIDADE FEDERAL DE SANTA CATARINA
PÓS-GRADUAÇÃO EM LETRAS/INGLÊS E LITERATURA CORRESPONDENTE

PERCEPTION AND PRODUCTION OF ENGLISH VOWELS BY BRAZILIAN EFL
SPEAKERS

ANDRÉIA SCHURT RAUBER

Tese submetida à Universidade Federal de Santa Catarina em cumprimento
parcial dos requisitos para obtenção do grau de

DOUTOR EM LETRAS

FLORIANÓPOLIS

Novembro de 2006

Esta tese de Andréia Schurt Rauber, intitulada *Perception and Production of English Vowels by Brazilian EFL Speakers*, foi julgada adequada e aprovada em sua forma final pelo Programa de Pós-Graduação em Letras/Inglês e Literatura Correspondente, da Universidade Federal de Santa Catarina, para fins de obtenção do grau de

DOUTOR EM LETRAS

Área de Concentração: Inglês e Literatura Correspondente
Opção: Língua Inglesa e Linguística Aplicada

Prof. Dr. José Luiz Meurer
Coordenador

BANCA EXAMINADORA:

Profa. Dra. Barbara O. Baptista
Orientadora e Presidente (UFSC)

Prof. Dr. Paul P. G. Boersma
Examinador (University of Amsterdam)

Prof. Dr. Michael A. Watkins
Examinador (UFPR)

Profa. Dra. Izabel C. Seara
Examinadora (UFSC)

Profa. Dra. Rosana D. Koerich
Examinadora (UFSC)

Florianópolis, 23 de novembro de 2006.

To my husband, Moacir Rauber;
my parents, Nelly Batista and Wiland Schurt;
my “adopted brother”, Ricardo Bion;
and my unforgettable friend, Ton Wempe.

ACKNOWLEDGMENTS

I would like to thank CAPES for the financial support during most of my doctoral course and for the grant to study at the University of Amsterdam for five months.

I would never have started the doctoral course without the encouragement of Professor Dr. Barbara Baptista, my advisor, friend, and “academic mother”.

I would also like to thank Ricardo Bion, for always being by my side, encouraging me to have ambitious dreams and helping me to make them come true. Besides being a real friend, his incredible intelligence and willingness to help were very important for the completion of this dissertation.

I would also like to thank Professor Dr. Paul Boersma, my foreign co-advisor, who contributed greatly with the data analysis procedures. His impressive knowledge and kindness were essential for my growth as both a person and researcher.

I would also like to thank Dr. Paola Escudero’s advice on the data collection procedure and on the dissertation as a whole. Both the data collection procedures and the stimuli used in this study are from her post-doctoral project, which I had the privilege of taking part of during the first three months of my stay in Amsterdam.

My special thanks to Ton Wempe, my dear friend and technician of the University of Amsterdam. His wisdom and patience were essential for me to discover the beauty of signal analysis.

I would also like to thank Professor Dr. Rosana Koerich for enlightening my days with her great sense of humor, and for always being extremely friendly and helpful.

I thank the members of the committee for their comments, suggestions and corrections: Professor Dr. Barbara Baptista, Professor Dr. Paul Boersma, Professor Dr.

Mike Watkins, Professor Dr. Izabel Seara, Professor Dr. Rosana Koerich, and Professor Dr. Rosane Silveira.

I also thank the staff of the LINSE Lab at UFSC, especially Fernando Pacheco, Dr. Izabel Seara, Monique Nicodem, Sandra Kafka and Simone Klein. At the lab I was able to apply the knowledge I have acquired on signal analysis, and learn more about text-to-speech conversion systems, my current area of investigation.

I also acknowledge the help of Dr. Cynthia Clopper, who sent me the means and standard deviations of her production data on vowels produced by speakers from six American English varieties.

I would never have started my graduate studies without the encouragement of my lovely husband, who has been patiently waiting for my attention for several years.

I would also like to thank my supportive and cheerful friends, who were also very important during this period of full dedication to the academia: Joara Bergsleither, Janete Machado, Denize Nobre Oliveira, Denise Kluge, Mara Reis, Margareth Perucci, Verónica Rivas, Jeannette Elsenburg, Luiz Armando Silveiro, Rosane Silveira, Jacir Baratieri, Melissa Bettoni, Fernanda Delatorre, Ane Girondi, João Carlos da Silva, and Priscilla Poleza. Unfortunately I cannot include everyone's names here, but I thank all my friends who help my life become brighter and more pleasant.

My special thanks to all the volunteer participants of this study.

And, of course, my thanks to God, who gives me strength to smile and go on at good and challenging times.

Abstract

Perception and production of English vowels by Brazilian EFL speakers

Andréia Schurt Rauber

Universidade Federal de Santa Catarina
2006

Supervisor Professor: Dr. Barbara Oughton Baptista

This study investigated the relationship between the perception and production of English vowels by 18 highly proficient Brazilian EFL speakers, most of them M.A. and doctoral students of the Graduate Program in English of the Universidade Federal de Santa Catarina. Two experiments were carried out: A production test to measure the first two formants of the learners' English and Brazilian Portuguese (BP) vowels, and an identification test with synthetic stimuli to investigate the L2 (second language) perception of English vowels. The production and perception results reveal that the Euclidean distance between the three English target pairs (/i/-/ɪ/, /ɛ/-/æ/, /ʊ/-/u/) was significantly larger for the American English monolinguals than for the L2 learners, thus indicating that the Brazilians have difficulty in both producing and perceiving these vowels in a native-like fashion. Importantly, some relationship between vowel perception and production was found because the target pairs which were better perceived were also the ones produced more accurately by the L2 learners. These results provide further evidence for the fact that L2 perception outperforms L2 production.

Number of pages: 157

Number of words: 40,393 (excluding appendixes)

Resumo

A percepção e produção das vogais do inglês por brasileiros falantes de inglês como língua estrangeira

Andréia Schurt Rauber

Universidade Federal de Santa Catarina
2006

Professora Orientadora: Dr. Barbara Oughton Baptista

Este estudo investigou a relação entre a percepção e produção das vogais do inglês por 18 falantes proficientes de inglês como língua estrangeira (ILE), a maioria mestrandos ou doutorandos do Programa de Pós-Graduação em Inglês da Universidade Federal de Santa Catarina. Dois testes foram aplicados: um teste de produção para medir os dois primeiros formantes das vogais do inglês dos aprendizes de ILE, e um teste de identificação com estímulos sintéticos para investigar a percepção das vogais do inglês por estes participantes. Os resultados de produção e percepção revelam que a distância euclidiana entre os três pares de vogais do inglês (/i/-/ɪ/, /ɛ/-/æ/, /ʊ/-/u/) é significativamente maior para os falantes monolíngües do inglês americano que para os aprendizes de inglês, o que indica que os brasileiros têm dificuldade tanto para perceber como para produzir estas vogais de forma nativa. Os resultados indicam que há uma relação entre a percepção e produção das vogais do inglês, já que os pares que foram mais bem percebidos foram também os produzidos mais corretamente pelos aprendizes de ILE. Estes resultados corroboram pesquisas anteriores que mostram que a percepção de sons da L2 precede a produção destes.

Número de páginas: 157

Número de palavras: 40,393 (excluindo os apêndices)

TABLE OF CONTENTS

LIST OF FIGURES	x
LIST OF TABLES	xiii
INTRODUCTION.....	1
CHAPTER 1. SPEECH PRODUCTION.....	6
1.1 The Portuguese and English vowel systems	12
1.1.1 Brazilian Portuguese vowels	12
1.1.2 American English vowels	23
1.2 Speech production in interphonology	35
CHAPTER 2. SPEECH PERCEPTION.....	40
2.1 Two L2 speech perception models.....	42
2.2 L2 perception studies	55
2.3 Studies on the interrelation between perception and production	58
CHAPTER 3. METHOD.....	63
3.1 Vowel production participants	63
3.1.1 American English (AE) monolingual participants.....	64
3.1.2 Brazilian Portuguese (BP) monolingual participants.....	65
3.1.3 L2 participants' (BP speakers of English as a foreign language) production	66
3.2 Corpus to collect vowel production	69
3.2.1 Corpus used to elicit the production of AE vowels	70
3.2.2 Corpus used to elicit the production of BP vowels	72
3.3 Data collection procedure of the production experiments	73
3.3.1 AE monolinguals' productions	74
3.3.2 BP vowel production by BP monolinguals and L2 speakers	74
3.3.3 L2 speakers' productions	75
3.4 Production measurements	76
3.4.1 Duration	77
3.4.2 Fundamental frequency (f ₀).....	78
3.4.3 The first three formants.....	79
3.5 Statistical analyses of the production results	80
3.6 Vowel perception participants.....	84
3.6.1 BP monolinguals	84
3.6.2 AE monolinguals.....	85
3.7 Perception stimuli	86
3.8 Perception test procedure	90

3.9 Analysis of the perception results	91
CHAPTER 4. RESULTS AND DISCUSSION	94
4.1 Speech production of BP monolingual speakers.....	94
4.2 Speech perception of BP monolinguals	100
4.3 Speech production of AE monolinguals	103
4.4 Speech perception of AE monolinguals.....	108
4.5 L2 speakers' production of BP vowels	111
4.6 L2 speakers' production and perception of L2 vowels	117
4.6.1 Female L2 speakers' production results.....	120
4.6.2 Female L2 speakers' perception results	126
4.6.3 Male L2 speakers' production results	130
4.6.4 Male L2 speakers' perception results.....	136
4.7 Discussion about the production and perception results	140
CHAPTER 5. CONCLUSION	147
REFERENCES.....	152
APPENDIXES	158
Appendix A - Praat script to plot EP/BP vowels (linear scale).....	159
Appendix B - Praat script to vowels (logarithmic scale - \log_{10})	160
Appendix C - Background information questionnaire (Portuguese version).....	162
Appendix D - Background information questionnaire (English version)	164
Appendix E - English sentences read by AE monolinguals and L2 speakers.....	165
Appendix F - Portuguese sentences read by BP monolinguals and L2 speakers.....	166
Appendix G - Pictures used in the English production test.	167
Appendix H - Set used in the production practice test.....	168
Appendix I - Pitch measurement script.....	169
Appendix J - Formant measurements script.....	170
Appendix K - Vowel generator script	172
Appendix L - Script to create perceptual vowel plots.....	175
Appendix M - Script to calculate the Euclidean distance in the perception test.....	179
Appendix N - Script to calculate the overlap between formant values of vowel pairs	183
Appendix O - Vowel normalization script.....	184
Appendix P - L2 female participants' L1 and L2 vowels	186
Appendix Q - L2 male participants' L1 and L2 vowels.....	197

LIST OF FIGURES

Figure 1. Drawings showing the different position of the tongue, jaw and lips for different vowels (Zue, 2000, p. 16).....	11
Figure 2. Delgado Martins's (1973) EP vowel triangle.....	15
Figure 3. F1 and F2 mean values from Lima (2001), in blue; Pereira (2001), in black; and Seara (2000), in red.....	19
Figure 4. F1 and F2 values from Moraes, Callou, and Leite (1996, p. 35) – Recife (black), Salvador (brown), Rio de Janeiro (green), São Paulo (red) and Porto Alegre (magenta).....	21
Figure 5. Distribution of AE monophthongs and semi-diphthongs in the vowel space.	23
Figure 6. Women's vowel centers from Peterson and Barney (1952) – in green, Hillenbrand et al. (1995) – in blue, and Hagiwara (1997) – in red.....	29
Figure 7. Men's vowel centers from Peterson and Barney (1952) – in green, Hillenbrand et al. (1995) – in blue, and Hagiwara (1997) – in red.....	29
Figure 8. Six US regions investigated by Clopper et al. (2005) for an acoustic description of American English vowels (figure printed with Clopper's permission).....	30
Figure 9. (a) Northern Cities Chain Shift, and (b) Southern Vowel Shift (Labov, 1998 in Clopper et al., 2005, p. 2, figures printed with Clopper's permission).....	31
Figure 10. Women's vowel centers from Clopper et al. (2005): black – New England, blue – Mid Atlantic, green – North, yellow – Midland, silver – South, and red – West.....	34
Figure 11. Men's vowel centers from Clopper et al. (2005): black – New England, blue – Mid Atlantic, green – North, yellow – Midland, silver – South, and red – West.....	34
Figure 12. The peripheral auditory system (Huckvale, 2006).....	42
Figure 13. The mapping of the continuum speech signal onto linguistic units (the waveform represents the word <i>seat</i> pronounced by a male native speaker of AE).....	50
Figure 14. Native language perception optimal states and L2 learning onset.....	51
Figure 15. Three possible scenarios for L2 speech perception, according to Escudero (2005, p. 124). CF = Canadian French and CE = Canadian English.....	52
Figure 16. Segmentation of the BP vowel /i/ produced in a sV.sV structure by a woman.....	78
Figure 17. Segmentation of the semi-diphthong /eɪ/ produced by an L2 female participant.....	80
Figure 18. Synthesized vowel continuum used to test vowel perception by the three....	87
Figure 19. Screen presented to BP monolinguals and L2 speakers to test their BP vowel perception.....	88
Figure 20. Screen presented to AE and L2 speakers to test their AE vowel perception.	89

Figure 21. BP female participants' mean and SD (in ellipses) of the F1 and F2 values in Hz.	96
Figure 22. BP male participants' mean and SD (in ellipses) of the F1 and F2 values in Hz.	97
Figure 23. Vowel duration values by monolingual BP speakers (/E/ and /O/ represent /ε/ and /ɔ/, respectively).	99
Figure 24. Perception results by BP monolinguals: One SD of the mean F1/F2 values of tokens identified as each vowel.	101
Figure 25. Number of vowels perceived with each of the two duration values by BP monolinguals.	102
Figure 26. Mean F1 and F2 values of vowels produced by the female AE monolinguals.	105
Figure 27. Mean F1 and F2 values of vowels produced by the male AE monolinguals.	105
Figure 28. Duration values of the vowels produced by the female and male AE monolinguals (the symbols I, E, V, A, O, and U correspond to the vowels /ɪ/, /ε/, /ʌ/, /ɑ/, /ɔ/, and /ʊ/, respectively).	108
Figure 29. Perception results by the AE monolinguals.	109
Figure 30. Number of vowels perceived with two duration values by AE monolinguals (the vowel I, E and U represent the vowels /ɪ/, /ε/ and /ʊ/, respectively)...	111
Figure 31. L2 female participants' mean and SD (in ellipses) of the F1 and F2 values of BP vowels in Hz.	113
Figure 32. L2 male participants' mean and SD (in ellipses) of the F1 and F2 values of BP vowels in Hz.	114
Figure 33. Female BP monolinguals' (in black) and L2 learners' (in grey) productions of BP vowels (mean F1 and F2 values). Note that the BP /o/ and L2 /u/ overlap.	114
Figure 34. Male BP monolinguals' (in black) and L2 learners' (in grey) productions of BP vowels (mean F1 and F2 values).	115
Figure 35. Participant 1's (woman) – in black – and female AE speakers' – in grey – means and SD (in ellipses), and the top and left marks indicating the minimum and maximum values used for normalizing the L2 vowels.	119
Figure 36. AE vowels (in grey) and L2 normalized vowels (in black) produced by the female speakers.	121
Figure 37. AE (in black) and BP (in grey) vowels produced by the female L2 speakers.	122
Figure 38. Mean duration values of English vowels produced by AE monolinguals and L2 speakers, and of Portuguese vowels produced by the L2 speakers (the vowels /I/, /E/ and /U/ represent the vowels /ɪ/, /ε/ and /ʊ/).	125
Figure 39. L2 perception results by the female participants (all durations included)...	127
Figure 40. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the female L2 participants.	129
Figure 41. L2 (in black) and AE (in grey) normalized vowels produced by the L2 male speakers.	131

Figure 42. L2 (in black) and BP (in grey) vowels produced by the male participants.	132
Figure 43. Mean duration values of English vowels produced by AE and L2 speakers, and of BP vowels produced by the L2 speakers.	135
Figure 44. L2 perception results by the male participants (all durations included).	137
Figure 45. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the male L2 participants.	139
Figure 46. The “new” scenario for the three L2 target vowel pairs.	144

LIST OF TABLES

Table 1. F1 and F2 values of vowels in stressed position from Lima (1991), Pereira (2001) and Seara (2001).....	18
Table 2. F1 and F2 values of vowels in stressed position from Moraes, Callou, and Leite (1996, p. 35).....	20
Table 3. Hagiwara's (1997) words for eliciting the 11 Southern Californian English vowels.	26
Table 4. Acoustic measurements of women's (W) and men's (M) productions of the AE vowels /i, ɪ, e, ε, æ/ analyzed in Peterson and Barney, (1952), Hillenbrand et al. (1995), and Hagiwara (1997).	27
Table 5. Acoustic measurements of women's (W) and men's (M) productions of the AE vowels /u, ʊ, o, ɔ, ɑ/ analyzed in Peterson and Barney, (1952), Hillenbrand et al. (1995), and Hagiwara (1997).	28
Table 6. Acoustic measurements of women's (W) and men's (M) productions of the AE vowels /ʌ, ɚ/ analyzed in Peterson and Barney, (1952), Hillenbrand et al. (1995), and Hagiwara (1997).	28
Table 7. Women's F1 and F2 frequencies measured by Clopper et al. (2005) in six AE dialects.	32
Table 8. Men's F1 and F2 frequencies measured by Clopper et al. (2005) in six AE dialects.	33
Table 9. American English participants' background.....	65
Table 10. Brazilian Portuguese female and male participants' background.....	66
Table 11. L2 female speakers' background.	68
Table 12. L2 male speakers' background.	69
Table 13. Target words read by the AE monolinguals and L2 speakers.....	71
Table 14. Target words read by the BP monolinguals.	73
Table 15. BP women's duration (in milliseconds), f0, F1, F2 and F3 values in Hertz...95	
Table 16. BP men's duration (in milliseconds), f0, F1, F2 and F3 values in Hertz.....	96
Table 17. Results of t-tests comparing F1 values within the groups of BP female and male monolinguals.	98
Table 18. BP women's and men's F1 and F2 values within 1 SD of the mean F1/F2 values of tokens identified as each vowel (N = number of vowels).	100
Table 19. Mean, median and SD of duration, f0, F1, F2 and F3 values of AE vowels produced by the female AE monolinguals.	103
Table 20. Mean, median and SD of duration, f0, F1, F2 and F3 values of AE vowels produced by the male AE monolinguals.	104

Table 21. Measurements comparison between the vowels produced by AE monolinguals in this study and those in Hagiwara (1997), Clopper et al. (2005), Hillenbrand et al. (1995), and Peterson and Barney (1952).....	106
Table 22. Euclidean distance and rate of overlap between the vowels of the 3 target pairs by the AE monolingual listeners.	109
Table 23. Number of tokens identified in the perception test within one standard deviation of the mean F1/F2 values of tokens identified as each vowel.....	110
Table 24. L2 women's duration, f0, F1, F2 and F3 values for BP vowels.	112
Table 25. L2 men's duration, f0, F1, F2 and F3 values for BP vowels.	113
Table 26. Results of two-tailed independent t-tests comparing the Euclidean distances between BP vowel pairs produced by the BP monolinguals and the L2 speakers.	115
Table 27. Median values of the Euclidean distances for AE and L2 vowels, percentage of L2 vs. AE similarity, and t-tests for the difference of each pair produced by the L2 female participants.	120
Table 28. Mean, median and SD of duration (D), f0, F1, F2 and F3 values of L2 vowels produced by the female speakers.	121
Table 29. Percentage of overlap between L2 and AE vowels by the female participants.	123
Table 30. Mean duration values (in ms) of AE vowels produced by the female AE monolinguals, and of L2 and BP vowels produced by the female L2 speakers.	124
Table 31. Rates of similarity between (i) the Euclidian distance (ED), and (ii) the mean duration values of AE and L2 vowel pairs; and paired t-tests of duration differences between the members of L2 pairs produced by the female L2 speakers.	125
Table 32. Euclidean distance (ED) values of AE and L2 vowel pairs perceived by the female participants (% of L2 similarity to AE values in parentheses), and overlap rate in the perception of AE monolinguals and L2 speakers.	126
Table 33. Percentage of short vowel (100 ms) identification by the female L2 speakers and AE monolinguals.	128
Table 34. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the female L2 participants.	128
Table 35. Median values of the Euclidean distances for AE and L2 vowels, rates of L2 vs. AE similarity, and t-tests for the difference of each pair produced by the L2 male participants.	130
Table 36. Mean, median and SD of duration (D) (in milliseconds), f0, F1, F2 and F3 values (in Hertz) of L2 vowels produced by the male speakers.	131
Table 37. Percentage of overlap between L2 and AE vowels by the male participants.	133
Table 38. Mean duration values (in ms) of AE, L2 and BP vowels produced by the male speakers.	134
Table 39. Rates of similarity between the Euclidian distance (ED) and mean duration values of AE and L2 vowel pairs, and t-tests of duration differences between the members of L2 pairs produced by the male L2 speakers.	135

Table 40. Euclidean distance (ED) values of AE and L2 vowel pairs perceived by the male participants (% of L2 similarity to AE values in parentheses), and overlap rate in the perception of L2 and AE speakers.	137
Table 41. Percentage of short vowel (100 ms) identification by the male L2 and AE speakers.	138
Table 42. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the male L2 participants.	139
Table 43. Rates of similarity of the Euclidean (ED) distances between the vowel pairs produced and perceived by the L2 female and male participants.	141
Table 44. Number of participants in each group: W = women, M = men.	147

INTRODUCTION

In the last decades, studies on the perception and production of English vowels by speakers of English as a foreign language (EFL) have shed some light on the difficulties these speakers have in perceiving vocalic sounds (e.g., Best, 1995; Bohn & Flege, 1992; Escudero, 2005; Flege, MacKay, & Meador, 1999; Flege, Munro, & Fox, 1994; Rochet, 1995) and producing them (e.g., Baptista, 2000, 2006; Flege, 1987a, 1987b; Major, 1987).

Concerning vowel quality, several studies investigating Brazilian Portuguese (BP) EFL speakers have already shown that these learners have difficulty in making a distinction between certain vowel pairs (e.g., Bion, Escudero, Rauber, & Baptista, 2006; Rauber, Escudero, Baptista, & Bion, 2005). In these studies, which analyzed the perception and production of American English (AE) vowels by proficient EFL speakers who had never lived in an English speaking country, it was observed that the participants tended not to make a distinction between the members of some English vowel contrasts, for example, /ɛ/-/æ/ (Bion et al., 2006; Rauber et al., 2005) or /u/-/ʊ/ (Rauber et al., 2005) in either perception or production. However, some distinction was made in the production of the vowel pairs /i/-/ɪ/ (Bion et al., 2006; Rauber et al., 2005) and /ɑ/-/ɔ/ (Rauber et al., 2005), although this distinction was not made in a native-like fashion.

Thus, with the aim of further investigating the pronunciation of AE vowels by BP advanced speakers of AE¹ who have never lived in an English speaking country, this study examined the acoustic properties of the vowels produced by three groups of speakers: Brazilian EFL speakers (henceforth L2 speakers), AE monolinguals, and BP

¹ Only participants who claimed to have American English as their target accent were selected for this study.

monolinguals. The AE vowel pairs analyzed were the three which are more likely to cause communication problems by BP speakers, since they tend to be mispronounced/misperceived: /i/-/ɪ/, /ɛ/-/æ/, and /ʊ/-/u/. Although only three AE vowel pairs were analyzed, the acoustic properties (duration, fundamental frequency and the first three formants) of eleven AE vowels ([i, ɪ, eɪ, ɛ, æ, ʌ, ɑ, ɔ, ɒ, u, ʊ]) were measured. Moreover, the same acoustic properties were also measured for the BP vowels ([i, e, ɛ, a, ɔ, o, u]), so as to test whether the L2 speakers rely on durational cues (the duration of a sound, a feature that differentiates it from another acoustically similar sound) or spectral cues (vowel quality, that is, formant values) to produce their native language (L1) vowels. The analysis will focus on the first two formants and on vowel duration, but the other measured properties will also be reported so as to serve as reference for forthcoming studies. All of the vowels were in a stressed syllable in the following structures: (a) in English: pVt², sVt, tVt, tVk, and kVt; and (b) in Portuguese: pV.pV, tV.kV, kV.kV, fV.fV, and sV.sV.

For the analysis of perception, the L2 speakers were asked to take a vowel identification test, which consisted of a vowel continuum formed by 339 vocalic sounds. These vocalic sounds were manipulated in terms of spectral quality and duration to investigate which of these two acoustic cues (spectral or durational cues) the L2 participants rely on to perceive L2 vowels. Vowel perception by two control groups, one formed by BP monolinguals and one by AE monolinguals, was also tested by means of the same vowel continuum used to test the L2 speakers; however, the labels³ used to test the BP monolinguals differed, since they contained only the BP vowels.

² V stands for vowel.

³ Labels in the perception tests are either vowels (in the case of the BP test) or words (in the case of the AE test) which were displayed on the computer screen to present the participants with the options they had when identifying the sounds they heard. After hearing a vocalic sound, the participant had to click on the label that corresponded to the best option of the vowel heard.

As regards the data analysis, it is important to state that in recent years advances in technology have allowed speech analyses to be performed on a larger corpus in a shorter time span. Acoustic analyses of vowels that used to be done manually and were really time-consuming can now be automated, resulting in more reliable and faster data analyses. In the Graduate Program in English (PGI) of the Universidade Federal de Santa Catarina, only a few studies have made use of acoustic phonetics in order to analyze production and/or perception data (Baptista, 2000; Baratieri, 2006; Bion et al., 2006; Rauber et al., 2005; Sada-Ribeiro, 2006). In all the studies on vowels (Baptista, 2000; Bion et al., 2006; Rauber et al., 2005), vowel measurement was done manually. The present research is the first study in the program to automate the analysis of vowel production and to make use of totally synthesized speech to investigate the perception of the whole vowel space of AE and BP. Taking into account the innovative data analysis techniques, almost all designed at the Institute of Phonetic Sciences of the University of Amsterdam by Dr. Paul Boersma and Dr. Paola Escudero, information about several basic terms used in acoustic analyses will be provided in order to facilitate comprehension of the text by a possible nonexpert reader.

In order to investigate how AE vowels were produced and perceived by the L2 speakers who participated in this study, the following research questions (RQs) were posed:

RQ1. Which AE target vowel pairs (/i/-/ɪ/, /ɛ/-/æ/, /ʊ/-/u/) will be more easily distinguished in both perception and production by the L2 speakers?

Hypothesis: Based on previous studies (Bion et al., 2006; Rauber et al., 2005), the following order of difficulty to distinguish the pairs in either perception and production will be found (least to most difficult): /i/-/ɪ/, /ʊ/-/u/, and /ɛ/-/æ/.

RQ2. What acoustic cues (spectral quality, duration) do the L2 speakers most rely on to perceive and produce L1 and L2 vowels?

Hypotheses: To perceive and produce the AE vowels, the L2 group will rely mostly on vowel duration. This hypothesis is based on Bohn (1995), who found that Spanish-speaking learners of English relied more on duration than on spectral quality to perceive English vowels. Since Spanish and Portuguese share similarities, Bohn's study was used as reference. Considering that all the participants are proficient English speakers, the difficulty in producing /ɪ/, /æ/ and /ʊ/ – vowels which do not exist in the BP vowel space – will lead them to make use of duration to differentiate between the members of each pair, especially because English L2 learners in Brazil are generally taught to distinguish between the vowels of a pair by identifying the *longer* and the *shorter* one. As for the BP monolinguals, they will rely primarily on spectral quality to identify the BP vowels. This hypothesis is based on Morrison (2006), who studied vowel perception and production by L1 Spanish speakers. Since the Portuguese and Spanish oral vowels have some resemblance, Morrison's study can be considered an appropriate reference to hypothesize how BP speakers will perceive L1 vowels. As regards the AE monolinguals, previous studies show that they rely primarily on spectral cues to perceive vowels (Bohn, 1995; Flege, Bohn, & Jang, 1997; Morrison, 2006).

RQ3. Is there an interrelation between L2 vowel perception and production?

Hypothesis: L2 vowels which are perceived in a native-like fashion will also be produced in a native-like fashion, and vowels which are misperceived will also be misproduced (Bion et al., 2006; Rauber et al., 2005).

In order to analyze how advanced Brazilian EFL speakers produce and perceive the target L2 vowel pairs, the dissertation is divided into 5 chapters. Chapter 1 reviews acoustic theories concerning vowel production and describes the AE and BP vowel

systems in phonetic terms by reporting the results of previous instrumental studies of vowels produced by monolingual speakers of the two languages. Finally, some studies on L2 production are also reported.

In Chapter 2, first a basic description of the human auditory system is provided, then theories of L2 perception are presented, and finally studies on L2 perception, as well as studies on the interrelation between L2 perception and production are reported.

In Chapter 3, a detailed description of the method adopted to collect and analyze the data is provided. Information is given about the participants, the recording procedures as well as the acoustic analysis employed to prepare the corpus for statistical analyses in the production experiments. The chapter also describes the procedures used to collect and analyze the perception data.

In Chapter 4, the results and discussion concerning the acoustic analysis of the vowels produced and perceived by the three groups of participants are presented. This is followed by a discussion of the relationship between vowel perception and production.

Finally, in Chapter 5 the conclusions about the results are provided, in light of the research questions and hypotheses elaborated for the present study. Some pedagogical implications are suggested, the limitations of this study are acknowledged, and suggestions for further research are presented.

CHAPTER 1

SPEECH PRODUCTION

An acoustic theory of speech production must be able to explain the characteristics of a given speech signal according to how this signal is generated. In articulatory terms, the source of a speech signal can be generated either by (i) the vibration of the vocal folds, as in voiced sounds; (ii) a noisy airstream, as in voiceless sounds; or (iii) a combination of the two, as in voiced fricatives (Harrington & Cassidy, 1999, p. 30⁴). With the vibration of the vocal folds, a series of complex periodic sound waves⁵ is produced. The number of repetitions of these waves in a second determines their fundamental frequency (f_0), which is measured in Hertz (Hz) (Ladefoged, 2003). Thus, an f_0 of 100 Hz means that the vocal folds make 100 complete opening and closing movements in a second. The f_0 changes according to the size, mass, and density of the vocal folds (Pickett, 1999, p. 57). Since children and women have smaller vocal folds than men, their f_0 is higher. This means that the smaller the vocal folds, the more they vibrate and thus the higher the f_0 . The f_0 is related to the pitch of a speech sound perceived by a listener and it also depends on the tension of the vocal folds: The higher the tension, the higher the pitch; conversely, the lower the tension, the lower the pitch (Pickett, 1999, p. 57). Subglottal pressure also determines the f_0 rate. In order for the vocal folds to vibrate, there must be enough air pressure in the lungs: If it is less than

⁴ Following the APA Publication Manual (2001, p. 121), some of the indirect citations will contain the page number of the source to help the readers locate it in the original text.

⁵ “A sound wave is a traveling pressure fluctuation that propagates through any medium that is elastic enough to allow molecules to crowd together and move apart” (Johnson, 2003, p. 4). The term “complex”, means that the wave has irregular variation in air pressure.

about 3 cm H₂O⁶, the air pressure in the lungs does not exceed that of the oral cavity and thus there is no vibration (Pickett, 1999, p. 57; Stevens, 1997, p. 464).

In the case of vowels, the vibration of the vocal folds produces the source of the sound which propagates through the vocal tract⁷ until reaching the outside air. Thus, since the vocal tract forms a resonating chamber, it can be considered a filter that amplifies some of the components of the source sound (Hayward, 2000; Pickett, 1999; Stevens, 1997). The view of the vocal tract serving as a filter for the glottal source for vowel production was first developed by Fant (1960) and is called the Source-Filter Theory of vowel production.

Depending on the shape of the vocal tract, its natural resonance frequencies form different vowel peaks which can be visualized in a spectrum⁸. Thus, formants are the natural resonance frequencies of the vocal tract (Johnson, 2003, p. 96). It is important to note that the fundamental frequency is not a formant, it is the source that causes resonances in the vocal tract. The picture of a spectrum shows the effect of the resonance peaks of the vocal tract on the glottal source. The location of the formant peaks affects both the location of the frequency of the vowel spectrum peaks as well as the amplitudes of the peaks in relation to each other. In a spectrum of vowels, the amplitude of the peaks near the region of the first formant is higher than the amplitude of the peaks which have higher frequencies (Pickett, 1999, p. 60). This means to say that there is greater energy at the fundamental frequency and at the first harmonics⁹ and this energy (or the amplitude of the harmonics) decreases gradually as the frequency goes up (Lieberman & Blumstein, 1988, pp. 34-35).

⁶ Air pressure is the force per unit area exerted by air molecules hitting a given surface. The cm H₂O is one of the most commonly used units to measure air pressure (Mateus, Andrade, Viana, & Villalva, 1990, p. 62).

⁷ The passages of the mouth, throat, and nose are collectively called the vocal tract (Ladefoged, 1993).

⁸ A spectrum is a plot that displays the amplitude versus the frequency of a sound wave (Johnson, 2003, p. 11).

⁹ "A harmonic is any whole-number multiple of the fundamental frequency" (Ladefoged, 1996, p. 38).

In order to measure the formant frequencies of oral vowels, many studies (including the present one) make use of the Linear Predictive Coding (LPC) analysis. The LPC analysis is a digital¹⁰ signal processing method that separates the source and filter components of a signal. With the decomposition of the speech signal into source and filter, the LPC provides a smoothed spectrum of the signal, the formant frequencies and bandwidths¹¹, which are necessary for formant tracking (Harrington & Cassidy, 1999, p. 211). The LPC analysis shows “an accurate representation about the vocal tract filter function”, since it finds the damped sine waves¹², and is thus used for locating broad peaks¹³ in the spectrum of a signal (Harrington & Cassidy, 1999, p. 99). Importantly, different pitch frequencies, or glottal pulse rates, do not affect the vowel spectrum envelope. The spectral envelope shows the overall shape of the spectrum, without the harmonics (Johnson, 2003, p. 97) and is only affected by the vocal tract shape (Pickett, 1999, p. 60). For an accurate analysis, the LPC order, that is, the number of peaks must be defined in advance, and the signal properties are the base of this definition. An inaccurate order selection, for instance, may lead to the emergence of peaks in wrong positions, since an incorrect setting of coefficients may result in too many or too few formant values. Moreover, the maximum formant frequency of the signal must be defined. For the analysis of oral vowels, the maximum formant frequency is generally set to 5000 Hz for men and 5500 Hz for women, if the number of peaks to look for is 5.

One model that explains the acoustic consequences of the vocal tract constrictions on the glottal source is the tube model of vowel production. Since the

¹⁰ Digital signal is the outcome of a conversion of the continuous speech signal into digits (Johnson, 2003, p. 20).

¹¹ Bandwidth is the width (in Hz) of the resonance peak (Johnson, 2003, p. 149). It is measured 3dB from the peak.

¹² Sine wave is a “simple periodic wave”, in other words, it has regular variation in air pressure (Johnson, 2003, pp. 7-8). If a sine wave is damped, it means that its spectral peak has a wider bandwidth, that is, the peak gets wider (Johnson, 2003, pp. 149-151).

¹³ A spectral peak is the frequency where the envelope has a local maximum (Johnson, 2003, pp. 31-32).

vocal tract is terminated by the vocal folds at one end and by the open air beyond the lips at the other end, phoneticians consider the vocal tract as a tube or a set of tubes. As stated previously, vowel production begins with the vibration of the vocal folds which produces the sound that propagates through the pharynx until reaching the outside air. The shape of the vocal tract determines which vowel will be produced. The relationship between the shape of the vocal tract and the location of the formants is particularly evident in the case of the first two formants, which are “closely tied to the shape of the vocal tract as the lips, tongue, pharynx, and jaw move to articulate the consonants and vowels” (Pickett, 1999, p. 38). It is important to stress that the first three formants are affected by the vocal tract shape, while F4, F5 and upper formants provide less linguistic information and vary according to the speaker (Borden, Harris, & Raphael, 2002 cited in Master, 2005, p. 9; Sundberg, 1987 cited in Master, Biase, Chiari, & Pedrosa, 2006, p. 113; Vieira, 2004, p. 71).

Since the length of the pharyngeal-oral tract affects the formant frequency locations, the longer the tract, the lower the formant frequencies. This explains why women and children have higher formant frequencies than men: Their pharyngeal-oral tract is smaller. An average male vocal tract is commonly estimated to have 17 cm (Hayward, 2000, p. 83). When the length of the straight tube is known, the first formant can be calculated by the formula $f = c/4l$, where c is the velocity of sound in air (340 meters/second) and l is the length of the tube. A vowel that is basically unstricted, thus having a single straight tube, is *schwa* [ə]. For this vowel the first formant is easily calculated:

$$f_1 = c/4l$$

where f = formant, c = 34,000 cm/sec (340 mts/sec), and l = 17 cm

$$f_1 = 34,000 / (4 \times 17)$$

$$f_1 = 500 \text{ Hz}$$

The other formants are odd multiples of F1, thus F2 and F3 values, for example, are:

$$f_2 = 3c/4l = 1500 \text{ Hz}$$

$$f_3 = 5c/4l = 2500 \text{ Hz}$$

As to the tongue, jaw and lips, their positions affect the formant locations because they form different points of constriction (Pickett, 1999, pp. 40-42), or different tubes. For vowels other than schwa, the articulations can be represented by joining several tubes; however, the calculations are more complicated because of the effects of coupling the tubes. Generally speaking, and disregarding formulas, it can be said that when there is any constriction in the front half of the oral part of the vocal tract, the frequency of the first formant (F1) is lowered. Consequently, the greater the constriction, the lower the F1. Low F1 frequencies are found in high vowels, such as [i] and [u]. Conversely, if there is constriction of the pharynx, the greater the constriction, the higher the F1. This is why low vowels such as [a] have higher F1 frequencies. As to the second formant (F2) frequencies, they have higher values the more constricted the oral tract becomes when the tongue is raised toward the palate, and the frequencies are lower when the tongue is raised toward the velum (Pickett, 1999, p. 42). That is why front vowels have the highest F2 frequencies and back vowels, the lowest. The position of the lips also affects formant frequencies: The more rounded, the more the constriction and the lower the formant frequencies. Thus, back vowels have the lowest formant frequencies, the values decreasing from [ɑ] to [u] due to the greater constriction in the front part of the oral tract and the more humped the tongue is toward the palate. In sum, high/low vowels are correlated with F1 and front/back vowels, with F2. The different

positions of the jaw, tongue and lips which directly affect the F1 and F2 frequencies can be more easily visualized in Figure 1.

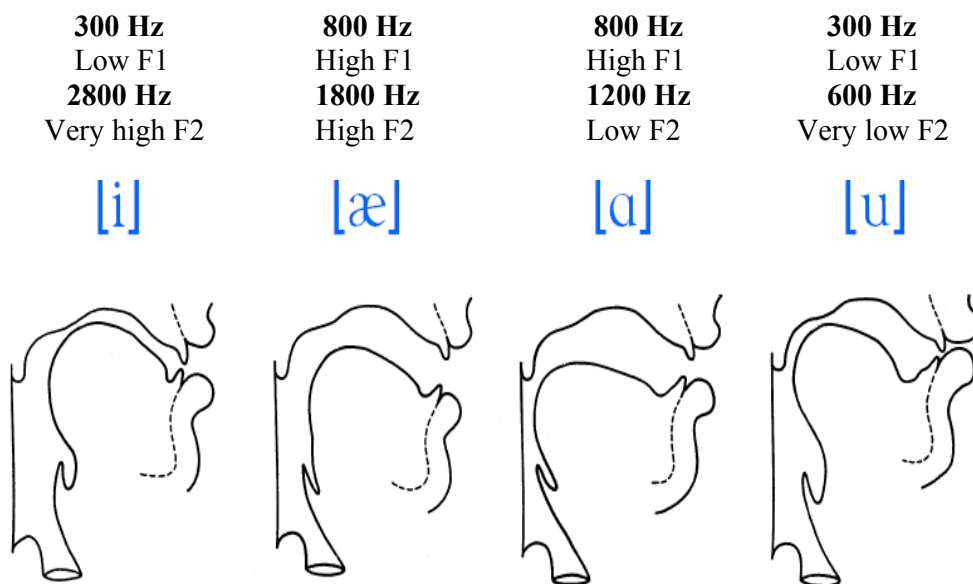


Figure 1. Drawings showing the different position of the tongue, jaw and lips for different vowels (Zue, 2000, p. 16).

In the case of vowels the description of how formants are related to the shape of the vocal tract that originates them can be a good attempt to describe the vowels' acoustic properties. The fundamental frequency, the formant frequencies, and the amplitude and bandwidth of a vowel waveform constitute some of the acoustic properties of a vowel. As stated in the introduction, in this study of oral vowels the data analysis will focus on the first two formants and on the duration of the BP and AE vowels.

This brief review of the physical aspects involving speech production gives support to the analysis of the vowels investigated in this study. Moreover, it facilitates

the understanding of the results obtained in other studies that analyzed acoustic properties of the target vowels, which will be reported in the following subsections.

1.1 The Portuguese and English vowel systems

This section reports on studies which measured the acoustic properties of oral vowels produced by BP and AE monolingual speakers. Tables and plots containing formant values of BP and AE vowels are included for the reader's reference.

1.1.1 Brazilian Portuguese vowels

The triangular BP vowel system consists of 12 vowels in stressed (tonic) position: Seven oral vowels with four degrees of height (/i, e, ε, a, o, u/), and five nasal vowels (/ĩ, ê, ã, õ, û/) with three degrees of height. In unstressed position, the oral vowels are reduced to five in pretonic position (/i, e, a, o, u/), and to three in posttonic position (/i, a, u/). The nasal vowels are kept to five in unstressed position, although they rarely occur in posttonic position (Moraes, 1999). Still concerning pretonic BP vowels, the mid vowels /e/ and /o/ may be pronounced as /ε/ and /ɔ/, or as /i/ and /u/, respectively, according to different BP dialects and syllable position (Cristófaró Silva, 2002).

BP vowels are not distinguished in terms of duration by BP speakers, as could be observed in this study. This may be the reason why most studies on BP vowels to the

present date disregard duration in their analysis. Vowel duration can be measured by marking the first and last periodic pulses on a waveform that have some considerable amplitude and resemble the vowel period. Although this procedure seems to be simple and straightforward, some consonants, mainly the voiced consonants, may hinder the precise identification of the points where the vowel begins and ends. In the present study only voiceless consonants were selected to form the phonological context where the vowels were inserted precisely to avoid this difficulty. However, even with these precautions there may be instances when the segments do not seem to have a clear beginning or end mainly because of coarticulation. In these cases, as suggested by Ladefoged (2003, p. 103), the segments must be measured consistently so that the duration of a segment be reported as accurately as possible. Apart from the problem caused by coarticulation, even a careful segmentation task may lead to mistakes because of its subjective and time-consuming nature, which eventually causes fatigue (Leung & Zue, 1984 cited in Barbosa, 1999, p. 24).

There are a limited number of studies which make use of acoustic phonetics to describe BP vowels (Moraes, Callou, & Leite, 1996; De Faveri, 1991; Lima, 1991; Pereira, 2001; Seara, 2000), and most of them have several limitations as regards the number of participants, the control of the phonological environment where the vowels were inserted, gender, and dialect.

Of the studies on BP phonetics carried out in the Linguistics Department of the Universidade Federal de Santa Catarina, two focused on vowel quality only (Lima, 1991; Pereira, 2001), one on vowel duration only (De Faveri, 1991), and one on vowel quality and duration (Seara, 2000). All these studies investigated participants who lived in the greater Florianópolis region¹⁴. A very well-known study which investigated speakers of different Brazilian dialects is Moraes et al. (1996). These studies make

¹⁴ Florianópolis is located in the South of Brazil, approximately 700 km from São Paulo, and 1700 km from Brasília, the capital of Brazil.

reference to Delgado Martins (1973), which was the first study that analyzed European Portuguese (EP) vowels acoustically. An overview of these studies on Portuguese vowels will be reported below.

Delgado Martins (1973) is the first study to make use of acoustic analysis to measure the first three formants of the EP vowels produced in Lisbon, Portugal. Her participants were eight men who had at least finished secondary school, and whose ages ranged from 18 to 40 years. All of them had always lived in Lisbon. The participants were asked to read words with CVC, CVCV, CVVCV, CVCCV, CVCVC frames inserted in the carrier sentence *Digo a palavra ... outra vez* (I say ... again). In the dissyllabic words the stress was always on the penultimate syllable. Three CVCVCVC words, also with penultimate stress, were inserted in the carrier sentences *Ontem papámos* ([pa^lpamuʃ]) *tudo* (Yesterday we ate everything), *Agora papamos* ([pa^lpemuʃ]) *tudo* (Now we eat everything), and *Para que papemos* ([pa^lpemuʃ]) *tudo* (So that we eat everything).

As regards phonological context, many consonants varied in place and manner of articulation, and the variable voicing was not controlled. This would have been no limitation if Delgado Martins had not averaged the F1 and F2 values from all the phonological contexts to obtain the acoustic vowel triangle seen in Figure 2 (see the Praat script used to plot EP/BP vowels in a linear scale in Appendix A).

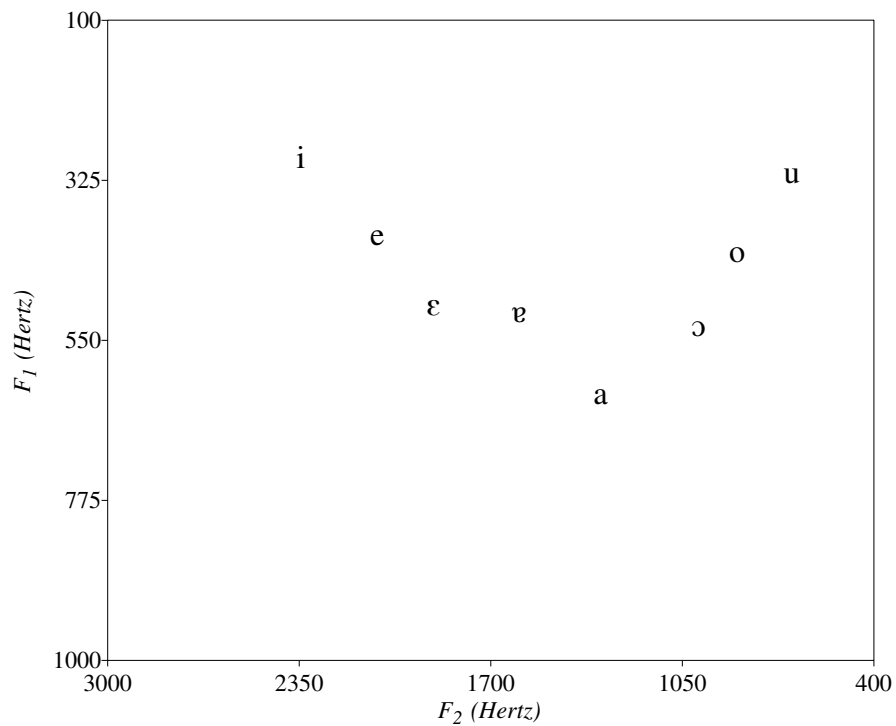


Figure 2. Delgado Martins's (1973) EP vowel triangle.

As to the Brazilian studies, Lima (1991) was the first to investigate vowel production in Florianópolis. The group of participants in his study consisted of 5 men who were born and had always lived in Florianópolis, and whose ages ranged from 21 to 35 years (mean = 27 years). Two participants had finished secondary school, one had graduated and the other two were undergraduate students at the time of the recordings. The main objective of Lima's study was to characterize acoustically the BP oral vowels in pretonic (280 items), tonic (560 items) and posttonic (210 items) positions. The data collection consisted of the reading of 98 non-artificial sentences in BP, which contained words with the target vowels preceded by the bilabials [p] or [b]. There was no control of the consonant which followed the target vowel. The sentences varied in length, and the position of the target words in the sentences was not controlled. Each CV item, always in different words, was produced three times in syllable-medial position, and the

number of productions varied in syllable-initial and final positions according to each vowel. The number of pretonic and posttonic productions also varied according to each vowel and syllable position. There was no statistical treatment and the author simply concluded that the effect of the preceding consonant was not strong in any syllable position of pretonic and tonic vowels, the variation being constant from vowel to vowel. The author observed more variation in posttonic vowels as regards syllable position and preceding consonant. However, no conclusive claims can be made due to the lack of statistical analysis.

Pereira (2001) aimed at investigating whether there would be any variation in the vocalic system of Florianópolis 10 years after Lima's (1991) study. The author hypothesized that there would be some variation both because of the constant natural changes in the language and also because of the increasing number of migrants mainly from Porto Alegre and São Paulo in recent years. The author recorded the productions of 5 men who were born and had always lived in Florianópolis, whose contact with people from other Brazilian regions varied from occasional (1 participant) to regular (1 participant) to intense (3 participants). Their ages ranged from 30 to 76 years (mean 49.8 years). The variable "education level" was not controlled: Three participants had finished primary school and two had graduated. Pereira focused on BP oral vowels in stressed position only. The participants were recorded while being interviewed and while reading a set of 50 sentences. As in Lima (1991), the length of the sentences read by the participants and the position of the target words within the sentence were not controlled. Although the sentences were given to the participants two days before the test for practice, many sentences are long and sound quite unusual, which might have caused the participants some difficulty. Each sentence was read three times in succession, which might have caused rising intonation of words in sentence-final

position. These particularities concerning the reading procedure are not mentioned in Pereira's text as possible limitations. The seven oral vowels were inserted in CV and CVC syllables, whose preceding context was one of the consonants /p, b, t, d, k, g/, and the following consonant was either /ʔ¹⁵, x, ɣ, h, ʌ, ʃ, ʒ/ or /w/. Due to the complexity of the words formed by the consonantal contexts, several tokens were disregarded because they would form unreal BP words. Despite the different data collection procedures and the limited number of participants, the author concluded that in the interval of ten years, /i, a, u/ in her study had become more anterior, /e, o/ more posterior, and /ɛ, ɔ/ higher than in Lima (1991). Two independent variables somewhat influenced F1 and F2 values – *preceding* and *following consonantal contexts* – but significant differences were found for the non-linguistic variables *participant*, *age*, and *amount of contact with migrants*. The variables *type of data collection* and *level of education* did not show statistically significant results. Although Pereira states that more participants should be investigated for more conclusive results, her findings corroborate the hypothesis that the different results from Lima (1991) may be due to the great number of migrants in Florianópolis. However, the different data collection and data analysis procedures in the two studies do not allow any safe conclusion.

Seara (2000) investigated the acoustic properties of BP nasal vowels, but in order to carry out her analyses the author also measured the formants and duration of the five oral vowels which have nasal counterparts: /i, e, a, o, u/. Seara analyzed the productions of five male participants: Four were born and had always lived in Florianópolis, and one was born in Tubarão-SC, but moved to Florianópolis when he was one year old and had lived there since then. Their ages ranged from 22 to 48 years (mean = 35.2 years). The vowels were inserted in either real or pseudo words in

¹⁵ [sic] The author does not give information about why the dark /l/ was selected as one of the phonological contexts. Given that this phoneme is not normally produced in the Florianópolis dialect, I assume that the author misused the symbol.

pretonic or tonic positions within the carrier sentence *Digo ____ pra ele* (I say ____ to him). The preceding context was always /p/ and the following context was either /p/, /t/ or /k/. The careful choice of phonological context avoided formant variation due to the preceding context and facilitated segmentation, since all the consonants are voiceless plosives and are more easily identified in the spectrogram than are voiced consonants. Each sentence was read 9 times, but only 7 productions were analyzed (the first and the last readings were discarded to avoid list reading effects). As regards vowel quality, Table 1 and Figure 3 show F1 and F2 values from Lima (1991), Pereira (2001) and Seara (2000).

Table 1. F1 and F2 values of vowels in stressed position from Lima (1991), Pereira (2001) and Seara (2001).

Formants		[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]
Lima	F1	332	424	550	620	550	437	328
	F2	2070	1860	1568	1288	1010	915	788
Pereira	F1	324	418	449	651	414	422	358
	F2	2252	1745	1618	1440	879	890	919
Seara	F1	263	400	--	740	--	427	307
	F2	2148	1964	--	1335	--	877	823

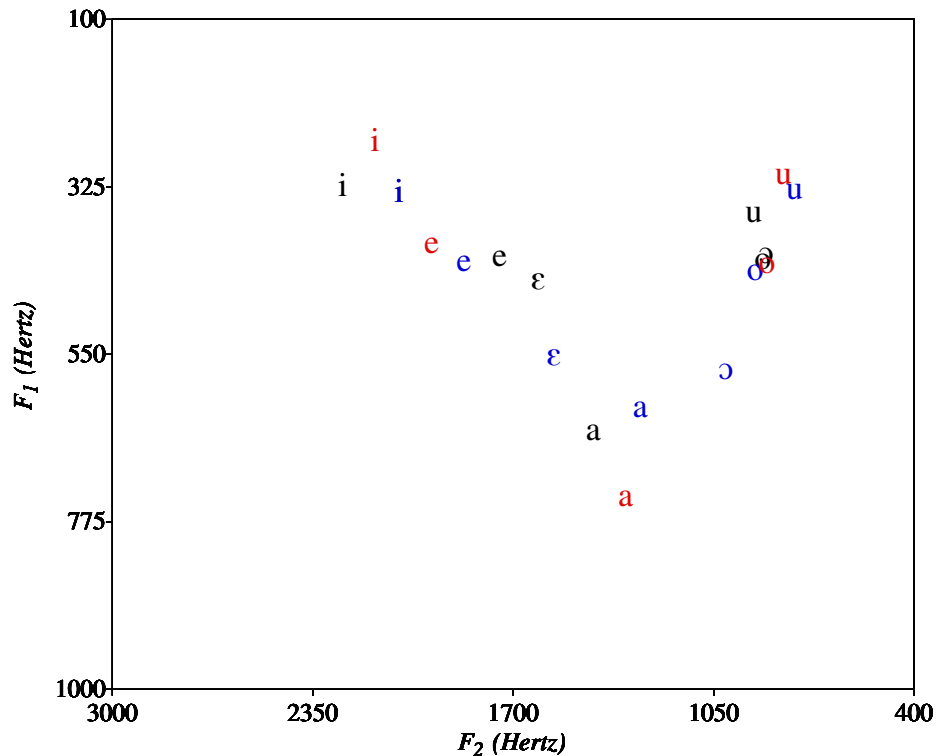


Figure 3. F1 and F2 mean values from Lima (2001), in blue; Pereira (2001), in black; and Seara (2000), in red.

As can be observed in Table 1 and Figure 3, the mean F1 and F2 values of the three studies show great variation in the series of front vowels, in the central vowel /a/, and in the back vowel /ɔ/. It is important to note that no conclusive claims can be made about the comparison of the results obtained in the three studies, since the data collection procedures and especially the contexts where the vowels were inserted differed greatly from each other. One observation that can be made is that, if each study is considered separately, the high, mid and low vowels /i, e, ɛ/ are aligned with their back vowel counterparts, following the tendency found for BP vowels in Moraes et al. (1996), and for EP vowels in Delgado Martins (1973).

In Moraes et al. (1996), the authors investigated vowel productions of 15 men from five capitals of three Brazilian regions: Recife and Salvador (Northeast); Rio de Janeiro and São Paulo (Southeast); and Porto Alegre (South). The participants from each city were divided into three age groups: 25-35 years, 36-56 years, and older than 56. This means that there was only one participant representing one age group in a given city. The age division was done in order to investigate whether there could be some phonetic change in progress in the BP vowel system. All the participants had already graduated. The tokens were extracted from spontaneous speech obtained by means of interviews. The authors do not mention the phonological contexts where the vowels were inserted. The oral vowels were analyzed in pretonic, tonic and posttonic positions. As regards the vowels in tonic position, the focus in the present study, the authors concluded that the BP system tends to be more compact when comparing the productions of the three age groups: The high vowels /i, u/ tend to be produced lower, and the central vowel /a/ tends to be produced higher the younger the participants. Table 2 and Figure 4 show the mean values obtained for each region.

Table 2. F1 and F2 values of vowels in stressed position from Moraes, Callou, and Leite (1996, p. 35).

Formants		[i]	[e]	[ɛ]	[a]	[ɔ]	[o]	[u]
Recife	F1	399	449	561	730	568	454	403
	F2	2235	2004	1850	1460	1110	1031	939
Salvador	F1	320	390	480	643	503	400	346
	F2	2106	1883	1716	1303	986	953	930
Rio de Janeiro	F1	336	400	533	653	546	410	350
	F2	2196	2016	1833	1426	1020	976	943
São Paulo	F1	336	403	550	706	570	410	336
	F2	2053	1953	1750	1396	990	913	933
Porto Alegre	F1	373	440	526	616	530	423	356
	F2	2213	1996	1816	1513	1056	990	896

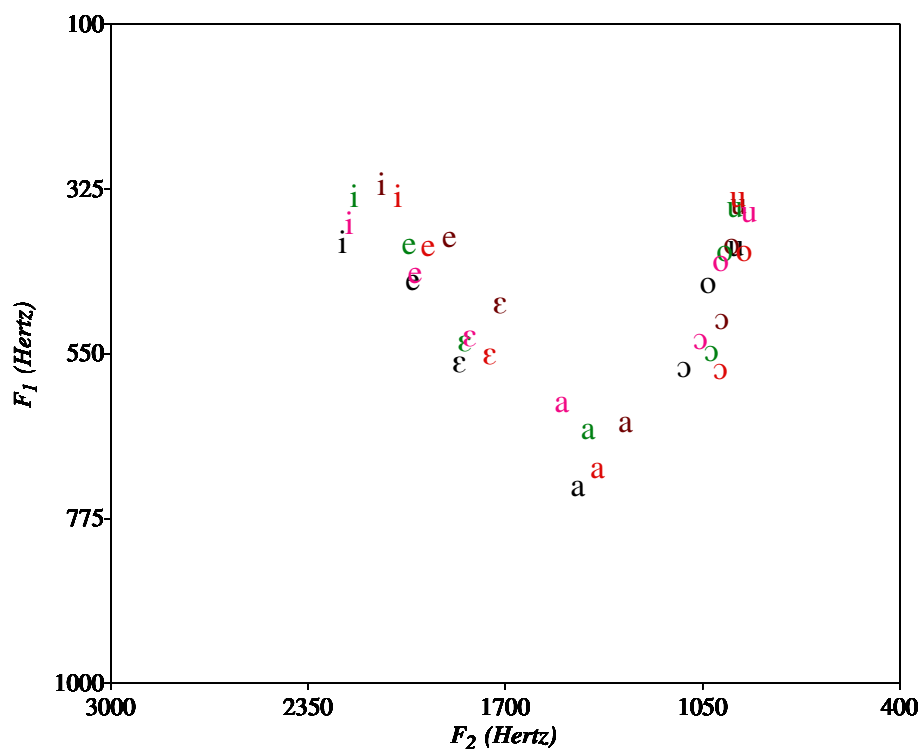


Figure 4. F1 and F2 values from Moraes, Callou, and Leite (1996, p. 35) – Recife (black), Salvador (brown), Rio de Janeiro (green), São Paulo (red) and Porto Alegre (magenta).

Again there is greater variation in the series of front vowels compared to that of back vowels. In general, the comparison between the five vowel systems shows that the vowels /i/, /a/, and /u/ from São Paulo are the farthest from each other; Recife has the lowest front and central vowels; Salvador has the highest vowels; and the vowels from Rio de Janeiro are in an intermediate position.

After this review of studies on vowel formant analysis, some interesting findings can be reported about two studies on BP vowel duration, both of which investigated speakers from Florianópolis. The first is De Faveri (1991), who analyzed the productions of the same participants as Lima (1991) – their background has been described previously. The main findings in De Faveri's study concerning stressed

vowels are that no statistically significant differences in duration were found between front and back vowels. As regards height, the difference between high and low vowels was statistically significant, low vowels being longer than high vowels. This difference is expected since the articulators have to open wider apart for the production of low vowels, which makes these vowels longer. De Faveri also analyzed the influence of the preceding (/p, b/) and following (/x, ɣ/) consonants on vowel duration. The author found that the vowels preceded by the voiced bilabial obstruent and followed by the voiced velar fricative were significantly longer than those preceded and followed by their voiceless counterpart. The position of the stressed syllable in the word was also observed and the findings revealed that the vowels were significantly longer in syllable-final position, followed by syllable-medial and then syllable-initial positions.

Similar findings were obtained by Seara (2000) as regards the five oral vowels analyzed. Seara also concluded that low vowels were significantly longer than high vowels in stressed position, the length decreasing from /a/ > /o/ > /e/ > /u/ > /i/. As described previously, the preceding context in her study was the voiceless bilabial plosive (/p/) and the following context was formed by one of the voiceless plosives /p, t, k/. The following context did not yield significant differences regarding length of the vowels.

These studies on formant analysis and duration provide some background knowledge for this study, which focuses on the vowels produced by participants from non-capital cities of the three Brazilian southern states. More information about the participants and data collection procedures will be reported in Chapter 3 (Method). The following sub-section will describe some studies on AE vowel production.

1.1.2 American English vowels

The American English (AE), or General American, vowel system is formed by 10 steady vowels or monophthongs (/i/, /ɪ/, /ɛ/, /æ/, /ə/¹⁶, /ʌ/, /ɑ/, /ɔ/, /ʊ/, /u/), two homogenous diphthongs or semi-diphthongs¹⁷ (/eɪ/, /oʊ/ or /eɪ/, /oʊ/), and three heterogeneous diphthongs¹⁸ (/aɪ/, /aʊ/, /ɔɪ/) (Roca & Johnson, 1999, pp. 169-203). Figure 5 shows how the monophthongs and semi-diphthongs are distributed in the vowel space.

	FRONT	CENTRAL	BACK
high	/i/ /ɪ/		/u/ /ʊ/
mid	/e/ /ɛ/	/ʌ/ /ə/	/o/
low			/ɔ/ /ɑ/

Figure 5. Distribution of AE monophthongs and semi-diphthongs in the vowel space.

As regards the acoustic properties of AE vowels, the most frequently cited paper is a study by Peterson and Barney (1952). The researchers recorded two repetitions of ten vowels (/i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u, ʌ, ə/), inserted in the /hVd/ words *heed*, *hid*, *head*, *had*, *hod*, *hawed*, *hood*, *who'd*, *hud*, *heard*, which were read by 33 men, 29 women, and 15 children. Their primary aim was to show that spectrographic analysis of speech was

¹⁶ This is the vowel found in the words *heard* [hɜːd] or *girl* [gɜːl], for instance.

¹⁷ The vowels /eɪ/ and /oʊ/ are called homogeneous diphthongs because both phases of the diphthongs are close in articulatory position and share the lip gesture (Roca & Johnson, 1999).

¹⁸ The two phases of the vowels /aɪ/, /aʊ/ and /ɔɪ/ are not close in articulatory position and do not share lip gesture, thus being called heterogeneous diphthongs.

useful to characterize vowel quality. The steady state portion of each vowel was measured so as to identify f_0 and the first three formants, and the /hVd/ tokens were later played to listeners for identification. It was possible to observe that there was a strong relationship between the measured values and the identification of the intended vowel. However, the study had several limitations. A considerable degree of variability between the formant frequencies was found between speakers from the same group and there was an overlap of values between adjacent vowels. Despite the variability, the listening test revealed that the vowels were identifiable. The overall error rate was low (5.6%) and errors generally occurred when adjacent vowels were confused. The great variability in formant frequency might have occurred because the participants were not separated by dialect (little information is provided about the participants' background). As regards their child participants, only a small group of children participated in the study and there is no information about their age or gender. As to the listening results, they were not reported separately for the men, women and children. The formant values obtained for the women and men will be reported below together with the results of other studies about AE vowels.

Taking into account the limitations of Peterson and Barney (1952), Hillenbrand, Getty, Clark, and Wheeler (1995) replicated and extended their study so as to measure the AE vowels /i, ɪ, e, ε, æ, ɑ, ɔ, o, u, ʌ, ə/, and the diphthongs [ɔɪ, aɪ, aʊ]. In their study, besides including the diphthongs and the vowels /e/ and /o/, vowel duration and the first four formants were measured. The raw frequencies of the diphthongs will not be reported in the present study, since the focus here is on monophthongs only. The number of participants in each group increased compared to Peterson and Barney's (1952) study: 45 men, 48 women, and 46 children (27 boys and 19 girls whose ages ranged from 10 to 12 years). The majority of participants were from a single region

(southeastern and southwestern parts of Michigan) and were selected from a larger group according to a careful dialect assessment procedure, whose main concern was to select participants who could make a distinction between [ɑ] and [ɔ]. Just as in Peterson and Barney (1952), the production test consisted of the reading of the /hVd/ words *heed, hid, hayed, head, had, hod, hawed, hoed, hood, who'd, hud, heard, hoyed, hide, hewed,* and *how'd*. Twelve lists containing the words in a randomized order were prepared and each participant had to read one of the 12 lists three times. Many children read the list only twice, because they got tired after the training session. The vowel formant values will be reported below. As regards duration, the results showed that the male participants produced significantly shorter vowels than women and children. Compared to Peterson and Barney's (1952) results, Hillenbrand et al. (1995) state that the great differences in formant values of certain vowels in the two studies cannot be explained only by differences in measurement procedures of the steady state portion of the vowels inserted in the /hVd/ frame. The authors consider that the lack of information about the participants' background in Peterson and Barney (1952) may explain the inconsistencies, since in the latter study the participants seemed to have formed a heterogeneous group. Moreover, the authors suggest that the productions were simply different in the two studies: A 40-year time span between the two studies is expected to reveal linguistic changes. Importantly, the listening results of the two studies did not yield significant differences; that is, the rates in the identification tests are quite similar.

Another study which aimed at measuring the acoustic properties of vowels from a specific dialect is Hagiwara (1997). The focus of his study was to analyze the 11 AE vowels /i, ɪ, e, ε, æ, ɑ, o, ʊ, u, ʌ, ə/ inserted in three phonological contexts /bVt/, /tVk/, and /hVd/ spoken by 15 undergraduate students at UCLA (University of California, Los Angeles), 9 women and 6 men, whose ages ranged from 18 to 26, all of

them speakers of *Southern Californian American English*. The words were 33 real English words and familiar proper nouns, as can be seen in Table 3. When a word with the target phonological context did not exist, another word whose consonants had a similar place of articulation was used. Each word was inserted in the carrier sentence *Cite ___ twice.*, and the participants produced each target word in random order three times.

Table 3. Hagiwara's (1997) words for eliciting the 11 Southern Californian English vowels.

	/i/	/ɪ/	/eɪ/	/ɛ/	/æ/	/u/	/ʊ/	/oʊ/	/ɔ/	/ʌ/	/ə/
b-t	beat	bit	bate	bet	bat	boot	put	boat	bought	but	Bert
t-k	teak	tick	take	tech	tack	duke	took	toke	tock	tuck	Turk
h-d	heed	hid	hate	head	had	hoot	hood	Hode	hod	hut	herd

Hagiwara provides the mean F1, F2 and F3 values for each vowel produced by each gender, but the means correspond to the three phonological contexts together, that is, no analysis of the influence of context on the vowels is reported. In comparison with Peterson and Barney's (1952) data, Hagiwara (1997) observes that (i) the back vowels /u/ and /ʊ/ are less rounded, thus acoustically more central, in the Southern Californian dialect; (ii) the central vowel /ʌ/ is higher; and (iii) the Californian women produced the low vowels /æ/ and /ɑ/ 200 Hz higher than did the women in Peterson and Barney; however, no significant differences were found for the male participants in the two studies.

The database of acoustic measurements of AE vowels provided by the three studies reviewed so far is shown in Tables 4-6 and plotted in Figures 6 and 7. The

vowels were plotted in a logarithmic scale (\log_{10}) in Praat. This scale, besides maintaining the values in Hz, represents to some extent the way the human ear perceives the differences in frequencies; that is, the higher the frequencies, the greater the distances between them are necessary for human beings to notice some change in pitch. That is why in the plot there is a greater difference between 200 Hz and 400 Hz values than between 800 Hz and 1000 Hz, for instance. From this point on, all the vowels will be plotted in the \log_{10} scale. The script to plot vowels can be found in Appendix B.

Table 4. Acoustic measurements of women's (W) and men's (M) productions of the AE vowels /i, ɪ, e, ε, æ/ analyzed in Peterson and Barney, (1952), Hillenbrand et al. (1995), and Hagiwara (1997).

Study		/i/		/ɪ/		/e/		/ε/		/æ/	
		W	M	W	M	W	M	W	M	W	M
Peterson & Barney	f0	235	136	232	135	--	--	223	130	210	127
	F1	310	270	430	390	--	--	610	530	860	660
Barney	F2	2790	2290	2480	1990	--	--	2330	1840	2050	1720
	F3	3310	3010	3070	2550			2990	2480	2850	2410
Hillenbrand et al.	f0	227	138	224	135	219	129	214	127	215	123
	F1	437	342	483	427	536	476	731	580	669	588
	F2	2761	2322	2365	2034	2530	2089	2058	1799	2349	1952
	F3	3372	3000	3053	2684	3047	2691	2979	2605	2972	2601
Hagiwara	F1	362	291	467	418	440	403	808	529	1017	685
	F2	2897	2338	2400	1807	2655	2059	2163	1670	1810	1601
	F3	3495	2920	3187	2589	3252	2690	3065	2528	2826	2524

Table 5. Acoustic measurements of women's (W) and men's (M) productions of the AE vowels /u, ʊ, o, ɔ, ɑ/ analyzed in Peterson and Barney, (1952), Hillenbrand et al. (1995), and Hagiwara (1997).

Study		/u/		/ʊ/		/o/		/ɔ/		/ɑ/	
		W	M	W	M	W	M	W	M	W	M
Peterson & Barney	f0	231	141	232	137	--	--	216	129	212	124
	F1	370	300	470	440	--	--	590	570	850	730
Barney	F2	950	870	1160	1020	--	--	920	840	1220	1090
	F3	2670	2240	2680	2240	--	--	2710	2410	2810	2440
Hillenbrand et al.	f0	235	143	230	133	217	129	210	121	215	123
	F1	459	378	519	469	555	497	781	652	936	768
	F2	1105	997	1225	1122	1035	910	1136	997	1551	1333
	F3	2735	2343	2827	2434	2828	2459	2824	2538	2815	2522
Hagiwara	F1	395	323	486	441	516	437	--	--	997	710
	F2	1700	1417	1665	1366	1391	1188	--	--	1390	1221
	F3	2866	2399	2926	2466	2904	2430	--	--	2743	2405

Table 6. Acoustic measurements of women's (W) and men's (M) productions of the AE vowels /ʌ, ə/ analyzed in Peterson and Barney, (1952), Hillenbrand et al. (1995), and Hagiwara (1997).

Study		/ʌ/		/ə/	
		W	M	W	M
Peterson & Barney	f0	221	130	218	133
	F1	760	640	500	490
Barney	F2	1400	1190	1640	1350
	F3	2780	2390	1960	1690
Hillenbrand et al.	f0	218	133	217	130
	F1	753	623	523	474
	F2	1426	1200	1588	1379
	F3	2933	2550	1929	1710
Hagiwara	F1	847	574	477	429
	F2	1753	1415	1558	1362
	F3	2989	2496	1995	1679

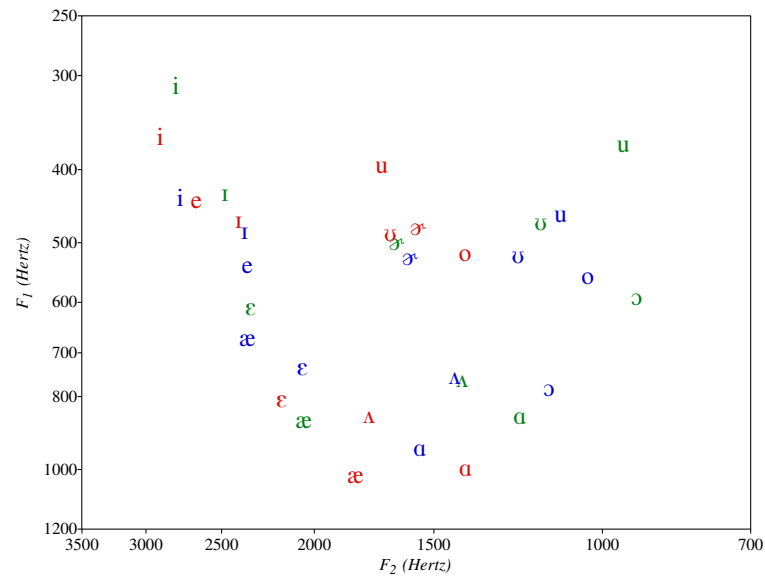


Figure 6. Women's vowel centers from Peterson and Barney (1952) – in green, Hillenbrand et al. (1995) – in blue, and Hagiwara (1997) – in red.

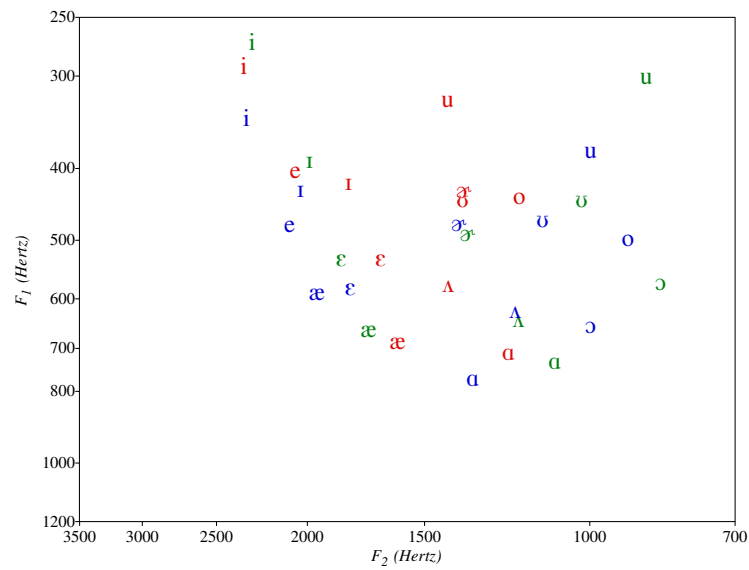


Figure 7. Men's vowel centers from Peterson and Barney (1952) – in green, Hillenbrand et al. (1995) – in blue, and Hagiwara (1997) – in red.

Although Hillenbrand et al. (1995) and Hagiwara (1997) were more careful than Peterson and Barney (1952) as regards the control of dialect for vowel analysis, the first two studies investigated only the northern Midwestern, and Southern Californian dialects, respectively. With the aim of providing acoustic measurements of vowels produced by speakers from six American regions, Clopper, Pisoni, and Jong (2005) recorded the productions of 48 AE monolinguals whose ages ranged from 18 to 25 years. The participants were four women and four men from each of the following US regions: New England, Mid Atlantic, North, Midland, South, and West (see Figure 8).

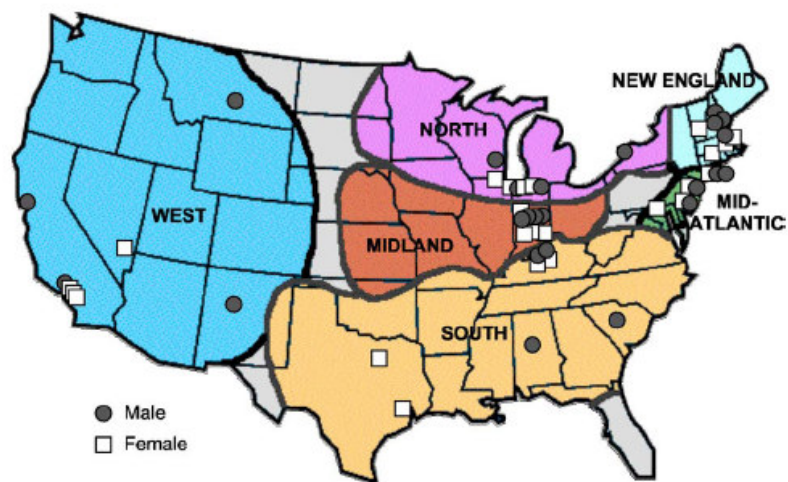


Figure 8. Six US regions investigated by Clopper et al. (2005) for an acoustic description of American English vowels (figure printed with Clopper’s permission).

Just as in the other three studies on AE vowels cited previously, Clopper et al. (2005) measured the acoustic properties of the 11 AE vowels /i, ɪ, e, ε, æ, ɑ, ɔ, ʌ, o, ʊ, u/ also inserted in the /hVd/ frame, except for /ɔ/, which was inserted in the words *frogs* and *logs* in sentence-final position. Each participant

produced a total of 56 tokens: Six tokens for the vowel /ɔ/ and five tokens for each of the other vowels. The authors show vowel plots for each region so as to illustrate the following conclusions: (i) the results of the Northern speakers show that there is a Northern Cities Chain Shift; that is, the speakers fronted /ɑ/ and raised and fronted /æ/ (see Figure 9a); (ii) the Southern speakers fronted /u/ and /o/, as in what is observed in the Southern Vowel Shift (see Figure 9b); (iii) the Midland and Western speakers also fronted /u/; (iv) the New England, Western, Mid-Atlantic, and Midland speakers partially merged /ɑ/ and /ɔ/, while the Northern and Southern speakers made a clear distinction between the two vowels; and (v) as regards vowel duration, overall the speakers from the South had significantly longer vowels than those from New England, Mid Atlantic and West, but these results do not mean that Southerners speak slower, they merely make a reduced durational distinction between lax and tense vowels.

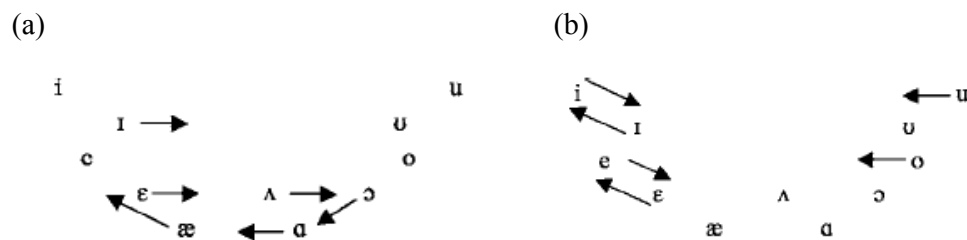


Figure 9. (a) Northern Cities Chain Shift, and (b) Southern Vowel Shift (Labov, 1998 in Clopper et al., 2005, p. 2, figures printed with Clopper's permission)¹⁹.

Although Clopper et al. (2005) do not provide a table with the raw frequencies in their article, the F1, F2 and SD values were obtained from Clopper and are shown in

¹⁹ The length of the arrow does not indicate position, just direction of the vowel change.

Tables 7 and 8. The mean F1 and F2 values of the vowels produced by the women and men from the six American regions are plotted in Figures 10 and 11, respectively.

Table 7. Women's F1 and F2 frequencies measured by Clopper et al. (2005) in six AE dialects.

	New England		Mid-Atlantic		North		Midland		South		West	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
i	370	2852	364	2885	331	2842	321	2821	379	2980	338	2994
SD	23	99	66	91	32	131	27	131	35	126	38	35
ɪ	557	2321	529	2416	497	2238	464	2258	567	2324	468	2356
SD	46	54	32	76	61	127	6	116	59	95	35	69
eɪ	490	2566	438	2757	469	2515	485	2459	581	2544	477	2708
SD	44	83	49	101	81	104	98	167	89	193	91	89
ɛ	819	2053	819	2162	815	1933	698	2076	776	2131	791	2108
SD	60	66	42	138	44	105	68	65	121	96	48	82
æ	941	1986	1022	2085	789	2132	874	1950	972	2013	984	1966
SD	77	113	37	127	70	152	115	43	95	91	62	119
ʌ	808	1580	782	1561	729	1394	713	1571	776	1764	742	1617
SD	58	73	70	113	90	33	68	115	163	124	60	87
ɑ	933	1380	1043	1548	947	1468	775	1244	939	1398	937	1292
SD	100	161	53	97	45	83	61	85	46	134	28	71
ɔ	861	1310	901	1414	835	1270	755	1229	838	1301	859	1316
SD	61	99	101	80	64	51	84	107	52	54	32	56
oʊ	581	1259	529	1259	547	1142	531	1210	628	1497	572	1325
SD	54	28	59	259	78	53	45	168	114	241	58	146
ʊ	640	1554	629	1549	550	1365	507	1472	625	1641	598	1564
SD	71	68	54	87	76	97	13	101	136	191	27	115
u	411	1275	407	1530	405	1288	406	1457	405	1586	430	1466
SD	40	93	30	376	53	264	60	145	32	340	29	317

Table 8. Men's F1 and F2 frequencies measured by Clopper et al. (2005) in six AE dialects.

	New England		Mid- Atlantic		North		Midland		South		West	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
i	299	2245	280	2292	292	2364	299	2288	276	2181	262	2251
SD	40	130	19	183	41	142	20	46	15	84	45	84
ɪ	431	1950	423	1909	451	2013	437	1948	397	1882	421	1926
SD	24	197	29	110	27	183	13	61	39	63	32	97
eɪ	413	2103	387	2138	432	2209	426	2091	422	1980	409	2095
SD	18	191	16	151	29	165	40	143	29	100	27	147
ɛ	570	1807	540	1763	607	1820	547	1797	496	1800	555	1747
SD	31	169	20	100	55	189	27	43	33	82	6	116
æ	666	1742	663	1669	639	1960	695	1773	655	1779	696	1681
SD	101	122	37	72	43	194	89	95	73	66	21	152
ʌ	591	1337	560	1209	605	1254	579	1329	546	1247	579	1307
SD	40	103	35	47	33	84	17	87	35	102	15	81
ɑ	659	1121	693	1220	798	1339	689	1155	676	1108	678	1075
SD	75	109	53	82	28	79	44	120	61	81	38	60
ɔ	612	1064	630	1175	638	1074	599	1049	574	953	637	1063
SD	89	109	27	73	16	39	24	86	47	71	31	26
oʊ	452	1074	448	1034	480	1001	469	1141	452	1177	440	1047
SD	25	109	8	69	37	85	33	105	42	170	41	223
ʊ	450	1305	460	1215	475	1180	473	1281	399	1238	432	1221
SD	20	68	39	101	23	46	22	73	26	83	33	269
u	346	1134	331	1093	334	1060	338	1246	320	1425	307	1237
SD	32	141	21	93	27	51	29	262	37	146	34	415

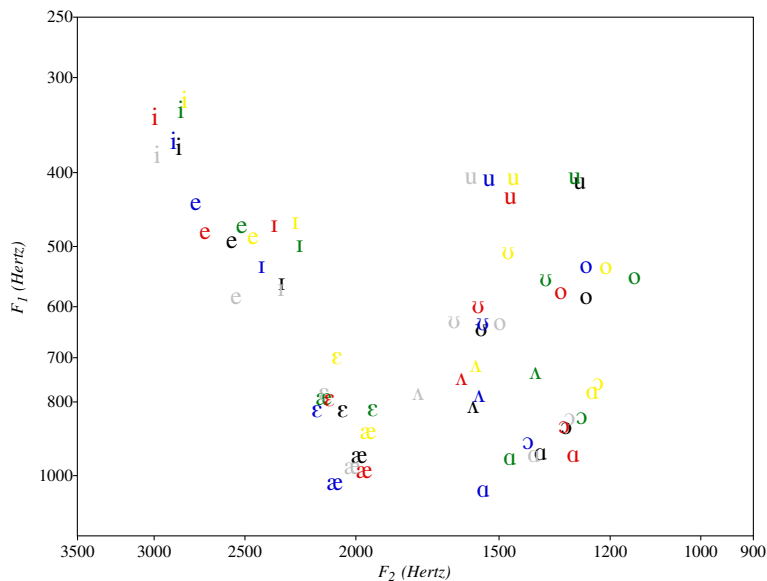


Figure 10. Women's vowel centers from Clopper et al. (2005): black – New England, blue – Mid Atlantic, green – North, yellow – Midland, silver – South, and red – West.

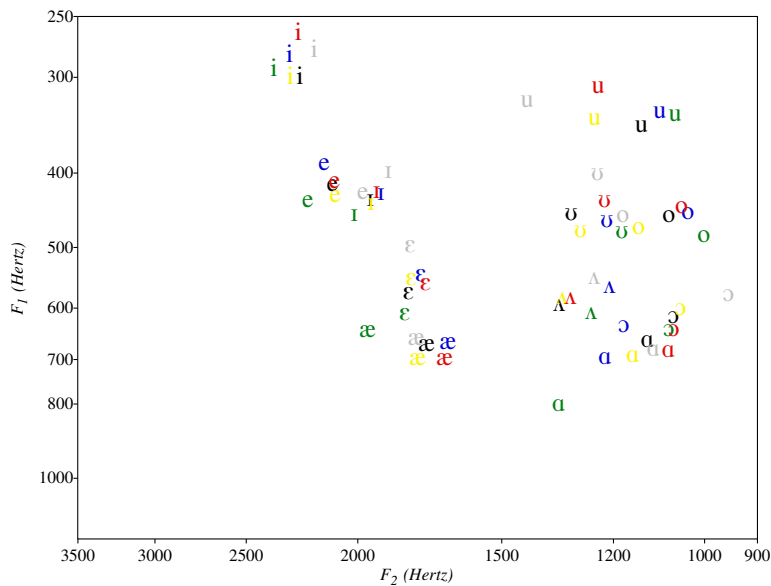


Figure 11. Men's vowel centers from Clopper et al. (2005): black – New England, blue – Mid Atlantic, green – North, yellow – Midland, silver – South, and red – West.

The formant frequency data from a wide variety of BP and AE dialects are useful in forming and testing theories on the relation between speech production and perception. The data obtained by means of monolinguals' productions serve as a good basis to investigate the acquisition²⁰ of the vowel system by L2 speakers, which is the aim of the present study. The next section will report on studies which investigated how L2 speakers produce L2 vowels.

1.2 Speech production in interphonology

Researchers in the Second Language Acquisition (SLA) field have carried out studies with different design and methods to investigate some factors which are claimed to affect the degree of L2 foreign accent (e.g., Flege, 1988, 1995; Flege, Munro, & MacKay, 1995; Meador, Flege, & MacKay, 2000; Piske, MacKay, & Flege, 2001). The most frequently investigated factors comprise age of L2 learning, length of residence (the amount of time an individual stays in a predominantly speaking L2 country, usually abbreviated as LOR), age of learning (the age when an individual is first exposed to the L2, generally abbreviated as AOL), gender, formal instruction²¹, motivation, language learning aptitude, and language use. Of all the factors just mentioned, one of the most important predictors of degree of L2 foreign accent is AOL, although the other factors also influence L2-accented speech to different degrees (Piske et al., 2001, p. 212).

²⁰ The terms *L2 acquisition* and *L2 learning* will be used interchangeably in this study, although in the Second Language Acquisition field a distinction between *acquisition* and *learning* is made, the former referring to the assimilation or subconscious learning of a second language in a natural environment, and the latter referring to the learning of a second (or foreign) language by means of formal instruction, that is, by means of consciously understanding and practicing structures and rules.

²¹ Formal instruction means that the L2 speakers learned English in a classroom, not in a natural environment, such as the environment of a predominantly L2 speaking country.

Taking into account the factors just mentioned, in the present study the participants had the following general characteristics: (i) all started learning English at puberty or later than that; (ii) neither had lived in an English speaking country for more than two months; (iii) all had received formal L2 instruction; and (iv) all had used English professionally, that is, they had taught English for at least 1.5 years. Considering the participants' characteristics and one of the aims of this study – to investigate the production of L2 vowels, the remainder of this section will focus on studies which examined how speakers from different first language backgrounds produce L2 vowels.

A number of studies have shown that the L1 vowel system directly influences the production of L2 vowels by late adolescent or early adult learners, who are called *late learners* in the literature (Flege, 1987b, 1995; Flege et al., 1997; Major, 1987). The term *late learners* implies that the learners started learning English after puberty, which is considered the period when the brain loses its plasticity, that is, when lateralization is completed, thus biologically explaining why learners who start learning an L2 at this point are more likely to speak an accented L2 (Scovel, 2000, p. 218). Although the formant frequency L2 values of late learners may be only intermediate between the values of their L1 and those of native speakers of the L2, the more exposure they have to the L2, the more likely it is that learners will produce vowels more accurately. In this respect, some findings reveal that experienced late learners (or learners with more exposure to the L2) were eventually able to accurately pronounce vowels that are located in a space in the vowel system which is unoccupied by an L1 vowel (Bohn & Flege, 1992; Flege, 1987a, 1987b; Major, 1987), a condition that will be further investigated in the present study.

With regard to vowel systems from different L1s, the differences in vowel representations in long-term memory across languages are closely related to both vowel

inventory size and structure. Three vowels were found to be present in 317 languages in Maddieson's (1984) investigation: /i/, /u/, and /a/. Moreover, two thirds of the languages in his study had five to seven different vowels. This indicates that there are preferred vowel articulations that might influence the organization of vowel representations in long-term memory in the world's languages.

As to the differences in the vowel inventories of two languages, Flege (1987b and elsewhere) claims that L2 speakers must learn how to produce L2 vowels that are either "new" with regard to their L1, that is, vowels which have no L1 counterpart and differ systematically from L1 vowels; or vowels "similar" to an existing L1 vowel, that is, vowels that differ only acoustically from an L1 counterpart.

Some evidence about the importance of exposure for the more native-like production of an L2 vowel was found in Flege (1987b). In this study, Flege examined how monolingual French speakers produced the French vowel /y/, and how several bilingual groups of native English speakers of French produced the sounds /u/ (similar) and /y/ (new). His results revealed that all three groups of native English participants, who differed considerably in amount of exposure to French, produced /y/ with formant values that approximated, to different degrees, the native French speakers' /y/, and only the least experienced group produced this sound with F2 frequency values which were significantly lower than those of the French monolinguals. As to the production of /u/, the least experienced English group produced a French /u/ that was even more anterior (closer to a French /y/) than their English /u/. This indicates that inexperienced learners confused the sounds /u/ and /y/, since they had not established a long-term memory representation for the new vowel. The other English groups produced /u/ similarly to that of their L1, and although the most experienced speakers were able to approximate somewhat the L2 norm for French /u/, their production was still English-like.

Bohn and Flege (1992) investigated native German speakers of English so that they could further examine whether amount of experience had an effect on the accuracy of adults' production of new and similar English vowels. The productions of /i/, /ɪ/, /ɛ/ (similar vowels) and /æ/ (different vowel) by a group of monolingual native English speakers and two groups of native German speakers varying in English proficiency were analyzed. The results indicate that the amount of experience did not improve the production of the similar English vowels /i/ and /ɪ/, but it affected the production of /ɛ/. The production of /ɛ/ by the group of German speakers with little English language experience did not differ acoustically from the production of native English speakers, while the group of experienced German speakers produced this sound shorter than both the native English and the inexperienced German group. These findings reveal that a similar sound is learned in the early stages of second language acquisition and does not progress much. As for the new vowel /æ/, the researchers found that only the experienced group produced this new sound with acoustic values that approximated native norms, corroborating what was found by Flege (1987b) as regards the importance of L2 experience to improve the production of a new sound.

Similar results were found by Major (1987), who investigated the production of /æ/ and /ɛ/ by Brazilian Portuguese speakers. In his study, the longer the exposure to the L2, the more accurate was the production of the new vowel /æ/. Thus, the findings in Major (1987) and in Bohn and Flege (1992), which investigated the productions of speakers from two fairly different native languages, indicate that experience may influence positively the production of the new vowel /æ/.

In a longitudinal study investigating the production of English vowels by 11 Brazilian Portuguese speakers of English living in the United States, Baptista (2000) examined to what extent L2 phonetic categories are established in long-term memory.

Concerning the acquisition of the acoustically similar vowel /ɪ/ and the acoustically new vowel /æ/, only two participants created a new long-term memory representation for /ɪ/, and only one for /æ/. To form the new category /ɪ/, the vowel /eɪ/ had to be lowered, and the acoustic vowel space had to be re-dimensioned for the establishment of the /æ/ category. Thus, in order for the interlanguage²² vowel schema to be organized, a restructuring of the schema was necessary. As to the production of /ʌ/, only three participants lowered this vowel, and they were the same participants who also lowered the neighboring vowels /æ/ and /ɑ/. These results indicate that although vowel similarity may be important for the formation of a long-term memory representation, the accuracy of this representation may depend on the neighboring interlanguage vowels and on the perception of the limits of the L2 vowel space. Thus, according to Baptista (2000), the acoustic perception and production of a vowel seems to depend on the accurate representation of the entire acoustic vowel space.

The studies described in this section provide some evidence for the hypothesis that the L1 blocks L2 phonetic acquisition, causing foreign accent, since the L2 sounds tend to be produced as the corresponding sounds of the L1. It is somewhat difficult to report on studies which investigated L2 production without mentioning the importance of perception for the acquisition of L2 sounds. Thus, the next chapter will focus on speech perception by first providing a description of how traveling sound waves are converted into neural information, then it will discuss some speech perception theories, and finally it will report on studies on interlanguage vowel perception as well as on studies on the interrelation between the perception and production of L2 vowels.

²² Interlanguage is the L2 in development.

CHAPTER 2

SPEECH PERCEPTION

In this chapter, I will briefly describe the peripheral auditory system as an introduction to the understanding of speech perception and then I will discuss two L2 speech perception theories: Flege's (1995) Speech Learning Model (SLM), a phonetically-oriented theory which has been widely used to help explain the results of many empirical L2 studies in the last decade, and Escudero's (2005) Second Language Linguistic Perception Model (L2LP), a carefully designed theory published recently which has predictions somewhat similar to those of the SLM, but has a more phonological basis. Although there are several other speech perception theories, such as Best's (1995) Perceptual Assimilation Model (PAM), Kuhl's (1991) Perceptual Magnet Effect, and Major's (2002) Ontogeny Phylogeny Model (OPM), only the SLM and the L2LP model will be considered in the present study, mainly because they are the two models which provide hypotheses about the perception of proficient L2 speakers, that is, they consider "ultimate attainment" in their models, not only beginning L2 learning phases (for a careful review of the perception models cited above see Escudero, 2005). After the description of the two models, I will report the results of some empirical studies focusing on L2 speech perception, and on the interrelation between L2 perception and production.

According to Trask (1996, p. 330), speech perception is a decoding activity which consists of the extraction of "identifiable linguistic elements from the continuous acoustic signal of speech". For this decoding activity to take place, the peripheral auditory system "translates acoustic signals into neural signals" (Johnson, 2003, p. 46).

This translation process starts when sound waves travel from the outer ear (the pinna) through the auditory canal (about 2.5 cm long, 0.65 cm wide) until reaching the eardrum, or tympanum, which is a thin membrane of skin that moves according to air pressure fluctuations. Low and high frequency sound waves produce slow and fast vibrations, respectively. The movements generated by the vibration of the eardrum travel through the air-filled middle inner ear by a chain of three tiny bones (the ossicles). These bones convert the lower-pressure eardrum sound vibrations into higher-pressure sound vibrations, which will then reach the oval window. Sound pressure amplification at this point is necessary because the membrane behind the oval window (the cochlea) is filled by a liquid (the endolymph fluid) and not by air, which makes vibration more difficult. The wave form information is then converted to nerve impulses in the cochlea, which is part of the inner ear. In the middle of the cochlea, a membrane called the basilar membrane vibrates according to the frequency received. It has two ends: (i) a narrow, thick end, close to the oval window, where high frequencies vibrate, and (ii) a wide, thin end, where low frequencies vibrate. This membrane is the base of the sensory cells of hearing, the hair cells, which are set in motion as the basilar membrane vibrates. Each hair cell, or nerve cell, has a natural sensitivity to a particular frequency of vibration; thus, when a given frequency matches the natural frequency of the nerve cell, it resonates with a larger amplitude of vibration. The larger vibration induces the cell to release an electrical impulse that is transmitted through the auditory nerve towards the brain. Finally, the brain interprets the sound received by the electric nerve impulses. How the brain interprets the nerve impulses has been debated by phoneticians and phonologists and is a topic to be discussed below. This basic explanation about the auditory system can be found in Johnson (2003), Huckvale (2006), and Larsen and Aarts (2004), for instance. The peripheral auditory system is illustrated in Figure 12.

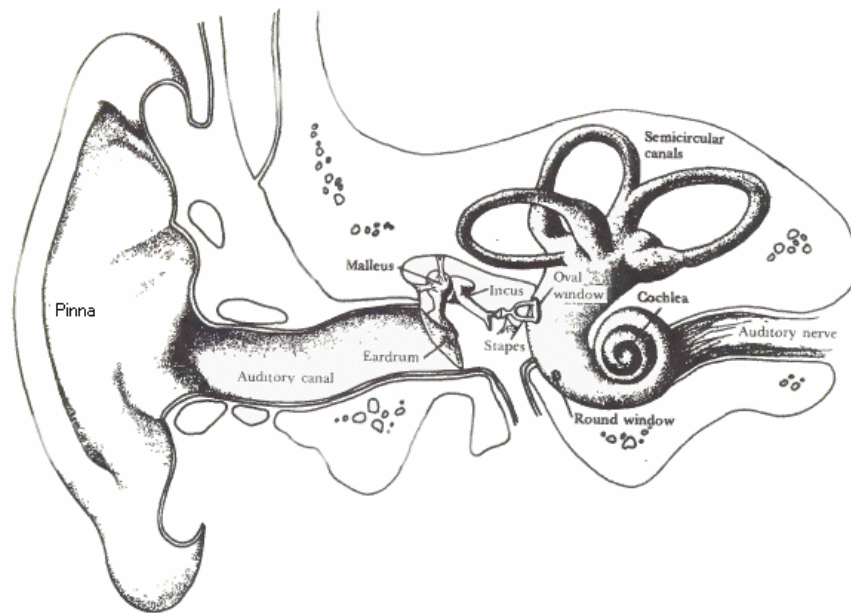


Figure 12. The peripheral auditory system (Huckvale, 2006).

As stated previously, the description of the physical speech perception process provided in this chapter is an introduction to the review of models that attempt to explain what happens after the neural networks send the perceived speech signal information to the brain. One of the interests of the present study is in second language perception; thus, the following section will report on two models that help predict and explain second language speech perception.

2.1 Two L2 speech perception models

Empirical studies on speech perception started to be published in the late 1940s, and they soon revealed that the segments present in the speech signal did not have a

simple corresponding perceptual unit, as hypothesized by phoneticians and phonologists (Strange, 1995, p. 4). The attempt to explain speech perception led researchers to establish the concept of categories, or long-term memory representations. To help explain the importance of categories in general, Boersma (1998, p. 163) states that humans make use of categories “to organize their views of the world,” thus “reducing cognitive load” and minimizing “mistakes in identifying groups of things that we had better treat in the same way”.

As regards speech perception, phonemes (abstract phonological representations) are phonetic categories, and their combination forms the words of the lexicon. When recognizing speech, “an acoustic representation is ultimately mapped to an underlying lexical form” (Boersma, 1998, p. 163). Since the speech signal is continuous, the problem concerning perception is that there is not a simple correspondence between a segment generated in production and a phoneme understood in perception (Strange, 1995, p. 5). In this sense, different segments may be categorized as the same phoneme or one segment may be categorized as different phonemes because they are produced in different contexts or because of within- and between-speaker production variability. Speech perception is thus defined as “the construction of a discrete phonological structure from raw acoustic material” (Boersma, 2000, p. 10).

Research investigating how children perceive speech, which started to emerge in the early 1970s (Polka, Jusczyk, & Rvachew, 1995, p. 49), has shown that infants as young as one month old can discriminate essentially all phonemes, not only those of their L1, although native sounds are more easily discriminated than nonnative sounds (Werker & Polka, 1993). Werker and Polka state that ease in perceiving native sounds changes significantly when children are five to six years old, and the ability to discriminate nonnative phonetic contrasts diminishes by adulthood with the more

complete acquisition of a particular native language. Boersma (1998) explains L1 acquisition by means of a phonological model of speech perception designed within the framework of Optimality Theory (Prince & Smolensky, 1993). By means of computer simulations, Boersma developed a natural learning algorithm, the Gradual Learning Algorithm (GLA), which provides a linguistic explanation of L1 perceptual acquisition. In Boersma, Escudero and Hayes (2003), the authors explain that this perceptual acquisition can be understood as a process that occurs by means of distributional learning, that is, the statistical processing that happens when infants hear the language that surrounds them. In other words, the more infants hear given acoustic properties of the speech signal, the more they will map these properties to specific categories, or the more they will turn the acoustic signal into discrete and language-dependent forms. As regards the less frequent acoustic properties heard by infants, distributional learning will allow these less frequent values to be perceived as pertaining to categories of the frequent values.

The GLA was empirically tested by Boersma et al. (2003) and consists of two stages: First the “auditory-driven” stage, when the auditory input the learner is exposed to eventually leads to the formation of categories, and then the “lexicon-driven” stage, when boundaries are shifted according to the mismatches between the perceived utterance and the lexicalized representation. An example of the “lexicon-driven” stage is the following situation: An infant or child perceives the word *pat* (/pæt/), but the word intended by an adult speaker is *pet* (/pɛt/). The English /æ/ has several acoustic cues, for example: It is a low front vowel, whose F1 and F2 values for female adults are of approximately 869 Hz and 2050 Hz, respectively. The vowel /ɛ/ has different acoustic cues: It is a mid-low front vowel, whose F1 and F2 values for female adults are of approximately 610 Hz and 2330 Hz, respectively (values by Peterson and Barney, 1952).

The meaning of the word which will be then understood because of the context where it was inserted will allow the child to notice that the correct perception should have been the mid-low vowel. This noticing will help the infant/child to solve the mismatch between the intended production and the actual perception by means of the reranking of cue constraints in the perception grammar²³, which in optimality theory terms means that the infant/child will lower the constraints against perceiving the F1 value as /ɛ/ (e.g., an F1 of 600Hz is not /ɛ/) and will raise the constraints against perceiving it as /æ/ (e.g., an F1 of 600Hz is not /æ/]. It is this reranking of constraints in the infant/child perception grammar that leads him/her to shift category boundaries. Thus, the simulated “baby” (the GLA) gradually learns his/her first language by exposure to “(i) acoustic events in the linguistic input, which give birth to ‘phonetic’ categories; and (ii) lexical representations, which lead to the development of ‘phonological’ categories” (Boersma et al., 2003, p. 1013).

Concerning second language acquisition, while children are able to modify their perception of nonnative contrasts, adults tend to rely on their native-language acoustic parameters. Thus, it can be said that although the speech signal is processed by general neural networks, as explained in the introduction to this chapter, and these networks are responsible for mapping the signal onto specific perceptual categories, no matter what the language, these “neural networks are trained on language-specific stimuli and therefore result in language-specific processing” (Escudero, 2005, p. 31).

Studies showing the influence of the L1 system on L2 perception led cross-linguistic speech perception researchers to characterize adult L2 learners as having a “perceptual foreign accent” (Strange, 1995, pp. 22, 39), a perceptual correspondent to the concept of foreign accent, that is, adults perceive L2 contrasts which do not exist in

²³ The term “perception grammar” was used by Boersma (1998) to explain speech perception as linguistic knowledge.

their L1 phonological space by relying on L1 acoustic parameters. As regards existing and non-existing contrasts in the L1, particularly *vowel* contrasts, since the speakers have already established patterns in their L1 for a vowel that has a similar L2 counterpart, there is a challenge to modify such established patterns so as to perceive the similar vowel in a native-like fashion. This difficulty had already been pointed out by Trubetzkoy (1939/1969), who hypothesized that L1 phonology causes L2 learners to “filter out” perceptual acoustic differences that are not relevant in the phonology of the L1. In other words, according to Flege (1987a), the “filter” would make similar sounds more difficult to be perceived by L2 learners due to a cognitive mechanism called “equivalence classification”. Flege states that this mechanism, which is very helpful for L1 learning because it allows children to identify sounds produced in different contexts or by different speakers as pertaining to the same category, might hinder L2 speakers’ formation of categories for similar sounds. Because of the equivalence classification mechanism, L2 learners may perceive an L2 category and an L1 category as sufficiently similar to consider them equivalent. In this respect, the amount of exposure to the L2 has been shown to improve L2 vowel quality to acoustic patterns similar to those of L2 native speakers.

Flege (1987a, 1991, 1995, 1996) makes use of the terms “new”, “similar”, and “identical” to better explain L1 influence in the *perception* of L2 sounds. Flege suggests three criteria to determine which sounds are new, similar or identical. The first criterion is the phonetic symbol itself. If the L1 and L2 sounds are represented by the same International Phonetic Alphabet (IPA) symbol, the L2 sound is supposed to be either identical or similar to an L1 sound.

The second criterion is acoustic difference. Considering that an L1 sound and an L2 sound are represented by the same IPA symbol, (i) if the acoustic properties of the

L2 sound do not differ significantly from those of the L1 sound, the L2 sound is considered to be identical; (ii) if the acoustic properties differ significantly, the L2 sound is considered to be similar. An L2 sound which does not resemble any L1 sound is considered to be new.

The third criterion is native listeners' judgments: The L2 sound will be considered (i) identical if the native listeners do not perceive any difference between the L1 and L2 sound; (ii) similar if they are able to discriminate the L2 from the L1 sound; and (iii) new if they recognize the L2 sound as not belonging to the L1 system (Flege, 1991).

Considering this conceptualization of what is "similar" and "new", Flege (1995) proposes the Speech Learning Model (SLM) as an attempt to explain that the lack of ability to create new categories for sounds perceived as similar is based on the fact that L2 speakers are not able to perceive the phonetic features in which the L2 and L1 sounds differ because they are perceptually equivalent. In the case of vowels, examples of such features are spectral cues and durational cues. For instance, speakers of a given language may rely primarily on spectral cues to perceive vocalic sounds, while L2 speakers of this language may rely on duration primarily and spectral cues only secondarily to perceive the vowels of this language. Several empirical studies which tested the importance of acoustic cues in L2 speech perception will be reported in the next section.

The SLM also hypothesizes about the influence of age of learning on L2 speech perception. Flege (1996) considers an *early* L2 learner someone who is exposed to the L2 before the age of 5-6 years, whereas *late* learners are those who have contact with the L2 later than that, after the onset of reading. The difference in amount of foreign accent between early and late learners is related to the interaction between L1 and L2

systems: Given that L1 and L2 sounds co-exist in the phonological space, the longer and more extensive experience in a predominantly (or unique) L1 context, the later learners' perception and production will be adjusted to their L1. However, the model predicts that late learners who have *extensive* L2 experience in a naturalistic environment may be able to perceive phonetic features, and also to modify the articulation of L2 sounds which had been previously established for the L1. This claim goes against the view that humans lose the motoric ability to produce sounds not found in the L1 after a certain age. The SLM's assumption is that only by perceiving sub-phonemic features can L2 speakers form a new category and achieve completely native-like perception, provided that the L2 sound is phonetically dissimilar from the closest L1 sound.

It is important to stress that the SLM is intended to predict the speech perception of experienced L2 speakers, that is, the model is not concerned with the pronunciation of beginners. Since "ultimate attainment" is the focus of the SLM, the model may be a useful tool to help explain the findings obtained with the L2 participants of the present study, who are all proficient L2 speakers. However, the participants had formal English instruction, a variable not predicted by the SLM, which was designed to explain L2 learning in natural settings, that is, in the country where the L2 is spoken as the native language.

What Flege (1995) calls "ultimate attainment" is called by Escudero (2005) the "end state" of L2 acquisition. Flege's SLM has a phonetic approach while Escudero's Second Language Linguistic Perception (L2LP) model has a phonological, phonetic and psycholinguistic approach to explain L2 sound acquisition, since the latter deals with phonetic constraints and models the process of mapping acoustic/auditory events onto categories. The phonetic/phonological distinction, briefly explained by Escudero (2005, pp. 128-129), basically implies that the main concern of a phonetic approach is not the

abstract systematizations that take place in L2 acquisition, but the properties of the acoustic signal and their relation to phonetic categories. This means that such an approach does not explain *how* the connection between the auditory input and the categories takes place, that is, how phonological categorization of phonetic input happens. The phonological approach, on the other hand, seeks to explain the acquisition of L2 sounds by considering learners' formal knowledge, or their "system of structures (rules, features, hierarchies, or constraints) that is represented in learners' minds" (Escudero, 2005, p. 128). Thus, the phonetic approach makes use of the term phonetic categories to mean "position-dependent" allophones (Flege, 1995), while the phonological approach considers "'distinctive' segments or phonological features to be the units of analysis for describing phonological systems" (Escudero, 2005, p. 128). The main differences between the two models will be pointed out after Escudero's model is presented.

Escudero's L2LP model describes three L2 speech perception phases: the initial state, development and the end state. In order to explain the acquisition process in each phase the model is composed of five ingredients, which are considered to be both a theoretical (it presents the theory adopted to predict and explain the acquisition of L2 sound perception) and a methodological (it provides a methodology guide to test the models' predictions) framework:

- Ingredient 1: *optimal L1 and optimal L2 target perception* – the description of the optimal perception of the sounds of the languages being investigated is important to predict the difficulties L2 speakers will have when learning the L2. Native speakers of a given language are the optimal perceivers of this language, thus the knowledge about the optimal perception of native speakers of the L2 provides information as regards how the L2 learner's perception needs to improve to become optimal, that is, what acoustic

cues (e.g., duration, spectral quality) are used by native L2 speakers, and which are used by L2 learners to perceive sounds. By optimal perception the model means the perception grammar that has been formed according to the acoustic properties of the production environment the listener is exposed to. This perception grammar implements the hypothesis of an optimal perception, which suggests that an optimal listener will construct the sounds (vowels and consonants) that most closely approximate what has been intended to be pronounced by the speaker. The model also makes a distinction between perceptual mappings and phonological categories. While mappings are *performed* by the perception grammar, categories are *constructed* by the perception grammar. Figure 13 illustrates what is meant by perceptual mappings, that is, the mapping of the raw acoustic signal onto linguistic units (vowels and consonants):

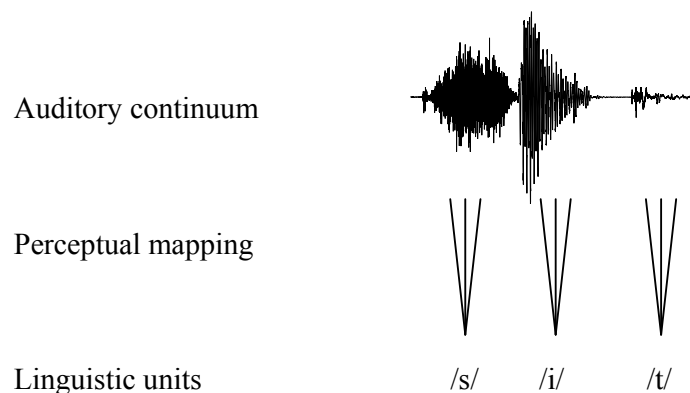


Figure 13. The mapping of the continuum speech signal onto linguistic units (the waveform represents the word *seat* pronounced by a male native speaker of AE).

- Ingredient 2: *L2 initial state* – if there is information about the cues used by the optimal listener to perceive sounds of his native language, it is possible to predict that the L2 initial state corresponds to the L1 optimal state. This means to say that L2

learners make use of categories, or perceptual mappings, established by their L1 to perceive the sounds of any foreign language, which is also hypothesized by the SLM. The L2LP model calls this initial L2 state “full copying”; that is, the learner transfers, “copies” or “duplicates” his L1 perception, considering it the starting point for L2 perception, since no previous knowledge of the L2 exists at this phase (see Figure 14 for an illustrative schema).

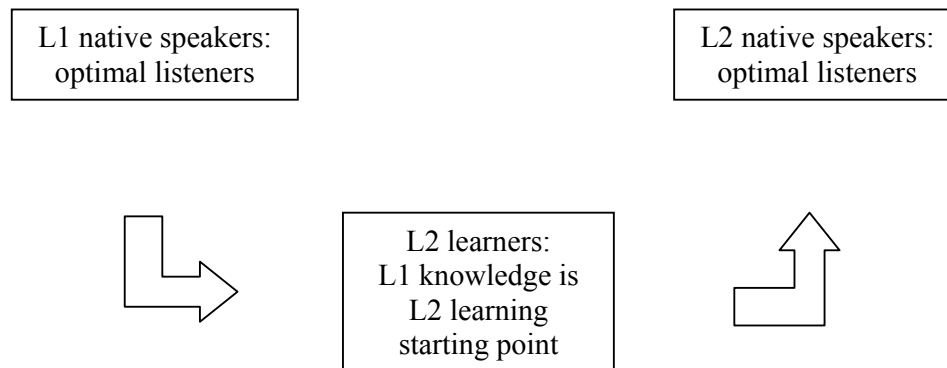


Figure 14. Native language perception optimal states and L2 learning onset.

In the L2LP model, at least three “scenarios” may occur at this point: (i) the “new” scenario, when the L2 has more categories than the L1, thus one L1 sound will be used to perceive two or more L2 sounds, (ii) the “similar” scenario, when the L2 has the same number of categories as the L1, thus only an adjustment of categories is predicted; or (iii) the “subset” scenario, when the L2 has fewer categories than the L1, thus L1 categories would form a subset of categories to perceive L2 sounds. One example of the subset scenario is a native speaker of Dutch learning Spanish: Dutch (the L1) has 12 vowels, while Spanish (the L2) has only five. Thus, two or even three L1 vowel

categories are available to perceive only one L2 vowel category (see Figure 15 for an illustration). The “subset” scenario has not been predicted by other L2 speech perception theories;

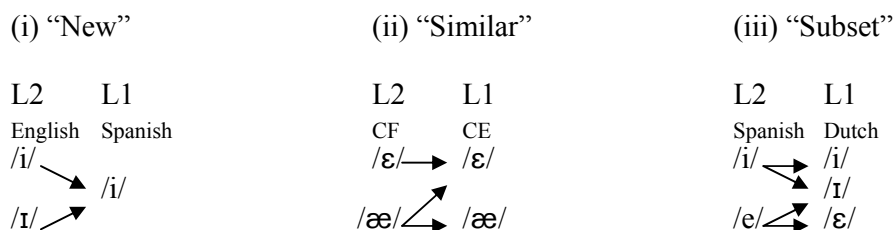


Figure 15. Three possible scenarios for L2 speech perception, according to Escudero (2005, p. 124). CF = Canadian French and CE = Canadian English.

- Ingredient 3: *the L2 learning task* – once the L2 initial state and the L2 optimal perception are known, it is possible to observe the differences in terms of phonetic categories of the L1 and the L2, or the “degree of mismatches” between the two perception grammars. As soon as the perceptual mismatches are identified, two types of tasks may be involved in the L2 perception learning process: a “perceptual task” and a “representational task”. If the mismatch is large the L2 learning task will involve both a perceptual and a representational task for the creation and integration of categories (“new” scenario), whereas when the mismatch is small, only a perceptual task will be involved for the adjustment of category boundaries. A perceptual task alone is expected to occur when the number of L1 and L2 categories are the same (“similar” scenario), or when the L1 has more categories than the L2 (“subset” scenario);
- Ingredient 4: *L2 development* – in order to create new L2 categories or adjust the already established ones, the L2LP model hypothesizes that the same process used to

acquire L1 perception will be adopted to acquire the L2. As stated earlier, infants perceive either sounds that exist in their native language or nonnative sounds, although there is an advantage in perceiving the former sounds. With the exposure to a given language, infants start tuning their perception to this specific language, since this is the constant input received. The L2LP model hypothesizes that a similar process takes place in L2 perception: The L2 learner gradually adjusts his perceptual grammar so as to match the L2 perception common to optimal listeners. In other words, the L2 learner creates new categories or adjusts his category boundaries similarly to infants who do so when perceiving L1 sounds;

- Ingredient 5: *the L2 end state* – the L2LP model hypothesizes that in order for the L2 not to influence L1 categorization, that is, in order for L1 and L2 perception to be optimal, the two need to be different systems. This means to say that if enough optimal L1 and L2 input is provided to similar extents, both L1 and L2 perception will remain optimal, that is, L2 perception will develop but it will not affect L1 optimal perception. The model also claims that if there is an intermediate L1-L2 perception, it will be because both perceptions will be activated simultaneously.

As regards the different learning phases, the two models differ in the following respects:

- 1) *L2 initial state*: while the L2LP model suggests that there is full copying of the L1's perception grammar and categories to his L2 perception grammar and categories when he starts learning an L2, the SLM does not hypothesize about beginning learners, but it considers that the L1 and L2 categories co-exist in a single phonological space;
- 2) *L2 development*: although the SLM focuses on *ultimate attainment*, adult L2 learners' development is supposed to take place due to learners' capacity to acquire the L2 similarly to the way infants acquire their L1. The SLM suggests that the formation of

L2 categories is more likely to take place the more the perceived L2 category differs from the closest L1 sound. On the other hand, the L2LP model suggests that the L2 learner will either create new perceptual mappings or adjust the already existing ones so as to create new phonological representations (categories). The hypothesis is that, in the L2 acquisition process, the L2 learner has the same access to the Gradual Learning Algorithm that led him to acquire his L1 perception; thus, the L2 learner adjusts his L2 perception to that of optimal L2 listeners, forming categories in dimensions in the phonological space that were never used to classify sounds in his L1.

3) *L2 end state*: the SLM claims that L1 phonetic categories will limit the possibility of L2 category formation because L1 and L2 sounds coexist in a single phonological space, and, as stated previously, L2 sounds are “filtered” through the learner’s L1 sounds. Category formation will be hindered by the mechanism of equivalence classification, which allows the establishment of additional categories for “new” sounds, but not for “similar” sounds. The SLM also predicts that the later in life learners start learning the L2, the less likely they will have L2 native-like perception. On the other hand, the L2LP model suggests that learners have two separate perceptual grammars, which allow L2 learners to create any new phonetic category by adjusting existing mappings in a process similar to the one which takes place during L1 acquisition, by means of distributional learning. However, for the two perceptual grammars to continue being activated learners must be exposed to both L1 and L2 to similar extents.

Second language perception theories are a good and necessary starting point for empirical research. The next section aims at reporting studies based on the SLM or the L2LP model so as to provide some examples of empirical research that corroborates many of the two models’ claims.

2.2 L2 perception studies

The aim of this section is not to provide a lengthy review of studies on L2 speech perception, but some examples of important empirical research that entirely or partially support the hypotheses of the two perception models reviewed in Section 2.1, the SLM and the L2LP model.

Several studies have found that L2 learners may use different “weightings” of acoustic parameters to differentiate what Flege calls “new” and “similar” sounds (Busà, 1992; Bohn, 2005; Bohn & Flege, 1992; Escudero, 2001, 2002; Escudero & Polka, 2003; Morrison, 2002).

Bohn (1995) investigated the perception of the English contrast /i/-/ɪ/ by native speakers of German, and the English contrast /ɛ/-/æ/ by native speakers of Mandarin and Spanish. A monolingual native English control group was also tested. The results show that the English speakers relied primarily on spectral quality to identify the vowels of the target contrasts. Conversely, the German participants relied both on spectral quality and duration to identify the vowels in the /ɛ/-/æ/ contrast, probably because vowels differ in both spectrum and duration in their native language; however, duration was much more frequently used by native Germans than by native English speakers to identify /ɛ/-/æ/. Bohn expected that Spanish and Mandarin speakers would rely on spectral cues primarily since these languages do not use vowel duration contrastively. Nevertheless, to identify the vowels of the /i/-/ɪ/ contrast, Spanish speakers made use of durational cues predominantly, and Mandarin speakers relied almost exclusively on vowel duration. The findings indicate that, differently from the English speakers, neither Spanish nor Mandarin speakers were able to use spectral cues to differentiate /i/ from /ɪ/. Bohn explains these findings by means of the “Desensitization Hypothesis”,

which suggests that whenever spectral cues are not sufficient for L2 speakers to differentiate between the vowels of a contrast, that is, whenever previous linguistic experience is not enough to sensitize learners about spectral differences, durational cues will be used so that L2 speakers differentiate the two vowels of an L2 contrast.

Similar results concerning acoustic cues were found for the production of German vowels. Bohn and Flege (1992) analyzed German speakers' production of the "similar" English vowels /i, ɪ, ε/ and the "new" vowel /æ/ and found that, differently from native English speakers, their participants relied more on durational cues than on spectral cues to produce the /i/-ɪ/ contrast. As for the /ε/-æ/ contrast, the inexperienced participants did not differentiate between the vowels in this pair, while the experienced participants did establish an L2 duration contrast between these two vowels and their spectral quality did not differ significantly from that of English vowels. Moreover, Busà (1992) analyzed the production of English /ʊ/ and /u/ by Italian speakers and also found that the Italians tended to rely on duration rather than on spectral quality of the sounds to produce the different vowels.

The same tendency occurred in Escudero's perception studies. In Escudero (2001, 2002), L2 Spanish speakers of Scottish English relied only or primarily on durational cues to perceive the Scottish English contrast /i/-ɪ/, differently from native Scottish speakers, who relied primarily on spectral cues. Escudero and Polka (2003) investigated the perception of Canadian French vowels by Canadian English listeners and found further evidence of the use of L1 cue weighting to discriminate L2 vowels. Canadian French speakers rely only on spectral cues; however, the authors found that the English listeners relied both on spectral cues and durational cues, the latter being a secondary cue. This strategy is used to identify L1 vowels, since English speakers also rely on duration to differentiate some vowel pairs.

Similar results were found in a longitudinal study by Morrison (2002), who investigated the perception of the English contrast /i/-/ɪ/ by Japanese and Spanish listeners. The participants were tested at two points in time: One month after their arrival in Canada, and again five months later. His findings revealed that the Japanese listeners relied primarily on duration to identify the vowel contrast at the two instances they were tested. As for the Spanish listeners, they also relied on durational cues in the initial test; however, one of the four participants managed to have a categorical boundary based on spectral cues in the second test. These results indicate that the Japanese listeners' increased exposure to English did not change their perceptual cue reliance; conversely, exposure to English did change one of the Spanish listeners' perception of the target contrast, indicating that this listener was able to establish a new category for the English /ɪ/. Due to the importance of cue reliance to help understand the difficulties in accurately perceiving and producing L2 sounds, the present study will also examine spectral and durational cue weightings.

The L2LP model explains the results of the studies just reviewed as the need L2 learners have to develop cue integration. All the instances when the L2 learners relied primarily on temporal cues instead of spectral cues to perceive the distinction between the target vowel pairs show evidence that in order for constraint reranking to occur, an integration between F1 and duration values needs to be established for all the L2 vowel categories, which can only be achieved by extensive L2 exposure, since L2 development is predicted to take place by means of lexicon driven or/and distributional learning. Flege's (1995) SLM also explains that L2 experience allows learners to perceive L2 vowels more accurately, although L2 accented input received due to contact with other nonnative speakers may influence L2 learners' performance.

Besides L2 experience, as stated in Section 1.2, age of L2 acquisition is another important factor that influences L2 speech perception/production. The results of a study by Flege, MacKay and Meador (1999) provide further evidence for the hypothesis that early bilinguals are able to establish new phonetic categories for some L2 vowels. These researchers examined the perception and production of English vowels by a group of 72 highly experienced native Italian speakers of English, taking into account the age of both their arrival in Canada and the beginning of their English studies. The findings revealed that accuracy in producing and perceiving English vowels was related to the age of first exposure to English, since “the early Italian/English bilinguals produced English vowels in a native-like fashion” (p. 2982).

In order to understand L2 acquisition, several studies carried out both perception and production experiments, which allowed them to discuss the relationship between the two abilities. Some relevant findings obtained by means of this type of study will be reported in the next section.

2.3 Studies on the interrelation between perception and production

As observed in both the SLM and the L2LP model, the difficulty to perceive L2 sounds that differ slightly or considerably from L1 sounds is one of the explanations for an accented pronunciation. Some evidence of the interrelation between perception and production was found by Rochet (1995). In his study, the participants were speakers of two different native languages (Portuguese and English) and were presented with French synthetic vowels. The two groups of speakers perceived and produced the same L2 sound (French /y/) differently: While the native Portuguese speakers of French

tended to perceive and imitate the vowel /y/ as /i/, the native English speakers of French tended to perceive and imitate the vowel /y/ as /u/.

Flege et al. (1997) also found some relationship between perception and production. The researchers tested the role of L2 experience in the perception and production of the English vowels /i, ɪ, ε, æ/ by native speakers of German, Spanish, Korean and Mandarin. In the perception experiment using synthetic stimuli, the results show that many participants relied on durational cues to identify the /i/-/ɪ/ and /ε/-/æ/ continua, while the native English participants relied solely or primarily on spectral cues to identify the contrasts. The amount of experience influenced the way English vowels were both perceived and produced, but differences between the participants' improvement in performance as a result of L2 experience depended on the L1, apparently because of differences in the perceived relation between the L1 and L2 vowels. In the perception test, the experienced L2 participants identified the members of the /ε/-/æ/ and /i/-/ɪ/ continua by making more use of spectral cues than did the inexperienced L2 participants. As for the production test, the more experienced participants produced English vowels more accurately than the inexperienced participants from the same L1 background. These findings reveal that L2 experience influences the way L2 speakers both perceive and produce L2 sounds, corroborating the SLM and the L2LP model's hypothesis that adults do not lose perception and production abilities. One of the conditions for the acquisition of L2 sounds to take place is extensive L2 exposure so that, according to the L2LP, lexicon-driven learning reranks cue constraints leading to optimal perception and eventually production.

Similar findings as regards L2 experience were obtained by Flege et al. (1999), who tested the discrimination of English vowels by experienced native Italian speakers of English living in Canada. Native English listeners evaluated the participants'

production through goodness ratings and forced-choice identifications. Their perception was assessed through a categorical discrimination test. The results show that age of arrival in Canada exerted an influence in both production and perception. The later the participants arrived in Canada, the less accurately they produced and perceived English vowels. As in Flege et al. (1997), some correlation was found between the production and perception scores: The higher the discrimination score, the more accurately the vowels were produced.

More evidence of the perception/production relationship was found by Bradlow (1996), who examined native American English and native Madrid Spanish speakers' production and perception of the /i/-/e/ and /u/-/o/ contrasts. The author investigated universal and language-specific aspects of the two contrasts by comparing the acoustic and perceptual characteristics of the contrasts within each of the two languages and also across languages. The results of the discrimination of the synthesized contrasts reveal that the L1 vowel categories were the reference in the identification of the acoustic stimuli, which means that the L1 vowel system had an effect on L2 perception. However, some flexibility of the perceptual vowel space was observed when the listeners were listening to stimuli located around nonnative vowel categories, which indicates that L2 listeners may "adjust their perceptual vowel system to match the stimuli" (Bradlow, 1996, p. 55).

In the specific case of Brazilian Portuguese- (BP) speaking learners of English as a foreign language, two studies on vowel perception and production show further evidence of the interrelation between the two abilities. The study by Rauber et al. (2005) investigated 16 advanced adult BP speakers of American English by means of a production test, which measured the participants' F1 and F2 values, and a categorical perception test, based on Flege, Munro, and Fox (1994). The perception test had an

oddy discrimination format and consisted of the presentation of three target tokens each inserted in the carrier sentence “This is a bVt”. The vowels tested in both perception and production tests were /i, ɪ, eɪ, ε, æ, ʌ, ɑ, ɔ, oʊ, ʊ, u/. The vowels /ɪ, æ, ʌ, ɑ, ʊ/ do not exist in the BP vowel inventory, and, with the exception of /ɪ/ and /ʌ/, they tended to be produced with the formant values of their L1 counterparts /ɔ, u/ or, in the case of /æ/, it tended to be produced with an F1 value which was higher than that of native speakers, that is, its production was neither native nor nonnative. As regards the perception test, the same vowels which were poorly produced (/æ, ɑ, ʊ) were also poorly discriminated.

The study by Bion et al. (2006) also measured the first two formants of vowels produced by BP speakers of American English, but the authors focused only on the front vowels (/i, ɪ, ε, æ/). To test perception, besides the same categorical test adopted by Rauber et al. (2005), synthesized front vowels were presented to 17 proficient EFL participants. As the focus of the experiment was to test the role of spectral quality in vowel perception, two speech continua, one for the /i/-/ɪ/ and one for the /ε/-/æ/ contrast, were created with F1 and F2 values modified in nine steps, while duration was kept constant. The results of the two perception tests revealed that the /i/-/ɪ/ contrast was more easily discriminated than the /ε/-/æ/ contrast, corroborating the results in Rauber et al. concerning front vowels. Seven out of the 17 participants did not discriminate the /ε/-/æ/ contrast at all, while only four participants did not make any discrimination between the /i/-/ɪ/ vowels, and those who did discriminate the vowel pairs did not do it in a native-like way; that is, they needed a greater distance between the two vowels of a contrast to make a distinction between them. The findings concerning the production test also corroborated Rauber et al.’s in that the participants produced a larger distance between /i/-/ɪ/ than between /ε/-/æ/. The results provide evidence that perception and

production are related, since greater discrimination in the perception test was related to better production results.

Flege (1995) also observed that, although adults are able to imitate L2 sounds, they tend not to produce them with native-like acoustic properties not only because they may perceive these sounds differently, but also because of some motoric output constraints, such as new articulatory patterns, or some phonological constraints, such as the permissible syllable types. Various studies on BP speakers of English show evidence of the influence of the L1 syllable structure on L2 production (e.g., Baptista & Silva Filho, 2006; Cornelian Jr., 2003; Koerich, 2002, 2006; Rauber, 2006; Rebello & Baptista, 2006). In these studies, the findings reveal that foreign language speakers whose native language has less marked syllable structures than those of the L2 may make use of some inappropriate strategies to produce particular syllables, such as addition of an extra vowel to the syllable onset or coda.

Thus, effects of age, motoric constraints, equivalence classification, different category boundaries across languages, syllable structure, and cue reliance are some factors that help explain learners' difficulty in perceiving L2 sounds accurately. The studies reviewed in Chapters 1 and 2 may stimulate future research on the perception and production of vowels by speakers of languages that have been little investigated. Moreover, more attention needs to be given to speakers who learn a second language in formal settings, which is the case of the present study.

The next chapter will provide information about the method adopted to collect and analyze the data so as to answer the research questions concerning vowel perception and production.

CHAPTER 3

METHOD

In order to investigate the perception and production of English vowels by advanced Brazilian EFL speakers, two experiments were carried out: One to measure the acoustic properties of the vowels produced by the participants, and the other to test the participants' vowel perception. The two experiments follow the same instrumentation and data collection procedures designed by Dr. Paola Escudero, of the University of Amsterdam, to test speech perception and production in her post-doctoral project entitled *Explaining L2 speech perception*. The stimuli used in the perception test are the same as those of Escudero's project. Information about the two experiments is provided in the following sections.

3.1 Vowel production participants

The productions of three groups of speakers were tested: American English monolinguals, Brazilian Portuguese monolinguals, and Brazilian EFL speakers. None of the participants were paid. The monolingual participants were selected after having answered a questionnaire about their background (see Appendixes C and D to check the Portuguese and English questionnaires, respectively). The questionnaires allowed me to make sure that the participants neither spoke any language other than BP or English nor

had contact with speakers of other languages in their daily routines. The following subsections will provide descriptions of the participants from each group.

3.1.1 American English (AE) monolingual participants

Ten Americans, five women and five men, were recorded. However, due to discrepancies between the formant values of the vowels by one of the women (a 51-year-old lady), the productions of only four of them were analyzed. The women's ages ranged from 25 to 44 years (mean = 33.5 years), and the men's ages ranged from 18 to 36 years (mean = 26.6 years). Only one of the participants was not born in the state of California. However, all the speakers had spent most of their lives in the city of Sacramento, the capital of California. The recordings were made by a volunteer native speaker of AE from Sacramento, who was trained to follow the data collection procedures described below. The choice for Sacramento was because of availability of participants. All participants reported having no knowledge of any language other than English, nor did they have any contact with speakers of other foreign languages in their daily routines. More detailed information about each participant's background can be seen in Table 9.

Table 9. American English participants' background.

Part.	Gender	Age	Place of birth	Place where spent most of life	Occupation
1	F	25	Sacramento-CA	Sacramento-CA	BA student: Geography
2	F	38	San Francisco-CA	Sacramento-CA	School teacher
3	F	27	Sacramento-CA	Sacramento-CA	BA student: History
4	F	44	Sacramento-CA	Sacramento-CA	Medical assistant
5	M	28	Sacramento-CA	Sacramento-CA	BA student: History
6	M	26	Salt Lake City-UT	Sacramento-CA	BA student: International Business
7	M	36	Berkeley-CA	Sacramento-CA	Social Science teacher
8	M	25	Berkeley-CA	Sacramento-CA	BA student: Computer Science
9	M	18	Sacramento-CA	Sacramento-CA	High school

3.1.2 Brazilian Portuguese (BP) monolingual participants

Twelve monolingual speakers of BP, six women and six men, were recorded, four from each of the Brazilian southern states Rio Grande do Sul (RS), Santa Catarina (SC) and Paraná (PR). The women's ages ranged from 20 to 31 years (mean = 27 years), and the men's ages ranged from 20 to 36 years (mean = 26 years). The speakers were from the following cities: Rio Grande-RS, Santa Maria-RS, Chapecó-SC, and Cascavel-PR. The choice of cities was to ensure that the vowels produced by the participants would be comparable to the L1 vowels of the English L2 speakers, all of whom were from non-capital cities of the three southern states, of similar location and size. In 2005, the estimated number of inhabitants in each city was: Rio Grande: 195,392; Santa Maria: 266,042; Chapecó: 169,256; and Cascavel: 278,185 (IBGE, 2005). None of the participants spoke any foreign language and all reported that they had had no contact

with speakers of other foreign languages in their daily routines. Table 10 shows each BP participant's background.

Table 10. Brazilian Portuguese female and male participants' background.

Part.	Gender	Age	Origin*	Education	Occupation
1	F	20	Cascavel-PR	BA student: Administration	Saleswoman
2	F	31	Cascavel-PR	BA student: History	Medical assistant
3	F	31	Rio Grande-RS	MA student: Portuguese	Portuguese teacher
4	F	25	Santa Maria-RS	BA student: Speech Therapy	Speech therapist
5	F	23	Chapecó-SC	BA student: Administration	Administrative assistant
6	F	29	Chapecó-SC	BA student: Administration	Administrative assistant
7	M	29	Cascavel-PR	BA student: Geography	Salesman
8	M	20	Cascavel-PR	BA student: Tourism	Tour agent
9	M	31	Rio Grande-RS	MA student: Portuguese	Portuguese teacher
10	M	36	Rio Grande-RS	BA: Journalism	Businessman
11	M	20	Chapecó-SC	BA student: Public Administration	Student
12	M	20	Chapecó-SC	BA student: Law	Administrative assistant

* Place of birth and where spent most of life.

3.1.3 L2 participants' (BP speakers of English as a foreign language) production

The L2 group consisted of 18 Brazilians: 11 women and 7 men. The women's ages ranged from 22 to 47 years (mean = 32.6 years), and the men's ages ranged from

26 to 41 years (mean = 32 years). The participants chosen for this group had spent no more than 8 weeks in an English speaking country and had had between one and fifteen years' experience of teaching English (mean = 8.1 years). All of them were late learners²⁴ of English and had AE as their target English variety. Most of them (14) were enrolled in the Graduate Program in English (PGI) at the Federal University of Santa Catarina (UFSC), which means to say that they were all highly proficient in English, since in order to enter the program they are interviewed in English and are also required to take exams that test both their linguistic/literary knowledge and their English proficiency. In order to pass the exams and the interview, their proficiency should be equivalent to a score of at least 550 points on the Test of English as a Foreign Language (TOEFL) test. All the doctoral students were from the PGI. As for the masters (MA) students, only Participant 12 was from the masters program in Linguistics of the Federal University of Passo Fundo-RS; all the other MA students were also from the PGI. In order to have a larger group of L2 speakers, 3 participants who had completed a Specialization (SP) course²⁵ in English at the University of the West of Santa Catarina (UNOESC) were recorded. The choice of UNOESC was the availability of participants and the easy access to them. The UNOESC students were selected on the basis of their English proficiency, all had scored at least 550 points at the TOEFL test by the time of the recordings. All 18 participants had taken an introductory course in English Phonetics/Phonology on their DO, MA or SP programs by the time of the recordings. Tables 11 and 12 show the female and male participants' background information, respectively.

²⁴ All the participants reported to have started learning English after the onset of reading, that is, later than 6-7 years old (for a more detailed explanation of the concept of *late* learners, see Section 1.2).

²⁵ A graduate-level course between a bachelor's degree and an MA.

Table 11. L2 female speakers' background.

Partic.	Age	City of origin	Education	Place and time abroad	English teaching experience
1	22	Ijuí-RS	MA Language	London-England, 4 weeks	4 years, language schools
2	28	Videira-SC	MA Language	Miami-USA, 8 weeks	7.5 years, language schools
3	45	Chopinzinho-PR	DO Language	--	15 years, public and private schools
4	25	Xanxerê-SC	SP Language	London, England, 3 weeks	6 years, language schools
5	32	Chapecó-SC	MA Language	--	1 year, private teacher
6	28	Brusque-SC	DO Language	--	7 years, public/private schools and language schools
7	38	Santo Ângelo-RS	DO Language	Boston-USA, 3 weeks	15 years, public/private schools and private universities
8	28	Matelândia-PR	MA Language	--	10 years, public/private schools, and language schools
9	34	Toledo-PR	SP Language	--	8 years, public/private schools and language schools
10	47	Passo Fundo-RS	MA Language	--	10 years, public/private schools and public university
11	32	Toledo-PR	SP Language	--	11 years, language schools

Table 12. L2 male speakers' background.

Partic.	Age	Place of birth and where spent most of life	Education	Place and time abroad	English teaching experience
12	37	Getúlio Vargas-RS	MA Language	--	9 years, private university and language schools
13	38	Santa Helena-PR	MA Language	--	4 years, language schools
14	31	Joinville-SC	MA Language	--	7 years, public and private schools, and language schools
15	29	Foz do Iguaçu-PR	MA Language	--	10 years, public and private schools, and language schools
16	28	Rio Grande-RS	MA Literature	--	3 years, language schools
17	26	Xanxerê-SC	MA Language	London-England, 8 weeks	9 years, public and private schools
18	35	Rio Grande-RS	DO Literature	London, England, 4 weeks	10 years, public and private schools, public university, language schools

3.2 Corpus to collect vowel production

This section describes the corpus used to collect the productions of AE and BP vowels. Note that the same English corpus was used to collect the productions of the AE monolinguals and those of the L2 speakers. The only difference was in the data collection procedure, as explained in Section 3.3.3.

3.2.1 Corpus used to elicit the production of AE vowels

The corpus consisted of 66 words, six for each of the eleven AE vowels /i, ɪ, eɪ, ε, æ, ʌ, ɑ, ɔ, oʊ, ʊ, u/, comprising the following six phonological structures:

bVt pVt sVt tVt tVk kVp

As it was impossible to find minimal sets of real words for all contexts, the same or almost the same consonantal contexts were maintained for all eleven vowels through the inclusion of one nonce words, five words with a different coda obstruent, and one with no onset (see Table 13). The onset and coda consonants were all voiceless obstruents, with the exception of those of the structure /bVt/, which appeared in isolation before many of the carrier sentences, as a model of the vowel of the words to come. Not all words that served as a model had the /bVt/ structure. The choice of the words was simply based on which would sound very familiar to the L2 participants. Since there were some rather rare words in the corpus, the participants were told that the target words of each carrier sentence should rhyme with the word in isolation which preceded it. These words were not included in the analysis, as the voicing of the obstruent in several of them can hinder the precise identification of the first periodic pulse of the vowel in the spectrogram, important information for duration measurements.

The words in isolation and the carrier sentences followed the model: *CVC. CVC and CVC sound like CVC*. Thus, the participants would read sentences like these: *Beat. Beat and Pete sound like seat*.

Table 13. Target words read by the AE monolinguals and L2 speakers.

Vowel	bVt	pVt	sVt	tVt	tVk	kVp/kVt
[i]	beat	pete	seat	teat	Teak	keep
[ɪ]	bit	pitt	sit	tit	Tick	kit
[eɪ]	bait	pate	sate	Tate	Take	Kate
[ɛ]	bet	pet	set	Tet	Tech	kept
[æ]	bat	pat	sat	tat	Tack	cat
[ʌ]	but	putt	shut**	tut	tuck	cut
[ɑ]	bot	pot	sot	tot	tock	cot
[ɔ]	bought	ought**	sought	taught	talk	caught
[oʊ]	boat	poach	soak**	tote	toke	coat
[ʊ]	book**	put	soot	--	took	cook**
[u]	boot	poop**	suit	toot	tuke*	coot

* Invented (nonce) word.

** Different phonological context.

The 22 target sentences were randomly ordered and appeared 3 times each (see Appendix E), resulting in 162 tokens (11 vowels x 5 contexts x 3 repetitions = 165 – 3 repetitions of the non-existent /tʊt/ = 162). The /tʊt/ token was preferred to be left out because BP speakers tend to have difficulty producing the vowel /ʊ/, and, although all tokens with the vowel /ʊ/ but /sʊt/ are very commonly used words, /sʊt/ (soot) is a real word while /tʊt/ is not. Thus, there were a total of 4,374 vowel tokens analyzed: 1,458 (162 x 9 participants) produced by the AE monolingual speakers and 2,916 (162 x 18 participants) produced by the L2 speakers.

3.2.2 Corpus used to elicit the production of BP vowels

The Portuguese corpus consisted of 70 words and nonce words, 10 for each of the BP vowels (/i, e, ε, a, ɔ, o, u/) in the following five phonological structures: pV.pV, tV.kV, kV.kV, fV.fV, sV.sV. The words and nonce words were contextualized in a carrier sentence preceded by one of the words: *CVCe/o. Em CVCe e CVCo temos V.* Differently from the English words, the Portuguese words were disyllabic, since there are no Portuguese CVC words formed by the consonants used in this study. All the words were stressed on the first syllable. Thus, the participants would read sentences like: *Pêpe. Em pêpe e pêpo temos ê.*, and *Pêpo. Em pêpe e pêpo temos ê* (Pêpe. In pêpe and pêpo there's an ê).

As in the case of the English corpus, the consonantal contexts were chosen by taking into account the place of articulation and the absence of voicing. Since each sentence contained two target words formed by the same vowel and consonantal context, the second syllable of the first and second words ended in the graphic vowels “e” and “o”, respectively, but were always read as [i] and [u] due to a BP production process of raising of the vowels /e/ and /o/ in word-final unstressed position. These two vowels were chosen because they have similar F1 values. The choice of having CV.Ce and CV.Co words (e.g., *fife* and *fifo*, *sasse* and *sasso*) was to have a more number of recorded sentences without asking the participants to repeat the very same sentence several times. For words containing the vowels /e/, /o/, /ε/ or /ɔ/ in the stressed syllable, the vowels were written as ê, ô, é or ó respectively to help the participants identify the vowel to be produced. Literate Brazilians are familiar with these diacritics. Table 14 shows the target BP words read by the participants.

Table 14. Target words read by the BP monolinguals.

Vowel	pVpV	tVkV	kVkV	fVfV	sVsV
[i]	pipe/pipo	tike/tiko	quique/quico	fife/fifo	sisse/sisso
[e]	pêpe/pêpo	têque/têco	quêque/quêco	fêfe/fêfo	sêsse/sêsson
[ɛ]	pépe/pépo	téque/téco	quéque/quéco	féfe/féfo	sésse/sésso
[a]	pape/papo	taque/taco	caque/caco	fafe/fafo	sasse/sasso
[ɔ]	pópe/pópo	tóque/tóco	cóque/coco	fófe/fófo	sósse/sósso
[o]	pôpe/pôpo	tôque/tôco	côque/coco	fôfe/fôfo	sôsse/sôsso
[u]	pupe/pupo	tuque/tuco	cuque/cuco	fufe/fufo	susse/susso

The word in isolation was always one of the target words and was not considered for analysis. Each BP monolingual read the 35 target sentences (see Appendix F) twice, resulting in 140 tokens (7 vowels x 2 tokens per sentence x 5 contexts x 2 repetitions), and each L2 speaker read the 35 target sentences once, resulting in 70 tokens (7 vowels x 2 tokens per sentence x 5 contexts). The total of BP tokens analyzed was 2940: 1680 (140 tokens x 12 participants) produced by BP monolinguals, and 1260 (70 tokens x 18 participants) produced by the L2 speakers.

3.3 Data collection procedure of the production experiments

This section reports on the data collection procedure used to collect both AE and BP vowels, either in Brazil or in the United States. It is subdivided into three sections so as to describe the procedure used to elicit vowel production by the AE monolinguals, the BP monolinguals and the L2 speakers. The productions analyzed in the present study were recorded with a Sony MZ-NHF800 minidisk recorder, with a Sony ECM-

MS907 condenser microphone. The data were converted to PCM (Pulse Code Modulation) at 22 kHz, with 16-bit accuracy, MONO.

3.3.1 AE monolinguals' productions

The native speaker in charge of recording the AE monolinguals' productions was instructed to ask the participants to read the sentences at normal speed. In order to help the participants to maintain a constant falling intonation, each sentence was shown on a different card. The recordings were made in a quiet room at the participants' homes and whenever there was any background noise during the recordings, the participant was asked to reread the affected sentence.

3.3.2 BP vowel production by BP monolinguals and L2 speakers

Similarly to the AE recording procedure, the recordings were made in a quiet room at the participants' homes, and in case there was any background noise during the recordings, the participants were asked to reread the affected item(s). All recordings but those made in the US were recorded by this researcher. The BP monolinguals were asked to read the target sentences at normal speed. Again, to help maintain a falling intonation, each sentence was shown on a different card. The participants were allowed to read the item on the next card only if the recording was satisfactory in terms of intonation and vowel height. By satisfactory in terms of intonation I mean that all the

sentences had to have a falling intonation at the end of the sentence. Attention to vowel height had to be given, because the speakers were informed that the two target words in the sentence should rhyme with the word in isolation, but even with the signaling in orthography (ê, é, ô, ó), in the sentences containing the vowels [e]-[ɛ] and [o]-[ɔ] many participants tended to mix vowels, pronouncing an open and a closed vowel in the same sentence. In the instances when this type of confusion occurred, the participants were asked to repeat the sentence immediately after the mistake until all the target vowels were pronounced with a similar quality.

3.3.3 L2 speakers' productions

For the recordings of the AE vowels by the L2 speakers, the same sentences read by the AE monolinguals were used. However, some pictures representing a CVC word that contained the target vowel (with the exception of the picture showing *oranges*, a VCVCCVC word) were presented before each sentence. The word *oranges* was chosen because it is very familiar to the participants, who were told that all the sentences that followed this specific word in isolation (or the picture representing *oranges*), had to rhyme with its first syllable. Thus, in this case the elicited vowel was /ɔ/. In this test, both the pictures and sentences were presented on a computer screen.

The pictures were included for two main reasons: (a) to minimize the influence of orthography, and (b) to facilitate the reading of the words with the correct target vowel, since the participants were told that the words in the sentence should rhyme with the word the picture represented. The eleven pictures used in the test are shown in Appendix G. During a training session which consisted of showing one picture, one set

of three isolated words and one target sentence, all containing a vowel embedded in a phonological context different from the target contexts, the participants were told that they should say what the picture represented and then keep the sound of that picture in mind, since the words inserted in the sentence presented after the picture should rhyme with the word displayed in the picture. The set used in the training session, with phonological contexts different from the target contexts, is shown in Appendix H.

Thus, after the speaker pronounced the word “egg”, the next slide was presented and they read (i) the first word in isolation and the sentence, (ii) the second word in isolation and the sentence, and finally (iii) the third word in isolation and the sentence. As previously stated, in case any of the sentences was read with a rising intonation, the participant was asked to reread it until a falling intonation was reached. The pictures were really helpful especially for the three tokens containing the vowel [ʊ] in the sentence *Book and put sound like soot*. Many participants tended to pronounce *put* with the central vowel [ʌ], and *soot*, a rather unfamiliar word, with the high back vowel [u]. In these cases, the picture of books was shown again and they were reminded that all the words should rhyme. Obviously, due to the difficulty of storing the /ʊ/ vowel, two kinds of productions took place: All of the vowels pronounced either as /u/ or /ʊ/, but at least the repetitions using only one vowel type were consistent.

3.4 Production measurements

In order to investigate how the BP and AE vowels are acoustically produced by the three groups of participants, the following acoustic properties of the vowels were

measured: f_0 , F1, F2, F3, and duration. The next sub-sections will describe the procedures used to measure these properties.

3.4.1 Duration

Before running a script to automatically and reliably measure formants, each vowel was manually segmented and labeled in the digitized sound wave by using the program Praat, version 4.4 (Boersma & Weenink, 2006). Either the beginning or the end of the selection was close to a zero crossing, that is, when the wave crosses zero amplitude. The start and end points were considered to be the first and last periodic pulses on the waveform that had considerable amplitude and resembled the vowel period. As stated in Section 3.2.1, the choice for voiceless consonantal contexts was exactly to facilitate the duration measurements, since these consonants allow a more precise identification of the first and last constant periodic pulses of the vowel. The duration of the AE semi-diphthongs /eɪ/ and /oʊ/ were not considered, since, as explained below, only the fundamental frequency and the formants of the first element of the semi-diphthong were measured. Figure 16 illustrates the segmentation of the BP vowel /i/ produced in a /s_sV/ context by a woman.

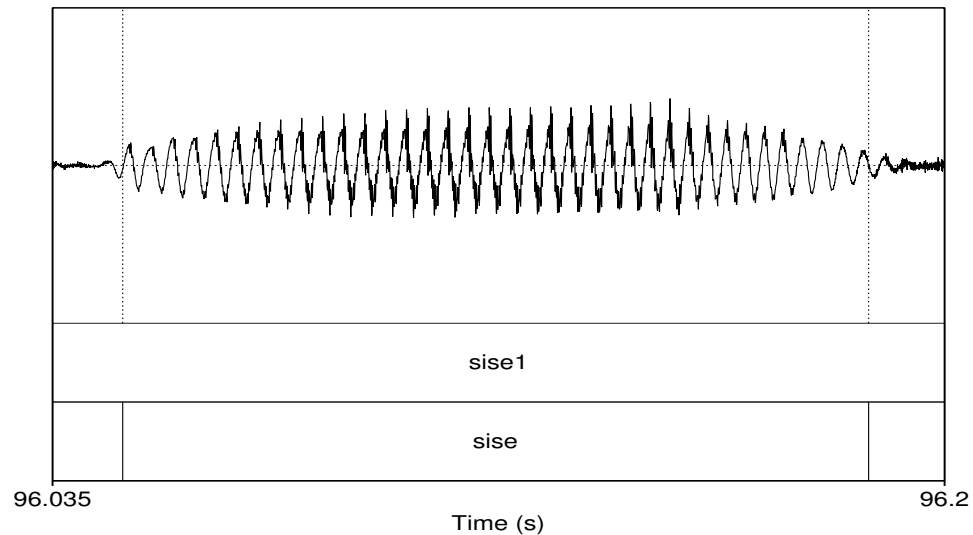


Figure 16. Segmentation of the BP vowel /i/ produced in a sV.sV structure by a woman.

3.4.2 Fundamental frequency (f₀)

In order to measure the fundamental frequency, the central 40% of the target vowels were measured automatically with Praat using cross-correlation (CC) analysis (see Appendix I). The pitch floor was set at 60 Hz for men and 120 Hz for women, and the pitch ceiling was set at 400 Hz for both. These values in Hz are appropriate for measuring f₀ of men and women, the former having lower f₀, thus requiring a lower pitch floor. After the pitch values of the center of the vowel were found, their median was calculated. The median is the middle of a distribution: Half the scores are above the median and half are below it. The median is a more robust measure than the mean because it is less sensitive to extreme scores, especially in highly skewed distributions. Thus, since f₀ had to be represented by a single value, the median offered the best figure because it avoided the influence of spurious numbers.

3.4.3 The first three formants

Despite its being widely used for formant measurements, Wempe (2001) and Wempe and Boersma (2003) point out that the LPC analysis may result in misleading results due to the need of carefully choosing the measurement parameters, such as the LPC order, that is, the number of formants to be observed. The LPC order must be defined in advance and the signal properties are the base of this definition. An inaccurate order selection, for instance, may lead to the emergence of spectral peaks in wrong positions, thus compromising automatic analyses.

The formant measurements for F1, F2 and F3 of each vowel were made by applying the burg algorithm (Anderson, 1978) built into Praat to calculate the LPC spectra with the number of formants per frame defined as 5. Due to the great differences between the vocal tract shapes, the formant ceiling was determined for each type of vowel produced by each participant. In order to find the optimal ceiling, all values from 4500 to 6500 Hz in 10-Hz steps were calculated for women, and all values from 4000 to 6000 Hz in 10-Hz steps for men. The optimal ceiling for each vowel for each speaker was the one which yielded the lowest standard deviation of F2 over the values measured, which came to a total of 20 for each vowel. The window length was set to 50 ms, and the time steps were defined as 25 ms. In the case of the AE semi-diphthongs /eɪ/ and /oʊ/, only the first element of the diphthong was considered and the vowel was segmented before the transition between the first and the second element, so that only the 40% central part of the first element was measured. An example of the segmentation of a semi-diphthong is shown in Figure 17. The script used to calculate the formants can be seen in Appendix J. It was written by Dr. Paul Boersma, of the University of Amsterdam, to analyze formant values of the seven BP and EP vowels

/i, e, ε, a, ə, o, u/. The study has not been published yet. Minor changes were made in the original script so as to adapt it to the present study.

To better visualize the formant measurements and make comparisons between languages or varieties, the F1 and F2 values were plotted with inverted scales to approximate traditional articulatory vowel charts. All the vowels were plotted in Hertz.

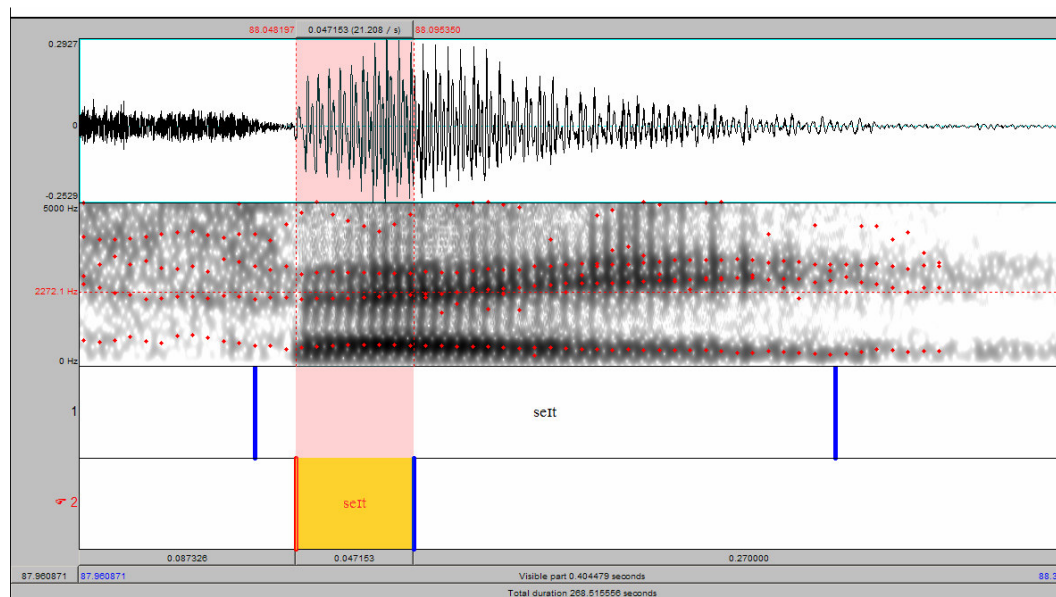


Figure 17. Segmentation of the semi-diphthong /eɪ/ produced by an L2 female participant.

3.5 Statistical analyses of the production results

In experimental phonetics, the use of statistical analysis is highly important so that the great amount of data can be interpreted satisfactorily. In the present study, with 3 groups of participants, 5 vowel characteristics measured (duration, f0, F1, F2, and F3),

and 4 groups of tokens produced (BP vowels by monolinguals, AE vowels by monolinguals, and BP and AE vowels by L2 speakers), a total of 36,570 values needed to be treated statistically (7,314 vowels x 5 vowel features).

In order to interpret the acoustic characteristics of the vowels under analysis, the first procedure was to calculate the mean (\bar{X}), median (Me) and standard deviation (SD) of the acoustic measurements of the vowel productions of each group of participants. The mean is obtained by summing the elements of a dataset and dividing them by the number of elements of this group (Brown, 1988, p. 66). The median, as explained in Section 3.4.2, is the value located in the middle of a distribution: 50% of the scores are above the median and 50% are below it (Brown, 1988, p. 67). The mean and the median are measures of central tendency and they are similar in symmetric distributions. The mean is more affected by extreme values than the median and is thus not a good measure of central tendency for extremely skewed or asymmetric distributions. The mean will be higher than the median for positively skewed distributions and less than the median for negatively skewed distributions (Barbetta, 2001, p. 110).

The means were plotted in most of the vowel plots, and in some of them the SD was also shown by means of ellipses around the vowel symbol, which represented the mean value.

The SD is a measure of variability and it thus indicates how spread the data are from the center, that is, how much they deviate from the mean. The formula to obtain the SD of a dataset (S) is:

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}}$$

Thus, the SD can be found by first computing the mean (\bar{X}) for the data set, then computing the deviation by subtracting the mean from each value (X), squaring each individual deviation, and adding the squared deviations. Next the sum of the squared

deviations is divided by the sample size (N). Finally, the square root is taken (Brown, 1988, p. 69).

In a normal distribution, 68% of the values are within one SD of the mean, and approximately 95% of the values are within 2 SD of the mean. Thus, a normal distribution is a symmetric distribution around the mean, that is, the values are more concentrated in the middle of the curve than at the tails (Barbetta, 2001, p. 158).

The mean, median and SD values show the central tendency of a dataset. However, in order to compare the mean or median results of several datasets and verify whether they differ or not in statistically significant ways, some tests of significance must be used. When an independent variable has two means/medians (two levels), and they have similar SD (similar variance), a t-test can be used to compare their means/medians (Brown, 1988, p. 176). If two datasets are independent from each other, the Independent-Samples t-test can be used. In the present study, the independent t-tests were two-tailed, that is, the hypotheses were not directional, since all possible outcomes were hypothesized. Examples of hypotheses considered in this study are the following:
 H_0 = there are no significant differences between the medians of L1 and L2 vowels produced by the L2 speakers.

H_1 = the medians of the L1 vowels are significantly lower than those of the L2 vowels.

H_2 = the medians of the L1 vowels are significantly higher than those of the L2 vowels.

If the means/medians between three or more datasets need to be compared, the Analysis of Variance (ANOVA) can be used. A one-way ANOVA is used to compare one nominal independent variable and one continuous (interval-scale) dependent variable. The independent variable can consist of any number of groups, also called levels. This test was used to investigate whether there were differences between the medians from three or more samples, each sample referring to a given population (e.g.,

when examining the medians between different vowels, different participants, different genders). When the medians of three or more samples showed a statistically significant difference, the Tukey post-hoc test was applied to identify which pairs of samples differed from each other because of factors other than chance (Brown, 1988, pp. 176-177).

These two statistical tests can only be applied if the data are normally distributed and there must be homogeneity of variance, that is, the variance of data in the observed groups should be similar (Woods, Fletcher, & Hughes, 1986).

The t-test and ANOVA were calculated in the SPSS 10.0 software and were used in the present study to verify (a) whether duration differed significantly between vowel pairs, (b) whether F1 and/or F2 values differed significantly between vowels, and (c) whether the Euclidean distance between vowels would differ significantly. The alpha level (p or significance value) was defined as $< .05$. In order to calculate the Euclidean distance between two vowels, that is, how much the F1/F2 values of a vowel would be distant from the F1/F2 values of another vowel within the productions of a particular group of participants, the following formula was used:

$$\sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$$

where p_x is the F1 value of the first and q_x is the F1 value of the second vowel to be contrasted, whereas p_y is the F2 value of the first and q_y is the F2 value of the second vowel to be contrasted. Thus, the F1 value of a vowel was subtracted from the F1 value of another vowel and the result was squared. Then the F2 value of a vowel was subtracted from the F2 value of another vowel and this result was also squared and then added to the F1 result. The square root of the sum of the two values was then taken. All values were calculated in Hertz.

As can be observed, several procedures to collect and analyze the production data needed to be followed. The next section will describe the procedures adopted to collect and analyze the perception data.

3.6 Vowel perception participants

Three groups of participants comprised the vowel perception experiment: L2 speakers, BP monolinguals, and AE monolinguals. The participants from the three groups took the perception test as volunteers and all of them claimed to have no hearing problems. The same group of L2 speakers which took the L2 production test also took the L2 perception test (see Section 3.1.3). Thus, the following subsections provide information about the AE and BP monolinguals only.

3.6.1 BP monolinguals

The group of BP monolinguals who were tested perceptually consisted of 10 participants, five women and five men, whose ages ranged from 19 to 23 years (mean 20.4 years). All participants were from cities in the three southern Brazilian states, as follows:

- a) Paraná: Curitiba (1), Maringá (1), Londrina (1);
- b) Rio Grande do Sul: Porto Alegre (2), Rio Grande (1);
- c) Santa Catarina: Concórdia (1), Chapecó (1), Florianópolis (2).

The participants were UFSC undergraduate students from several courses: Administration (5), History (1), Mathematics (2), Physics (2), and Psychology (1). They were selected after having answered a questionnaire about their background information, the same questionnaire completed by the BP monolinguals of the production experiment (see Appendix C).

The perception test with the BP monolinguals was carried out with participants other than the ones involved in the production test. The choice of having different participants was to get completely naïve data. The perception test was applied in Florianópolis.

3.6.2 AE monolinguals

Four AE monolinguals performed the perception test. The participants were two women and two men, whose ages ranged from 22 to 29 years (mean 25.6 years). Two were undergraduate students in Arts and the other two had already graduated, one in Linguistics and the other in Chemical Engineering. The participants had spent most of their lives in California (1), Ohio (1), and Washington state (2). The perception test with the AE monolinguals was carried out with participants other than the ones involved in the production test so as to get completely naïve data.

3.7 Perception stimuli

The script to generate the synthetic vowels used as stimuli was written by Dr. Paul Boersma and Ton Wempe, from the Institute of Phonetic Sciences of the University of Amsterdam, and can be seen in Appendix K.

The vowels were not inserted in frames or carrier sentences, but presented in isolation. The use of synthetic stimuli instead of natural stimuli is important because the formant values and duration can be controlled precisely, which would be impossible with natural stimuli. Moreover, when natural stimuli are used, the speaker's accent may be a limitation, since formant and duration values are pronounced differently from person to person, or even her own productions may vary each time they are pronounced. One disadvantage of synthetic stimuli, however, is that some vowels may sound too unnatural. Interestingly, when finishing the test several participants were asked to describe who they thought might have pronounced the vowels. By their responses, one can make sure most of the synthetic vowels were natural enough, because several participants answered they thought a "thin dark-haired man" or a "young blond man" had recorded the sentences. All of them were surprised when told the stimuli were nonhuman.

The stimuli consisted of a single continuum of synthesized vowels which was used to test the perception of the three groups of participants, in other words, the same stimuli were heard by the AE monolinguals, the BP monolinguals and the L2 learners. The stimuli had 339 synthesized vowels with 14 F1 values, 10 F2 values, and 3 different durations (100 ms, 141 ms, and 200 ms)²⁶. The scales of numbers follow the sequences of a corresponding articulatory vowel chart. As can be seen in Figure 18, on the

²⁶ The stimuli were the same as those designed in Escudero's (2005a) post-doc research.

horizontal axis, the F2 values ranged from 580 to 2700 Hz, and on the vertical axis the F1 values ranged from 240 to 900 Hz. The two areas of empty spaces above the 448 Hz F1 area are necessary and represent vowels that are highly unusual or even impossible to be articulated by humans, since these areas would correspond to low vowels farther forward (left of chart) and farther back (right of chart) than the articulatory tract allows us to produce.

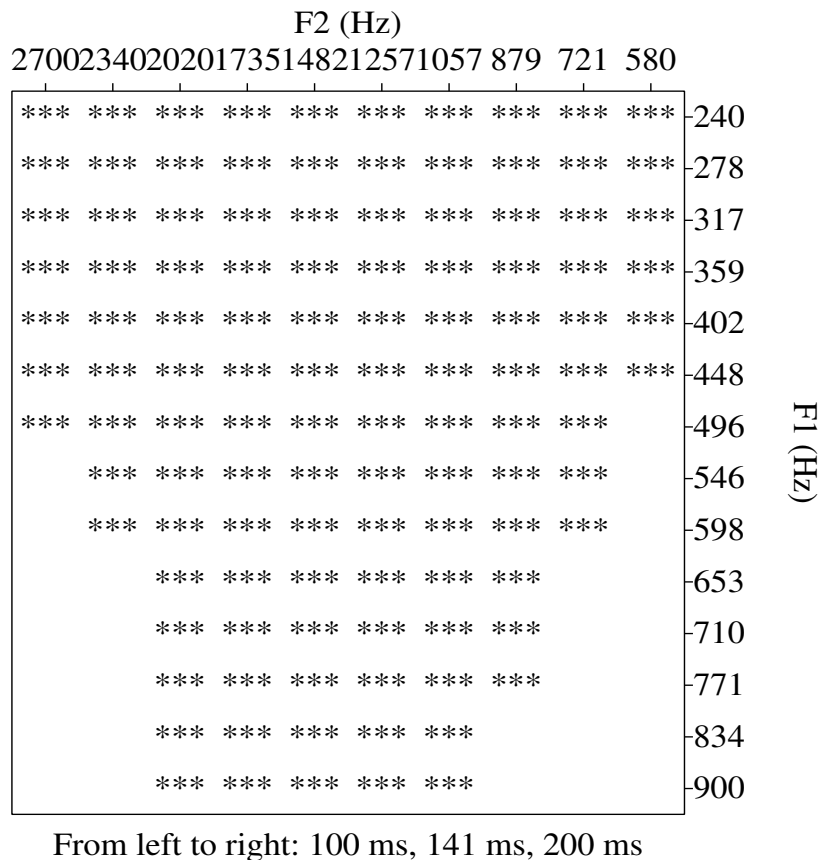


Figure 18. Synthesized vowel continuum used to test vowel perception by the three groups of participants.

Since the same stimuli were used to test BP and AE perception, what differed in each test was the way the stimuli were presented. In the experiment testing BP perception, the monolinguals and L2 speakers were shown the screen in Figure 19, with Portuguese instruction and all the labels containing the seven target BP vowels. Note that the vowels /ɛ/ and /ɔ/ were orthographically written as *é* and *ó* so that the participants could differentiate between /e/-/ɛ/ and /o/-/ɔ/.

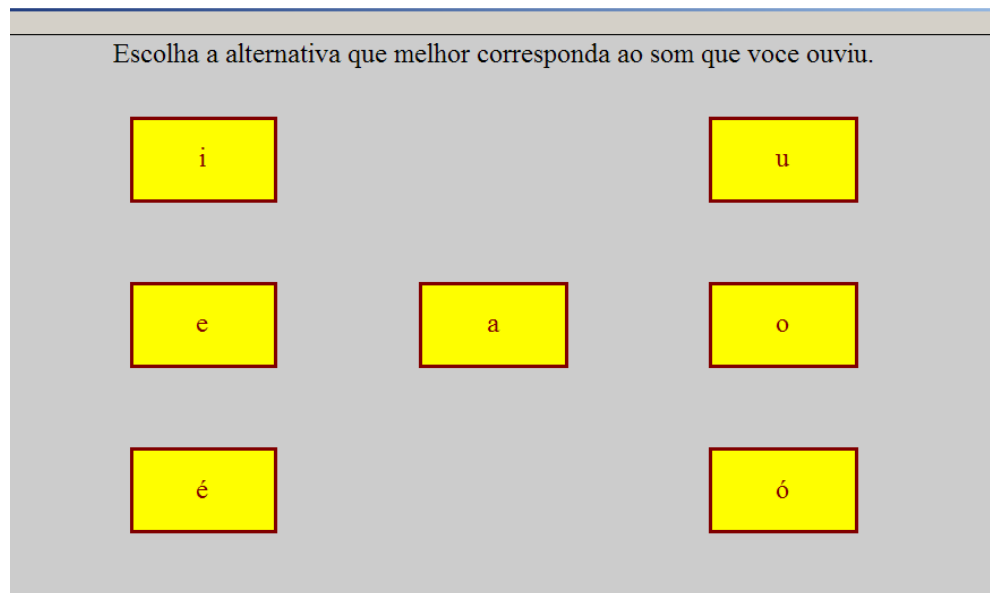


Figure 19. Screen presented to BP monolinguals and L2 speakers to test their BP vowel perception.

When testing AE perception, the participants were presented with the screen shown in Figure 20.

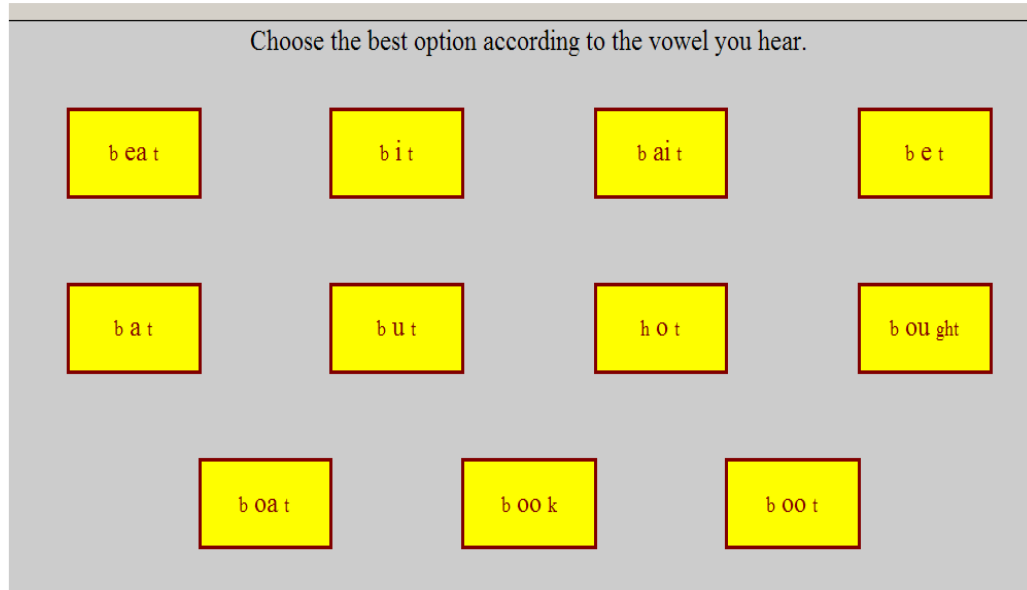


Figure 20. Screen presented to AE and L2 speakers to test their AE vowel perception.

This time the instruction was in English and the labels did not contain only a vowel, but an English word representing each vowel sound. Since the English vowels cannot be adequately identified by the vocalic graphemes, 11 words were chosen to represent each of the 11 vowel sounds: *beat*, *bit*, *bait*, *bet*, *bat*, *but*, *hot*, *bought*, *boat*, *book*, *boot*. The choice of these specific words was because the bVt frame can be used to insert 9 of the 11 English vowels and form very well known words. The choice of *hot* and *book* was because they are very familiar to all the participants.

3.8 Perception test procedure

The perception test was run on a computer and the participants used earphones to hear the stimuli. The L2 speakers and the BP monolinguals were individually tested by this researcher in a quiet room at the Federal University of Santa Catarina. The AE monolinguals were tested in a quiet room at Georgetown University, in Washington, by another Brazilian researcher. Although the experimenter is highly proficient in English, when giving the instructions in the US he tried to speak as little as possible in order not to influence the participants' results because of his nonnative English.

The perception test was a forced-choice labeling test²⁷ which consisted of the participants' listening to one synthetic vowel and clicking on the label which most resembled the vowel heard, according to Figures 19 and 20. A short break was suggested after the presentation of 34 vowels, when the sentence *Descanse um pouco. Em seguida clique na tela com o mouse para continuar.* (Take a short break. Then click on the screen with the mouse to continue) was displayed in the BP perception test, or when the sentence *Pause. Rest for some seconds and then click to continue.* was presented in the AE perception test. The participants had the choice to ignore the break and continue the test.

Before the test started, the participants were asked to read aloud what each label represented and they listened to the first 34 stimuli as a training session so as to become familiarized with the testing procedure. The stimuli were randomly organized and every time the test was restarted a new order of vowels was generated. The participants were told that the vowels had been cut from real words from running speech, and for this

²⁷ A forced-choice labeling test is a perception test in which the participants are presented with a number of labels from which they have to choose one when they hear a stimulus.

reason some of the vowels would sound weird, but in these cases they should still click on the label that most closely resembled the vowel they heard.

3.9 Analysis of the perception results

The analysis of the perception experiments was carried out by plotting the participants' answers to the stimuli they were exposed to (see the script to create perceptual vowel plots in Appendix L). The stimuli consisted of vowels with three duration values; however, only the difference between the extreme values was analyzed: 100 ms vs. 200 ms. For this plot, only the values within one standard deviation of the mean F1 and F2 values of tokens identified as each vowel were considered. Then, the number of times each vowel was identified as short (100 ms) was compared to the number of times it was identified as long (200 ms). The comparison between the reliance on duration by the L2 and AE speakers was important to investigate whether the participants relied more on durational cues or spectral cues.

In order to examine the reliance on spectral cues, the Euclidean distance (ED) between the vowel contrasts perceived by each group of participants was compared (see Appendix M).

Besides computing the ED between the two vowels of a pair, another interest was to examine the percentage of overlap in both the perception and production of vowels by the AE monolinguals and the L2 speakers. This computation was important to identify how much the two vowels of a pair were distinguished by monolinguals and L2 speakers, and how much the latter participants' results differed from those of the former. In order to compare the percentage of overlap for each formant for each vowel

pair, first the extreme formant values of a vowel were calculated: The maximum formant value was found by adding the SD value of that vowel to its mean, and the minimum formant value was found by subtracting the SD value of that vowel from its mean. This computation was done for F1 and F2 values of each vowel pair. Then, to calculate the range between the minimum formant value of one vowel of the pair and the maximum formant value of the other vowel of the pair, the minimum formant value of one vowel was subtracted from the maximum formant value of the other vowel of the pair, and the same procedure was repeated for the other minimum and maximum formant values of the two vowels. Finally, these minimum and maximum values were divided by each other and then multiplied by 100, providing this way the percentage of overlap between the formant values of a vowel pair. In order to illustrate this explanation, the minimum and maximum F1 values of /i/ and /ɪ/ are written below, followed by the formula that was explained in this paragraph (the letters a-d were included to help understand the formula):

	Min	Max
/i/ F1	317 (a)	449 (b)
/ɪ/ F1	237 (c)	341 (d)

$$\text{Percentage of overlap for F1} = \left(\frac{d - a}{b - c} \right) \cdot 100$$

The same formula was applied to calculate the percentage of overlap of F2 values. The script used to calculate the vowel overlap can be seen in Appendix N. Note that the minimum and maximum formant values were included manually in the script.

As regards other statistical tests, besides the t-test to calculate the difference between Euclidean distance results by the groups of speakers, the chi-square statistical test (χ^2) was also used to test the percentage of the participants' preference to identify

tokens which varied in duration. The chi-square tests the differences between nominal data, taking into account observed and expected values (Barbetta, 2001, p. 247).

CHAPTER 4

RESULTS AND DISCUSSION

This chapter will report the results of the production and perception tests carried out in this study. In order to test vowel perception and production, the data were collected with monolingual Brazilian Portuguese (BP) speakers, monolingual American English (AE) speakers, and second language (L2) speakers. Although only duration, F1 and F2 values were analyzed, the tables also show f0 and F3 values, information that might be useful for other studies on BP and AE vowels.

The present chapter will first report the production and perception results from each of the three groups of speakers. After the presentation of the results obtained from each experiment, a discussion of the interrelation between production and perception will be provided.

4.1 Speech production of BP monolingual speakers

As described in Section 3.1.2, the BP monolinguals (6 women and 6 men) were from three different Brazilian states: Rio Grande do Sul, Santa Catarina and Paraná. An analysis of variance (ANOVA) of the independent variables F1 and F2 revealed that there are significant differences among participants within the group of men and within the group of women. Tukey post-hoc tests showed that there is no homogeneous tendency that allows the participants to be grouped by dialect, because, for instance, one

participant from one dialect did not have any statistically significant formant value difference from participants from the other two dialects, but for some vowels this participant's formant values differed significantly from the other participant from the same dialect. This lack of consistency in the results can be explained by the limited number of participants from each region (only two per gender per dialect). Thus, henceforth the participants will be considered to pertain to only one dialect: The dialect spoken in the non-capital cities of the three southern Brazilian states. The participants will be grouped only by gender.

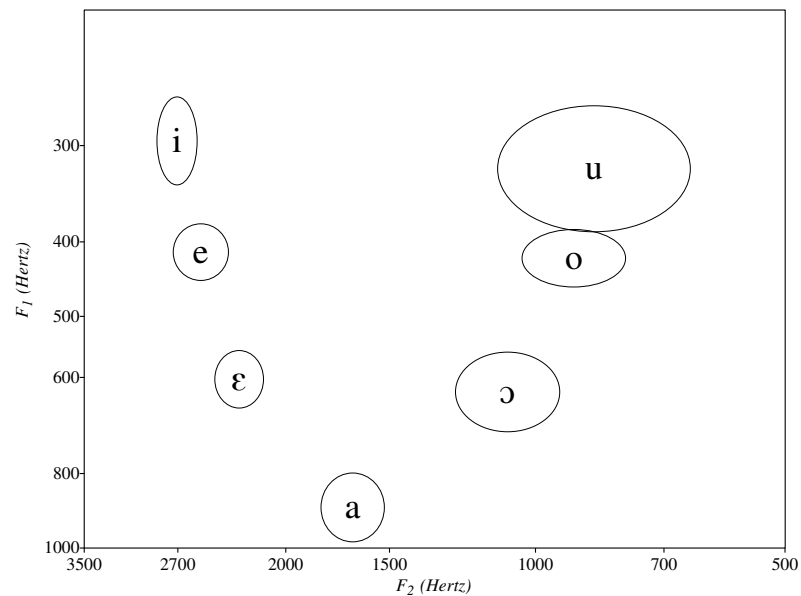
The mean, median and standard deviation (SD) of the formant values of all the vowels produced by the BP female and male monolinguals in the five phonological contexts are shown in Tables 15 and 16, respectively. The mean values are plotted in Figures 21 and 22.

Table 15. BP women's duration (in milliseconds), f0, F1, F2 and F3 values in Hertz.

		i	e	ɛ	a	ɔ	o	u
Dur.	Mean	92	111	127	127	123	111	93
	Median	94	111	128	127	124	112	94
	SD	19	24	26	24	23	23	19
F0	Mean	241	222	206	202	206	221	245
	Median	234	214	202	200	204	212	240
	SD	37	34	31	32	31	23	40
F1	Mean	298	414	606	890	631	422	326
	Median	286	408	611	896	636	416	320
	SD	41	36	51	87	74	38	49
F2	Mean	2710	2540	2282	1667	1091	908	880
	Median	2694	2558	2283	1682	1098	904	825
	SD	151	192	152	143	150	128	235
F3	Mean	3200	3021	2912	2580	2693	2880	2875
	Median	3248	3058	2964	2627	2676	2902	2904
	SD	328	279	286	321	217	184	211

Table 16. BP men's duration (in milliseconds), f0, F1, F2 and F3 values in Hertz.

		i	e	ɛ	a	ɔ	o	u
Dur.	Mean	95	114	130	130	130	113	96
	Median	95	116	134	132	131	114	98
	SD	17	19	21	17	20	18	16
F0	Mean	151	146	136	136	135	146	157
	Median	153	143	136	134	134	146	154
	SD	41	26	30	28	30	35	35
F1	Mean	292	344	494	651	542	379	308
	Median	293	339	497	686	574	390	308
	SD	23	35	64	109	69	37	21
F2	Mean	2212	2080	1908	1405	971	874	834
	Median	2199	2061	1888	1383	967	849	790
	SD	130	170	133	157	105	128	174
F3	Mean	2950	2755	2614	2346	2334	2466	2526
	Median	2972	2734	2620	2314	2346	2452	2468
	SD	214	200	156	224	266	186	261

**Figure 21.** BP female participants' mean and SD (in ellipses) of the F1 and F2 values in Hz.

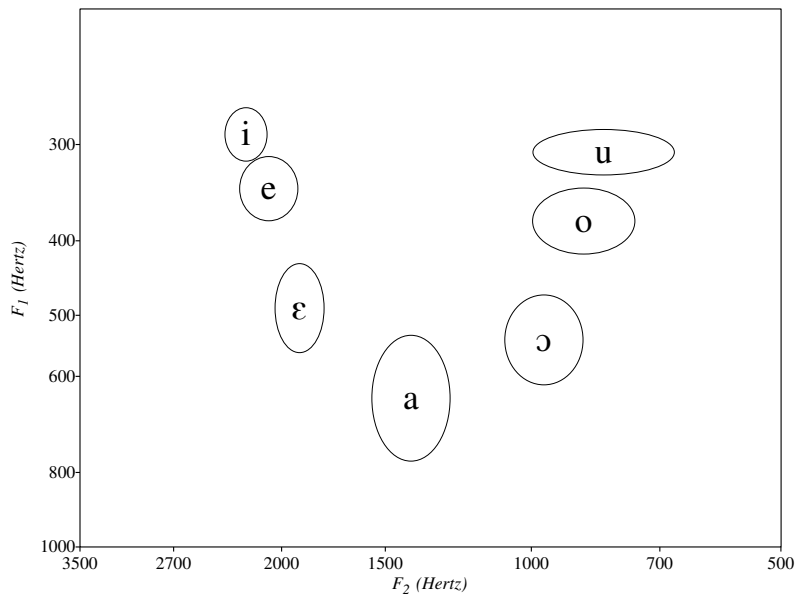


Figure 22. BP male participants' mean and SD (in ellipses) of the F1 and F2 values in Hz.

The vowels plotted in Figures 21 and 22 show that the women's vowel system is more symmetrical than the men's if the series of front and back vowels are taken into account. The median F1 values of each of the seven BP vowels were calculated for each speaker, and then paired-samples t-tests were run for the two high, mid and low vowels. Since there were three computations using the same data, a Bonferroni correction²⁸ was carried out, setting the significance level to $\alpha = .017$. The results show that the men from the Brazilian Southern states have higher front than back vowels, but none of the women from the Brazilian Southern states has the front vowels significantly higher than their back vowel counterparts, as can be seen in Table 17.

²⁸ Bonferroni correction: the alpha level (in this study $\alpha = .05$) divided by the number of times that variables are compared.

Table 17. Results of t-tests comparing F1 values within the groups of BP female and male monolinguals.

Vowels	Women	Men
/i/ - /u/	t(5) = -2.611, p < .05	t(5) = -4.503, p = .0003
/e/ - /o/	t(5) = -0.653, p = .54	t(5) = -3.661, p = .007
/ɛ/ - /ɔ/	t(5) = -1.200, p = .15	t(5) = 5.726, p = .0003

The results indicate that the Southern Brazilian women have larger vowel spaces than the Southern Brazilian men, as would be expected. The distance between the high vowels and the low central vowel is significantly smaller for men than for women, as revealed by the two-tailed independent-samples t-tests of the Euclidean distances between the vowels /i/ - /a/ - /u/ pronounced by the two groups:

- 1) /i/-/a/: t(10) = 5.353, p < .0001;
- 2) /u/-/a/: t(10) = 4.567, p < .0001; and
- 3) /i/-/u/: t(10) = 4.738, p < .0001.

The Euclidean distances were calculated in the following way: First the median of each vowel produced by each participant was found, then the Euclidean distance between the two vowels of a target pair produced by each of the 12 participants was measured, and finally t-tests were calculated to examine whether the women's Euclidean distances between the vowels differed significantly from those of men.

Although there were statistically significant differences between formant values of vowels inserted in some of the different phonological contexts, the variable *context* was not the focus of analysis in this study, whose main aim was to further investigate how BP speakers of English produce and perceive English vowels.

The results concerning duration differences between the target vowels confirmed that the lower the vowel the greater its duration, which is a typical intrinsic vowel

characteristic, since the more the jaw needs to open to articulate a vowel, the longer its duration. As regards the women's and men's productions, no statistically significant length difference was found between the duration values between women and men for (a) the low vowels /ɛ/, /a/ and /ɔ/; (b) the mid vowels /e/ and /o/; or (c) the high vowels /i/ and /u/. An ANOVA revealed a nonsignificant duration vs. gender interaction ($F = .457, p = .841$). Figure 23 shows the mean values for each vowel produced by each gender in all the phonological contexts.

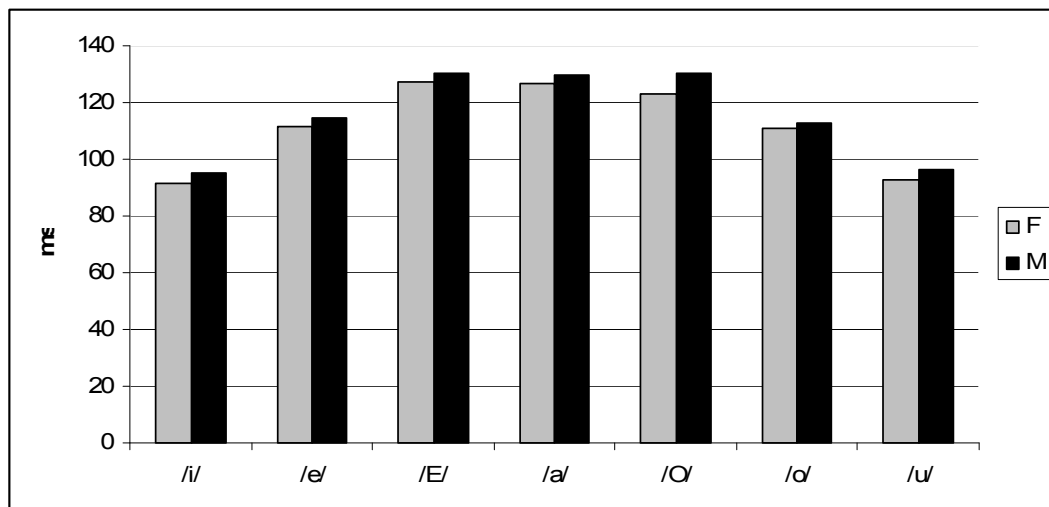


Figure 23. Vowel duration values by monolingual BP speakers (/E/ and /O/ represent /ɛ/ and /ɔ/, respectively).

4.2 Speech perception of BP monolinguals

As described in Section 3.6.1, the BP monolinguals (five women and five men) were from three different Brazilian states: Rio Grande do Sul, Santa Catarina and Paraná.

The values considered for analysis were the F1 and F2 values within one standard deviation of the mean F1/F2 values of the tokens identified as each vowel (see Table 18 and Figure 24).

Table 18. BP women's and men's F1 and F2 values within 1 SD of the mean F1/F2 values of tokens identified as each vowel (N = number of vowels).

Dur = 100 ms		i	e	ɛ	a	ɔ	o	u
	N	113	199	179	115	142	159	223
F1	Mean	297	411	654	794	641	450	309
(Hz)	Median	278	402	653	834	653	448	278
F2	Mean	2196	1985	1921	1357	997	934	1025
(Hz)	Median	2340	2020	2020	1257	1057	879	1057
Dur = 200 ms								
	N	99	192	219	131	134	172	193
F1	Mean	284	396	636	769	636	431	298
(Hz)	Median	278	402	598	771	598	448	278
F2	Mean	2205	1982	1942	1340	969	928	1015
(Hz)	Median	2340	2020	2020	1257	879	879	1057

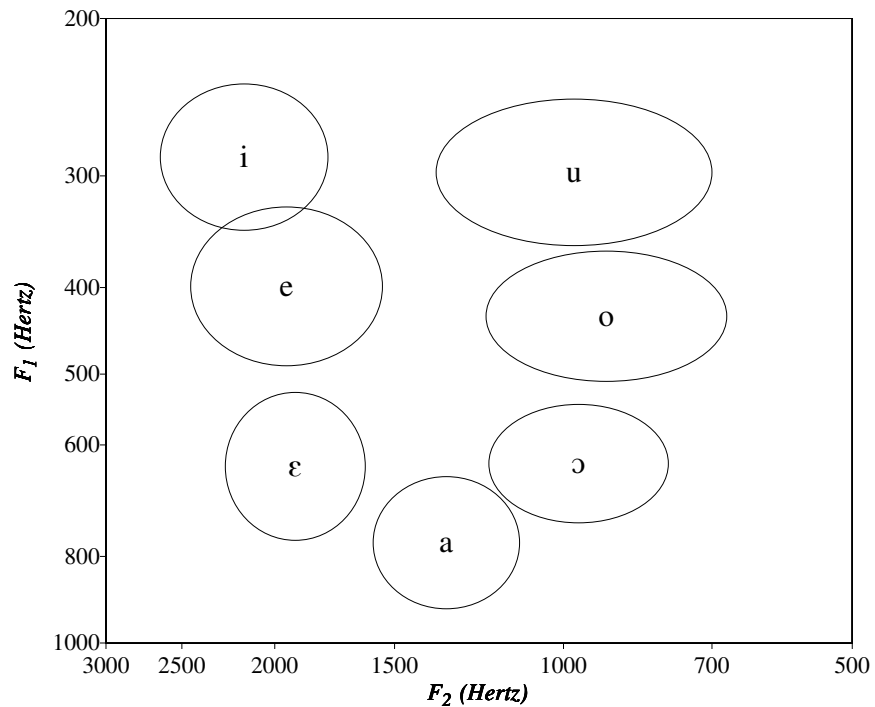


Figure 24. Perception results by BP monolinguals: One SD of the mean F1/F2 values of tokens identified as each vowel.

The plot in Figure 24 shows that the BP vowels perceived by the BP monolinguals are arranged more or less evenly in the acoustic vowel space, with little overlap between adjacent vowels, similarly to what was observed in Figures 21 and 22 which show the plotted vowels of the production results.

The analysis of the perception results reveals that the BP speakers do not rely on duration to perceive their native language vowels (see Figure 25). For the present analysis, the 141-ms vowels were disregarded, and the only comparison made was between the perception of vowels with the extreme duration values: 100 ms and 200 ms. Paired-samples t-tests revealed that the differences between 100-ms vowels and 200-ms vowels were not a determinant factor to discriminate any of the vowels. Chi-square tests showed that the number of times the participants correctly identified a vowel with short

or long duration did not differ significantly for most of the vowels, except for two: /ɛ/, which was more frequently chosen when presented with longer duration ($\chi^2(1, N = 398) = 8.04, p = .004$), and /u/, which was more frequently chosen when presented with a shorter duration ($\chi^2(1, N = 416) = 4.32, p = .03$).

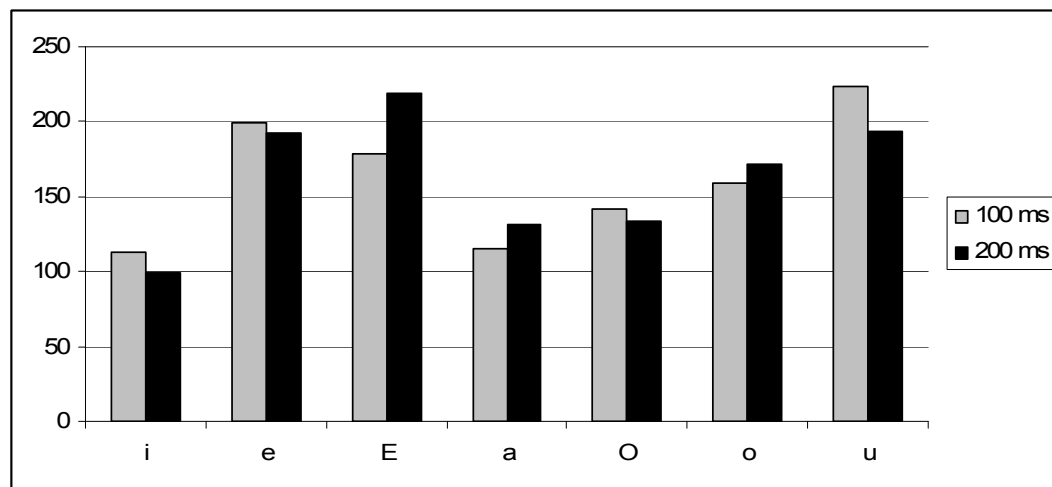


Figure 25. Number of vowels perceived with each of the two duration values by BP monolinguals.

Compared to the production results, the significant differences in terms of duration for /ɛ/ and /u/ corroborate the fact that the mid vowel is more easily perceived if its duration is longer, since it is also longer in production, the same happening to the high back vowel, which is shorter in production and is also more easily perceived with shorter duration values. The claim that duration is relatively unimportant as a distinctive feature for BP vowels is confirmed by the lack of significance found for 5 of the 7 vowels when duration differences were taken into account in the forced-choice labeling test. Therefore, it is possible to conclude that the participants made use primarily of vowel quality to perceive BP vowels, duration being a secondary or irrelevant cue.

The next section will report the results of the AE monolinguals. The figures obtained by the acoustic analysis of the vowels produced and perceived by the two groups of monolinguals will provide data about what Escudero (2005) calls “optimal production”, some useful information when the L2 participants’ data are analyzed.

4.3 Speech production of AE monolinguals

Differently from the BP monolinguals, who were from different Brazilian states, all the AE monolinguals were born and/or had lived most of their lives in the city of Sacramento, California. The mean, median and standard deviation (SD) of the formant values of all the vowels produced by the female and male AE participants in the five phonological contexts are shown in Tables 19 and 20, respectively.

Table 19. Mean, median and SD of duration, f0, F1, F2 and F3 values of AE vowels produced by the female AE monolinguals.

		/i/	/ɪ/	/ei/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/ou/*	/u/	/u/
	N	60	60	60	60	60	60	60	60	60	48	60
D	Mea	130	103	49	116	167	110	152	168	48	114	127
	Me	125	105	46	112	167	106	144	164	44	110	125
	SD	28	22	17	31	36	24	36	43	18	25	25
F0	Mea	195	192	208	182	178	177	176	182	200	166	203
	Me	192	191	217	182	181	188	178	181	194	174	206
	SD	37	36	27	45	50	55	45	38	30	50	35
F1	Mea	308	501	450	704	820	718	749	705	519	540	335
	Me	306	518	448	696	807	720	739	704	526	540	332
	SD	35	55	81	58	89	71	83	62	95	38	36
F2	Mea	2766	2121	2386	1910	1808	1695	1293	1239	1492	1554	1782
	Me	2753	2110	2423	1910	1807	1711	1267	1218	1505	1592	1791
	SD	117	95	227	113	128	137	137	139	205	156	245
F3	Mea	3310	2975	3024	2839	2668	2747	2654	2659	2735	2750	2730
	Me	3322	2989	3075	2846	2735	2768	2638	2658	2749	2727	2699
	SD	191	146	226	142	231	182	115	146	121	105	128

* Measurement values of the first element of the semi-diphthong.

Table 20. Mean, median and SD of duration, f0, F1, F2 and F3 values of AE vowels produced by the male AE monolinguals.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/ou/*	/u/	/u/
N		75	75	75	75	75	75	75	75	75	60	75
D	Mea	140	118	52	134	179	131	169	174	49	128	135
	Me	134	115	49	135	183	129	166	179	46	130	135
	SD	24	20	23	26	29	27	32	36	17	25	29
F0	Mea	135	131	128	125	118	123	117	119	135	127	134
	Me	124	127	126	121	116	119	116	118	135	123	133
	SD	32	20	21	18	19	24	22	18	16	18	21
F1	Mea	280	412	395	559	668	580	604	616	434	451	307
	Me	276	423	398	582	671	605	642	628	439	454	306
	SD	22	43	47	69	59	77	99	66	65	51	26
F2	Mea	2331	1884	2056	1729	1669	1406	1106	1083	1254	1371	1556
	Me	2346	1931	2068	1742	1693	1406	1108	1078	1260	1398	1598
	SD	152	172	189	124	123	94	120	102	166	97	226
F3	Mea	2918	2593	2667	2562	2431	2493	2439	2468	2349	2375	2269
	Me	2934	2648	2706	2568	2423	2491	2433	2479	2358	2424	2319
	SD	147	189	208	162	162	173	127	112	229	194	216

*Measurement values of the first element of the semi-diphthong.

The vowels plotted in Figures 26 and 27 show that the vowel systems of the two genders are organized similarly. As regards height, a one-way ANOVA revealed that there were no statistically significant differences between the F1 values of the following vowels for the women: (a) /i/-/u/, (b) /ɪ/-/ou/, (c) /ʊ/-/ou/, (d) /ɛ/-/ɔ/-/ʌ/, and (e) /ɑ/-/ɔ/-/ʌ/, and for the men: (a) /i/-/u/, (b) /ɪ/-/eɪ/-/ou/, (c) /ɪ/-/ʊ/-/ou/, (d) /ɛ/-/ɑ/-/ʌ/, and (e) /ɑ/-/ɔ/-/ʌ/. These results indicate that /i/ and /u/ clearly form the group of high vowels, whereas the great SD of the F1 values of /e/ shows that this vowel is higher than the other mid vowels /ɪ/ and /ou/ for women, but not for men, whose results show that /ɪ/, /eɪ/ and /ou/ have similar degrees of height. The vowels /ɛ/, /ʌ/, /ɑ/ and /ɔ/ form the group of low vowels, which differ in F1 from /æ/ for both female and male AE monolinguals.

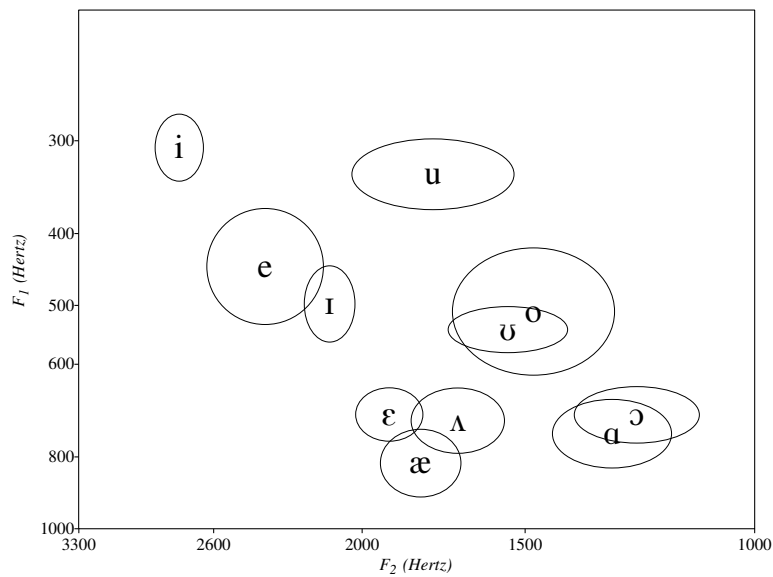


Figure 26. Mean F1 and F2 values of vowels produced by the female AE monolinguals.

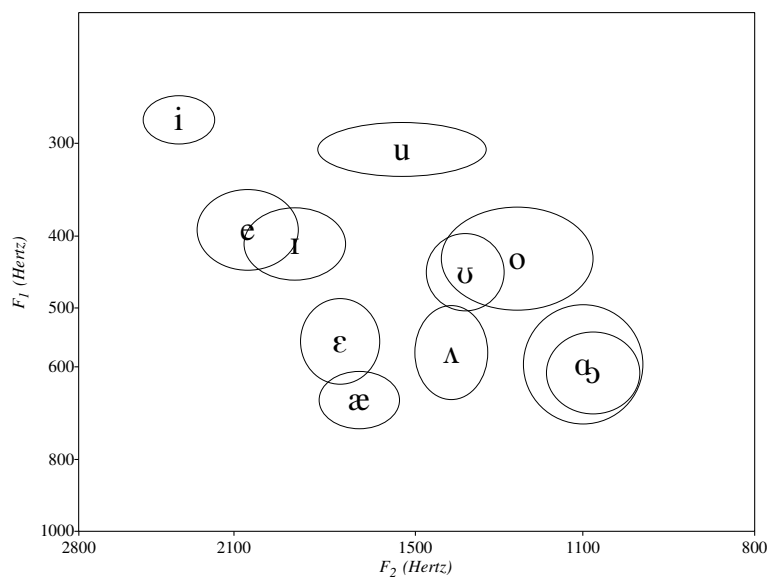


Figure 27. Mean F1 and F2 values of vowels produced by the male AE monolinguals.

The AE monolinguals' data analyzed in this study show similar tendencies to the data collected with Californians by Hagiwara (1997) and with Western US speakers by

Clopper et al. (2005). Table 21 shows the differences and similarities between this study and the other studies in the following aspects:

Table 21. Measurements comparison between the vowels produced by AE monolinguals in this study and those in Hagiwara (1997), Clopper et al. (2005), Hillenbrand et al. (1995), and Peterson and Barney (1952).

Vowel	This study	Hagiwara	Clopper et al.: US Western speakers	Hillenbrand et al.	Peterson & Barney
/i/	More fronted and higher than /ɪ/	Similar results	Similar results	Similar results	Similar results
/ɪ/	Lower and farther back than /eɪ/	Similar results	Similar height and farther back than /eɪ/	Different results: higher and farther back than /eɪ/	The study did not analyze /eɪ/
/ɛ/	More fronted and higher than /æ/	Similar results	Similar results	Different results: lower and farther back than /æ/	Similar results, but women's distance between /ɛ/ and /æ/ is greater
/ɑ/	Great overlap for men, but slightly lower and more fronted than /ɔ/ for women	The study did not analyze /ɔ/	Similar results, although the women's /ɑ/ is slightly farther back than /ɔ/	Different results: more fronted and lower than /ɔ/, great distance between the two vowels	Different results: more fronted and lower than /ɔ/, great distance between the two vowels
/oʊ/	Slightly higher and farther back than /ʊ/	Similar results: farther back than /ʊ/ for the 2 genders; however, slightly lower than /ʊ/ for women and virtually at the same height for men	Similar results: farther back than /ʊ/ for the 2 genders; however, higher than /ʊ/ for women and virtually at the same height for men	Slightly different results: slightly lower and farther back than /ʊ/	Not analyzed
/u/	More fronted and considerably higher than /ʊ/	Similar results: higher and slightly more fronted than /ʊ/	Similar results: higher and slightly farther back than /ʊ/ for women, and with virtually the same F2 values as /ʊ/ for men	Different results: higher and farther back than /ʊ/	Different results: higher and farther back than /ʊ/

As can be observed when comparing the vowel measurements from different studies, the variable *dialect* influences the organization of the vowel system within a given language. Escudero's (2005) L2LP model suggests that perception data should be collected with monolingual speakers of the language under investigation so that it is possible to have information about the optimal perceiver. Monolinguals' production data should also provide information about optimal production. The question is: In the specific case of formal EFL English learners, what type of data, that is, which dialect, should be taken into account to describe optimal L2 perception and/or production? In general, EFL learners are mostly exposed to native-like production by means of audio material recorded by native speakers on cassettes, CDs or on sound files available on the web, and this material generally disregards dialectal differences and makes use of what can be considered standard American English, or General American. Moreover, most materials present EFL learners with very unnatural language input, since the audio material generally comprises recited citation forms. Thus, one limitation of the present study is that the L2 vowels were compared to a single AE variety, the vowels from California²⁹. The comparison to a single AE variety is justified by (i) the impossibility of collecting data with speakers of other varieties; (ii) the lack of access to the raw data of studies which have already investigated several other AE varieties, these data being essential for running statistical tests; (iii) and even if it were possible to have access to other researchers' database, the corpus and data collection procedure would be different, which could be another limitation.

As regards duration, similarly to the results obtained in the analysis of the BP vowels, the lower vowels had the greater duration values. Figure 28 shows the duration values of the AE vowels produced by the female and male AE monolinguals.

²⁹ The Californian accent can be considered a somewhat neutral (or close to General American) accent in the United States.

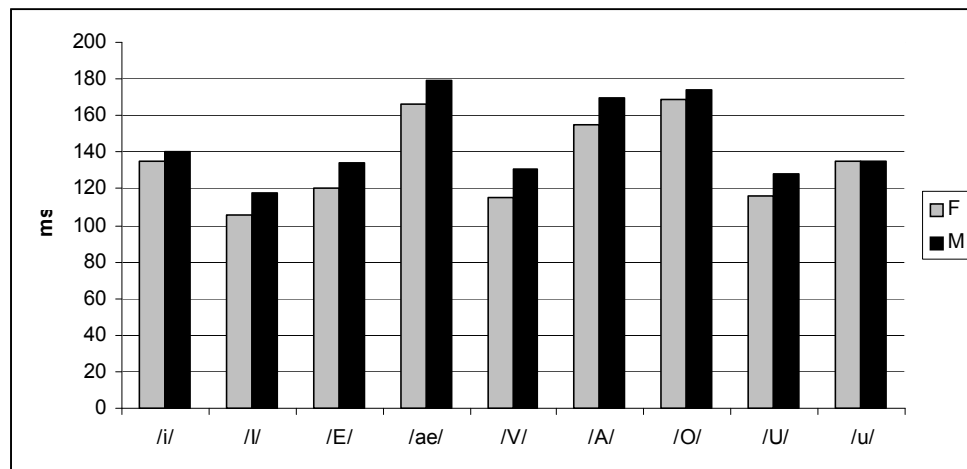


Figure 28. Duration values of the vowels produced by the female and male AE monolinguals (the symbols I, E, V, A, O, and U correspond to the vowels /ɪ/, /ɛ/, /ʌ/, /ɑ/, /ɔ/, and /ʊ/, respectively).

For both women's and men's productions, no statistically significant difference was found between (a) the low vowels /æ/, /ɑ/ and /ɔ/; and (b) the mid and high vowels /ɛ/, /ɪ/, /ʌ/, /ʊ/, /u/, and /i/. An ANOVA revealed a nonsignificant duration vs. gender interaction ($F = 1.674$, $p = .081$). Differently from the BP monolinguals, whose productions of the vowels /i/ and /u/ were the shortest, the tense feature of these two vowels results in greater duration than their BP counterparts, which explains the nonsignificant difference between the high and the mid vowels in terms of duration.

4.4 Speech perception of AE monolinguals

Similarly to the perception results by the BP monolinguals, the perception results by the AE monolingual listeners show that the percentage of overlap between the

vowels of each pair was low, thus indicating, as expected, that the participants made a good distinction in the forced-choice labeling test between the pairs, as can be seen by the Euclidean distance and overlap results in Table 22 and Figure 29.

Table 22. Euclidean distance and rate of overlap between the vowels of the 3 target pairs by the AE monolingual listeners.

Vowel pair	ED (Hz)	Overlap
/i/-ɪ/	235	F1: 11% F2: 42%
/ɛ/-/æ/	590	F1: 0% F2: 11%
/ʊ/-/u/	145	F1: 0% F2: 71%

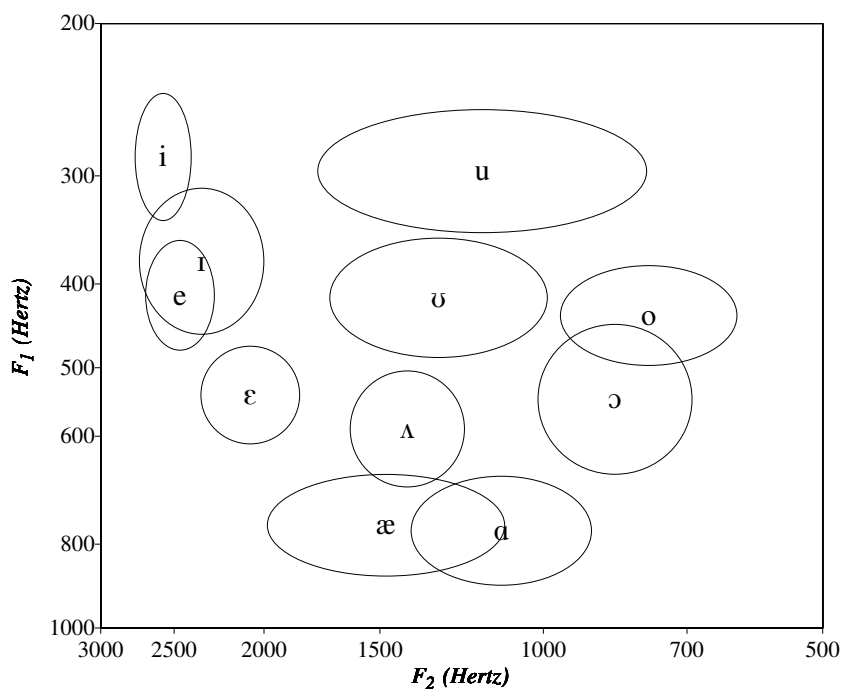


Figure 29. Perception results by the AE monolinguals.

Taking into account the number of tokens identified as a certain vowel within one standard deviation of the mean F1/F2 values of tokens identified as each vowel, Table 23 and Figure 30 show that there was no great difference between the number of times the vowels were identified as /i/, /æ/ or /u/ in the two durations (100 ms and 200 ms). However, the participants identified a significantly limited number of tokens when they heard a long /ɪ/ ($\chi^2(1, N = 24) = 21.3, p < .0001$), /ɛ/ ($\chi^2(1, N = 78) = 5.02, p = .02$), and /ʊ/ ($\chi^2(1, N = 82) = 28.2, p < .0001$). The small number of AE monolinguals does not allow safe conclusions to be made regarding the reliance on temporal cues, but the preference for short vowels when the token had 100 ms, and the small difference rate of identification of long vowels in the two durations indicate that, besides spectral quality, the participants also rely on duration to identify 3 (the shortest) of the 6 vowels, but duration does not seem to be the primary cue to identify vowels.

Table 23. Number of tokens identified in the perception test within one standard deviation of the mean F1/F2 values of tokens identified as each vowel.

Dur	i	ɪ	ɛ	æ	ʊ	u
100 ms	23	20	46	82	58	128
200 ms	24	4	32	72	24	130
Similarity	95.8%	20%	69.6%	87.8%	41.4%	98.5%

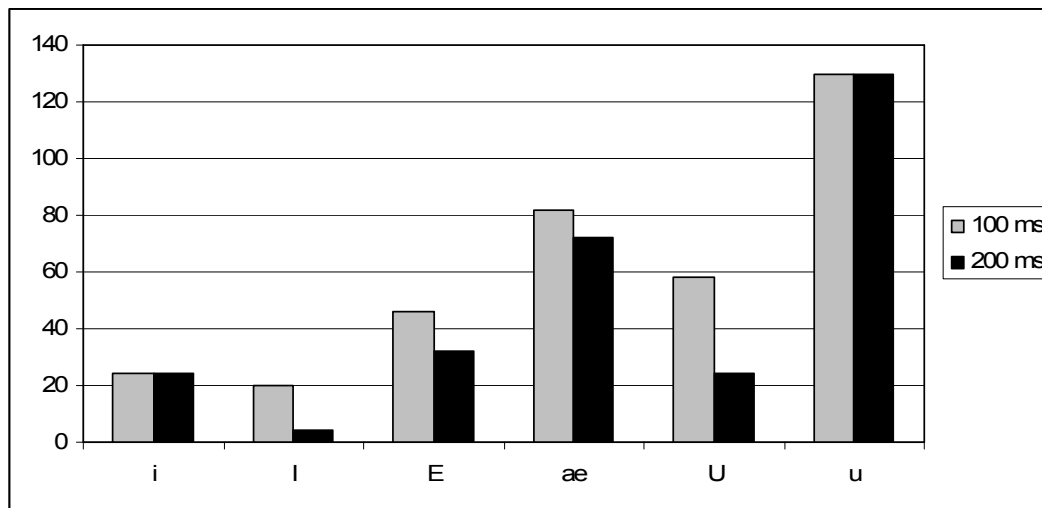


Figure 30. Number of vowels perceived with two duration values by AE monolinguals (the vowel I, E and U represent the vowels /ɪ/, /ɛ/ and /ʊ/, respectively).

With the monolinguals' data analyzed, the next section will show the results of the L2 speakers' production of BP and L2 vowels.

4.5 L2 speakers' production of BP vowels

The same instrumentation used to collect data with the BP monolinguals was adopted in the data collection of the L2 speakers' BP vowels. The only difference was the number of tokens: The L2 speakers produced each vowel only ten times (twice in each of the five phonological contexts), while the BP monolinguals produced each vowel 20 times. Again the variable *context* was disregarded.

The mean, median and standard deviation (SD) of the formant values of all the BP vowels produced by the L2 female and male participants in the five phonological

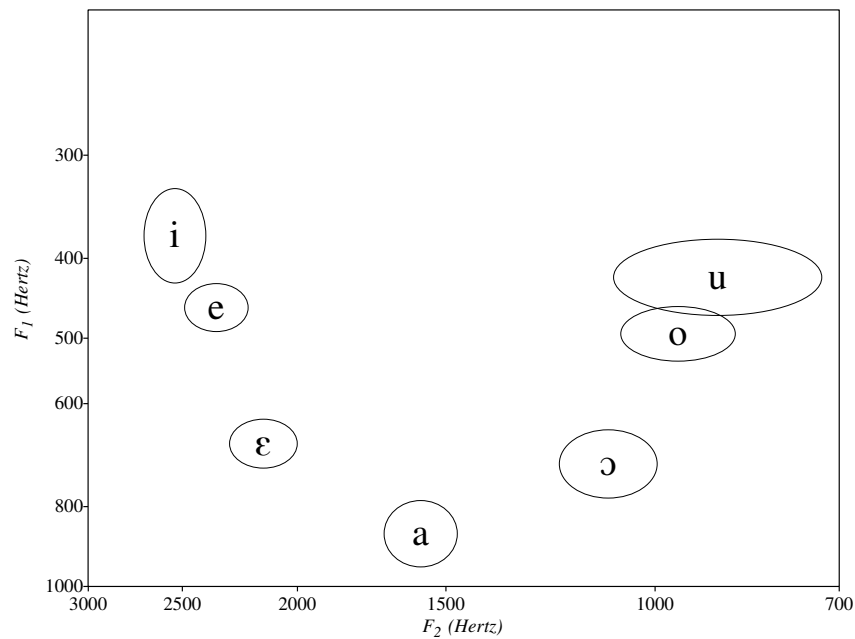
contexts are shown in Tables 24 and 25, respectively. The mean values are plotted in Figures 31 and 32.

Table 24. L2 women's duration, f0, F1, F2 and F3 values for BP vowels.

		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/
Dur.	N	110	110	110	110	110	110	110
	Mean	91	117	136	139	141	116	95
	Median	90	114	135	134	136	116	90
	SD	26	28	31	34	34	28	28
F0	Mean	241	219	204	204	208	224	243
	Median	243	221	206	204	209	226	249
	SD	22	21	25	18	18	18	33
F1	Mean	379	460	673	866	713	495	424
	Median	388	456	670	860	707	496	425
	SD	49	31	46	80	68	38	44
F2	Mean	2540	2345	2141	1579	1100	962	904
	Median	2548	2356	2162	1580	1101	932	874
	SD	152	143	140	111	104	112	185
F3	Mean	3042	2806	2699	2512	2541	2642	2672
	Median	3066	2840	2815	2504	2545	2648	2686
	SD	280	227	313	288	274	200	194

Table 25. L2 men's duration, f0, F1, F2 and F3 values for BP vowels.

		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/
Dur.	N	110	110	110	110	110	110	110
	Mean	79	91	106	107	107	95	83
	Median	77	90	102	105	102	95	82
	SD	19	22	26	24	24	21	28
F0	Mean	155	146	137	131	135	146	159
	Median	156	142	133	126	134	147	151
	SD	38	26	23	24	29	34	34
F1	Mean	313	384	557	667	573	414	342
	Median	310	384	560	672	584	413	344
	SD	34	37	53	46	72	42	28
F2	Mean	2262	2090	1848	1340	983	852	834
	Median	2260	2104	1878	1326	952	844	808
	SD	162	193	237	141	147	89	126
F3	Mean	2831	2614	2530	2296	2342	2534	2490
	Median	2840	2640	2530	2303	2308	2454	2422
	SD	235	275	202	228	256	224	253

**Figure 31.** L2 female participants' mean and SD (in ellipses) of the F1 and F2 values of BP vowels in Hz.

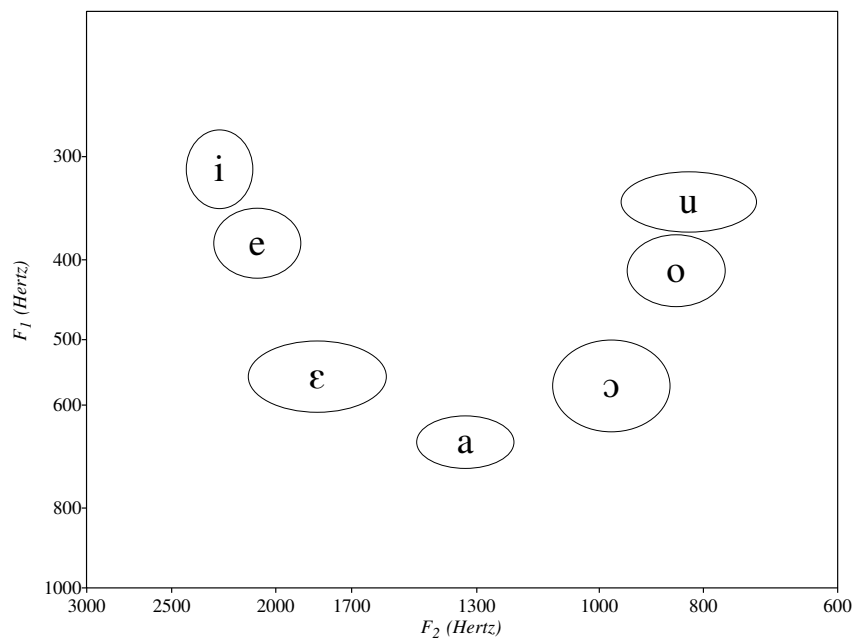


Figure 32. L2 male participants' mean and SD (in ellipses) of the F1 and F2 values of BP vowels in Hz.

Figures 33 and 34 show the BP monolinguals and the L2 speakers' vowels in a single plot.

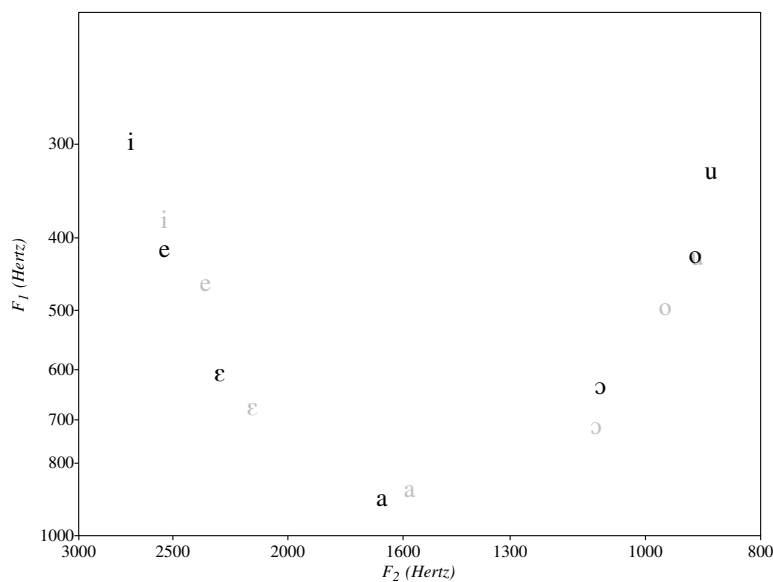


Figure 33. Female BP monolinguals' (in black) and L2 learners' (in grey) productions of BP vowels (mean F1 and F2 values). Note that the BP /o/ and L2 /u/ overlap.

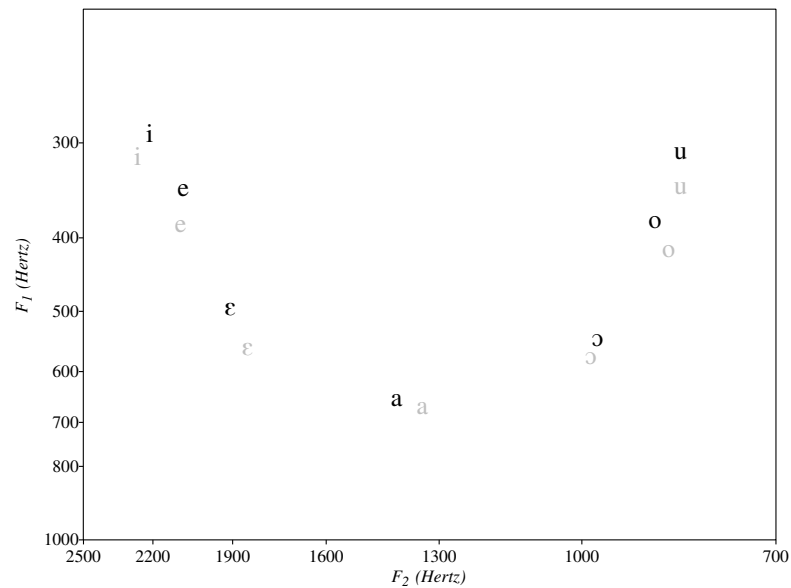


Figure 34. Male BP monolinguals' (in black) and L2 learners' (in grey) productions of BP vowels (mean F1 and F2 values).

The plots show that the female monolinguals' vowel system is more spread than that of the female L2 speakers, but only in terms of height. Independent sample t-tests (two-tailed) show that the Euclidean distances between the extreme vowels /i/-/a/ and /u/-/a/ produced by the female BP monolinguals are significantly greater than that of the female L2 speakers. As regards the men's results, no significant difference between the BP monolinguals' and the L2 speakers' results were found, as can be observed in Table 26.

Table 26. Results of two-tailed independent t-tests comparing the Euclidean distances between BP vowel pairs produced by the BP monolinguals and the L2 speakers.

Vowels	Women	Men
/a/ - /i/	t(15) = 2.271, p = .038	t(11) = -1.546, p = .15
/a/ - /u/	t(15) = 2.503, p = .024	t(11) = 1.032, p = .32
/i/ - /u/	t(15) = 2.087, p = .054	t(11) = -.547, p = .59

The results of the comparison between the L1 vowel systems of the male monolinguals and bilinguals indicate that L1 production was not affected by L2 learning. The lack of significant differences between the Euclidean distances of the BP vowels produced by the male monolinguals and L2 speakers for the extreme vowels /i/-/a/-/u/ supports Escudero's (2005) L2LP model's hypothesis that learners have two separate perceptual grammars. However, the Euclidean distance difference between the females' high vowels (/i/-/u/) to the central vowel (/a/) corroborate Flege's (1995) SLM's hypothesis that L1 and L2 sounds coexist in a single phonological space. Empirical research by Flege (1987) shows that similar findings were found as regards L2 influencing L1 sounds; however, they were not about vowel production, but about VOT values. The results of Flege's study revealed that experienced English and French L2 speakers produced /t/ in their L1 with significantly different values than did monolingual speakers of these two languages: The French participants produced French /t/ with a longer VOT than monolinguals, while the English participants produced English /t/ with a shorter VOT than monolinguals. Thus, Flege's conclusion is that the exposure to acoustically different phones in L1 and L2 led the participants to modify or restructure the phonetic representation they had established for their L1 /t/.

As regards the contradictory findings in the present study, the limited number of participants and the cross-sectional nature of the analysis do not lead to any safe conclusion as regards the L2 influence on L1 sounds.

The results of this section show that, for one group of participants, L1 may have been influenced by L2 acquisition. The question now is whether L2 is influenced by L1. The following subsection will show the vowel measurements obtained of the L2 speakers' English production and compare these participants' production with that of monolinguals.

4.6 L2 speakers' production and perception of L2 vowels

In this section, first an explanation about how the L2 production data were analyzed is presented. Then, four subsections report the production and perception results obtained by the L2 speakers: The first two report the L2 women's results; and the two others, the L2 men's. A comparison between the production and perception results is made for the females' results at the end of Section 4.6.2, and for the male's results at the end of Section 4.6.4.

As regards vowel production, the extent to which the L2 speakers' production of a vowel differed from the phonetic norm of English was estimated by computing the difference between the Euclidean distance of the participants' F1/F2 values and that of the AE speakers F1/F2 values. In order to have comparable vowel spaces, the L2 speakers' productions were normalized taking into account the minimum and maximum F1 and F2 values of the AE monolinguals for each gender. First, the L2 speakers' values are converted to 0 (zero) and 1. In order to do that, the first step is to find the minimum and maximum F1 and F2 values of the L2 speakers. The minimum F1 value is the mean of the highest vowel minus 1 standard deviation (SD). The maximum F1 value is the mean of the lowest vowel plus 1 SD. The minimum F2 value is the mean of the farthest back L2 vowel minus 1 SD. Finally, the maximum F2 value is the mean of the most fronted L2 vowel + 1 SD. In sum:

F1 min = lowest F1 mean – 1 SD

F2 min = lowest F2 mean – 1 SD

F1 max = highest F1 mean + 1 SD

F2 max = highest F2 mean + 1 SD

After the minimum and maximum F1 and F2 values of the L2 speakers' productions were found, the vowels were converted to 0 and 1 by using the formula:

$$x_j = \frac{v_i - \min v_i}{\max v_i - \min v_i}$$

where x_j is the normalized value to be calculated, v_i is the formant value to be normalized, and $\min v_i$ and $\max v_i$ are the minimum and maximum mean F1 and F2 values ± 1 SD observed in the L2 data. This computation converts the minimum value to zero and the maximum value to 1. The values greater than 1 SD have negative values if lower than the mean $-$ SD, and values greater than 1 if they are higher than the mean $+$ SD. Some possible negative numbers or numbers higher than 1 may occur due to very low or very high values to be normalized, but these are the figures which are not within one standard deviation. After that, the results of these computations are mapped to those of the native speakers by applying this formula:

$$x = \min_f + \text{norm}_y (\max_v - \min_v)$$

where x is the normalized value, \min_v is the minimum F1 or F2 mean minus 1SD value found in the AE monolinguals' database, norm_y is the result of the computation for transforming numbers from 0 to 1, and \max_v is the maximum F1 or F2 mean plus 1 SD value found in the AE monolinguals' database. With these computations, the edges of the extreme values of the ellipses (SD) of the L2 speakers were found, and the edges of the extreme values of the L2 speakers were aligned according to those of the AE monolinguals, thus allowing for a more reliable comparison between the vowel spaces of the two groups to be made. The normalization was done by running the script shown in Appendix H. Figure 35 shows the plot of the female AE monolinguals and Participant

nonsignificant results were not included in the text; thus, the words *similar* and *overlap* imply that a t-test was applied and nonsignificant differences were found.

4.6.1 Female L2 speakers' production results

If the normalization can be accepted as making the vowel spaces really comparable, the measurements of the female vowels show that the median Euclidean distance between the vowels of the three target pairs (/i/-/ɪ/, /ɛ/-/æ/, and /u/-/ʊ/) differed significantly compared to that of the AE monolinguals (see Table 27). The L2 /ɛ/-/æ/ were the least separated vowels by the L2 participants (22.5%). The vowels of the /i/-/ɪ/ pair which were produced with the greatest Euclidean distance by the AE monolinguals (690 Hz) were poorly separated by the L2 participants (184 Hz, or 27.1%). The L2 /ʊ/-/u/ had the closest Euclidean distance to that of AE /ʊ/-/u/, but still the rate was very low (35.3%). For a better visualization of the results, Figure 36 shows the means and standard deviations (in ellipses) of the AE and the normalized L2 vowels by the L2 female speakers. Table 28 and Figure 37 show the values of the BP and L2 vowels (with no normalization). The tables with the acoustic measurements of each female participant's BP and L2 vowels are in Appendix P.

Table 27. Median values of the Euclidean distances for AE and L2 vowels, percentage of L2 vs. AE similarity, and t-tests for the difference of each pair produced by the L2 female participants.

Vowel pair	AE (Hz)	L2 (Hz)	L2 vs. AE similarity	t-test
/i/-/ɪ/	678	184	27.1%	t(13) = -5.529, p < .0001
/ɛ/-/æ/	151	34	22.5%	t(13) = -2.742, p = .017
/u/-/ʊ/	289	102	35.3%	t(13) = -2.372, p = .034

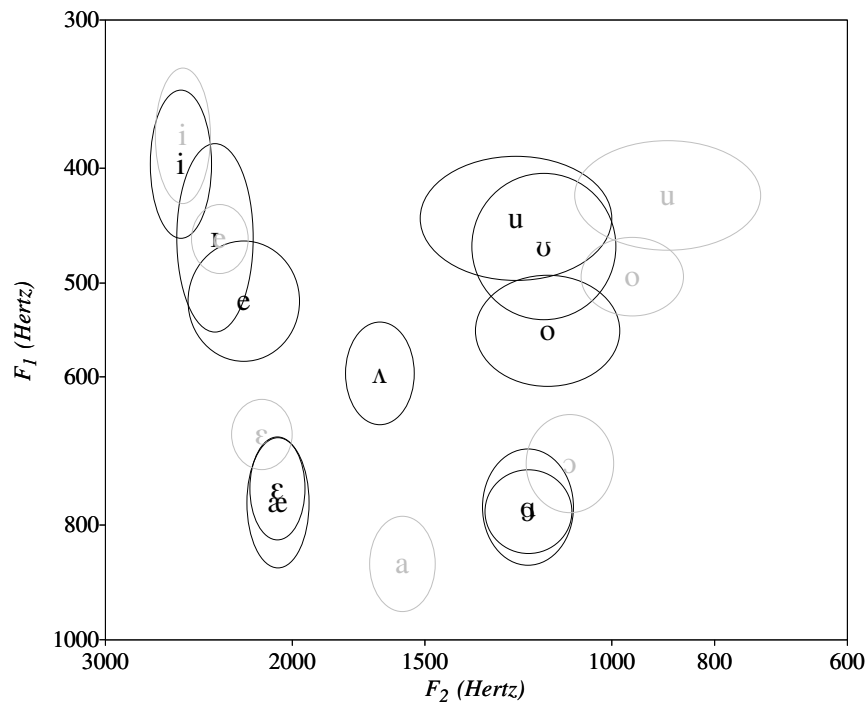


Figure 37. AE (in black) and BP (in grey) vowels produced by the female L2 speakers.

As can be seen in Figures 38 and 39, there is great overlap between the vowels of the three target pairs. The greatest percentage of overlap is between the vowels of the L2 / ϵ /-/ æ / pair, since 72% and 89% of the F1 and F2 of / æ /, respectively, overlap with the F1 and F2 values of / ϵ /. The lowest percentage of overlap is between the vowels of the / i /-/ ɪ / pair, followed by the / $ʊ$ /-/ u / pair, as can be seen in Table 29. Thus, the greatest difference in percentage of overlap between the L2 speakers and the AE monolinguals in the three pairs is in the following order: / $ʊ$ /-/ u / less overlapped than / i /-/ ɪ / than / ϵ /-/ æ /. Interestingly, as shown in Table 27, compared to the AE pairs, the L2 / $ʊ$ /-/ u / has the highest rates of Euclidean distance similarity (35.3%), while the rate for the L2 / i /-/ ɪ / is of 27.1%. However, the percentage of overlap between the vowels of the L2 / $ʊ$ /-/ u / is higher than that of L2 / i /-/ ɪ /, especially for the F2 values. This indicates that the participants make use of F1 more than of F2 to distinguish between the high back

vowels, and they make use of both F1 and F2 at similar degrees to distinguish between the high front vowels. Thus, although the Euclidean distance difference between the L2 /ʊ/-/u/ was more similar to the AE speakers than that of /i/-/ɪ/, the relative positions of the L2 and AE vowels if F1 and F2 values are taken into account differ more greatly for the back than for the front vowels. As regards the /ɛ/-/æ/ vowels, this contrast is the one produced with the highest overlap rates by both the AE monolinguals and the L2 speakers.

Table 29. Percentage of overlap between L2 and AE vowels by the female participants.

Vowel pair	AE speakers		L2 speakers	
	F1	F2	F1	F2
/i/-/ɪ/	0%	0%	42%	37%
/ɛ/-/æ/	14%	40%	72%	89%
/ʊ/-/u/	0%	34%	64%	64%

Compared to the BP vowels, paired-samples t-tests show that the L2 female speakers' /i/ does not differ in height or backness from the BP /i/. As regards the L2 /ɪ/, it is significantly lower ($t(10) = 3.684$, $p = .004$) and farther back than the BP /i/ ($t(10) = -2.719$, $p = .022$). It can be said that this vowel is in between an L2 /i/ and an AE /ɪ/.

The L2 vowels of the /ɛ/-/æ/ pair overlap almost totally. The two L2 vowels are significantly lower (/ɛ/: $t(10) = 6.000$, $p < .0001$; /æ/: $t(10) = 5.698$, $p < .0001$) than the BP /ɛ/, but only the L2 /ɛ/ is significantly farther back than the BP /ɛ/ (/ɛ/: $t(10) = -3.503$, $p = .006$; /æ/: $t(10) = -2.188$, $p = .053$).

No difference in height, but a significant difference in backness ($t(10) = -2.351$, $p = .041$) was found between the L2 vowels of the /u/-/ʊ/ pair: /u/ being more fronted than /ʊ/. The participants' L2 /u/ did not differ in terms of height from the BP /u/, but it

was significantly more fronted than the BP /u/ (F2: $t(10) = 9.527$, $p < .0001$). As regards the L2 /u/, this vowel was significantly lower ($t(10) = 2.372$, $p = .03$), and more fronted ($t(10) = 6.768$, $p < .0001$) than the BP /u/. Concerning the AE speakers, the Californians in this study produced a very central /u/, and, not surprisingly, the difference between the AE and L2 /u/ is considerable.

In sum, the female L2 participants did not make a native-like distinction in terms of spectral quality between the three target vowel pairs, the Euclidean distance between the L2 pairs being significantly smaller than those of the AE pairs. The ellipses of all the target pairs overlapped to different degrees, but some distinction in the correct direction was made between the high front and back vowels.

As regards duration, Table 30 and Figure 38 show that only the short vowel /ɛ/ was produced with significantly longer durations by the L2 speakers than by the AE monolinguals. The difference rate between the duration values of the members of the L2 and AE pairs is shown in Table 31. Compared to the difference in duration between the vowels of each pair produced by the AE monolinguals, the L2 speakers' duration distinction differs considerably for each pair: 22.2% for the /i/-/ɪ/ contrast, 11.8% for the /ɛ/-/æ/ contrast, and 61.5% for the /u/-/ʊ/ contrast.

Table 30. Mean duration values (in ms) of AE vowels produced by the female AE monolinguals, and of L2 and BP vowels produced by the female L2 speakers.

Vowel	AE	L2	BP	t-test AE vs. L2	t-test L2 vs. BP
/i/	130	129	91	$t(13) = -.538$, $p = .60$	$t(10) = 3.902$, $p = .003$
/ɪ/	103	123		$t(13) = 1.246$, $p = .23$	
/ɛ/	116	158	136	$t(13) = 2.728$, $p = .017$	$t(10) = 2.300$, $p = .044$
/æ/	167	164		$t(13) = -.165$, $p = .87$	
/u/	127	134	95	$t(13) = .420$, $p = .68$	$t(10) = 3.851$, $p = .003$
/ʊ/	114	126		$t(13) = .812$, $p = .43$	

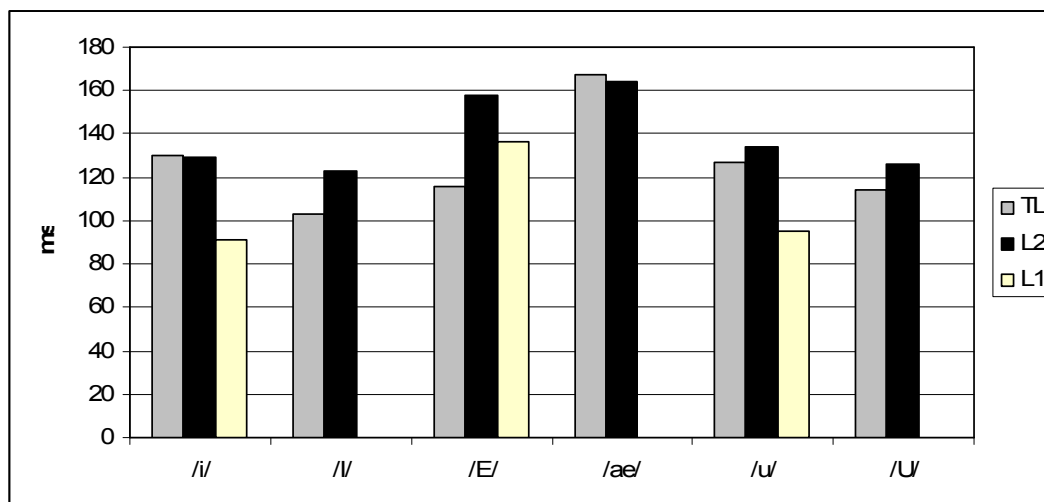


Figure 38. Mean duration values of English vowels produced by AE monolinguals and L2 speakers, and of Portuguese vowels produced by the L2 speakers (the vowels /ɪ/, /E/ and /U/ represent the vowels /ɪ/, /ɛ/ and /ʊ/).

Table 31. Rates of similarity between (i) the Euclidian distance (ED), and (ii) the mean duration values of AE and L2 vowel pairs; and paired t-tests of duration differences between the members of L2 pairs produced by the female L2 speakers.

Vowel pair	ED AE-L2	Duration AE-L2	t-test: duration between the vowels of the L2 pairs
/i/-ɪ/	27.1%	22.2%	t(10) = .441, p = .66
/ɛ/-æ/	22.5%	11.8%	t(10) = -.745, p = .47
/ʊ/-u/	35.3%	61.5%	t(10) = -1.986, p = .07

As previously stated, the Euclidean distances by the L2 speakers are much smaller than those by the AE monolinguals; thus, the former could have made use of duration to distinguish between the vowels of a pair. However, the difference in duration values between the vowels of each pair was not statistically significant, which indicates that they do not rely on duration to distinguish between the vowels.

Thus, although poor, the use of F1 and F2 to make distinctions between the L2 vowel pairs leads to the assumption that the participants rely *primarily* on spectral cues, since if rates of similarity to the AE monolinguals' production are taken into account,

the Euclidean distance rates were higher than the duration rates, although neither cue was used in a native-like fashion.

The lack of native-like reliance on temporal or spectral cues to produce vowels might be related to how the participants perceive vowels. The following subsection will report the results of the perceptual analysis.

4.6.2 Female L2 speakers' perception results

Table 32 shows the Euclidean distances in perception for both L2 and AE listeners and also the percentage of overlap in the perception of each of the three target pairs. Compared to the AE listeners' results, the L2 Euclidean distance in perception was 58.3% of that of the AE listeners for the /ɛ/-/æ/ pair. The results closest to those of the AE listeners were regarding the Euclidean distance between the L2 /i/-/ɪ/ (97.8%), followed by the /u/-/ʊ/ (83.9%). As regards the rate of overlap in perception, Figure 39 shows that the female L2 listeners' rates follow the same order as that of the AE listeners: /ɛ/-/æ/ has the lowest overlap rate, followed by /i/-/ɪ/ and /u/-/ʊ/.

Table 32. Euclidean distance (ED) values of AE and L2 vowel pairs perceived by the female participants (% of L2 similarity to AE values in parentheses), and overlap rate in the perception of AE monolinguals and L2 speakers.

Vowel pair	ED L2 (Hz)	ED AE (Hz)	Overlap L2	Overlap AE
/i/-/ɪ/	230 (97.8%)	235	F1: 49% F2: 55%	F1: 11% F2: 42%
/ɛ/-/æ/	344 (58.3%)	590	F1: 43% F2: 35%	F1: 0% F2: 11%
/u/-/ʊ/	122 (83.9%)	145	F1: 66% F2: 72%	F1: 0% F2: 71%

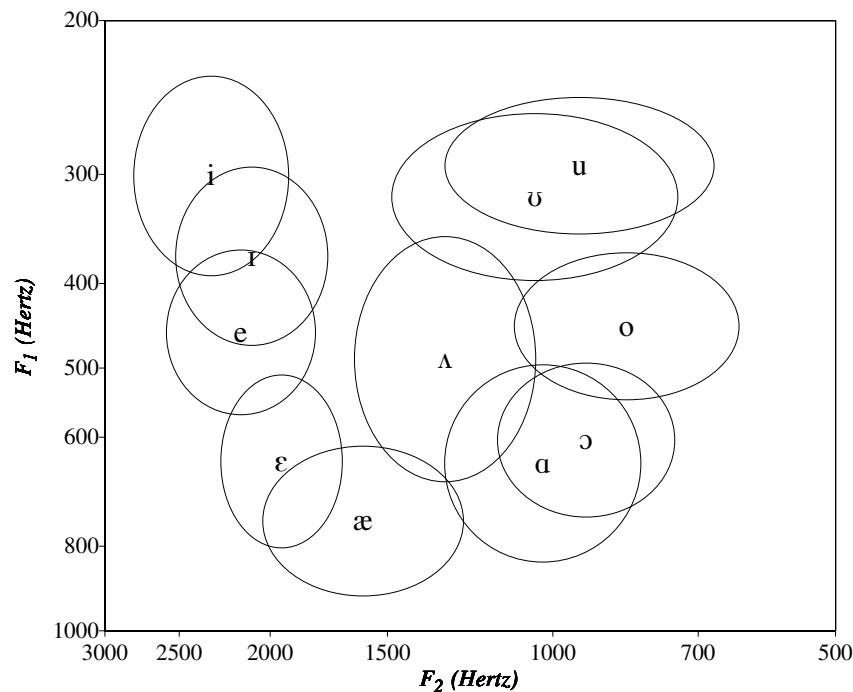


Figure 39. L2 perception results by the female participants (all durations included).

As regards the use of duration to perceive vowels, Table 33 indicates that the participants tended to make use of duration in a way similar to that of the AE listeners. Duration seems to be a cue used to distinguish between the L2 /i/-/ɪ/, since only 44% of the 100-ms /i/, and 61% of the 100-ms /ɪ/ were identified as /i/ and /ɪ/, respectively. This means that, since /i/ is longer than /ɪ/, the participants preferred to identify the 200-ms /i/ and the 100-ms /ɪ/ as /i/ and /ɪ/, respectively. As regards /ɛ/-/æ/, the participants tended to perceive 100-ms and 200-ms /ɛ/ at similar degrees (approximately 50% each); however, the 200-ms /æ/ was more often identified (60%) than its shorter version. The /ʊ/-/u/ vowels also seemed to have been perceptually distinguished by the L2 female participants by means of durational cues: They tended to identify the L2 100-ms /u/ less frequently (45%) than they identified the 100-ms /ʊ/ (57%).

Table 33. Percentage of short vowel (100 ms) identification by the female L2 speakers and AE monolinguals.

Vowel pair	L2 (%)	AE (%)
/i/	44	43
/ɪ/	61	68
/ɛ/	49	49
/æ/	40	49
/ʊ/	57	64
/u/	45	46

The production and perception results compared show that the Euclidean distances by the L2 female participants were smaller than those by the AE monolinguals for both; however, the Euclidean distance in perception was greater than in production (see Table 34 and Figure 40). Taking Euclidean distances into account, the vowel pairs were distinguished in the following order in production: /ʊ/-/u/ > /i/-/ɪ/ > /ɛ/-/æ/, and in perception /i/-/ɪ/ > /ʊ/-/u/ > /ɛ/-/æ/. Some asymmetry was observed when comparing the Euclidean distance differences in production and perception, since the closest to native-like pair in production (/ʊ/-/u/) was not the same in perception (/i/-/ɪ/). The L2 /ɛ/-/æ/ had the most symmetrical results, since it was the pair with the least native-like Euclidean distances in both production and perception.

Table 34. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the female L2 participants.

Vowel pair	Production ED AE-L2	Perception ED L2 (Hz)
/i/-/ɪ/	27.1%	97.8%
/ɛ/-/æ/	22.5%	58.3%
/ʊ/-/u/	35.3%	83.9%

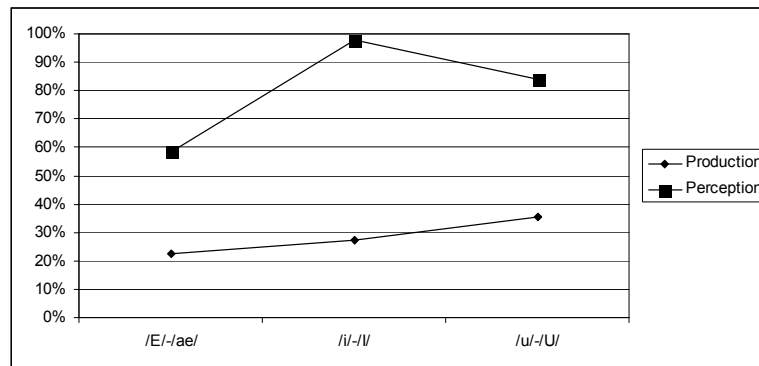


Figure 40. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the female L2 participants.

Thus, the results of the L2 female participants reveal that (i) the Euclidean distance of the L2 /ʊ/-/u/ pair in production is the one that most approximates that of the AE monolinguals (35.3%); (ii) the Euclidean distance of the L2 /i/-/ɪ/ pair in perception is the one that mostly approximates that of the AE listeners (97.8%); (iii) the Euclidean distance of the L2 /ɛ/-/æ/ pair was the least native-like for both production and perception; (iv) the participants make use of both temporal and durational cues to distinguish between the vowels of the L2 pairs in perception, but spectral cues are the primary cues in production, although they are poorly used. Thus, the results lead to the conclusion that perception outperforms production and that there is some interrelation between both: The least distinguished pair in production (/ɛ/-/æ/) was also the least distinguished in perception, and the most distinguished pair in production (/ʊ/-/u/) had high rates of distinction in perception (83.9%).

4.6.3 Male L2 speakers' production results

Differently from the results of the female participants, a significant difference between the Euclidean distance of vowels produced by the L2 speakers and AE monolinguals was found only for the /i/-/ɪ/ pair (see Table 35). The nonsignificant difference for /ɛ/-/æ/ was because Participants 4 and 7 produced a Euclidean distance greater than that of the AE monolinguals, and, in the case of /ʊ/-/u/, Participant 7 produced a Euclidean distance greater than 4 of the 5 AE monolinguals. Thus, the analysis of variance resulted in nonsignificant findings due to the outliers.

Similarly to the L2 women's results, the L2 /ɛ/-/æ/ were the least separated vowels (19.8%), a percentage even smaller than that of the L2 female speakers. The L2 /i/-/ɪ/ were the vowels closest to a native-like distinction (59.5%); however, the Euclidean distance was also significantly smaller than that of the AE monolinguals. No great distinction was made between the productions of the L2 /ʊ/-/u/, 25.7% of that of AE monolinguals. Again the L2 /i/-/ɪ/ were the vowels which most approximated the relative positions of those of the AE monolinguals; that is, /ɪ/ is lower and farther back than /i/ (see Figure 41). Table 36 and Figure 42 show the values of the BP and L2 vowels (with no normalization). The tables with the acoustic measurements of each male participant's BP and L2 vowels are in Appendix Q.

Table 35. Median values of the Euclidean distances for AE and L2 vowels, rates of L2 vs. AE similarity, and t-tests for the difference of each pair produced by the L2 male participants.

Vowel pair	AE	L2	L2 vs. AE similarity	t-test
/i/-/ɪ/	440	262	59.5%	t(10) = -3.199, p = .01
/ɛ/-/æ/	101	20	19.8%	t(10) = -1.293, p = .22
/ʊ/-/u/	249	64	25.7%	t(10) = -1.673, p = .12

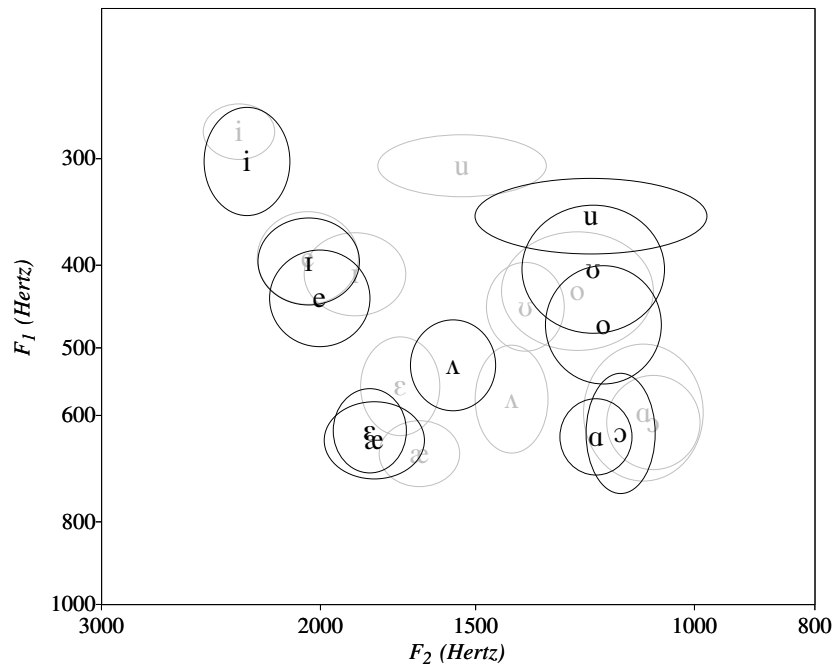


Figure 41. L2 (in black) and AE (in grey) normalized vowels produced by the L2 male speakers.

Table 36. Mean, median and SD of duration (D) (in milliseconds), f0, F1, F2 and F3 values (in Hertz) of L2 vowels produced by the male speakers.

		/i/	/ɪ/	/ei/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/oɪ*	/ʊ/	/u/
D	Mea	126	102	45	120	130	114	126	143	51	112	129
	Me	124	98	44	119	130	113	127	140	51	111	131
	SD	34	28	10	30	30	27	30	32	14	23	35
F0	Mea	135	131	128	123	115	130	120	134	138	132	142
	Me	130	128	132	122	117	129	120	125	132	128	134
	SD	42	41	39	46	39	42	36	62	25	33	36
F1	Mea	321	405	443	614	628	522	623	621	475	416	363
	Me	317	407	451	614	631	531	629	643	482	412	362
	SD	43	42	49	63	58	56	57	81	63	63	33
F2	Mea	2310	2041	1996	1797	1784	1512	1112	1053	1098	1127	1151
	Me	2296	2031	1994	1779	1772	1506	1110	1055	1104	1102	1138
	SD	202	209	206	136	184	136	88	80	138	175	282
F3	Mea	2848	2659	2648	2516	2512	2540	2464	2468	2528	2485	2453
	Me	2834	2648	2647	2505	2529	2518	2470	2495	2495	2471	2423
	SD	216	237	204	184	228	170	244	246	242	193	196

* Measurement values of the first element of the semi-diphthong.

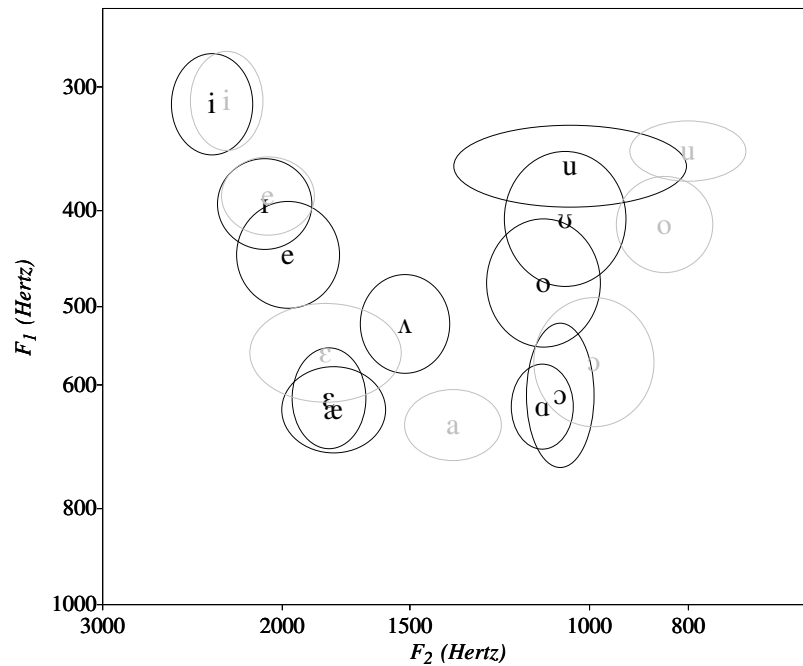


Figure 42. L2 (in black) and BP (in grey) vowels produced by the male participants.

As can be seen in Figures 41 and 42, there is no overlap between the vowels of the L2 /i/-/ɪ/ within 1 standard deviation from the mean, but as shown in Table 37, there is 22% of overlap of F2 values if all the values of the pair are taken into account. There is partial overlap between the L2 /ʊ/-/u/, but the greatest percentage of overlap is between the vowels of the L2 /ɛ/-/æ/ pair, since 75% and 74% of the F1 and F2 of /æ/, respectively, overlap with the F1 and F2 values of /ɛ/. Thus, similarly to the women's results, the greatest difference in percentage of overlap between the L2 speakers and the AE monolinguals in the three pairs is in the following order: /i/-/ɪ/ less overlapped than /ʊ/-/u/ than /ɛ/-/æ/. The male L2 speakers make use of F1 more than of F2 to distinguish between the high front and back vowels, and they make poor use of both F1 and F2 at similar degrees to distinguish between the low front vowels.

Table 37. Percentage of overlap between L2 and AE vowels by the male participants.

Vowel pair	AE speakers		L2 speakers	
	F1	F2	F1	F2
/i/-/ɪ/	0%	0%	0%	22%
/ɛ/-/æ/	18%	67%	75%	74%
/u/-/ʊ/	0%	23%	32%	62%

Compared to the BP vowels, paired-samples t-tests show that the L2 male speakers' /i/ is similar in height and backness to their BP /i/. As regards the L2 /ɪ/, it is significantly lower ($t(6) = 4.043$, $p = .007$) and farther back than the BP /i/ ($t(6) = -3.481$, $p = .013$). Similarly to the women, the L2 males' /ɪ/ is in between an L2 /i/ and an AE /ɪ/.

The L2 vowels of the /ɛ/-/æ/ pair overlap almost totally. The two L2 vowels are significantly lower (/ɛ/: $t(6) = 3.630$, $p = .011$; /æ/: $t(6) = 3.780$, $p < .01$) and farther back (/ɛ/: $t(6) = -2.563$, $p = .043$; /æ/: $t(6) = -3.805$, $p < .01$) than the BP /ɛ/.

A significant difference in height ($t(6) = -3.418$, $p = .013$), but no difference in backness was found between the L2 vowels of the /u/-/ʊ/ pair. The participants' L2 /u/ did not differ significantly in height from their BP /u/; however, the L2 /u/ was significantly more fronted than the BP /u/ ($t(6) = 4.543$, $p = .004$). Moreover, the L2 /ʊ/ was significantly lower and more fronted than the BP /u/ (F1: $t(6) = 3.716$, $p = .01$; F2: $t(6) = 9.468$, $p < .0001$).

As the L2 females, the L2 male participants did not make any native-like distinction in terms of spectral quality between the three target vowel pairs, the distance between the L2 pairs being significantly smaller than those of the AE pairs. There was great overlap of the low front vowels, and partial overlap of the high back vowels. However, the male participants' distinction between the members of the /i/-/ɪ/ pair was greater than that of females.

As regards duration, Table 38 and Figure 43 show that no significant duration differences were found between the L2 and AE vowels. The values of the spectral analysis indicate that the Euclidean distances by the L2 speakers are much smaller than those by the AE monolinguals, especially for /ɛ/-/æ/ and /ʊ/-/u/; thus, they could make use of durational cues to distinguish between the vowels of a pair. The difference rate of the duration values between the vowels of the L2 and the AE pairs is shown in Table 39. The vowels /ɪ/, /ɛ/ and /ʊ/ were indeed significantly shorter than /i/, /æ/ and /u/, respectively, following the AE tendency. Interestingly, compared to the AE monolinguals the L2 male speakers produced a somewhat native-like duration distinction for /i/-/ɪ/, a very different duration distinction for /ɛ/-/æ/, and the L2 /ʊ/-/u/ vowels were produced with more than double the duration values of those produced by the AE monolinguals.

Table 38. Mean duration values (in ms) of AE, L2 and BP vowels produced by the male speakers.

Vowel	AE	L2	BP	t-test AE vs. L2	t-test L2 vs. BP
/i/	140	126	79	t(10) = -.531, p = .60	t(6) = 3.964, p = .007
/ɪ/	118	102		t(10) = -1.338, p = .21	
/ɛ/	134	120	106	t(10) = -1.280, p = .22	t(6) = 1.592, p = .16
/æ/	179	130		t(10) = -3.598, p = .06	
/u/	135	129	83	t(10) = -.807, p = .43	t(6) = 4.962, p = .003
/ʊ/	128	112		t(10) = 1.405, p = .19	

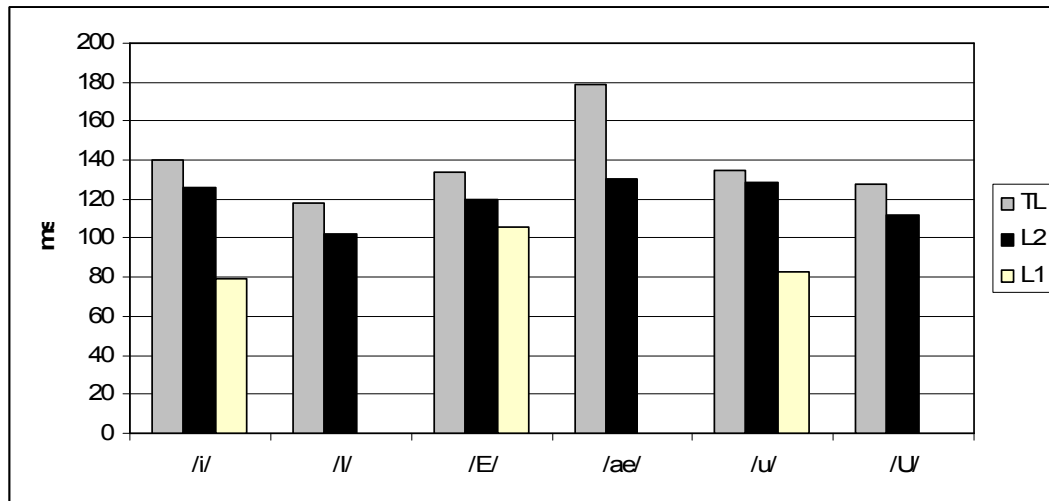


Figure 43. Mean duration values of English vowels produced by AE and L2 speakers, and of BP vowels produced by the L2 speakers.

Table 39. Rates of similarity between the Euclidian distance (ED) and mean duration values of AE and L2 vowel pairs, and t-tests of duration differences between the members of L2 pairs produced by the male L2 speakers.

Vowel pair	ED AE-L2	Duration AE-L2	t-test : duration between the vowels of L2 pairs
/i/-/ɪ/	56.4%	109.1%	$t(6) = 2.443, p < .05$
/ɛ/-/æ/	15.4%	22.2%	$t(6) = -3.316, p = .016$
/ʊ/-/u/	26.6%	242.8%	$t(6) = 2.651, p = .038$

Thus, the comparison of the Euclidean distance and duration differences between the production of the vowels of the target pairs by the L2 speakers and AE monolinguals indicates that the L2 /i/-/ɪ/ is the pair which mostly approximates native-like production in both spectral and durational cues. The partial overlap of the F1 and F2 values of the L2 /ʊ/-/u/, and the significant difference in duration between them indicate that the male L2 speakers make use of duration to differentiate between the members of this pair, but are not consistent in terms of spectral cues, since there is great standard deviation and some overlap in the productions of /ʊ/ and /u/. As regards /ɛ/-

/æ/, the great overlap between these vowels and the significant duration differences between them indicate that the L2 participants make use of durational cues *primarily* to distinguish between them. Thus, it is possible to conclude that the male L2 speakers do not rely mainly on spectral cues or durational cues to distinguish between the three target pairs: Cue reliance varies according to the pair, the two cues being used somewhat similarly to distinguish between high vowels, while duration is the main cue to distinguish between low vowels. Compared to BP vowels, all the L2 vowels were produced significantly longer. This may indicate two possibilities: (i) the L2 sentences were read slightly slower than the BP sentences, or (ii) the participants produce the L2 /i/ and /u/ significantly longer than their BP counterparts to contrast them with the short L2 vowels /ɪ/ and /ʊ/.

The next subsection will report the perception results by the male L2 speakers.

4.6.4 Male L2 speakers' perception results

As regards perception, the male results were similar to those of the females, but the Euclidean distance between the L2 vowels was slightly smaller than that of the females. Table 40 shows the Euclidean distances for both L2 and AE male listeners and also the percentage of overlap in the perception of each of the three target pairs. Compared to the AE monolinguals' results, the L2 Euclidean distance in perception was 50% of that of the AE listeners for the /ɛ/-/æ/ pair. The results closest to those of the AE listeners were regarding the Euclidean distance between the L2 /i/-/ɪ/ (94%), followed by the L2 /ʊ/-/u/ (67.5%). As regards the rate of overlap in perception, Figure 44 shows that, overall, the male L2 listeners' rates follow the same order as that of the

AE monolinguals: /ɛ/-/æ/ has the lowest overlap rate, followed by /i/-/ɪ/ and /u/-/ʊ/. However, there was slightly more overlap of the F1 values of the L2 /i/-/ɪ/ and more overlap of the F2 values of the L2 /ʊ/-/u/.

Table 40. Euclidean distance (ED) values of AE and L2 vowel pairs perceived by the male participants (% of L2 similarity to AE values in parentheses), and overlap rate in the perception of L2 and AE speakers.

Vowel pair	ED L2 (Hz)	ED AE (Hz)	Overlap L2	Overlap AE
/i/-/ɪ/	221 (94.0%)	235	F1: 68% F2: 51%	F1: 11% F2: 42%
/ɛ/-/æ/	295 (50.0%)	590	F1: 35% F2: 43%	F1: 0% F2: 11%
/ʊ/-/u/	98 (67.5%)	145	F1: 50% F2: 82%	F1: 0% F2: 71%

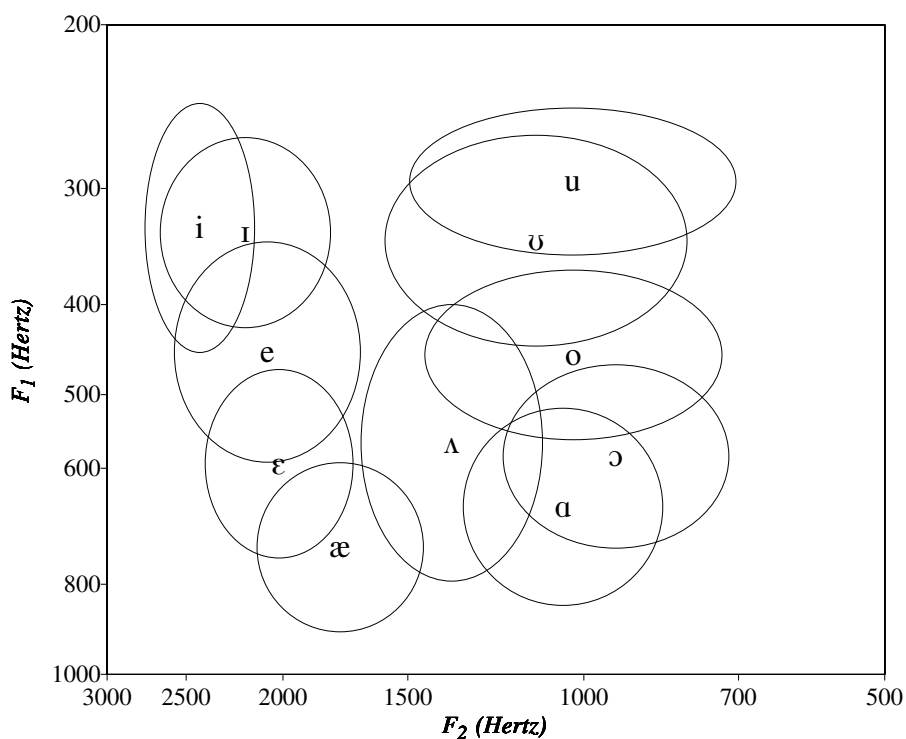


Figure 44. L2 perception results by the male participants (all durations included).

As regards the use of duration to perceive vowels, Table 41 indicates that the participants tend to make use of duration in a direction opposite to that of the AE monolinguals for the L2 /ʊ/-/u/ pair. The L2 males identified the 100-ms /u/ 67% of the times, while the 100-ms /ʊ/ was identified only 41% of the times. However, the participants tended to use duration as a cue to distinguish between /i/-/ɪ/, since only 33% of the 100-ms /i/, and 59% of the 100-ms /ɪ/ were identified as /i/ and /ɪ/, respectively. This means that, since /i/ is longer than /ɪ/, the participants preferred to identify the 200-ms /i/ and the 100-ms /ɪ/ as /i/ and /ɪ/, respectively. As regards /ɛ/-/æ/, the participants identified 56% of the 100-ms /ɛ/, and only 39% of the 100-ms /æ/. The participants seemed not to have relied on duration to distinguish between the L2 /ʊ/-/u/, since the L2 100-ms /u/ was more frequently (67%) identified than the 100-ms /ʊ/ (41%).

Table 41. Percentage of short vowel (100 ms) identification by the male L2 and AE speakers.

Vowel pair	L2 (%)	AE (%)
/i/	33	43
/ɪ/	59	68
/ɛ/	56	49
/æ/	39	49
/ʊ/	41	64
/u/	67	46

Similarly to the results of the L2 females, the production and perception results compared show that the Euclidean distances by the male L2 participants were smaller than those by the AE monolinguals for both; however, the Euclidean distance in perception was greater than that in production. Taking Euclidean distances into account,

there was more symmetry between the production and perception results by the male than by the female participants. The vowel pairs were distinguished in the following order in both production and perception: /i/-ɪ/ > /ʊ/-u/ > /ɛ/-æ/, as can be seen in Table 42 and Figure 45.

Table 42. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the male L2 participants.

Vowel pair	Production ED AE-L2	Perception ED L2 (Hz)
/i/-ɪ/	59.5%	94.0%
/ɛ/-/æ/	19.8%	50.0%
/ʊ/-/u/	25.7%	67.5%

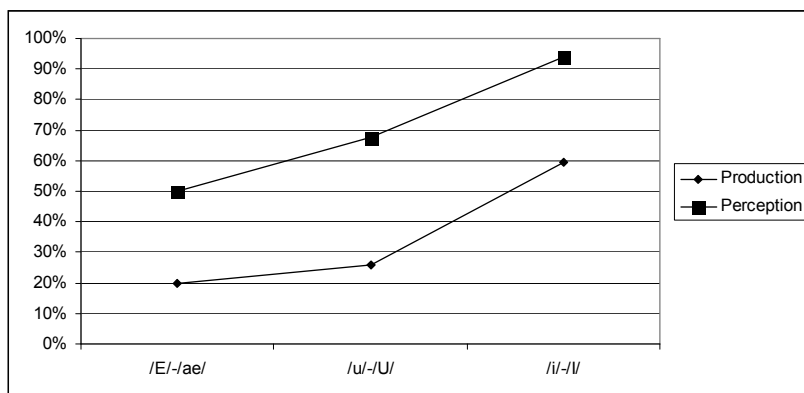


Figure 45. Rates of similarity of the Euclidean distances between the vowel pairs produced and perceived by the male L2 participants.

Thus, the results of the L2 male participants reveal that (i) the Euclidean distance of the L2 /i/-ɪ/ pair in both production and perception is the one that most approximates that of the AE monolinguals (59.5% in production and 94% in perception); (ii) the Euclidean distance of the L2 /ɛ/-æ/ pair was the least native-like for both production and perception; (iii) the participants make use of both temporal and spectral cues to perceive the L2 pairs, but only the perception results approximate those

of the AE monolinguals. Neither spectral nor temporal cues are used in a native-like fashion to produce the L2 vowel pairs. Thus, the better performance in perception corroborate the results found in the analysis of the female speakers and provides further indication that perception outperforms production. The male results were more consistent and showed that the vowels which were best distinguished in production were also best distinguished in perception.

4.7 Discussion about the production and perception results

Three research questions guided this study. The questions, hypotheses and the results were the following:

RQ1. Which L2 vowel pairs (/i/-/ɪ/, /ɛ/-/æ/, /ʊ/-/u/) will be more easily distinguished in both perception and production by the L2 speakers?

Bion et al. (2006) and Rauber et al. (2005) had already carried out a small-scale study that investigated the perception and production of English vowels by Brazilian advanced EFL speakers. The two studies found that the Brazilian participants neither perceived nor produced the L2 vowel pairs in a native-like fashion, but the L2 /i/-/ɪ/ was the pair whose distinction most approximated that of AE monolinguals in both perception and production, while /ɛ/-/æ/ was poorly perceived and produced. Based on these findings, it was hypothesized that the L2 participants in the present study would have the following order of difficulty to distinguish the pairs in either perception or production (least to most difficult): /i/-/ɪ/ < /ʊ/-/u/ < /ɛ/-/æ/. The hypothesis was partially corroborated by the L2 females' results, who produced the distinctions in the following order /ʊ/-/u/ < /i/-/ɪ/ < /ɛ/-/æ/. However, the hypothesis was totally

corroborated by the L2 males' results (see Table 43). As regards the perception results, the L2 females and males had a near native-like distinction between the vowels of the /i/-ɪ/ pair. For both perception and production, /ɛ/-/æ/ was the most poorly produced and perceived pair. The L2 male participants' results were more consistent, since they distinguished the vowel pairs in both perception and production in the following order: /i/-ɪ/ was better distinguished than /ʊ/-/u/, which was better distinguished than /ɛ/-/æ/.

Table 43. Rates of similarity of the Euclidean (ED) distances between the vowel pairs produced and perceived by the L2 female and male participants.

Vowel pair	Females		Males	
	Production ED AE-L2	Perception ED L2	Production ED AE-L2	Perception ED L2
/i/-ɪ/	27.1%	97.8%	59.5%	94.0%
/ɛ/-/æ/	22.5%	58.3%	19.8%	50.0%
/ʊ/-/u/	35.3%	83.9%	25.7%	67.5%

RQ2. What acoustic cues (spectral quality, duration) do the L2 speakers most rely on to perceive and produce BP and L2 vowels?

Based on Escudero (2001, 2002, 2005) and on Bohn (1995), it was hypothesized that the L2 speakers would rely more on duration than on spectral quality to perceive and produce the L2 vowels, while they would rely on spectral quality alone to both produce and perceive BP vowels. Concerning L2 vowel perception and production, the hypothesis was not corroborated. The female and male L2 participants made poor use of both temporal and spectral cues to produce the L2 pairs, only the perception results approximating those of the AE monolinguals. The results show that the participants do not rely primarily on either duration or spectral cues: Both cues are used, but in a

nonnative-like fashion. As regards BP perception/production, the participants relied primarily on spectral quality to both produce and perceive the seven BP oral vowels.

RQ3. Is there an interrelation between L2 vowel perception and production?

Based on Bion et al. (2006), Escudero (2005), Flege (1995) and Rauber et al. (2005), it was hypothesized that the vowel pairs which were perceived in a native-like fashion would also be produced in a native-like fashion, and the vowels which were misperceived would also be misproduced. The results indicate that perception outperforms production for both female and male L2 participants and that there is some interrelation between them. Some slight asymmetries involving the two were found for the female participants: The pair least distinguished in production was also the pair least distinguished in perception (/ɛ/-/æ/); however, the most distinguished pair in production (/u/-/ʊ/) was not the most distinguished pair in perception (/i/-/ɪ/). By contrast, the male results were more consistent and showed that the vowels which were distinguished better in production were also distinguished better in perception.

Although all the participants are experienced late learners and have taught English for about 8 years on average, the Euclidean distance for all three vowel pairs differed greatly from that of AE monolinguals in production. The poor results concerning the /ɛ/-/æ/ pair indicate that the L2 participants were not able to accurately pronounce vowels that are located in a space in the vowel system which is unoccupied by an L1 vowel (Bohn & Flege, 1992; Flege, 1987a, 1987b; Major, 1987). Thus, although the participants did not learn English in a natural setting, as was the case in the studies just mentioned, the findings in the present study do not corroborate those of Major (1987) or Bohn and Flege (1992), who concluded that the longer the exposure to the L2, the more accurate the production of the new vowel /æ/. The experienced L2

participants in the present study either overgeneralized the L2 /ɛ/ and /æ/ to /æ/, or merged the two L2 categories.

As regards the differences in the vowel inventories of two languages, Flege (1987b and elsewhere) claims that L2 speakers must learn how to produce L2 vowels that are either “new” or “similar” to an existing L1 vowel. His claim is that, since the speakers have already established patterns in their L1 of a vowel that has a similar L2 counterpart, it may be more difficult to modify such established patterns so as to produce the similar vowel in a native-like fashion due to the mechanism of equivalence classification. In the present study, although the “similar” vowels /ɪ/ and /ʊ/ were not produced in a native-like fashion, they were more easily distinguished from /i/ and /u/ than the “new” vowel /æ/ could be distinguished from /ɛ/. Major (1987) considered the English /æ/ as a “new” vowel for Brazilian EFL speakers, but the AE plots in this study show that /ɛ/-/æ/ is actually the vowel pair whose Euclidean distance is the smallest. This means that the F1 and F2 values of /ɛ/ and /æ/ differ less than those of /i/-/ɪ/ and /u/-/ʊ/. Thus, it seems controversial to consider /æ/ as “new”, and /ɪ/ and /ʊ/ as “similar”. The greatest difficulty to both perceive and produce the /ɛ/-/æ/ by the Brazilian L2 speakers provide evidence that /æ/ is in fact a “similar” and not a “new” vowel if the Portuguese and English vowel systems are compared.

The present study also provides evidence that the experienced L2 speakers did not tend to rely on their L1 acoustic parameters for all the pairs. They did not perceive or classify the L2 sounds according to the L1 system in the case of the L2 /i/-/ɪ/, whose Euclidean distance in perception by the L2 listeners was similar to that of the AE listeners. The L2 listeners did not perceive the L2 /u/-/ʊ/ as well as the L2 /i/-/ɪ/, but the Euclidean distances for these pairs approximated those of the AE monolinguals in more than 67%. However, the L2 participants seemed to have a “perceptual foreign accent”

(Strange, 1995, pp. 22, 39) in perceiving the vowels of the /ɛ/-/æ/ pair, since the Euclidean distances between the vowels of this pair was much smaller than that of the AE monolinguals, which may explain the difficulty to perceive the L2 contrast which does not exist in their L1 phonological space. In the present study, the vowels /ɪ/ and /ʊ/ tended not to be distinguished in a near native-like fashion in production, but they were in perception. This indicates that they were not perceptually confused with their BP /i/ or /u/ counterparts.

Taking into account Escudero's (2005) "scenarios" to help explain L2 speech perception, Brazilian EFL speakers are faced with the "new" scenario: The L2 has more categories than the L1, thus one L1 sound may have influenced the perception of two or more L2 sounds, as illustrated in Figure 46.

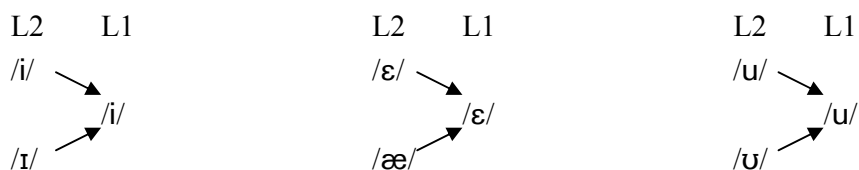


Figure 46. The "new" scenario for the three L2 target vowel pairs.

Some reflection about single category assimilation in production must be made. The L2 participants did not make a native-like distinction within any of the pairs, and neither of the vowels was produced as an L1 sound. The L2 values tended to be produced half way between an AE vowel and a BP vowel. Thus, it does not seem that two L2 vowels were assimilated as one BP vowel, since some distinction was made

between the /i/-/ɪ/ and /u/-/ʊ/ pairs. As regards the /ɛ/-/æ/ pair, the two vowels overlapped, but the overlapped values were significantly different from the BP /ɛ/.

With the analysis of the perception and production data of the BP and AE monolinguals, it was possible to have information about Escudero's (2005) *L2 initial state* and *L2 optimal perception/production*. The "new" scenario showed the differences in terms of phonetic categories of the L1 and the L2, or, as stated by Escudero, the "degree of mismatches" between the two perception grammars. In order to create new L2 categories or adjust the already established ones, the L2LP model hypothesizes that a process similar to that of L1 perception acquisition takes place in L2 perception acquisition: The L2 learner gradually adjusts his perceptual grammar so as to match the L2 perception common to optimal listeners. In the case of the L2 participants of the present study, most of them interact only with other nonnative speakers, which means that they are not exposed to "optimal" productions. This may explain why even highly proficient English speakers have difficulty to perceive and produce English contrasts in a native-like fashion. As regards the *L2 end state*, the L2LP model hypothesizes that in order for the L2 not to influence L1 categorization, both the L1 and L2 systems need to be different systems. Escudero claims that if enough optimal L1 and L2 input is provided to similar extents, both L1 and L2 perception will remain optimal; however, if there is an intermediate L1-L2 perception, it will be because both perceptions are activated simultaneously. The perception results by the L2 participants in this study show that their perception was close to optimal for the /i/-/ɪ/ pair, but it was intermediate in the other pairs. As previously stated, the L2 participants were not exposed to optimal input, a situation that is not discussed by the L2LP model, and they were not exposed to L1 and L2 to similar degrees, since they do not live in an English

speaking country. Thus, it is not surprising to find results that show their intermediate L1-L2 perception.

As regards cue weighting, in Escudero (2001, 2002), L2 Spanish speakers of Scottish English relied solely or primarily on durational cues to perceive the Scottish English contrast /i/-ɪ/, unlike native Scottish speakers, who relied primarily on spectral cues. The results of the present study show that neither durational nor spectral cues were being used in a native-like way, and there is not enough evidence to state that one is preferred to the other.

Thus, the perception and production results of the present study strongly corroborate those found by Bion et al. (2006) and Rauber et al. (2005), since in the three studies the perception tests revealed that the /i/-ɪ/ contrast was more easily discriminated than the /ɛ/-æ/ contrast, although the participants did not discriminate any contrast in a native-like way, that is, they needed a greater distance between the two vowels of a contrast to make a distinction between them. The production results in the three studies show that the participants produced a greater distance between /i/-ɪ/ than between /ɛ/-æ/. Thus, the results provide evidence that perception outperforms production, and also that perception and production are related, since greater discrimination in the perception test was related to better production results.

The next chapter will provide the conclusions of the present study, as well as some limitations and pedagogical implications.

CHAPTER 5

CONCLUSION

The aim of this study was to investigate how advanced Brazilian speakers of English perceive and produce L2 English vowels. Differently from other studies investigating vowel production and perception by EFL speakers (e.g., Baptista, 2000, 2006; Escudero, 2001, 2002; Flege et al., 1997, 1999), none of the participants in the present research had spent more than 8 weeks in an English speaking country; they all learned English in formal settings in Brazil.

In order to test L2 vowel production and perception, three groups of participants took production and perception tests, as shown in Table 44.

Table 44. Number of participants in each group: W = women, M = men.

Group	Production test	Perception test
Monolingual BP	12 (6 W, 6 M)	10 (5 W, 5 M)
Monolingual AE	9 (4 W, 5 M)	4 (2 W, 2 M)
L2 participants*	18 (11 W, 7 M)	18 (11 W, 7 M)

* Only this group of participants was the same to take production and perception tests.

In the perception test, the participants read sentences containing the vowels of the target pairs analyzed in this study: /i/-/ɪ/, /ɛ/-/æ/ and /u/-/ʊ/. In the perception test, the participants listened to a continuum of synthesized stimuli and had to identify the vowel heard in a forced-choice identification test.

The data analysis was carried out in order to answer the following questions:
RQ1. Which L2 vowel pairs (/i/-/ɪ/, /ɛ/-/æ/, /u/-/ʊ/) will be more easily distinguished in both perception and production by the L2 speakers?

Answer: Overall, the vowel pairs were distinguished in both perception and production in the following descending order: /i/-/ɪ/ > /ʊ/-/u/ > /ɛ/-/æ/.

RQ2. What acoustic cues (spectral quality, duration) do the L2 speakers most rely on to perceive and produce BP and L2 vowels?

Answer: It was not possible to conclude that the L2 participants relied mainly on spectral quality or duration to distinguish between the vowels of the pairs. The analyses simply indicate that for some vowels (especially /ɛ/-/æ/), the L2 participants relied on duration to distinguish between the vowels, while for the other vowels both spectral quality and duration were used. The results concerning cue reliance are inconclusive according to the analysis of the females' and males' data. What it is possible to conclude is that they do not use either durational or spectral cues to distinguish between the vowel pairs in a native-like fashion.

RQ3. Is there an interrelation between L2 vowel perception and production?

Answer: The results indicate that there seems to be an interrelation between perception and production and that the former precedes the latter, since the L2 participants' results approximated more closely those of the AE monolinguals in perception than in production (see Table 43). Overall, it is possible to state that the vowels which were well distinguished in perception (/i/-/ɪ/) were the ones also well distinguished in production, whereas the vowels poorly distinguished in perception (/ɛ/-/æ/) were also the ones produced with Euclidean distances that differed significantly from those of the AE monolinguals.

The findings in the present study lead to the conclusion that, in general, L2 perception/production theories (Flege, 1995; Escudero, 2005) are rather optimistic when

saying that with extensive exposure to the target language native-like pronunciation/perception may be acquired. Native-like performance seems to take place only in isolated cases, and maybe it is not possible or very unlikely to be acquired if one does not live in a country where the target language is the predominantly spoken language, as is the case of the Brazilian EFL speakers of this study.

All the L2 participants analyzed are graduate students in English, and they have been teaching English for a mean of about 8 years. Although very unlikely, an EFL student might not be interested in acquiring native-like pronunciation, but an EFL teacher should be aware that there are spectral quality/duration differences between English vowels that form minimal pairs, and call their learners' attention to these differences. Thus, since communication using English as an international language is unavoidable, both materials writers and EFL pronunciation teachers should focus on specific aspects that are difficult to acquire and are likely to cause communication problems. The mispronunciation of vowels is one aspect that may hinder or delay communication between interlocutors, thus some attention to vowel training should be given. A study that analyzed the effect of training on the perception and production of English vowels by Brazilian EFL speakers could be an interesting starting point towards a better understanding of how vowel instruction could be given for improvements in the learners' vowel perception and production to take place.

This study had several limitations. Maybe to have even more reliable vowel measurements, the target word should always be at the same position within a sentence, since this would result in production with a more constant pitch. Although the participants' productions were controlled and they were asked to reread sentences as many times as necessary so that their productions had constant speech rate and pitch, the difficulty in reading the sentences, especially the tricky BP sentences, led some

participants to feel bored easily. In many instances the participants were not consistent in reading the same vowel for the two target words in a sentence, so they would read, for instance: *Em pépe* ([pepe]) *e pépo* ([pepo]) *temos ê* ([e]). Some participants could not understand what the problem in their pronunciation was when they were corrected and asked to reread items, since they did not notice they were producing two different vowels within the same sentence. Thus, less tricky sentences could have facilitated the production data collection of BP vowels. The participants who read the English sentences were more consistent and rarely made confusions between the vowels of the target words, but again one target word per sentence would have resulted in less variation in pitch and duration, and would have facilitated the participants' reading of the carrier sentences.

Another limitation is that the L2 vowel production was compared to a single AE variety, the vowels from northern California. Future research could compare L2 vowel production with findings from other US regions, since it is not possible to control the AE language input the participants are exposed to in a non-English-speaking country.

As regards perception, one improvement in the perception test could be if the participants are given an option to click on a label that says "none of the vowels". Some of the vowels of the synthesized continuum are rather strange or nonexistent in either or both of the L1 or L2 inventories, so more reliable results could have been found if the participant had an option to choose a "none of the alternatives" label.

In sum, this study can be considered an interesting starting point to better understand the difficulties Brazilian EFL speakers have when perceiving and producing English vowels. Moreover, although it was not the main objective of the present research, the production and perception experiments carried out with BP

monolinguals show evidence that these speakers rely mainly on spectral cues to both perceive and produce vowels, information that, to my knowledge, had not been published yet. Still concerning the English and Portuguese monolinguals' results, the present study provides a rich database that can be used by other researchers in studies on vowels, since duration, f_0 , F1, F2, and F3 values are provided.

REFERENCES

- Anderson, N. (1978). On the calculation of filter coefficients for maximum entropy spectral analysis. *Modern Spectral Analysis, IEEE Press*, 252-255.
- Baptista, B. O. (2000). *The acquisition of English vowels by Brazilian Portuguese speakers*. Florianópolis: Universidade Federal de Santa Catarina.
- Baptista, B. O. (2006). Adult phonetic learning of a second language vowel system. In B. O. Baptista and M. A. Watkins (Eds.), *English with a Latin Beat: Studies in Portuguese/Spanish – English Interphonology* (pp. 19-40). Amsterdam: John Benjamins.
- Baptista, B. O., & Silva Filho, J. L. A. da (2006). The influence of voicing and sonority relationships on the production of English final consonants. In B. O. Baptista and M. A. Watkins (Eds.), *English with a Latin Beat: Studies in Portuguese/Spanish – English Interphonology* (pp. 73-89). Amsterdam: John Benjamins.
- Baratieri, J. P. (2006). Production of /l/ in the English coda by Brazilian EFL learners: An acoustic-articulatory analysis. Unpublished master's thesis. Universidade Federal de Santa Catarina.
- Barbetta, P. A. (2001) Estatística aplicada às ciências sociais. Florianópolis: Editora da UFSC.
- Barbosa, P. A. (1999). Revelar a estrutura rítmica de uma língua construindo máquinas falantes: pela integração de ciência e tecnologia da fala. In E. Scarpa (Ed.), *Estudos da prosódia* (pp. 21-52). Campinas: Editora da Unicamp.
- Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-203). Timonium, MD: York Press.
- Bion, R. A. H., Escudero, P., Rauber, A. S., & Baptista, B. O. (2006). Category formation and the role of spectral quality in the perception and production of English front vowels. Paper presented at the INTERSPEECH 2006, Pittsburgh.
- Boersma, P. (1998). *Functional phonology*. The Hague: Holland Academic Graphics.
- Boersma, P. (2000). The OCP in the perception grammar. *Rutgers Optimality Archive* 435.
- Boersma, P., P. Escudero, & R. Hayes (2003). Learning abstract phonological from auditory phonetic categories: An integrated model for the acquisition of language-specific sound categories. In M. J. Sole, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*, 1013-1016. Barcelona: Causal Productions.
- Boersma, P., & Weenink, D. (2006). *PRAAT: doing phonetics by computer (Version 4.4.23)* (Computer program), retrieved 12 June 2006, from <http://www.praat.org>.

- Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304). Timonium, MD: York Press.
- Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14, 131-158.
- Borden G. J., Harris, K. S., & Raphael L. J. (2002). *Speech science primer*. 4th ed. Baltimore: Williams & Wilkins.
- Bradlow, A. R. (1996). A perceptual comparison of the /i/-/e/ and /u/-/o/ contrasts in English and in Spanish: Universal and language specific aspects. *Phonetica*, 53, 55-85.
- Brown, J. D. (1988). *Understanding research in second language learning*. Cambridge University Press.
- Busà, M. G. (1992). *On the production of English vowels by Italian speakers with different degrees of accent*. Paper presented at New Sounds 92: 1992 Amsterdam Symposium of the Acquisition of Second-language Speech, Amsterdam.
- Clopper, C. G., Pisoni, D. P., & Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America* 118, (3), 1661-1676.
- Cornelian Jr., D. (2003). *Brazilian learners' production of initial /s/ clusters: phonological structure and environment*. Unpublished master's thesis. Universidade Federal de Santa Catarina.
- Cristófaros Silva, T. (2002). *Fonética e fonologia do português*. São Paulo: Contexto.
- De Faveri, C. (1991). *Análise da duração das vogais orais do português de Florianópolis-Santa Catarina*. Unpublished master's thesis. Universidade Federal de Santa Catarina.
- Delgado Martins, M. R. (1973). Análise acústica das vogais tónicas em português. *Separata do Boletim de Filologia*, 22, 303-314.
- Escudero, P. (2001). *The role of the input in the development of L1 and L2 sound contrasts: Language-specific cue weighting for vowels*. Paper presented at the Proceedings of the 25th Annual Boston University Conference on Language Development, Somerville, MA: Cascadilla Press.
- Escudero, P. (2002). *The perception of English vowel contrasts: Acoustic cue reliance in the development of new contrasts*. Paper presented at New Sounds 2000: Proceedings of the fourth international symposium on the acquisition of second language speech. Klagenfurt, Austria: University of Klagenfurt.
- Escudero, P. (2005). *Linguistic perception and second language acquisition* (Vol. 113). Utrecht: LOT.
- Escudero, P. (2005a). *Explaining L2 speech perception*. Unpublished post-doctoral project. University of Amsterdam.
- Escudero, P., & Polka, L. (2003). A cross-language study of vowel categorization and vowel acoustics: Canadian English vs. Canadian French. In M. J. Sole, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 861-864). Barcelona.

- Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton.
- Flege, J. E. (1987a). Effects of equivalence classification on the production of foreign language speech sounds. In A. James & J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 9-39). Dordrecht: Foris.
- Flege, J. E. (1987b). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15, 47-65.
- Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, 84, 70-79.
- Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theories* (pp. 249-290). Philadelphia: John Benjamins.
- Flege, J. E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.
- Flege, J. E. (1996). English vowel production by Dutch talkers: More evidence for the "similar" vs "new" distinction. In A. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 11-52). Berlin: Moulton de Gruyer.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.
- Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *Journal of the Acoustical Society of America*, 105 (5), 2973-2987.
- Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *Journal of the Acoustical Society of America*, 95(6), 3623-3641.
- Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125-3134.
- Hagiwara, R. (1997). Dialect variation and formant frequency: The American English vowels revisited. *Journal of the Acoustical Society of America*, 102 (1), 655-658.
- Harrington, J., & Cassidy, S. (1999). *Techniques in speech acoustics*. Dordrecht/Boston/London: Kluwer Academic Publishers.
- Hayward, K. (2000). *Experimental phonetics*. Harlow: Pearson Education.
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97 (5), 3099-3111.
- Huckvale, M. (2006). UCL Department of Phonetics and Linguistics web tutorial: Hearing – loudness. <<http://www.phon.ucl.ac.uk/resource/tutorials.html#phon>> retrieved on July 9, 2006.

- IBGE (2005). *Instituto Brasileiro de Geografia e Estatística - censos demográficos*. (Retrieved on July 5, 2006 from < <http://www.ibge.gov.br/cidadesat/default.php>>)
- Johnson, K. (2003). *Acoustic and auditory phonetics* (2 ed.). Malden/Oxford/Victoria: Blackwell.
- Koerich, R. D. (2002). Perception and production of word-final vowel epenthesis by Brazilian EFL students. Unpublished doctoral dissertation. Universidade Federal de Santa Catarina.
- Koerich, R. D. (2006). Perception and production of vowel paragoge by Brazilian EFL students. In B. O. Baptista & M. A. Watkins (Eds.), *English with a Latin Beat: Studies in Portuguese/Spanish – English Interphonology* (pp. 91-104). Amsterdam: John Benjamins.
- Kuhl, P. K. (1991). Human adults and human infants show a ‘perceptual magnetic effect’ for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics* 50, 93-107.
- Labov, W. (1998). The three dialects of English. In *Handbook of Dialects and Language Variation* (pp. 39–81), M. D. San Diego: Linn Academic Press.
- Ladefoged, P. (1993). *A course in phonetics*. 3rd edition. Fort Worth, Texas: Harcourt, Brace.
- Ladefoged, P. (1996). *Elements of acoustic phonetics*. 2nd edition. Chicago: University of Chicago Press.
- Ladefoged, P. (2003). *Phonetic data analysis: An introduction to field work and instrumental techniques*. Oxford: Blackwell Publishing Ltd.
- Larsen, E., & Aarts, R. M. (2004). *Audio bandwidth extension: Application of psychoacoustics, signal processing and loudspeaker design*. Wiley: West Sussex, England. (pp. 29-32)
- Leung, H. C., & Zue, V. W. (1984). A procedure for automatic alignment of phonetic transcriptions with continuous speech. In *Proceedings of the IEE ICASSP 1* (pp. 271-274). San Diego.
- Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology, speech perception, and acoustic phonetics*. Cambridge Studies in Speech Science and Communication. Cambridge: Cambridge University Press.
- Lima, R. (1991). *Análise acústica das vogais orais do português de Florianópolis*. Unpublished master’s thesis. Universidade Federal de Santa Catarina.
- Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.
- Major, R. C. (1987). Phonological similarity, markedness and rate of L2 acquisition. *Studies in Second Language Acquisition*, 9, 63-82.
- Major, R. C. (2002). The ontogeny and phylogeny of second language phonology. In J. Leather & A. James (Eds.), *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second Language Speech* (pp. 223-235). Klagenfurt: University of Klagenfurt.
- Master, S. (2005). *Análise acústica e perceptivo-auditiva da voz de atores e não atores masculinos: long term average spectrum e o “formante do ator”*. Unpublished doctoral dissertation, Universidade Federal de São Paulo.

- Master, S., Biase, N. D., Chiari, B. M., & Pedrosa, V. (2006). O espectro médio de longo termo na pesquisa e na clínica fonoaudiológica. *Pró-Fono Revista de Atualização Científica*, 18 (1), 111-120.
- Mateus, M. H. M., Andrade, A., Viana, M. C., & Villalva, A. (1990). Fonética, fonologia e morfologia do português. Lisboa: Universidade Aberta.
- Meador, D., Flege, J. E., & MacKay, I. R. A. (2000). Factors affecting the recognition of words in a second language, *Bilingualism: Language and Cognition*, 3, 55-67.
- Moraes, J. A. (1999). Um algoritmo para a correção/simulação da duração dos segmentos vocálicos em português. In E. Scarpa (Ed.), *Estudos da prosódia* (pp. 69-84). Campinas: Editora da Unicamp.
- Moraes, J. A., Callou, D., & Leite, Y. (1996). O sistema vocálico do português do Brasil: caracterização acústica. In M. Kato (Ed.), *Gramática do Português Falado*, 5 (pp. 33-53). Campinas: Editora da Unicamp.
- Morrison, G. S. (2002). Perception of English /i/ and /ɪ/ by Japanese and Spanish listeners: Longitudinal results. In G. S. Morrison & L. Zsoldos (Eds.), *Proceedings of the North West Linguistics Conference 2002* (pp. 29-48). Burnaby, BC, Canada: Simon Fraser University Linguistics Graduate Student Association.
- Morrison, G. S. (2006). L1 & L2 production and perception of English and Spanish vowels: A statistical modeling approach. Unpublished doctoral dissertation. University of Alberta, Edmonton, Alberta, Canada.
- Pereira, A. L. D. (2001). *Caracterização acústica do sistema vocálico tônico oral florianopolitano: alguns indícios de mudança*. Unpublished master's thesis. Universidade Federal de Santa Catarina.
- Peterson, G. E., & Barney H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America*, 24, 175-184.
- Pickett, J. M. (1999). *The acoustics of speech communication: Fundamentals, speech perception theory, and technology*. Needham Heights, MA: Allyn & Bacon.
- Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29, 191-215.
- Polka, L., Jusczyk, P. W., & Rvachew, S. (1995). Methods for studying speech perception in infants and children. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 49-89). Timonium, MD: York Press.
- Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint interaction in generative grammar*. Rutgers University Center for Cognitive Science Technical Report 2.
- Rauber, A. S. (2006). Production of English initial /s/-clusters by speakers of Brazilian Portuguese and Argentine Spanish. In B. O. Baptista and M. A. Watkins (Eds.), *English with a Latin Beat: Studies in Portuguese/Spanish – English Interphonology* (pp. 155-167). Amsterdam: John Benjamins.
- Rauber, A. S., Escudero, P., Bion, R. A. H., & Baptista, B. O. (2005). *The interrelation between the perception and production of English vowels by native speakers of Brazilian Portuguese*. Paper presented at the EUROSPEECH-INTERSPEECH 2005 9th European Conference on Speech Communication and Technology, Lisbon.

- Rebello, J. R., & Baptista, B. O. (2006). The influence of voicing on the production of initial /s/- clusters by Brazilian learners. In B. O. Baptista and M. A. Watkins (Eds.), *English with a Latin Beat: Studies in Portuguese/Spanish – English Interphonology* (pp. 139-154). Amsterdam: John Benjamins.
- Roca, I., & Johnson, W. (1999). *A course in phonology*. Massachusetts: Blackwell.
- Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Timonium, MD: York Press.
- Sada-Ribeiro, T. (2006). *Acoustic correlates of word stress production in the connected speech of American English and Brazilian Portuguese speakers*. Unpublished master's thesis. Universidade Federal de Santa Catarina.
- Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, 20, 213–223.
- Seara, I. C. (2000). *Estudo acústico-perceptual da nasalidade das vogais do português brasileiro*. Unpublished doctoral dissertation. Universidade Federal de Santa Catarina.
- SPSS for Windows 10.0.1 (1999). Computer program.
- Stevens, K. N. (1997). Articulatory-acoustic-auditory relationships. In W. J. Hardcastle, & L. Laver (Eds.), *The handbook of phonetic sciences*. Oxford: Blackwell.
- Strange, W. (1995). 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.
- Sundberg, J. (1987). *The science of the singing voice*. Illinois: Northern Illinois University Press.
- Trask, R. L. (1996). *A dictionary of phonetics and phonology*. London: Routledge.
- Trubetzkoy, N. (1939/1969). *Principles of Phonology*. Berkeley/ Los Angeles: University of California Press [1^a ed. in German: 1939].
- Vieira, M. N. (2004). Uma introdução à acústica da voz cantada. Paper presented at the *I Seminário Música Ciência Tecnologia: Acústica Musical* (pp. 70-79). São Paulo: Universidade de São Paulo.
- Wempe, T. (2001). F0-related formant measurements. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 24, 167-187.
- Wempe, T., & Boersma, P. (2003). The interactive design of an f0-related spectral analyzer. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam*, 25, 163-170.
- Werker, J. F., & Polka, L. (1993). Developmental changes in speech perception: New challenges and new directions. *Journal of Phonetics*, 21, 83-101.
- Woods, A., Fletcher, P., & Hughes, A. (1986). *Statistics in language studies*. Cambridge: Cambridge University Press.
- Zue, V. (2000). Automatic speech recognition, *Workshop 2000: Introduction to Speech and Language Processing*. Johns Hopkins University. Available at <http://www.clsp.jhu.edu/ws2000/presentations/preliminary/victor_zue/Zue-lecture2.pdf>. Retrieved on June 11, 2006.

APPENDIXES

Appendix A - Praat script to plot EP/BP vowels (linear scale)

#By Ricardo Bion

Select outer viewport... 0 10 0 8

Black

Font size... 18

Erase all

Axes... 3000 400 1000 100

Draw inner box

Marks bottom... 5 yes yes no

Marks left... 5 yes yes no

Text left... yes %F_%1 %(%H%e%r%t%z%)

Text bottom... yes %F_%2 %(%H%e%r%t%z%)

Text special... 2343 Centre 294 Half Times 24 0 i

Text special... 2084 Centre 403 Half Times 24 0 e

Text special... 1893 Centre 501 Half Times 24 0 \ef

Text special... 1602 Centre 511 Half Times 24 0 \at

Text special... 1326 Centre 626 Half Times 24 0 a

Text special... 994 Centre 531 Half Times 24 0 \ct

Text special... 864 Centre 426 Half Times 24 0 o

Text special... 678 Centre 315 Half Times 24 0 u

Appendix B - Praat script to vowels (logarithmic scale - log₁₀)

```
#Script to plot vowels (logarithmic scale): participants with different colors  
#By Ricardo Bion
```

```
clearinfo  
Erase all  
Select outer viewport... 0 8 0 6  
12  
  
max_F2 = log10(3500)  
min_F2 = log10(700)  
max_F1 = log10(1200)  
min_F1 = log10(250)  
  
12  
Black  
Line width... 1  
Plain line  
Axes... max_F2 min_F2 max_F1 min_F1
```

```
f = log10(250)  
One mark left... f no yes no 250  
f = log10(300 )  
One mark left... f no yes no 300  
f = log10(400 )  
One mark left... f no yes no 400  
f = log10(500)  
One mark left... f no yes no 500  
f = log10(600 )  
One mark left... f no yes no 600  
f = log10(700 )  
One mark left... f no yes no 700  
f = log10(800)  
One mark left... f no yes no 800  
f = log10(1000)  
One mark left... f no yes no 1000  
f = log10(1200)  
One mark left... f no yes no 1200  
f = log10(700)  
One mark bottom... f no yes no 700  
f = log10(1000)  
One mark bottom... f no yes no 1000  
f = log10(1500)  
One mark bottom... f no yes no 1500  
f = log10(2000)  
One mark bottom... f no yes no 2000  
f = log10(2500)  
One mark bottom... f no yes no 2500  
f = log10(3000)  
One mark bottom... f no yes no 3000  
f = log10(3500)  
One mark bottom... f no yes no 3500
```

```
Draw inner box
```

```
Text left... yes %F_%1 %(%H%e%r%t%z%)  
Text bottom... yes %F_%2 %(%H%e%r%t%z%)  
nr = Get number of rows  
for i to nr
```

```
study$ = Get value... i study
label$ = Get value... i vowel
f1 = Get value... i F1
f2 = Get value... i F2

if study$ = "1"
color_of_the_vowel$ = "Green"
elseif study$ = "2"
color_of_the_vowel$ = "Blue"
elseif study$ = "3"
color_of_the_vowel$ = "Red"
endif

call plot

endfor
procedure plot

f1 = log10(f1)
f2 = log10(f2)

'color_of_the_vowel$'
#Paint circle... 'color_of_the_vowel$' 'f2' 'f1' 120
#Draw circle... 'f2' 'f1' 60
#Draw circle... 'f2' 'f1' 61

Text special... 'f2' Centre 'f1' Half Times 18 0 'label$'
Plain line
Line width... 1
endproc
```


Appendix C - Background information questionnaire (Portuguese version)

Questionário para selecionar participantes para um estudo sobre o português brasileiro

Data: ____ / ____ / ____

Nome: _____

Fone: _____

E-mail: _____

Idade: _____ Local e data de nascimento: _____

Profissão: _____

Se estudante universitário, em que fase está: _____

Nome do curso: _____

1) Relacione as cidades e países para os quais você tenha viajado ou nos quais tenha morado por mais de duas semanas desde que nasceu:

Cidade e país: _____, Duração da estadia: _____

Cidade e país: _____, Duração da estadia: _____

Cidade e país: _____, Duração da estadia: _____

Cidade e país: _____, Duração da estadia: _____

Cidade e país: _____, Duração da estadia: _____

2) Onde os seus pais nasceram? Mencione a cidade.

a) Mãe: _____ b) Pai: _____

3) Na sua casa se fala outro(s) idioma(s) além do português? _____

Especifique qual(is) idioma(s): _____

4) No momento, você estuda algum idioma?

Especifique o(s) idioma(s) e nível (iniciante, intermediário, avançado):

Idioma: _____, Nível: _____

Idioma: _____, Nível: _____

5) Onde estuda o(s) idioma(s)? (Por exemplo: colégio, cursinho de idiomas, aulas particulares, etc.)

Idioma: _____, Lugar: _____

Idioma: _____, Lugar: _____

6) Quantas horas por semana você estuda o(s) idioma(s)?

Idioma: _____, Horas por semana: _____

Idioma: _____, Horas por semana: _____

7) Já estudou outro(s) idioma(s) anteriormente? _____

Especifique qual(is) idioma(s): _____

8) Que idade tinha quando começou a estudar outro(s) idioma(s)?

Idioma: _____, Idade: _____

Idioma: _____, Idade: _____

Idioma: _____, Idade: _____

9) Onde estudou outro(s) idioma(s)? (por exemplo: colégio, cursinho de idiomas, aulas particulares)

Idioma: _____, Lugar: _____

Idioma: _____, Lugar: _____

Idioma: _____, Lugar: _____

10) Quantas horas por semana você estudava o(s) idioma(s)?

Idioma: _____, Horas por semana: _____

Idioma: _____, Horas por semana: _____

Idioma: _____, Horas por semana: _____

11) Por quanto tempo estudou outro(s) idioma(s)?

Idioma: _____, Anos: _____

Idioma: _____, Anos: _____

Idioma: _____, Anos: _____

12) Se estudou em um cursinho de idiomas, até que nível chegou? _____

13) Indique, marcando o número correspondente, seu nível de **compreensão** no(s) idioma(s) que você estuda ou estudou. (0 significa que não entende nada; 5 significa que entende absolutamente tudo.)

Idioma: _____, 1 2 3 4 5

Idioma: _____, 1 2 3 4 5

Idioma: _____, 1 2 3 4 5

14) Indique, marcando o número correspondente, o quanto você pode **falar** no(s) idioma(s) que estudou/está estudando. (0 significa que não fala nada; 7 significa que fala perfeitamente, como um falante nativo deste idioma.)

Idioma: _____, 1 2 3 4 5

Idioma: _____, 1 2 3 4 5

Idioma: _____, 1 2 3 4 5

15) Você tem algum conhecido com quem fala em outro idioma fora das aulas?

Especifique a sua relação com essa pessoa (por exemplo: um amigo, uma tia, irmão, etc.):

Idioma: _____, Pessoa: _____

16) Quanto tempo em horas ou minutos por semana você fala em outro idioma fora de suas aulas de idiomas?

Idioma: _____, Horas ou minutos por semana: _____

17) Você assiste a programas de televisão em outros idiomas? _____

Especifique em qual(is) idiomas: _____

18) Quantas horas por semana você assiste televisão em outros idiomas?

Idioma: _____, Horas por semana: _____

19) Você escuta rádio ou música em outros idiomas? _____

Especifique _____ em _____ quais _____ idiomas:

20) Quantas horas por semana você escuta rádio ou música em outros idiomas?

Idioma: _____, Horas por semana: _____

Appendix D - Background information questionnaire (English version)

Questionnaire to select participants for a study on American English vowels

1. Name:
2. Age:
3. Place and date of birth:
4. Occupation:
5. (Under)graduate course:
6. In what cities have you lived/stayed for more than 2 weeks?
7. Where were your parents born?
8. Do you speak any foreign language (FL) with your family at home?
9. Are you studying any FL now?
IF YES,
 - a. Which one(s)?
 - b. What's your proficiency level (e.g., beginner, intermediate, advanced)?
 - c. Where do you study the FL(s) (e.g., school, private language school)?
 - d. How many hours a week do you study the FL(s)?
10. Have you already studied any foreign language?
IF YES,
 - a. What was your age when you started studying it (them)?
 - b. Where did you study the FL(s)?
 - c. How many hours a week did you study the FL(s)?
 - d. If you studied at a private language school, until what level did you study (beginner, intermediate, advanced)?
 - e. In a scale from 1 to 7 (one being little and 7, much), what's your oral comprehension level in the FL(s) you've studied?
 - f. In a scale from 1 to 7 (one being little and 7, much), what's your oral production level in the FL(s) you've studied?
 - g. Do you speak any FL language with someone outside the classroom? How often?
 - h. Do you watch TV in any FL? If yes, what languages and how many hours a week?
 - i. Do you listen to the radio or music in any FL? If yes, what languages and how many hours a week?

Appendix E - English sentences read by AE monolinguals and L2 speakers

Beat and Pete sound like seat.

Bot and pot sound like sot.

Tet and tech sound like kept.

Book and put sound like soot.

Tat and tack sound like cat.

Tot and tock sound like cot.

Boat and poach sound like soak.

Bate and pate sound like sate.

Bought and ought sound like sought.

Bat and pat sound like sat.

Book and took sound like cook.

Tote and toke sound like coat.

Bet and pet sound like set.

Bit and Pitt sound like sit.

Tate and take sound like Kate.

But and putt sound like shut.

Tit and tick sound like kit.

Boot and poop sound like suit.

Taught and talk sound like caught.

Tut and tuck sound like cut.

Teat and teak sound like keep.

Toot and tuke sound like coot.

Appendix F - Portuguese sentences read by BP monolinguals and L2 speakers

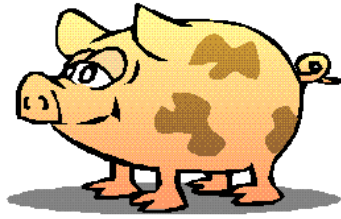
Em tique e tico temos i.
Em pêpe e pêpo temos ê.
Em cóque e cóco temos ó.
Em pupe e pupo temos u.
Em quique e quico temos i.
Em táque e táco temos a.
Em pôpe e pôpo temos ô.
Em susse e susso temos u.
Em fáfe e fáfo temos a.
Em cuque e cuco temos u.
Em téque e téco temos é.
Em fife e fifo temos i.
Em pépe e pépo temos é.
Em fófe e fófo temos ó.
Em sôsse e sósso temos ô.
Em quéque e quéco temos é.
Em fufe e fufo temos u.

Em têque e têco temos ê.
Em cáque e cáco temos a.
Em tóque e tóco temos ó.
Em pópe e pópo temos ó.
Em fêfe e fêfo temos ê.
Em côque e côco temos ô.
Em sésse e séssó temos é.
Em pápe e pápo temos a.
Em fôfe e fôfo temos ô.
Em sósse e sósso temos ó.
Em tôque e tôco temos ô.
Em quêque e quêco temos ê.
Em sisse e sisso temos i.
Em pipe e pipo temos i.
Em tuque e tuco temos u.
Em sásse e sásso temos a.
Em féfe e féfo temos é.

Appendix G - Pictures used in the English production test³⁰.



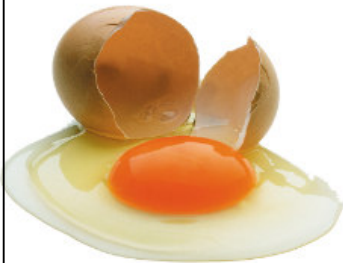
Bee [i]



Pig [ɪ]



Maid [eɪ]



Egg [ɛ]



Cat [æ]



Cup [ʌ]



Dog [ɑ]



Oranges [ɔ]



Boat [oʊ]



Books [ʊ]

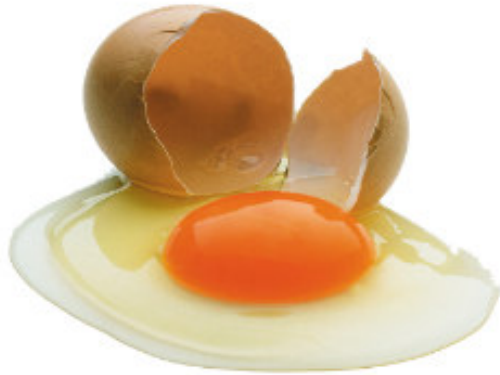


Boot [u]

³⁰ The word *dog* is pronounced with [ɑ] in Californian English, but it is pronounced as [ɔ] in other AE varieties.

Appendix H - Set used in the production practice test

First screen:



Following screen:

Red.

Web. Red and web sound like led.

Led.

Appendix I - Pitch measurement script

```
# By Paul Boersma

Read Table from table file... table5600.txt
numberOfRows = Get number of rows
assert numberOfRows = 1782
previousSpeaker$ = ""
for row to numberOfRows
    speaker$ = Get value... row speaker
    gender$ = Get value... row gender
    start = Get value... row start
    end = Get value... row end
    #
    # Be a bit economical with memory space.
    #
    if speaker$ <> previousSpeaker$
        if previousSpeaker$ <> ""
            select Sound 'previousSpeaker$'
            plus Pitch 'previousSpeaker$'
            Remove
        endif
        Read from file... 'speaker$'.wav
        previousSpeaker$ = speaker$
        pitchFloor = if gender$ = "M" then 60 else 120 fi
        To Pitch (ac)... 0 pitchFloor 15 no 0.03 0.45 0.01 0.35 0.14 400

    endif
    duration = end - start
    mid = start + duration / 2
    startpart = mid - duration / 5
    endpart = mid + duration / 5
    select Pitch 'speaker$'
    medianPitch = Get quantile... startpart endpart 0.5 Hertz
    #
    # Save results in table5600.
    #
    select Table table5600
    if medianPitch = undefined
        medianPitch = 0
    endif
    Set string value... row F0 'medianPitch:3'
endfor
Write to table file... table5600.txt
select Sound 'previousSpeaker$'
plus Pitch 'previousSpeaker$'
Remove
```


Appendix J - Formant measurements script

```
# Written by Paul Boersma and adapted by Andréia Rauber
# L2 English speakers - females

form Add reliable formants
    real Maximum_warping_(Hz) 1000
    positive Ceiling_step_(Hz) 10
endform
Read Table from tab-separated file... table5600multi.txt
Read Table from tab-separated file... table280.txt
Append column... ceiling
echo Results:
call doGender F
#call doGender M
procedure doGender gender$
    select Table table5600multi
    Extract rows where column (text)... gender "is equal to" 'gender$'
    Rename... gender
    call doDialect L2
    select Table gender
    Remove
endproc
procedure doDialect dialect$
    select Table gender
    Extract rows where column (text)... dialect "is equal to" 'dialect$'
    Rename... dialect
    for speaker to 11
        call doSpeaker speaker
    endfor
    select Table dialect
    Remove
endproc
procedure doSpeaker .speaker
    select Table dialect
    Extract rows where column (text)... speaker "is equal to" 'dialect$'_'gender$'_'speaker'
    Rename... speaker
    call doVowel \ae
    call doVowel \as
    call doVowel \ct
    call doVowel \ef
    call doVowel \hs
    call doVowel \ic
    call doVowel \vt
    call doVowel e
    call doVowel i
    call doVowel o
    call doVowel u
    select Table speaker
    Remove
endproc
procedure doVowel vowel$
    select Table speaker
    Extract rows where column (text)... vowel "is equal to" 'vowel$'
    Rename... vowel
    guessedFormantCeiling = if gender$ = "F" then 5500 else 5000 fi
    formantCeiling = guessedFormantCeiling - maximum_warping
    stdevBest = 1e300
    while formantCeiling <= guessedFormantCeiling + maximum_warping
```

```

    stdev = Get standard deviation... F2_'formantCeiling'
    if stdev < stdevBest
        formantCeilingBest = formantCeiling
        stdevBest = stdev
    endif
    formantCeiling += ceiling_step
endwhile
f1 = Get quantile... F1_'formantCeilingBest' 0.5
f2 = Get quantile... F2_'formantCeilingBest' 0.5
printline 'gender$' 'dialect$' 'speaker' 'vowel$' 'formantCeilingBest' 'stdevBest:1' 'f2:0'
select Table table280
    row = Search column... speaker 'dialect$' 'gender$' 'speaker'
    row += if vowel$ = "\as" then 11 else if vowel$ = "\ct" then 22 else
... if vowel$ = "\ef" then 33 else if vowel$ = "\hs" then 44 else if vowel$ = "\ic" then 55 else
... if vowel$ = "\vt" then 66 else if vowel$ = "e" then 77 else if vowel$ = "i" then 88 else
... if vowel$ = "o" then 99 else if vowel$ = "u" then 110 else 0 fi fi fi fi fi fi fi fi
    Set string value... row F1 'f1:3'
    Set string value... row F2 'f2:3'
    Set numeric value... row ceiling 'formantCeilingBest'
select Table vowel
Remove
endproc
select Table table280
Write to table file... table280_reliable.txt

```

Appendix K - Vowel generator script

```
# By Ton Wempe and Paul Boersma
# Generate synthetic vowels with duration, F1 and F2 steps
# Stores resulting sounds in specified directory

form Generate vowels (cascade mode) with duration, F1 and F2 steps
  positive Initial_F0_(Hz) 150
  positive Final_F0_(Hz) 100
  sentence Directory_to_write_to C:\vowels
  positive Minimum_duration_(ms) 100
  positive Maximum_duration_(ms) 200
  positive Number_of_duration_values 3
  positive Minimum_F1_(Hz) 240
  positive Maximum_F1_(Hz) 900
  positive Number_of_F1_values 14
  comment If F1 values are equal to or higher than F2 values the sounds are
  comment marked as "1" in the column "rep". The marked sounds are not generated!
  comment
  positive Minimum_F2_(Hz) 580
  positive Maximum_F2_(Hz) 2700
  positive Number_of_F2_values 10
endform

# calculate duration steps
logrange = log10(maximum_duration / minimum_duration)
logstep = logrange / (number_of_duration_values - 1)
for i to number_of_duration_values
  d'i' = minimum_duration * 10^((i-1)*logstep)
endfor

# calculate F1 values
if number_of_F1_values > 1
  maxmel = hertzToMel(maximum_F1)
  minmel = hertzToMel(minimum_F1)
  melrange = maxmel - minmel
  melstep = melrange / (number_of_F1_values - 1)
  for i to number_of_F1_values
    melvalue = minmel + (i-1) * melstep
    first'i' = melToHertz(melvalue)
  endfor
else
  first1 = minimum_F1
endif

# calculate F2 values
if number_of_F2_values > 1
  maxmel = hertzToMel(maximum_F2)
  minmel = hertzToMel(minimum_F2)
  melrange = maxmel - minmel
  melstep = melrange / (number_of_F2_values - 1)
  for i to number_of_F2_values
    melvalue = minmel + (i-1) * melstep
    second'i' = melToHertz(melvalue)
  endfor
else
  second1 = minimum_F2
endif

# initialize duration and formants table
```

```

numsounds = number_of_duration_values * number_of_F1_values * number_of_F2_values
Create TableOfReal... params numsounds 4
Set column label (index)... 1 rep
Set column label (index)... 2 f1
Set column label (index)... 3 f2
Set column label (index)... 4 dur

# generate sounds & update table
row = 0

for d to number_of_duration_values
  dur = d*d/1000
  for second to number_of_F2_values
    f2 = second*second'
    for first to number_of_F1_values
      rep = 0
      f1 = first*first'
      if f1 >= f2 - 100
        rep = 1
      endif
      select TableOfReal params
      row += 1
      Set row label (index)... 'row' 'first'_'second'_'d'
      Set value... row 2 f1
      Set value... row 3 f2
      Set value... row 4 dur
      if rep = 1
        Set value... row 1 rep
      endif
      call generate
      if rep = 0
        Write to WAV file... 'directory_to_write_to$\'first'_'second'_'d'.wav
      endif
      Remove
    endfor
  endfor
endfor
select TableOfReal params
Write to binary file... 'directory_to_write_to$\vowelparams.TableOfReal
Write to headerless spreadsheet file... 'directory_to_write_to$\vowelparams.txt

procedure generate

# Create voice source signal
Create PitchTier... sweep 0.0 dur
Add point... 0 initial_F0
Add point... dur final_F0
To PointProcess
Remove points between... 'dur'-0.005 'dur'
To Sound (phonation)... 44100 1 0.01 0.7 0.01 3 4

# Add some extra formants to get a flatter spectrum.
f3 = max (2500, f2 + 500)
f4 = max (3500, f3 + 400)
f5 = max (4000, f4 + 600)
f6 = f5 + 1000
f7 = f6 + 1000
f8 = f7 + 1000
f9 = f8 + 1000

```

```
f10 = f9 + 1000
for i to 10
  Filter with one formant (in-line)... f'i' sqrt(80^2+(f'i/20)^2)
endfor

# clear up
select PitchTier sweep
plus PointProcess sweep
Remove

select Sound sweep
Scale... 0.99

Endproc
```

Appendix L - Script to create perceptual vowel plots

```
#By Ricardo Bion

form Info
integer How_many_SDs: 1
choice Plot: 1
button Hz
button Barks
sentence Directory_to_read_from: C:\a\Perception_L2 speakers
endform

#####
max_F2 = 3000
min_F2 = 500
max_F1 = 1000
min_F1 = 200
Erase all
Select outer viewport... 0 10 0 8
Black
Line width... 1
Plain line
Font size... 18
Axes... log10(max_F2) log10(min_F2) log10(max_F1) log10(min_F1)

One logarithmic mark bottom... 500 yes yes no
One logarithmic mark bottom... 700 yes yes no
One logarithmic mark bottom... 1000 yes yes no
One logarithmic mark bottom... 1500 yes yes no
One logarithmic mark bottom... 2000 yes yes no
One logarithmic mark bottom... 2500 yes yes no
One logarithmic mark bottom... 3000 yes yes no
One logarithmic mark left... 200 yes yes no
One logarithmic mark left... 300 yes yes no
One logarithmic mark left... 400 yes yes no
One logarithmic mark left... 500 yes yes no
One logarithmic mark left... 600 yes yes no
One logarithmic mark left... 800 yes yes no
One logarithmic mark left... 1000 yes yes no

Draw inner box

Text left... yes %F_%1 %(%H%e%r%t%z%)
Text bottom... yes %F_%2 %(%H%e%r%t%z%)

#####

Create Strings as file list... filelist 'directory_to_read_from$\'*. *
nfiles = Get number of strings

select Strings filelist
file$ = Get string... 1
Read from file... 'directory_to_read_from$\'file$'

mfc$ = selected$("ResultsMFC", 1)

select ResultsMFC 'mfc$'
trials = Get number of trials

# get number of diferent labels

trials = Get number of trials
clearinfo

c_resp=1
response1$ = ""
for label to trials
  response$ = Get response... label
  new = 1
  for resp to c_resp
    if response$ = response'resp'$
      new = 0
```

```

endif
endifor
if new = 1
  c_resp = c_resp + 1
  response'c_resp'$ = response$
  res$ = response'c_resp'$
  printline 'res$'
endif
endifor
# this part on the top got all the labels the participant used in the MFC experiment

select ResultsMFC 'mfc$'
Remove

for file_i to nfiles
  select Strings filelist
  file$ = Get string... file_i
  Read from file...'directory_to_read_from$\file$'
endifor

select Strings filelist
Remove

# Get a number for each sound file
select all
for n_object to numberOfSelected("ResultsMFC")
object'n_object' = selected("ResultsMFC", n_object)
endifor

# this part initializes some dumb variables which will be used in the next part of the

for difresp from 2 to (c_resp)
  resp$ = response'difresp'$
  c'resp$' = 0
  f1'resp$' = 0
  f2'resp$' = 0
endifor

for difresp2 from 2 to (c_resp)
  resp$ = response'difresp2'$
  x=0
  if x=0
  for file to nfiles
    object = object'file'
    select 'object'
  for trial_c to trials

    filename$ = Get stimulus... trial_c
    response$ = Get response... trial_c
    p$ = filename$ - ".wav"
    if response$ = resp$
      c'resp$' = c'resp$' + 1
      c = c'resp$'
      call formantvalues
      f1'c' = log10(f1)
      f2'c' = log10(f2)
      f1'resp$' = f1'resp$' + f1'c'
      f2'resp$' = f2'resp$' + f2'c'
    v1 = f1'c'
    v2 = f2'c'
    v3 = c'resp$'
    printline 'response$' 'v1:1' 'v2:1'
    endif
  endfor
endifor

call get_mean_and_sd f1'resp$' f2'resp$' c'resp$'
call labels

if difresp2 = 2
  draw_grid = 1
else
  draw_grid = 0
endif

```

```

f1 = meanf1
f2 = meanf2
stdev_f1 = how_many_SDs*stdvf1
stdev_f2 = how_many_SDs*stdvf2

color_of_the_vowel$ = "Black"
line_of_the_sd$ = "Plain line"

call plot
endif
#####
endfor
endif
select all
Remove

procedure get_mean_and_sd totalf1 totalf2 numberv
  meanf1 = (totalf1/numberv)
  meanf2 = (totalf2/numberv)
  for sd to numberv
    for formant to 2
      sd'formant'sd' = (f'formant'sd' - meanf'formant')^2
    endfor
  endfor
  temp1 = 0
  temp2 = 0
  for sdn to numberv
    for formantn to 2
      temp'formantn' = temp'formantn' + sd'formantn'sdn'
    endfor
  endfor
  for formantx to 2
    stdvf'formantx' = (sqrt (temp'formantx'/(numberv-1)))
  endfor
endproc

procedure plot

Draw inner box

if plot = 2
Text left... yes %F_%1 %(%B%a%r%k%)
Text bottom... yes %F_%2 %(%B%a%r%k%)
else
Text left... yes %F_%1 %(%H%e%r%t%z%)
Text bottom... yes %F_%2 %(%H%e%r%t%z%)
endif

if f1 != undefined and f2 != undefined
'color_of_the_vowel$'
Text special... 'f2' Centre 'f1' Half Times 24 0 'label$'
Plain line
Line width... 1

if stdev_f2 = undefined
stdev_f2 = 0
endif
if stdev_f1 = undefined
stdev_f1 = 0
endif

x1 = 'f2'-'stdev_f2'
x2 = 'f2'+stdev_f2'
y1 = 'f1'+stdev_f1'
y2 = 'f1'-'stdev_f1'

'line_of_the_sd$'
Line width... 1
Draw ellipse... 'x1' 'x2' 'y1' 'y2'

endif

```



```

endproc

procedure labels
if resp$ = "E"
  label$ = "\ef"
elseif resp$ = "O"
  label$ = "\ct"
elseif resp$ = "a"
  label$ = "\ae"
elseif resp$ = "e"
  label$ = "e"
elseif resp$ = "i"
  label$ = "i"
elseif resp$ = "o"
  label$ = "o"
elseif resp$ = "u"
  label$ = "u"
elseif resp$ = "A"
  label$ = "\as"
elseif resp$ = "I"
  label$ = "\ic"
elseif resp$ = "U"
  label$ = "\hs"
elseif resp$ = "v"
  label$ = "\vt"
else
  label$ = resp$
endif
endproc

```

```

procedure formantvalues
if finename$ = "1_1_1.wav"
  f1= 239.99999999999997
  f2= 580.0000000000001

```

(And goes on with the specifications of the other 337 stimuli)

```

elseif finename$ = "14_10_3.wav"
  f1= 900.0000000000002
  f2= 2699.9999999999996
endif
endproc

```

Appendix M - Script to calculate the Euclidean distance in the perception test

```
#By Ricardo Bion

form Info
integer How_many_SDs: 1
choice Plot: 1
button Hz
button Barks
sentence Directory_to_read_from: C:\a\Perception_L2 males
endform

Erase all

Create Strings as file list... filelist 'directory_to_read_from$\*.*'
nfiles = Get number of strings

select Strings filelist
file$ = Get string... 1
Read from file... 'directory_to_read_from$\file$'

mfc$ = selected$("ResultsMFC", 1)

select ResultsMFC 'mfc$'
trials = Get number of trials

# get number of diferent labels

trials = Get number of trials
clearinfo

c_resp=1
response1$ = ""
for label to trials
  response$ = Get response... label
  new = 1
  for resp to c_resp
    if response$ = response'resp'$
      new = 0
    endif
  endfor
  if new = 1
    c_resp = c_resp + 1
    response'c_resp'$ = response$
    res$ = response'c_resp'$
  endif
endfor
# this part on the top got all the labels the participant used in the MFC

select ResultsMFC 'mfc$'
Remove

for file_i to nfiles
  select Strings filelist
  file$ = Get string... file_i
  Read from file... 'directory_to_read_from$\file$'
endfor

select Strings filelist
Remove

# Get a number for each sound file
select all
for n_object to numberOfSelected("ResultsMFC")
object'n_object' = selected("ResultsMFC", n_object)
endfor
printline resposta'tab$'F1'tab$'F2'tab$'duracao

for dur_x to 3
if dur_x = 1 or dur_x = 3 or dur_x = 2

# this part initializes some dumb variables which will be used in the next

for difresp from 2 to (c_resp)
```

```

resp$ = response'difresp$
c'resp$' = 0
f1'resp$' = 0
f2'resp$' = 0
endfor

for difresp2 from 2 to (c_resp)
resp$ = response'difresp2$
x=0
if x=0
for file to nfiles
object = object'file'
select 'object'
for trial_c to trials

filename$ = Get stimulus... trial_c
response$ = Get response... trial_c
p$ = filename$ - ".wav"
if response$ = resp$ and right$(p$, 1) = "dur_x"
c'resp$' = c'resp$' + 1
c = c'resp$'
call formantvalues
if plot = 1
f1'c' = f1
f2'c' = f2
f1'resp$' = f1'resp$' + f1'c'
f2'resp$' = f2'resp$' + f2'c'
else
f1'c' = hertzToBark(f1)
f2'c' = hertzToBark(f2)
f1'resp$' = f1'resp$' + f1'c'
f2'resp$' = f2'resp$' + f2'c'
endif
v1 = f1'c'
v2 = f2'c'
v3 = c'resp$'

printline 'response$'tab$v1:1"tab$v2:1"tab$dur_x'
endif
endfor
endfor

endif
#####
endfor
endif
endfor
select all
Remove

filedelete c:\ed.txt
fappendinfo c:\ed.txt
Read from file... c:\ed.txt
filedelete c:\ed.txt

clearinfo

Pool... resposta "" "F1 F2" "" ""
number_vowels = Get number of rows
Sort rows... resposta
for vo to 11
for formt to 2
vowel'vo$' = Get value... 'vo' resposta
f'formt'vo' = Get value... 'vo' F'formt'
endfor
endfor

printline pair'tab$'EDL2'tab$'EDNS'tab$%'
i1= sqrt(((f13 - f18)^2)+((f23 - f28)^2))
percent = 100*i1/235
printline i-1'tab$"i1:0"tab$'235'tab$"percent%"
eae= sqrt(((f12 - f16)^2)+((f22 - f26)^2))
percent = 100*eae/590
printline E-ae'tab$"eae:0"tab$'590'tab$"percent%"
uU= sqrt(((f110 - f15)^2)+((f210 - f25)^2))
percent = 100*uU/145

```

```

printline u-U'tab$'uU:0"tab$'145'tab$"percent"%

printline
printline vowel'tab$'responses'tab$'duracao'tab$'F1'tab$'F2'tab$'SDF1'tab$'SDF2'tab$'
for duracao to 3
select Table ed
Extract rows where column (number)... duracao "equal to" duracao
te = selected("Table")

for cvowel to number _vowels
select te
vt$ = vowel'cvowel'$
Extract rows where column (text)... resposta "is equal to" 'vt$'
tempv = Get number of rows
for ft to 2
ftemp'ft' = Get mean... F'ft'
sdtemp'ft' = Get standard deviation... F'ft'

endfor

tempv'vt$"duracao' = tempv
if duracao !=2
if vt$ = "i" or vt$ = "I" or vt$ = "E" or vt$ = "a" or vt$ = "U" or vt$ = "u"
printline 'vt$"tab$"tempv"tab$"duracao"tab$"ftemp1:0"tab$"ftemp2:0"tab$"sdtemp1:0"tab$"sdtemp2:0'
endif
endif
endif

endfor

select all
minus Table ed
Remove

#printline pair'tab$"difference

iI= (tempvi3 - tempvi1)+(tempvI1 - tempvI3)
#printline i-I'tab$"iI'
eae= (tempva3 - tempva1)+(tempvE1 - tempvE3)
#printline E-ae'tab$"eae'
uU= (tempvu3 - tempvu1)+(tempvU1 - tempvU3)
#printline U-u'tab$"uU'

printline
printline vowel'tab$"percentage_in_short_duration'tab$"natives
percent1 = 100*tempvi1/(tempvi3 + tempvi1)
#iI= (tempvi3 - tempvi1)+(tempvI1 - tempvI3)
printline i'tab$"percent1:0%"'tab$"43%
percent2 = 100*tempvI1/(tempvI1 + tempvI3)
printline I'tab$"percent2:0%"'tab$"68%
percent3 = 100*tempvE1/(tempvE1 + tempvE3)
printline E'tab$"percent3:0%"'tab$"49%
percent4 = 100*tempva1/(tempva3 + tempva1)
printline a'tab$"percent4:0%"'tab$"49%
percent5 = 100*tempvu1/(tempvu3 + tempvu1)
printline u'tab$"percent5:0%"'tab$"64%
percent6 = 100*tempvU1/(tempvU3 + tempvU1)
printline U'tab$"percent6:0%"'tab$"46%

printline
printline maaaaaybe one can calculate the use of duration for the i=I contrast
printline as the percentage of /i/ in the long duration plus
printline the percentage of /I/ in the short duration...
printline if duration is used, this number should be higher than 100
printline
printline cause in the end, comparing F1 and F2 does not say if participants used duration
printline rather, it says whether the vowel needs to be higher/lower/fronted/back
printline when it is short, and when it is long
printline sooooo....

iI = percent2+(100-percent1)
printline i-I 'iI'
eae = percent3+(100-percent4)
printline E-ae 'eae'
uU = percent6+(100-percent5)
printline U-u 'uU'

```

```

# i-I      16
# E-ae    0
# U-u     20

procedure get_mean_and_sd totalf1 totalf2 numberv
  meanf1 = (totalf1/numberv)
  meanf2 = (totalf2/numberv)
  for sd to numberv
    for formant to 2
      sd'formant'sd' = (f'formant'sd' - meanf'formant')^2
    endfor
  endfor
  temp1 = 0
  temp2 = 0
  for sdn to numberv
    for formantn to 2
      temp'formantn' = temp'formantn' + sd'formantn'sdn'
    endfor
  endfor
  for formantx to 2
    stdvf'formantx' = (sqrt (temp'formantx'/(numberv-1)))
  endfor
endproc

procedure labels
if resp$ = "E"
  label$ = "\e"
elseif resp$ = "O"
  label$ = "\o"
elseif resp$ = "a"
  label$ = "\ae"
elseif resp$ = "e"
  label$ = "\e"
elseif resp$ = "i"
  label$ = "\i"
elseif resp$ = "o"
  label$ = "\o"
elseif resp$ = "u"
  label$ = "\u"
elseif resp$ = "A"
  label$ = "\as"
elseif resp$ = "I"
  label$ = "\ic"
elseif resp$ = "U"
  label$ = "\us"
elseif resp$ = "v"
  label$ = "\vt"
else
  label$ = resp$
endif
endproc

procedure formantvalues

if finename$ = "1_1_1.wav"
  f1= 239.99999999999997
  f2= 580.0000000000001
elseif finename$ = "2_1_1.wav"
  f1= 277.77985604139104
  f2= 580.0000000000001

(And goes on with the specifications of the other 337 stimuli)
endif
endproc

```

Appendix N - Script to calculate the overlap between formant values of vowel pairs

```
#By Ricardo Bion
clearinfo

#I F1 317 449
#i F1 237 341
overlap = 341 - 317
rangeofF1 = 449 - 237
percentage = (100*overlap)/rangeofF1
printline The overlap of F1/I-i is 'percentage:0%'
printline

#E F1 472 612
#a F1 667 867
distance = 667 - 612

printline The overlap of F1/E-ae is 0%
printline They are 'distance'Hz distant
printline

#U F1 356 484
#u F1 250 350

printline The overlap of F1/U-u is 0%
printline They are 6Hz distant
printline

#i F2 2400 2750
#I F2 2013 2707
overlap = 2707 - 2400
rangeofF1 = 2750 - 2013
percentage = (100*overlap)/rangeofF1
printline The overlap of F2/I-i is 'percentage:0%'
printline

#E F2 1825 2345
#a F2 1119 1959
overlap = 1959 - 1825
rangeofF1 = 2345 - 1119
percentage = (100*overlap)/rangeofF1
printline The overlap of F2/E-ae is 'percentage:0%'
printline

#U F2 996 1688
#u F2 771 1749
overlap = 1688 - 996
rangeofF1 = 1749 - 771
percentage = (100*overlap)/rangeofF1
printline The overlap of F2/U-u is 'percentage:0%'
printline
```

Appendix O - Vowel normalization script

```
#By Ricardo Bion
form normalize
comment normalize each participant to a new max and min value
integer nmaxF1: 727
integer nminF1: 258
integer nmaxF2: 2483
integer nminF2: 981
endform

select all

Pool... speaker "" F1 "" ""
participants = Get number of rows

select all
tablex = selected("Table")

for i from 1 to participants
select tablex
Extract rows where column (number)... speaker "equal to" i
t'i = selected("Table")
endfor

count = 0
for y from 1 to participants
select t'y'

call other

for i from 1 to 162
count = count + 1
f1v'count' = Get value... 'i' F1
f2v'count' = Get value... 'i' F2
vo'count'$ = Get value... 'i' vowel
endfor
endfor

select tablex
count = 0
for y from 1 to 5
for i from 1 to 162
count = count + 1
vo$ = vo'count'$
Set string value... 'count' vowel 'vo$'
f1v = f1v'count'
Set numeric value... 'count' F1 'f1v'
f2v = f2v'count'
Set numeric value... 'count' F2 'f2v'
endfor
endfor

select all
minus tablex
Remove

procedure other
nvalues = Get number of rows

for formant to 2
```

```

for i from 1 to nvalues
v'i' = Get value... i F'formant'
endfor

# define max e min based on the max and min vowel_mean-+vowel_SD
table1 = selected("Table")
Pool... vowel "" "F1 F2" "" ""
nrows = Get number of rows
table2 = selected("Table")

for i from 1 to nrows
select table2
label$ = Get value... i vowel
printline 'label$'
value'i' = Get value... i F'formant'
t = value'i'
printline 't'
select table1
Extract rows where column (text)... vowel "is equal to" 'label$'
sd'i' = Get standard deviation... F'formant'
t = sd'i'
printline 't'
endfor

# define max e min
min = 999999
max = 0

for i from 1 to nrows

# max
temp = value'i' + sd'i'
if temp > max
max = temp
endif

# min
temp = value'i' - sd'i'
if temp < min
min = temp
endif
endfor

##### defined max and min
# convert these values to a scale from 0 to 1
for i from 1 to nvalues
normalized'i' = (v'i' - min)/(max-min)
endfor
printline 'newline$'

# convert to a new max and min
for i from 1 to nvalues
new = (nminF'formant')+(normalized'i'*(nmaxF'formant'-nminF'formant'))
select table1
Set numeric value... 'i' F'formant' 'new'
endfor

endfor
endproc

```


Appendix P - L2 female participants' L1 and L2 vowels

Table 1. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 1.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/u/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	113	85	31	117	130	98	120	163	54	103	124
	Me	126	90	29	124	132	103	132	178	52	103	128
	SD	26	16	9	23	28	18	23	32	15	16	20
f0	Mea	192	212	205	183	174	198	180	180	201	200	213
	Me	219	224	207	204	199	211	204	192	204	203	225
	SD	61	30	16	54	74	23	54	51	10	29	33
F1	Mea	373	417	508	679	684	571	730	745	542	428	417
	Me	367	421	507	674	677	574	734	754	544	427	432
	SD	35	23	44	25	35	33	43	38	33	27	48
F2	Mea	2595	2457	2048	1975	1989	1702	1350	1333	1332	1266	1328
	Me	2599	2460	1960	1966	1993	1708	1336	1321	1335	1254	1358
	SD	64	74	245	46	50	46	50	47	154	137	150
F3	Mea	3037	2951	2895	2916	2903	2834	2840	2892	2728	2565	2618
	Me	3062	2955	2958	2929	2898	2795	2860	2923	2774	2619	2600
	SD	118	114	198	64	52	138	152	149	193	154	86
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	89	116	132	135	136	114	95				
	Me	90	114	135	132	134	115	96				
	SD	5	12	12	12	7	7	7				
f0	Mea	252	224	212	210	208	228	253				
	Me	251	224	210	212	208	226	248				
	SD	12	8	9	9	8	7	11				
F1	Mea	388	465	636	751	696	511	442				
	Me	393	470	632	758	703	512	442				
	SD	22	18	22	21	25	35	15				
F2	Mea	2430	2228	2086	1638	1271	1090	1080				
	Me	2435	2238	2082	1634	1264	1004	1007				
	SD	88	100	67	85	77	155	133				
F3	Mea	2971	2819	2856	2733	2812	2561	2498				
	Me	2992	2845	2858	2792	2818	2520	2458				
	SD	274	82	28	148	111	134	107				

* Only the first element of the semi-diphthong was measured.

Table 2. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 2.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	150	166	60	184	199	166	190	197	67	140	157
	Me	149	161	60	185	201	162	179	194	66	142	155
	SD	19	21	8	27	22	26	24	32	15	20	26
F0	Mea	212	205	216	192	187	213	192	200	222	197	219
	Me	208	203	218	188	190	216	189	190	226	198	209
	SD	15	15	17	14	10	20	12	18	17	10	24
F1	Mea	430	601	639	780	907	687	783	845	623	548	429
	Me	426	592	638	778	924	680	778	860	622	545	421
	SD	25	40	22	41	58	42	31	49	26	23	35
F2	Mea	2685	2206	2162	2073	1947	1634	1139	1286	1303	1398	1581
	Me	2669	2204	2169	2082	1963	1647	1134	1310	1361	1396	1647
	SD	143	112	122	88	100	85	54	117	162	108	186
F3	Mea	3132	2704	2845	2429	2217	2771	2394	2462	2534	2572	2739
	Me	3156	2838	2858	2238	2071	2751	2349	2499	2582	2533	2766
	SD	377	269	236	342	315	161	135	133	140	81	111
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	92	120	148	151	148	114	95				
	Me	92	127	147	156	142	111	92				
	SD	11	22	14	24	20	18	16				
F0	Mea	240	227	210	211	207	221	234				
	Me	244	224	203	208	204	220	240				
	SD	23	14	13	13	18	19	20				
F1	Mea	405	461	720	932	734	532	440				
	Me	402	449	722	936	736	529	438				
	SD	28	26	27	36	31	38	38				
F2	Mea	2750	2513	2242	1642	1077	1030	1072				
	Me	2734	2527	2214	1631	1045	972	984				
	SD	48	82	91	96	102	122	211				
F3	Mea	3242	2645	2537	2266	2460	2555	2736				
	Me	3296	2608	2394	2318	2426	2600	2792				
	SD	389	155	314	245	150	175	156				

* Only the first element of the semi-diphthong was measured.

Table 3. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 3.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	113	127	56	144	153	140	159	179	57	114	131
	Me	106	127	57	138	146	136	157	185	56	116	132
	SD	25	15	8	19	22	20	26	32	12	16	17
F0	Mea	234	243	235	225	204	210	222	228	248	241	250
	Me	242	250	232	224	224	235	222	237	251	250	265
	SD	52	60	15	32	65	68	26	26	18	59	62
F1	Mea	375	429	476	654	629	521	650	722	501	443	450
	Me	372	450	469	640	637	518	648	724	506	436	451
	SD	38	79	29	48	55	26	38	25	32	54	53
F2	Mea	2613	2422	2499	2332	2367	1586	1006	1092	1070	1001	1086
	Me	2597	2420	2515	2345	2384	1636	1016	1097	1038	985	1065
	SD	71	243	146	63	56	150	76	56	109	113	197
F3	Mea	3164	2978	3078	2863	2898	2701	2654	2633	2620	2674	2654
	Me	3117	2989	3047	2847	2895	2686	2689	2689	2631	2668	2652
	SD	102	202	139	98	102	88	137	144	149	97	44
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	86	114	120	134	140	116	83				
	Me	84	110	122	140	142	118	79				
	SD	19	12	23	22	18	16	21				
F0	Mea	237	216	192	192	200	220	237				
	Me	232	212	190	190	200	218	232				
	SD	26	21	13	15	12	20	31				
F1	Mea	384	433	609	850	659	449	408				
	Me	390	430	604	858	670	441	412				
	SD	27	40	29	77	38	30	31				
F2	Mea	2578	2454	2326	1476	990	925	681				
	Me	2574	2432	2331	1468	992	876	698				
	SD	52	38	35	111	60	125	121				
F3	Mea	3142	2991	2903	2466	2492	2542	2619				
	Me	3161	3002	2902	2514	2484	2539	2594				
	SD	183	112	51	213	60	111	157				

* Only the first element of the semi-diphthong was measured.

Table 4. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 4.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	74	82	35	110	143	88	109	110	42	102	71
	Me	73	75	34	102	151	85	100	108	35	104	73
	SD	21	24	9	28	25	16	31	36	15	16	15
F0	Mea	183	225	222	211	178	196	167	162	226	235	147
	Me	231	243	225	228	212	231	220	223	228	235	149
	SD	91	67	14	62	94	86	106	106	15	43	86
F1	Mea	423	509	525	695	799	560	767	723	545	508	444
	Me	440	496	551	706	797	552	768	713	544	522	442
	SD	64	50	64	31	38	37	56	32	30	64	41
F2	Mea	2326	2166	2171	2016	1938	1652	1253	1174	1005	1150	1277
	Me	2292	2152	2140	2037	1945	1680	1259	1190	998	1169	1331
	SD	111	143	116	64	42	64	57	80	63	117	204
F3	Mea	2858	2849	2889	2705	2763	2824	2468	2743	2905	2755	2846
	Me	2838	2840	3008	2820	2849	2855	2674	2742	2899	2716	2862
	SD	224	151	254	412	346	142	362	351	212	202	95
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	65	94	107	118	111	89	69				
	Me	64	93	110	117	112	85	64				
	SD	12	12	12	17	12	14	12				
F0	Mea	243	209	213	214	215	220	238				
	Me	246	213	212	214	216	220	248				
	SD	21	34	11	14	10	15	42				
F1	Mea	321	451	682	802	605	489	358				
	Me	319	441	686	797	600	498	361				
	SD	19	41	23	31	46	42	19				
F2	Mea	2443	2300	2172	1584	959	905	853				
	Me	2440	2309	2183	1582	942	890	802				
	SD	107	91	55	105	62	92	137				
F3	Mea	2900	2980	2902	2736	2786	2876	2687				
	Me	2843	2994	2886	2802	2864	2860	2736				
	SD	195	168	105	130	264	130	259				

* Only the first element of the semi-diphthong was measured.

Table 5. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 5.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	146	170	63	174	194	180	193	207	64	152	169
	Me	143	162	56	173	193	173	201	213	60	151	172
	SD	32	23	20	24	16	28	21	22	16	17	20
F0	Mea	236	225	245	214	202	229	207	210	242	219	239
	Me	239	233	246	215	205	232	210	207	242	218	242
	SD	32	37	9	19	20	26	28	27	13	22	38
F1	Mea	421	527	520	815	890	621	779	800	521	492	471
	Me	410	524	511	802	881	619	768	785	518	489	492
	SD	66	45	42	82	62	44	61	36	29	39	58
F2	Mea	2345	2103	1845	1963	1969	1529	1169	1171	1142	1208	1255
	Me	2355	2074	1874	1969	1986	1495	1181	1176	1169	1208	1264
	SD	99	101	347	82	88	135	41	55	135	93	120
F3	Mea	2552	2481	2514	2348	2429	2351	2448	2347	2634	2524	2354
	Me	2376	2608	2436	2251	2596	2413	2427	2386	2640	2502	2357
	SD	361	274	296	261	290	209	84	275	81	127	94
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	127	154	173	195	185	168	144				
	Me	130	154	163	199	188	166	150				
	SD	32	16	27	22	27	20	22				
F0	Mea	252	232	218	216	220	236	255				
	Me	255	232	218	214	219	234	254				
	SD	14	9	10	11	6	10	8				
F1	Mea	305	460	687	813	689	488	379				
	Me	305	456	673	798	682	491	372				
	SD	14	15	61	60	25	19	34				
F2	Mea	2356	2163	1974	1415	1071	957	849				
	Me	2354	2177	1992	1424	1070	930	798				
	SD	30	98	59	72	30	91	119				
F3	Mea	2748	2629	2165	2272	2206	2363	2395				
	Me	2657	2672	2033	2286	2224	2366	2404				
	SD	304	192	272	62	98	42	71				

* Only the first element of the semi-diphthong was measured.

Table 6. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 6.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	140	151	53	172	171	149	183	157	62	151	171
	Me	139	148	57	174	176	153	187	142	58	146	179
	SD	30	22	13	33	22	25	32	33	20	35	27
F0	Mea	239	226	243	212	202	233	202	215	246	228	240
	Me	243	230	248	213	212	238	213	217	243	228	256
	SD	36	24	13	31	60	23	60	24	17	31	32
F1	Mea	402	559	524	828	853	640	940	855	650	505	469
	Me	412	556	518	813	854	635	951	829	651	496	493
	SD	48	45	50	60	52	25	43	75	26	115	47
F2	Mea	2708	2336	2391	2200	2210	1822	1356	1304	1414	1351	1164
	Me	2734	2316	2376	2202	2209	1837	1359	1305	1375	1395	1148
	SD	105	49	145	53	47	118	62	32	156	207	114
F3	Mea	3300	3019	2828	2938	2886	2887	2839	2850	2927	2763	2744
	Me	3322	3018	2927	2986	2947	2873	2826	2827	2924	2740	2726
	SD	167	58	304	218	225	103	92	180	130	92	69
BP												
D	N	10	10	10	10	10	10	10	10	10	10	10
	Mea	104	123	143	141	142	122	113				
	Me	106	124	143	140	140	120	110				
	SD	14	10	12	7	14	11	12				
F0	Mea	255	238	223	224	230	238	258				
	Me	254	238	226	224	232	240	259				
	SD	16	9	6	9	12	6	8				
F1	Mea	434	482	701	975	754	517	423				
	Me	438	480	681	976	746	510	419				
	SD	27	17	41	41	51	23	37				
F2	Mea	2730	2475	2299	1639	1134	965	810				
	Me	2714	2456	2306	1627	1132	930	786				
	SD	69	90	49	96	74	99	81				
F3	Mea	3337	3003	2994	2790	2644	2850	2749				
	Me	3335	3040	3088	2804	2650	2780	2760				
	SD	101	209	214	114	102	172	70				

* Only the first element of the semi-diphthong was measured.

Table 7. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 7.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	143	132	49	196	202	154	206	208	60	156	164
	Me	154	133	49	190	197	155	192	205	62	152	154
	SD	24	21	8	25	25	16	27	22	10	26	33
F0	Mea	213	206	205	192	186	206	187	187	210	198	208
	Me	207	202	203	181	181	201	181	184	208	196	209
	SD	18	24	10	19	18	16	17	16	18	14	26
F1	Mea	391	390	589	828	790	656	826	770	628	403	428
	Me	405	392	586	835	817	651	822	775	625	404	432
	SD	38	21	35	57	64	36	31	25	44	27	38
F2	Mea	2499	2500	2131	2005	2059	1636	1187	1068	1192	1049	1069
	Me	2507	2509	2106	2005	2054	1662	1181	1074	1227	1026	1086
	SD	97	98	153	46	80	58	53	66	188	101	157
F3	Mea	3040	2980	2912	2767	2668	2723	2722	2734	2798	2681	2691
	Me	3075	3000	2907	2762	2735	2743	2688	2687	2783	2686	2682
	SD	158	225	77	66	229	97	100	208	152	71	122
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	81	109	134	115	123	108	82				
	Me	80	112	128	113	122	103	88				
	SD	11	16	14	13	13	11	12				
F0	Mea	216	196	189	186	188	207	231				
	Me	218	198	189	184	188	206	236				
	SD	12	9	8	8	7	13	18				
F1	Mea	399	486	687	823	727	525	444				
	Me	398	488	691	820	731	522	446				
	SD	15	28	28	19	28	26	27				
F2	Mea	2381	2303	1982	1560	1105	916	856				
	Me	2365	2292	1986	1560	1106	890	817				
	SD	65	47	66	66	41	98	163				
F3	Mea	2892	2857	2683	2313	2622	2664	2666				
	Me	2954	2869	2680	2348	2554	2648	2668				
	SD	240	89	57	207	223	79	113				

* Only the first element of the semi-diphthong was measured.

Table 8. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 8.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	115	120	61	172	167	137	146	165	52	123	142
	Me	118	103	46	173	154	116	144	149	49	126	157
	SD	32	45	38	44	36	43	36	62	13	30	44
F0	Mea	192	190	202	168	192	189	179	162	208	202	183
	Me	208	207	199	180	182	184	178	179	208	197	206
	SD	64	59	24	73	28	25	57	70	22	41	79
F1	Mea	383	464	484	744	766	563	793	752	504	424	390
	Me	391	471	473	738	756	561	784	752	509	420	385
	SD	47	32	36	36	45	26	47	49	24	45	40
F2	Mea	2679	2303	2363	2070	2054	1563	1181	1138	1052	946	1017
	Me	2701	2322	2338	2057	2055	1587	1173	1138	1075	960	985
	SD	138	128	153	80	71	112	57	39	114	86	206
F3	Mea	3118	2791	2628	2309	2414	2824	2456	2444	2782	2790	2901
	Me	3172	2851	2554	2249	2227	2837	2483	2431	2745	2830	2816
	SD	217	170	324	295	347	60	163	165	107	161	227
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	58	73	92	99	97	73	56				
	Me	60	70	90	98	96	70	54				
	SD	13	9	6	9	8	10	10				
F0	Mea	255	226	206	205	210	228	259				
	Me	250	222	206	204	212	226	258				
	SD	14	12	12	15	12	17	15				
F1	Mea	386	455	670	910	723	471	428				
	Me	376	447	662	914	714	466	426				
	SD	40	21	26	31	40	34	39				
F2	Mea	2593	2324	2044	1598	1100	960	892				
	Me	2574	2344	2054	1614	1100	926	866				
	SD	148	151	82	68	43	86	228				
F3	Mea	2931	2620	2321	2330	2227	2478	2719				
	Me	2967	2670	2343	2353	2224	2473	2739				
	SD	257	188	200	150	142	119	256				

* Only the first element of the semi-diphthong was measured.

Table 9. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 9.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	121	98	60	161	154	132	153	170	60	108	118
	Me	116	89	63	172	157	131	151	163	63	107	117
	SD	32	21	10	38	16	24	14	36	10	16	20
F0	Mea	224	243	208	196	194	210	189	189	208	238	261
	Me	238	262	205	200	214	202	205	202	205	240	275
	SD	49	49	13	25	34	32	28	26	13	37	48
F1	Mea	354	360	473	757	724	573	778	796	473	471	455
	Me	334	344	488	762	723	581	781	815	488	475	461
	SD	73	41	30	54	41	60	38	68	30	56	44
F2	Mea	2558	2567	2376	1972	1988	1718	1229	1173	2376	1210	1385
	Me	2628	2571	2387	1961	1988	1735	1237	1222	2387	1139	1436
	SD	209	68	106	67	47	55	39	109	106	148	214
F3	Mea	2965	2953	2792	2551	2450	2623	2395	2426	2792	2629	2652
	Me	3004	2939	2804	2528	2575	2593	2394	2402	2807	2632	2654
	SD	164	94	177	55	270	94	86	162	177	98	72
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	85	98	112	103	116	95	80				
	Me	84	99	114	105	121	98	81				
	SD	12	8	14	9	18	13	7				
F0	Mea	253	219	213	202	207	234	252				
	Me	253	224	215	198	207	234	254				
	SD	10	16	11	10	8	12	13				
F1	Mea	320	453	656	846	702	493	449				
	Me	316	456	659	854	694	490	461				
	SD	13	27	23	43	41	18	44				
F2	Mea	2537	2239	2000	1590	1164	1000	1027				
	Me	2558	2251	1991	1565	1178	982	996				
	SD	124	135	85	99	48	87	152				
F3	Mea	3087	2637	2678	2326	2242	2567	2685				
	Me	3122	2568	2682	2379	2260	2574	2684				
	SD	190	143	154	194	70	109	152				

* Only the first element of the semi-diphthong was measured.

Table 10. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 10.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	109	127	53	149	153	147	149	162	59	135	132
	Me	110	125	52	153	154	147	150	154	59	132	132
	SD	14	23	10	31	22	25	24	31	8	16	28
F0	Mea	209	213	194	180	180	196	151	179	205	203	216
	Me	214	221	196	182	176	206	169	183	203	202	226
	SD	38	33	12	28	21	19	63	22	11	36	32
F1	Mea	383	403	480	700	695	603	676	768	511	457	473
	Me	393	407	479	709	710	607	682	771	513	460	474
	SD	38	42	30	34	44	35	47	54	31	26	25
F2	Mea	2554	2571	2254	2068	2122	1694	1187	1266	1165	1132	1402
	Me	2564	2585	2259	2071	2122	1694	1203	1253	1203	1138	1380
	SD	101	111	120	94	71	53	95	68	126	139	234
F3	Mea	3201	3175	2763	2674	2657	2818	2552	2545	2856	2749	2813
	Me	3140	3119	2874	2705	2772	2803	2567	2518	2860	2738	2823
	SD	172	157	378	275	329	78	162	231	78	87	58
BP												
D	N	10	10	10	10	10	10	10	10	10	10	10
	Mea	96	130	156	145	176	137	105				
	Me	86	130	158	150	179	136	100				
	SD	28	24	20	20	39	17	20				
F0	Mea	215	206	187	179	179	205	226				
	Me	216	207	184	174	174	198	229				
	SD	15	28	14	21	20	17	21				
F1	Mea	390	448	637	874	714	468	427				
	Me	392	442	631	894	694	476	430				
	SD	21	37	30	56	50	19	31				
F2	Mea	2549	2410	2191	1571	1066	887	967				
	Me	2558	2433	2194	1563	1040	908	950				
	SD	100	126	82	67	107	59	155				
F3	Mea	3110	2869	2809	2592	2548	2731	2730				
	Me	3142	2940	2839	2594	2582	2716	2720				
	SD	233	240	171	77	143	99	110				

* Only the first element of the semi-diphthong was measured.

Table 11. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 11.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	101	97	46	164	143	112	158	152	44	98	100
	Me	102	94	47	172	146	111	159	151	43	99	97
	SD	19	13	10	31	19	17	28	30	7	17	17
F0	Mea	230	221	239	220	216	216	216	198	238	257	238
	Me	252	256	239	218	211	222	215	220	243	254	234
	SD	71	98	16	23	26	29	29	62	34	37	50
F1	Mea	478	463	519	757	757	590	832	827	517	497	458
	Me	492	466	515	749	756	573	844	829	516	500	456
	SD	50	54	29	53	54	37	62	56	39	58	73
F2	Mea	2504	2472	2368	2084	2103	1705	1199	1233	1088	1190	1258
	Me	2492	2459	2389	2081	2068	1692	1211	1261	1057	1201	1315
	SD	55	65	156	66	100	86	145	112	141	140	297
F3	Mea	3011	2980	3004	2867	2708	2956	2778	3035	2949	2878	2903
	Me	3029	2985	2998	2968	2974	2953	2895	3022	2967	2885	2892
	SD	101	55	77	322	434	58	406	77	46	36	60
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10	10	10	10	10
	Mea	119	158	183	188	182	143	121				
	Me	123	160	189	187	186	144	128				
	SD	16	17	18	17	21	7	18				
F0	Mea	230	216	209	211	215	227	217				
	Me	232	213	214	213	215	227	245				
	SD	17	17	18	17	11	18	89				
F1	Mea	436	466	714	954	845	506	467				
	Me	433	463	706	938	834	509	475				
	SD	32	28	42	66	47	32	43				
F2	Mea	2597	2384	2234	1654	1164	953	855				
	Me	2622	2372	2218	1644	1167	961	821				
	SD	103	62	48	100	46	54	102				
F3	Mea	3106	2819	2837	2813	2916	2873	2905				
	Me	3090	2985	2993	2918	2899	2895	2922				
	SD	94	307	320	423	140	94	48				

* Only the first element of the semi-diphthong was measured.

Appendix Q - L2 male participants' L1 and L2 vowels

Table 1. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 12.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/u/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	93	94	36	107	130	108	108	129	47	115	111
	Me	98	94	33	109	131	112	113	118	42	116	104
	SD	19	27	9	31	37	23	21	28	15	20	39
F0	Mea	106	104	125	108	108	128	98	171	127	102	121
	Me	112	112	127	104	111	116	99	119	127	108	120
	SD	41	24	13	69	13	52	21	104	11	19	18
F1	Mea	366	434	432	586	625	515	609	657	458	433	362
	Me	349	437	419	597	634	540	630	659	471	450	369
	SD	52	38	52	41	52	69	54	23	76	81	35
F2	Mea	2116	1828	2031	1713	1686	1494	1193	1006	1011	1317	1366
	Me	2126	1822	2019	1705	1681	1489	1167	1028	1005	1307	1424
	SD	76	80	150	70	74	71	81	68	141	133	227
F3	Mea	2621	2512	2568	2409	2418	2384	2362	2478	2448	2370	2416
	Me	2634	2503	2588	2392	2391	2367	2331	2459	2415	2379	2400
	SD	84	75	166	77	97	110	93	76	117	154	119
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	78	91	109	108	100	91	65				
	Me	76	88	98	112	98	88	60				
	SD	7	26	21	10	15	26	11				
F0	Mea	158	148	135	126	138	148	157				
	Me	156	147	136	134	138	147	156				
	SD	9	11	10	22	5	11	7				
F1	Mea	291	342	543	688	552	342	298				
	Me	291	345	549	688	549	345	302				
	SD	20	30	23	20	14	30	16				
F2	Mea	2259	2152	1837	1262	901	2152	897				
	Me	2260	2147	1868	1271	885	2147	851				
	SD	58	129	137	96	86	129	182				
F3	Mea	2750	2477	2493	2132	2219	2477	2422				
	Me	2719	2363	2459	2130	2256	2363	2422				
	SD	206	278	160	160	100	278	140				

* Only the first element of the semi-diphthong was measured.

Table 2. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 13.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	137	95	46	124	130	110	134	159	51	109	134
	Me	144	96	41	120	129	112	128	159	51	108	131
	SD	23	14	10	13	17	11	14	19	14	10	24
F0	Mea	122	113	107	136	104	118	118	148	124	124	123
	Me	133	128	122	122	114	130	121	124	124	123	128
	SD	18	36	45	53	34	17	12	95	13	14	18
F1	Mea	323	423	453	670	676	563	688	700	489	482	375
	Me	330	419	473	666	680	555	690	690	500	488	371
	SD	19	43	64	25	43	32	32	42	50	43	14
F2	Mea	2182	1967	1898	1725	1699	1488	1154	1078	1037	1126	846
	Me	2154	1978	1892	1729	1699	1504	1172	1086	992	1158	925
	SD	71	85	99	60	35	79	86	31	162	108	159
F3	Mea	2820	2579	2646	2557	2559	2548	2449	2454	2595	2468	2517
	Me	2841	2611	2683	2551	2555	2585	2399	2453	2511	2505	2456
	SD	208	82	87	68	71	111	169	161	265	121	222
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	56	66	70	76	79	67	65				
	Me	57	66	68	76	75	68	65				
	SD	8	6	8	8	14	6	9				
F0	Mea	119	124	116	113	115	116	129				
	Me	128	121	117	114	112	114	129				
	SD	43	8	8	5	11	8	9				
F1	Mea	298	394	569	688	591	438	347				
	Me	299	396	568	686	585	433	346				
	SD	8	27	16	28	15	41	12				
F2	Mea	2196	1932	1720	1356	1007	892	871				
	Me	2178	1928	1710	1330	1008	890	827				
	SD	65	93	72	92	88	89	161				
F3	Mea	2725	2643	2500	2342	2264	2450	2452				
	Me	2710	2654	2477	2326	2266	2390	2410				
	SD	209	214	63	159	174	226	233				

* Only the first element of the semi-diphthong was measured.

Table 3. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 14.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	106	88	41	140	142	111	143	141	50	107	112
	Me	111	91	42	139	142	114	144	139	50	107	114
	SD	20	14	7	14	14	17	15	18	12	12	29
F0	Mea	146	140	149	127	117	146	124	123	155	143	156
	Me	145	153	148	130	123	154	121	125	155	150	181
	SD	32	36	15	22	25	28	23	28	18	22	40
F1	Mea	322	363	473	662	677	477	657	668	476	347	351
	Me	330	353	477	667	677	475	658	665	485	352	356
	SD	43	30	31	12	19	18	28	27	40	34	20
F2	Mea	2382	2263	2102	1978	2006	1697	1108	1123	1098	921	914
	Me	2381	2272	2098	1979	2007	1731	1100	1156	1110	951	921
	SD	67	62	216	64	39	156	68	71	135	108	179
F3	Mea	2906	2769	2601	2467	2443	2486	2192	2103	2481	2557	2386
	Me	2848	2735	2691	2508	2582	2495	2200	2047	2482	2535	2398
	SD	154	143	240	192	297	61	88	117	170	138	88
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	102	113	145	140	145	126	101				
	Me	103	114	142	144	151	127	102				
	SD	18	18	20	13	16	11	15				
F0	Mea	180	161	146	142	148	168	178				
	Me	180	156	146	139	140	167	174				
	SD	22	19	17	15	21	18	24				
F1	Mea	354	402	619	664	633	412	345				
	Me	358	398	624	660	629	412	336				
	SD	28	20	18	23	28	16	27				
F2	Mea	2352	2214	2043	1523	1006	782	747				
	Me	2344	2205	2058	1569	1002	760	709				
	SD	53	40	68	95	63	87	83				
F3	Mea	2989	2644	2488	2038	2105	2460	2391				
	Me	3012	2692	2550	2022	2132	2450	2413				
	SD	169	191	200	78	102	26	110				

* Only the first element of the semi-diphthong was measured.

Table 4. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 15.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/u/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	150	95	47	105	110	105	108	127	50	105	143
	Me	150	92	48	102	107	105	109	123	46	108	147
	SD	25	14	6	22	22	16	21	24	10	21	29
F0	Mea	106	121	106	98	95	103	96	104	115	103	110
	Me	105	108	107	97	94	103	95	105	114	108	109
	SD	12	36	4	7	2	7	5	9	13	15	12
F1	Mea	298	412	460	509	656	528	612	554	468	406	334
	Me	296	420	463	509	665	526	618	543	485	428	330
	SD	12	27	21	16	40	31	26	58	52	43	16
F2	Mea	2102	1762	1720	1624	1491	1332	1078	978	1168	1002	1042
	Me	2093	1734	1735	1629	1501	1346	1072	981	1183	1004	976
	SD	82	63	85	52	42	52	63	67	142	59	169
F3	Mea	2770	2328	2433	2390	2416	2424	2689	2532	2373	2260	2215
	Me	2756	2374	2422	2408	2480	2430	2665	2511	2271	2265	2199
	SD	97	173	93	127	281	110	99	93	270	70	147
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	88	104	107	105	103	104	82				
	Me	91	105	108	104	100	101	84				
	SD	20	12	10	14	15	12	10				
F0	Mea	124	118	118	115	113	126	126				
	Me	124	105	118	115	113	126	126				
	SD	12	7	9	7	8	4	6				
F1	Mea	274	352	487	634	487	372	322				
	Me	276	352	490	636	482	372	322				
	SD	10	15	21	32	20	11	20				
F2	Mea	2009	1815	1477	1203	1046	819	831				
	Me	2014	1819	1604	1178	901	795	802				
	SD	45	77	276	86	352	70	83				
F3	Mea	2793	2265	2451	2526	2412	2434	2410				
	Me	2744	2277	2403	2540	2426	2418	2323				
	SD	133	178	176	68	96	72	303				

* Only the first element of the semi-diphthong was measured.

Table 5. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 16.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	142	148	53	156	165	156	152	170	59	132	153
	Me	136	152	52	153	161	151	148	174	57	129	149
	SD	14	20	5	21	19	26	16	21	7	19	12
F0	Mea	122	120	130	110	111	121	116	118	126	119	123
	Me	127	122	131	116	115	125	117	118	125	119	126
	SD	12	13	9	16	16	10	10	9	8	10	11
F1	Mea	307	421	438	598	580	521	582	579	473	413	367
	Me	296	418	447	599	583	514	577	576	472	407	363
	SD	22	28	27	23	20	20	19	44	19	39	16
F2	Mea	2292	1997	2004	1836	1837	1474	1016	1009	1087	1241	1289
	Me	2297	2019	1990	1835	1845	1502	1023	1027	1079	1304	1337
	SD	40	97	74	62	62	72	47	47	77	162	136
F3	Mea	2857	2689	2757	2497	2490	2548	2205	2306	2345	2407	2391
	Me	2843	2695	2778	2561	2529	2551	2197	2318	2353	2426	2379
	SD	72	72	137	185	179	74	130	249	154	75	83
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	89	109	124	132	122	110	93				
	Me	91	107	124	134	119	111	94				
	SD	10	10	11	10	10	8	9				
F0	Mea	121	118	116	112	116	118	121				
	Me	122	117	116	112	115	112	119				
	SD	7	9	9	6	7	11	8				
F1	Mea	338	402	554	678	585	431	347				
	Me	334	399	551	674	585	434	345				
	SD	18	19	20	23	12	10	14				
F2	Mea	2201	2091	1907	1271	978	863	903				
	Me	2196	2070	1907	1263	963	836	891				
	SD	68	87	67	77	58	66	96				
F3	Mea	2823	2795	2455	2150	2177	2504	2332				
	Me	2897	2775	2497	2117	2218	2485	2326				
	SD	217	123	206	120	248	140	112				

* Only the first element of the semi-diphthong was measured.

Table 6. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 17.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	165	110	52	114	136	126	154	170	59	135	162
	Me	170	111	50	109	130	128	149	165	60	134	158
	SD	25	23	11	27	28	17	27	33	18	14	27
F0	Mea	165	147	142	135	132	148	132	129	149	149	173
	Me	163	156	141	135	132	142	132	133	148	154	174
	SD	26	46	12	14	9	16	19	18	19	13	20
F1	Mea	318	376	457	606	622	579	637	649	538	466	386
	Me	313	374	481	594	609	580	649	646	542	463	387
	SD	24	32	49	36	30	26	34	17	22	21	24
F2	Mea	2527	2271	1943	1752	1774	1511	1073	1064	1096	1081	1140
	Me	2530	2310	1925	1753	1771	1517	1064	1073	1104	1060	1191
	SD	48	97	109	73	76	27	61	68	89	107	135
F3	Mea	2904	2714	2615	2474	2482	2533	2576	2646	2566	2515	2539
	Me	2900	2721	2584	2474	2498	2514	2569	2596	2570	2546	2506
	SD	202	82	66	50	60	44	84	109	63	96	101
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	77	75	87	87	97	84	78				
	Me	74	78	88	89	93	82	81				
	SD	12	8	10	12	16	8	11				
F0	Mea	189	167	155	147	138	161	202				
	Me	192	166	153	153	152	168	200				
	SD	13	8	8	25	49	36	15				
F1	Mea	331	432	585	682	591	475	366				
	Me	342	444	580	690	592	478	366				
	SD	25	25	36	31	18	17	18				
F2	Mea	2293	2045	1833	1383	973	848	776				
	Me	2275	2060	1836	1345	939	862	761				
	SD	93	113	131	142	67	80	118				
F3	Mea	2623	2552	2458	2313	2486	2514	2530				
	Me	2646	2577	2414	2384	2552	2557	2524				
	SD	170	218	114	146	136	194	146				

* Only the first element of the semi-diphthong was measured.

Table 7. Mean, median and SD of duration (D, in milliseconds), f0, F1, F2 and F3 (in Hertz) values of AE and BP vowels produced by Participant 18.

		/i/	/ɪ/	/eɪ/*	/ɛ/	/æ/	/ʌ/	/ɑ/	/ɔ/	/o/*	/ʊ/	/u/
D	N	15	15	15	15	15	15	15	15	15	12	15
	Mea	88	79	38	91	98	83	86	108	40	78	91
	Me	89	75	39	90	93	83	81	110	38	80	91
	SD	13	15	7	19	19	7	18	21	8	11	10
F0	Mea	184	177	175	160	170	177	174	166	180	185	186
	Me	199	180	178	172	171	184	179	171	182	182	190
	SD	42	20	8	47	34	24	24	15	10	30	28
F1	Mea	311	406	389	667	562	473	576	541	420	363	366
	Me	302	395	379	680	595	442	568	606	393	362	352
	SD	68	46	44	59	65	70	81	125	92	27	58
F2	Mea	2571	2200	2274	1951	1995	1592	1163	1112	1192	1203	1460
	Me	2649	2220	2235	1965	1966	1626	1121	1095	1152	1211	1489
	SD	208	93	140	64	138	114	72	70	126	157	267
F3	Mea	3057	3022	2916	2816	2778	2854	2772	2754	2888	2815	2709
	Me	3049	3009	2924	2807	2847	2845	2763	2783	2880	2810	2710
	SD	317	192	186	138	256	130	181	164	82	59	149
BP		/i/	/e/	/ɛ/	/a/	/ɔ/	/o/	/u/				
D	N	10	10	10	10	10	10	10				
	Mea	63	80	98	101	102	82	98				
	Me	60	82	100	100	100	84	76				
	SD	12	17	12	10	6	14	60				
F0	Mea	196	183	172	151	172	185	194				
	Me	192	184	174	165	172	184	192				
	SD	15	7	7	32	4	7	9				
F1	Mea	303	361	541	636	574	382	365				
	Me	309	360	552	645	631	382	362				
	SD	36	13	84	90	155	44	14				
F2	Mea	2524	2379	2116	1379	970	905	814				
	Me	2522	2370	2132	1434	966	917	810				
	SD	90	65	82	133	46	82	45				
F3	Mea	3115	2919	2867	2569	2730	2960	2893				
	Me	3104	2954	2849	2602	2774	2910	2929				
	SD	165	159	80	171	223	138	208				

* Only the first element of the semi-diphthong was measured.