UNIVERSIDADE FEDERAL DE SANTA CATARINA

PÓS-GRADUAÇÃO EM LETRAS/INGLÊS E LITERATURA CORRESPONDENTE

THE EFFECT OF PERCEPTUAL TRAINING ON THE LEARNING OF ENGLISH VOWELS BY BRAZILIAN PORTUGUESE SPEAKERS

DENIZE NOBRE-OLIVEIRA

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

DOUTOR EM LETRAS

Florianópolis

Maio de 2007

Esta tese de Denize Nobre-Oliveira, intitulada *The effect of perceptual training on the learning of English vowels by Brazilian Portuguese 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 Lingüística Aplicada

____

Prof. Dr. José Luiz Meurer
Coordenador

BANCA EXAMINADORA:

____

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

____

Prof. Dra. Thaïs Cristófaro Silva (UFMG)
Examinadora

____

Prof. Dra. Giovana F. G. Bonilha (UFSM)
Examinadora

____

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

____

Prof. Dra. Rosane Silveira (UFSC)
Examinadora

Florianópolis, 25 de maio de 2007.

ACKNOWLEDGMENTS

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

I would like to thank my advisor, Professor Dr. Barbara Baptista, for her support during the last four years. Since the first time we have met, she has believed in me and has been a source of great encouragement and inspiration.

I would also like to thank Professor Dr. Paul Boersma, my foreign co-advisor, and Dr. Paola Escudero, for her advice on the design of the experiment.

I thank the members of the committee for their comments, suggestions and corrections: Professor Dr. Barbara Baptista, Professor Dr. Thaïs Cristófaro-Silva, Professor Dr. Giovana Bonilha, Professor Dr. Rosana Koerich, and Professor Dr. Rosane Silveira.

I also thank Professor Dr. Rosana Koerich and Professor Armando Silveiro, who had conceded me some classes with their undergraduate groups for data collection. My special thanks to all the participants of this study.

I would also like to thank some fabulous people for their friendship and support: Andréia Rauber, Mara Reis, Denise Kluge, Melissa Bettoni-Techio, Joara Bergsleither, Ricardo Bion, Verónica Rivas, Jeannette Elsenburg, Ton Wempe, João Carlos da Silv Priscilla Poleza, and all those who shared their friendship with me.

Above all, I thank my mother, who has always been a model of strength, love and serenity and has always done whatever she can to make my dreams come true, and I thank my father, who has been always guiding my steps and watching for me wherever I am. I also thank my brothers for all their love and support. Finally, I thank my darling husband Angelo for always believing in me and for showing me that there is more to life than vowels.

ABSTRACT

THE EFFECT OF PERCEPTUAL TRAINING ON THE LEARNING OF ENGLISH VOWELS BY BRAZILIAN PORTUGUESE SPEAKERS

DENIZE NOBRE-OLIVEIRA

UNIVERSIDADE FEDERAL DE SANTA CATARINA

2007

Supervising Professor: Barbara Oughton Baptista

Recent research has shown perceptual training to be an effective tool for improving L2 learners' ability to perceive certain non-native sounds, especially when done with enhanced acoustic-perceptual cues. This study investigates the effect of perceptual training on the learning of the English vowels /i/, /ɪ/, /ɛ/, /æ/, /ʊ/ and /u/, whose misperception can potentially cause comprehension problems. Secondary objectives include (i) the effect of training with enhanced stimuli, (ii) generalization of the acquired knowledge to new contexts and speakers, (iii) transfer of the perceptual improvement to the production domain, and (iv) long-term effects. The training on these vowels was given over a three-week period to twenty-nine Brazilian EFL learners, who were distributed within two groups: fifteen trained with natural stimuli (NatS group) and fourteen with synthesized stimuli (SynS group). The synthesized stimuli consisted of computer-generated utterances with enhanced spectral cues and no variation in duration, whereas the natural stimuli were recorded normally by native speakers of American English. Results show that the experimental groups improved significantly after training, and there was more improvement in the SynS group than in the NatS group. Considering that the training given the SynS group involved only synthesized stimuli and the tests involved only natural stimuli, this finding suggests also that the knowledge acquired with artificially enhanced stimuli is transferred to stimuli produced naturally. The improvement was also maintained one month after the training was over. These findings support the claim that perceptual training may serve as an effective tool for teachers to help learners overcome potential perceptual difficulties, and thus prevent potential miscomprehension.

Number of pages: 158
Number of words: 34,853 (excluding appendixes)

RESUMO

O EFEITO DO TREINAMENTO PERCEPTUAL NO APRENDIZADO DAS VOGAIS DO INGLÊS POR FALANTES DE PORTUGUÊS BRASILEIRO

DENIZE NOBRE-OLIVEIRA

UNIVERSIDADE FEDERAL DE SANTA CATARINA

2007

Professora Orientadora: Barbara Oughton Baptista

Pesquisas recentes mostram que o treinamento perceptual é uma ferramenta eficaz para melhorar a habilidade de perceber certos sons não-nativos de aprendizes de uma L2, especialmente quando esse treinamento é feito com a manipulação das pistas acústico-perceptuais. O presente estudo investigou o efeito do treinamento perceptual no aprendizado das vogais do inglês /i/, /ɪ/, /ɛ/, /æ/, /ʊ/ e /u/, cuja percepção deficiente pode causar problemas de compreensão. Os objetivos específicos incluem investigar (i) o efeito do treinamento com estímulo artificial, (ii) a generalização do novo conhecimento para novos contextos e novos falantes, (iii) a transferência da melhora na percepção auditiva para produção oral e (iv) os efeitos de longo prazo. O treinamento das vogais foi ministrado durante o período de três semanas para 29 aprendizes brasileiros distribuídos em dois grupos: 15 treinaram com estímulo natural (grupo *NatS*) e 14 com estímulo sintetizado (grupo *SynS*). O estímulo sintetizado consistiu em elocuções com pistas espectrais enfatizadas e sem variação de duração e foram geradas por computador, enquanto que o estímulo natural foi gravado por falantes nativos de inglês americano. Os resultados apontam para uma melhora significativa dos grupos experimentais após o treinamento, sendo que houve uma melhora maior no grupo SynS do que no grupo NatS. Considerando que o treinamento ministrado para o grupo SynS consistiu apenas de estímulos sintetizados e que os testes incluíam apenas estímulos naturais, esse resultado também sugere que houve uma transferência do conhecimento adquirido com estímulo artificial para estímulos produzidos naturalmente. A melhora na performance dos alunos também foi mantida durante um mês após o final do treinamento. Estes resultados mostram que o treinamento perceptual pode servir como uma ferramenta eficaz para professores auxiliarem seus alunos a superar dificuldades perceptuais, evitando possíveis mal entendidos.

Número de páginas: 158
Número de palavras: 34.853 (excluindo os apêndices)

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background to the study

Much research has shown training to be effective in the improvement of one's ability to perceive and to produce foreign language sounds (Strange & Dittmann, 1984; Jamieson & Morosan, 1986; Rochet, 1995; Logan & Pruitt, 1995; Yamada, Tohkura, Bradlow & Pisoni, 1996; Pisoni, Lively & Logan, 1994; Bradlow, Pisoni, Yamada & Tohkura, 1997; Ceñoz Iragui & Garcia Lecumberri, 1999; Hardison, 2002; Wang, 2002; Wang & Munro, 2004; Hawkey, Amitay & Moore, 2004; among others). Many of the studies carried out so far approach a number of difficulties nonnative speakers of English may have depending on their L1. For instance, a considerable number of studies on the effects of training have focused on the learning of the /l/-/r/ contrast by Japanese learners of English, others on the learning of nonnative vowel contrasts by native speakers of Spanish, Mandarin or Cantonese. Most of those studies not only investigated the effect of training on the trainees' perceptual abilities, but also assessed possible improvements in their production. Although there have been so many studies on this issue involving learners with various native-language backgrounds, no studies have focused on perceptual training involving Brazilian learners.

In this study I investigated the effect of perceptual training on the perception and production of three English vowel contrasts (/i-ɪ/, /ɛ-æ/, and /ʊ-u/) by Brazilian EFL students. Issues such as the use of synthesized stimuli with cue enhancement (see

Section 4.2), generalization of the acquired knowledge, the relation between perception training and production improvement, and long-term retention of a possible improvement on the perception and production of the targets were investigated.

The present study is innovative in a number of aspects. First, there has been no research on training involving synthesized speech stimuli with cue weighting (see Section 3.3) or its use for improving English perception and production in the Brazilian undergraduate context. Most of the previous research involving synthesized speech stimuli with cue weighting dealt with consonants (Strange & Dittmann, 1984; Jamieson & Morosan, 1986; Ortega & Hazan, 1999; Hazan & Simpson, 2000; Ortega-Llebaria, Faulkner & Hazan, 2001). Furthermore, most studies were carried out in the context of an English-speaking country, and only some of them aimed at testing the perceptual learning transfer to the production domain. To my knowledge, there have been only three of this kind of study dealing with vowels: Wang (2002) and Wang and Munro (2004) with native speakers of Mandarin and Cantonese, and Fox and Maeda (1999), in which only native speakers of Japanese participated. In these three studies the participants were trained in and tested on perception only and in the context of an English-speaking country. There are also some studies linking perception training to production improvement (Rochet, 1995; Yamada et al., 1996; Bradlow et al., 1997, 1999; Frieda, Walley, Flege & Sloane, 2000; Hardison, 2002, 2003), but these studies did not involve cue enhancement and were also carried out in the context of an English-speaking country. Thus, the present study is novel in bringing together the choice of the L1 of the L2 learners, the classroom as L2 learning context (it is not being carried out in the context of an English-speaking country), and the expected transfer of perception training to production improvement.

1.2. Objectives and Hypotheses

The main objective of this research was to investigate whether perception training would have a positive effect on the learning of the English vowels contrasts /i-ɪ/, /ɛ-æ/, and /ʊ-u/ by Brazilian EFL learners. More specifically, I investigated whether – before and after training – learners would categorize the L2 vowels /ɪ/, /æ/ and /ʊ/ into the same preexisting L1 categories of the vowels /i/, /ɛ/ and /u/, respectively, or whether they would create new categories for the L2 sounds, avoiding category overlap.

I also investigated whether Brazilian learners would benefit or not from training involving synthesized speech stimuli with cue enhancement – a specific kind of training in which learners would be exposed to artificial stimuli in which crucial portions of the signal would be emphasized. Previous studies involving synthesized stimuli (Wang, 2002; Wang & Munro, 2004) have shown that vowel synthesis can be useful to enhance subtle but crucial L2 properties not usually perceived by a nonnative listener, and that this enhancement can facilitate perception. In this study, I investigated how the effects of training with artificial stimuli compared to the effects of training with natural stimuli for the purpose of perception improvement of the targets, that is, a better identification rate.

I also checked whether the knowledge acquired by means of synthesized stimuli was transferred to natural listening settings, that is, if training with synthesized stimuli led to improvement in listening to natural speech. Evidence of this kind of generalization was found by Jamieson and Morosan (1986), Yamada et al. (1996), Bradlow et al. (1997), Wang et al. (1999), Fox and Maeda (1999), Wang (2002), Hardison (2003), Wang and Munro (2004), Hazan et al. (2005), and Pruitt et al. (2006).

Another aspect that I investigated was whether perception training in a non-English speaking country (Brazil) would lead to production improvement, as it was found to do in an English speaking environment in previous studies (Strange & Dittmann, 1984; Yamada et al., 1996; Hardison, 2002, 2003; and others). The fact that most Brazilian EFL learners do not interact with native speakers of English on a daily basis prevents them from profiting from authentic input, and this lack of exposure to native speakers may interfere in L2 learners' degree of improvement (Hazan et al, 2005). A pilot study showed natural-stimuli-based perception training to be effective in improving the vowel perception of Brazilian EFL learners, but no carryover to production improvement was evidenced (Nobre-Oliveira, 2005). The use of synthesized stimuli involving cue enhancement, already shown in previous research to improve the perception of English consonantal contrasts by nonnative listeners (Ortega & Hazan, 1999; Fox & Maeda, 1999; Hazan & Simpson, 2000; Ortega-Llebaria et al., 2001), is hypothesized to lead to improvement in both skills.

Finally, Brazilian EFL learners were assessed – in terms of perception and production – one month after the end of training in order to check whether training had a long-term effect, as found in previous studies (Yamada et al., 1996; Wang, Spence, Jongman & Sereno, 1999; Ceñoz Iragui & Garcia Lecumberri, 2002; Wang, 2002; Wang & Munro, 2004).

Five hypotheses were formulated. The first hypothesis, related to the first objective of this study, was that training would have a positive effect on the perception of L2 vowels. That is, while Brazilian learners before perception training would tend to assimilate similar L2 vowels to L1 categories – English /ɪ/ to Portuguese /i/, English /æ/ to Portuguese as /ɛ/, and English /ʊ/ to Portuguese /u/ – after the training sessions their ability to identify these L2 sounds would improve. As to the English vowels /i/, /ɛ/, and

/u/, although these vowels are not acoustically identical to Portuguese /i/, /ɛ/, and /u/, the participants were not expected to have difficulty identifying them.

The second hypothesis was that the participants would benefit more from training with synthesized stimuli: that is, training with this kind of stimuli would be more effective than training with natural stimuli. This would be justified because speech stimuli involving cue enhancement would provide learners with more prominent cues that could not otherwise be noticed by nonnative listeners.

The last three hypotheses were the following. As regards the generalization of the acquired knowledge, the third hypothesis was that improvement in the categorization of the synthesized targets would be transferred to natural listening settings. The fourth hypothesis was that perceptual training would lead to production improvement even without any specific production training. This hypothesis is related to the fourth objective, which was to investigate the link between perception and production. Finally, the fifth hypothesis was that the learners' perceptual improvement would be maintained one month after perception training was over. This hypothesis was based on the results of previous studies carried out with Mandarin and Cantonese (Wang, 2002; Wang & Munro, 2004) and Spanish and Basque (Ceñoz Iragui & Garcia Lecumberri, 2002) speakers of English.

An outline of the objectives and hypotheses of this study are shown in Table 1.

Table 1. Summary of objectives and hypotheses

Issue	Objectives	Hypotheses	Theoretical background
	General. To find out if perception training helps to improve the learning of the English vowels by Brazilian EFL learners.	General. Training would have a positive effect on the perception of L2 vowels.	
Appropriate categorization of the targets	Obj1. To check if training has positive effects on the adequate categorization of the target L2 sounds;	H1. Brazilian learners without training would identify the L2 vowels /ɪ, æ, ʊ/ as L1 /i, ɛ, u/, respectively; after the training sessions, their ability to identify L2 sounds would improve;	Strange & Dittmann, (1984); Jamieson & Morosan (1986); Rochet (1995); Yamada et al. (1996); Bradlow et al. (1997); Ceñoz Iragui & Garcia Lecumberri (1999); Hardison (2002); Wang (2002); Wang & Munro (2004).
Comparing two training methods	Obj2. To find out which kind of training would lead to better results, training with natural stimuli or with synthesized stimuli with cue enhancement;	H2. Training involving synthesized stimuli would be more effective than training with natural stimuli;	Strange & Dittmann (1984); Jamieson & Morosan (1986); Ortega & Hazan (1999); Fox & Maeda (1999); Kuhl (1991, 2000); Hazan & Simpson (2000); Ortega-Llebaria et al. (2001); Wang (2002); Wang & Munro (2004).
Generalization of the acquired knowledge	Obj3. To check whether knowledge acquired by means of synthesized stimuli is generalized to natural listening settings;	H3. Improvement in synthesized speech would be transferred to natural listening settings;	Jamieson & Morosan (1986); Yamada et al. (1996); Bradlow et al.(1997); Wang et al. (1999); Fox & Maeda (1999); Wang (2002); Wang & Munro (2004); Hazan et al. (2005); Pruitt et al. (2006).
Perception/ production interface	Obj4. To find out whether perception training leads to production improvement;	H4. Perceptual training would lead to production improvement;	Rochet (1995); Yamada et al. (1996); Bradlow et al. (1997, 1999); Hardison (2002, 2003); Hazan et al. (2005).
Long-term effects	Obj5. To check whether training has a long-term effect.	H5. The learners' improvement would be maintained some time after perception training is over.	Yamada et al. (1996); Cenoz Iragui & Garcia Lecumberri (2002); Wang et al. (1999); Wang (2002); Wang & Munro (2004).

1.3 Organization of the dissertation

This dissertation is divided into six chapters. Chapter 2 contains a description of the Brazilian Portuguese and the American English vowel systems, a report of early studies on the perception and production of vowels, a brief explanation of key concepts of acoustic phonetics, and some relevant cross-language issues such as vowel similarity and cue weighting. Chapter 3 consists of a review of current models of speech perception and production, and of previous studies on cross-language perception and production. In Chapter 4 the results of previous perceptual training studies are summarized, as well as findings related to use of cue-enhancement, generalization, effects of production, and retention of learning. Chapter 5, the Method, contains a detailed description of how the material for the data collection was designed and of how the experiments were carried out. The results of the experiments are reported and discussed in Chapter 6. In Chapter 7 I draw conclusions about the main findings of the study and discuss its limitations and contributions, as well as suggestions for future research.

Chapter 2

Vowel inventories

2.1 Early investigations in vowel perception and production

In order to provide a historical background for the studies involving vowel perception and production, Jenkins (1987) wrote “A selective history of issues in vowel perception”. According to this paper, the first references specifically to articulatory phonetics date from the 17th century, with the creation of the *English School of Phonetics*. However, a modern articulatory classification of vowels was not conceived until 1867, when Alexander Melville Bell first characterized vowels in terms of height and backness/frontness, as a result of the movements of the tongue. References to the acoustics of speech sounds were made by Henry Sweet first, in 1911, although these references were rather superficial, due to the lack of technical instrumentation (they were mostly on the nature of the sounds).

Jenkins (1987) reports that in the late 19th century, phoneticians believed that vowels differed acoustically in terms of perceptual *intrinsic pitch*; That is, even if a speaker holds his tone of voice to produce three different vowels, there seemed to be a variation of perceived pitch among the three. Thus, phoneticians concluded that each vowel had a different pitch, and they sought to determine the different vocal pitches for each of the vowels. A physicist called Helmholtz decided to test that assumption by carrying out experiments with tuning forks, and he found quite the opposite: It was not the vocal pitch itself that characterized the vowels, but the specific reinforced harmonics of the fundamental frequency in the vocal tract (i. e., F1, F2, etc.) that determined the vowels.

After the phonograph[1] (Figure 1) was invented, two engineers, Jenkin and Ewing, were able to analyze the major components of the sounds, that is, the reinforced harmonics and the pitch. However, they found difficulty in modeling recognizable vowels when there was variation in pitch and in voice quality.

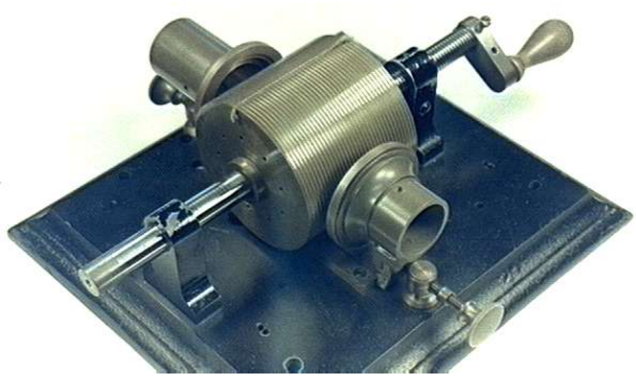


Figure 1. The phonograph (www.google.com.br).

Still according to Jenkins (1987), in 1879, after carrying out several experiments with the phonograph, Alexander Graham Bell was convinced that vowels did not depend on specific harmonics to be determined, but rather on the enhancement of specific resonances produced by cavities of the vocal tract. In 1890, phonetician Lloyd took Bell's argument further by saying that it was not only the enhanced resonances that determined vowel quality, but also the ratio of the pitches of their resonances. This idea would cope with the problem of speaker variability and the size of the vocal tract, but it did not convince the critics, who argued that the same ratios of resonance pitches would not always lead to the same perceived vowels.

---

[1] The phonograph, an instrument for recording and reproducing sounds, was invented by Thomas Edison in 1877.

The lack of appropriate instrumentation was the main reason for such disagreement in the field. Jenkins (1987, p. 545) points out that in the 19th century vowels were studied as a "static posture on the articulatory side and as a sustained sound on the acoustic side", and speaker and linguistic variables (such as age, sex, speech rate, and contextual effects) were not considered to be of much relevance. Only in the 20th century, after the invention of oscilloscopes[2] (Figure 2), which display the waveforms, and spectrographs[3], which show the spectra, it was possible to have more precise pictures of the acoustic components of speech sounds.

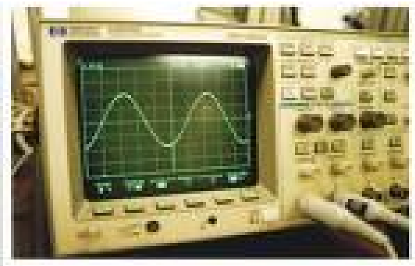


Figure 2. An oscilloscope (**www.google.com.br**).

At the end of his article, Jenkins (1987) reports that, in 1948, the linguist Martin Joos was able to design the first postulates of a modern vowel theory by confirming – and going beyond – some of the early findings of Helmholtz's and Bell's. Joos finally confirmed the importance of speaker variables, which were not considered in previous studies. Joos proposed that vowel sounds are perceived based on the values of the first two formants, and that these values vary from speaker to speaker but, in spite of this variation, listeners somehow adjust to the speaker's particular formant values. A series

[2] The oscilloscope, invented by Karl Ferdinand Braun in 1897, is a machine that allows the visualization of signal voltages (as visible wave forms) in two-dimensional graphs.
[3] The spectrograph was developed by Bell Telephone Laboratories during World War II. It converts a sound wave into a sound spectrogram.

of subsequent studies involving speech synthesis confirmed Joos's theory, although some speaker normalization procedures are still under debate.

Some more recent empirical studies on speech perception have aimed at finding ways to cope with problems, such as speaker normalization and speaking rate, as well as providing evidence for cross-language speech perception models. The following sections summarize the main findings of these studies and relate them to the perception models described in this chapter.

2.2 Vowel acoustics

To describe vowels acoustically it is necessary to understand some basic concepts and procedures, such as what formants are and how to look at a spectrogram. Whenever any vowel is produced, they are characterized by different resonance frequencies which vary according to the position of the articulators[4] (Ashby & Maidment, 2005). These frequencies are called formant frequencies. Formant frequencies are resonances of the vocal tract that are represented under the form of peaks in the spectrum (Figure 3), and they characterize vowels acoustically in that they are determined by the shape of the mouth and pharynx cavities, for instance (Pickett, 1999).

[4] The tongue, lips and pharynx are some examples of articulators.

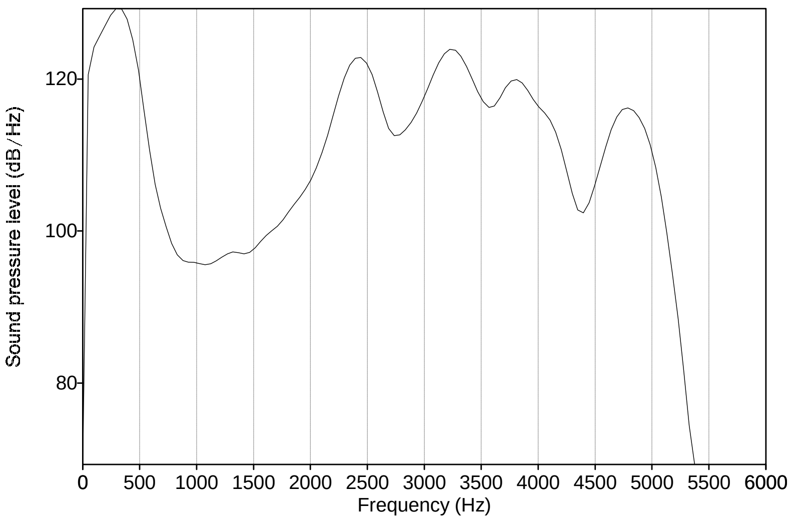


Figure 3. Spectrum of the vowel /i/ produced by a female American speaker[5]. The three first formant peaks are, approximately, at 306Hz (F1), 2351Hz (F2), and 3226HZ (F3).

The most important frequencies to characterize vowels are the first three formant frequencies, but usually using only the first two is enough (Ladefoged, 2005). The first formant frequency (F1), which varies according to the height of the body of the tongue, provides information about vowel height. The second formant frequency (F2) gives us information about vowel backness (backness or frontness of the tongue) and lip rounding[6]. With this information in mind, it is possible to infer a vowel based only on formant frequency values. For instance, low vowels are characterized by high F1 values, while low F1 values characterize high vowels; back vowels have low F2 values whereas front vowels have high values for the second formant.

It is generally agreed that the third formant frequency (F3) is related to rhoticized vowels[7] (Stevens, 1997; Hayward, 2000; Ladefoged, 2005). Stevens (1997) argues that

---

[5] All figures without reference were created by the author of the dissertation.
[6] It seems that, in general, tongue backness and lip rounding are redundant aspects in terms of formant frequency, since lip rounding causes the lowering of formant frequencies (see Ladefoged, 2005, p. 42). Thus, the more back the body of the tongue is, the more rounded the lips are – and the lower F2 will be.
[7] In this respect, Ladefoged (2005) wrote that "the only vowel in which the third formant plays a significant role is the vowel in *bird* of as pronounced in General American English." (p.49)

F3 and F4 frequencies do not vary substantially with the changes of the body of the tongue, thus making F3 and F4 unessential to characterize the oral vowels.

In order to measure formant frequencies, we use a spectrogram. Figure 4 shows a spectrogram[8] of the vowel /i/ spoken by a native American man. The top-half of the picture shows the waveform of /i/ and the bottom-half shows the spectrogram itself. Note that the four first formant frequencies are shown by the dotted tracks in the spectrogram. The two arrows show the first and the second formants, F1 being the lowest of the two.

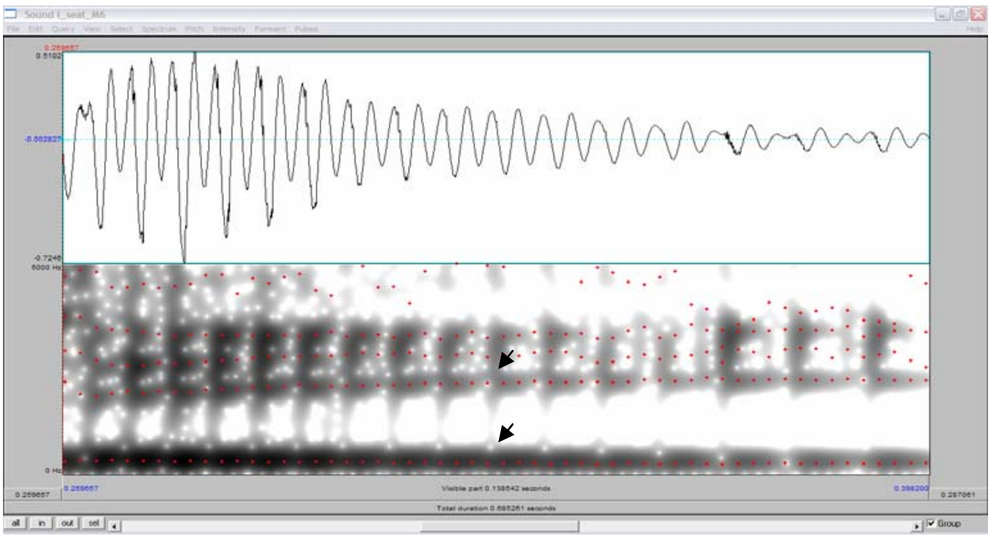


Figure 4. Visualization of the vowel /i/ spoken by a native American man. On the top-half of the picture there is the waveform of the vowel /i/ and on the bottom-half there is the spectrogram of /i/.

The waveform gives us information about the amplitude[9] (vertical axis) and the timing of a sound (horizontal axis), whereas the spectrogram shows the interaction between frequency, amplitude and time. Figure 5 shows a waveform and a spectrogram

[8] There are many sound editors available in the market, but I strongly recommend *Praat*, which I used for analyzing the data in the present study. This free software can be downloaded from http://www.praat.org/.

[9] Amplitude is the amount of pressure variations (Pickett, 1999).

separately of the vowel /i/. Note that the amplitude in the spectrogram is visualized as the darkest shadowing in the figure.

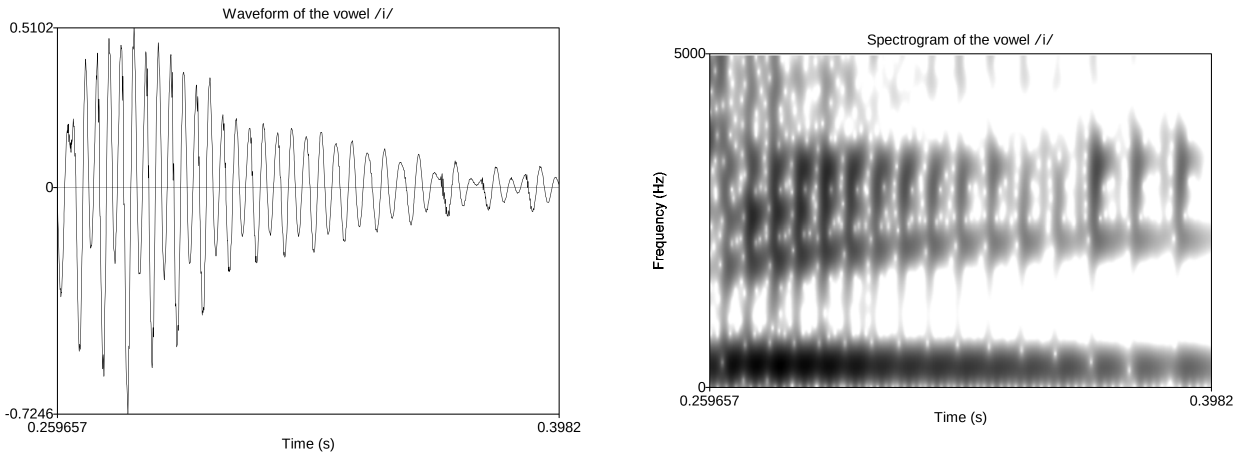


Figure 5. Example of a waveform on the left (amplitude vs. time) and of a spectrogram on the right (frequency vs. time vs. amplitude).

Once one has the values of the vowel formants, one can draw vowel plots and vowel charts. Vowel plots are *pictures* of the relative location of vowels within the vowel space, and they are useful in that "the [plotted] distance between any two vowels reflects how far apart they sound" (Ladefoged, 2003, p. 130). We say that vowel plots and vowel charts show the *relative location* of a sound because they only indicate location tendencies of one sound in relation to another; it is always important to remember that vowels form a continuum. Another important aspect to bear in mind is that vowel plots and vowel charts represent relative vowel locations in terms of perception and production. Some examples of vowel charts are those shown in Sections 2.3 and 2.4 (Figures 6, 7 and 8). Examples of vowel plots are shown in Section 2.4 (Figures 9, 11 and 12).

A last issue on the acoustics of vowels is vowel normalization. It is definitely not appropriate to write raw values of formant frequencies to characterize and compare vowels of different speakers because there will always be variation (i. e. pitch), and variation in formant frequencies between speakers "can obscure the phonetic differences between vowels in a formant space" (Harrington & Cassidy, 1999, p. 74). In order to cope with this situation, *normalization procedures* are necessary to make accurate comparisons. To *normalize* speech is to "abstract phonetic sameness from the diversity which exists across speakers" (Hayward, 2000, p. 169). Therefore, by applying normalization procedures it is possible to represent acoustically vowels produced by various speakers in only one chart.

There are a number of normalization procedures. One of the first consists of calculating the ratio between F1 and F2 values and, according to Peterson (1952, 1961), this is a reliable procedure for front vowels. However, formant ratios are not a reliable procedure to compare back vowels, since formant ratios may be quite similar among back vowels (Harrington & Cassidy, 1999). For this reason, a different normalization procedure was adopted, which is described in Section 5.2.2.3.

2.3 The Brazilian Portuguese vowels

The Brazilian Portuguese (BP) oral vowel inventory is composed of seven phonemes in stressed syllable position (Câmara Jr., 1970, followed by Callou & Leite, 1990; Bisol, 1999; Cristófaro Silva, 2001), which are distributed as shown in Figure 6.

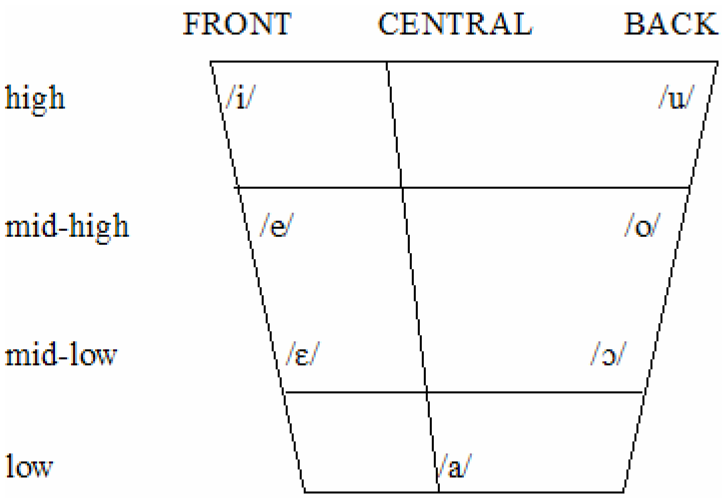


Figure 6. Distribution of BP oral vowels in stressed position (adapted from Câmara Jr., 1970, p. 41).

According to Câmara Jr.'s (1970) syllable-stress-based description of the BP vowels, besides the stressed syllable oral vowels shown in Figure 6, there are also three other sub-inventories for unstressed syllable vowels: a pretonic (with five vowels), a mid posttonic (with four vowels) and a final posttonic (with three vowels), represented in Figure 7. Thus, the BP system has all seven vowels only when they appear in the tonic position.

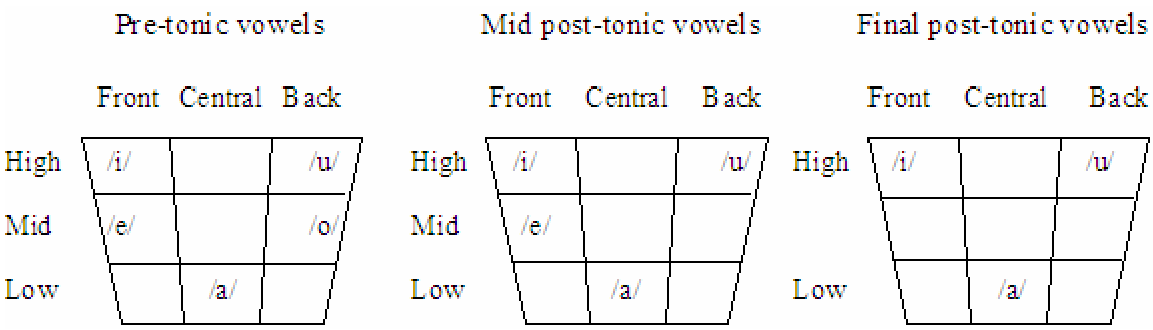


Figure 7. BP vowels in unstressed syllable positions (adapted from Câmara Jr., 1970, p. 44).

Although Câmara Jr.'s (1970) classification of the BP vowels according to syllable-stress is taken as the standard, regional variation may occur. Therefore, the vowel /e/ is allowed in post-tonic position in some regions of Paraná (a southern Brazilian state), and one may hear word "leite" pronounced as [ˈlej.ˌte]. Similarly, in some parts of the northeast of Brazil, one may hear the word "colégio" pronounced as [kɔ.ˈlɛ.ʒɪw], as the vowels /ɛ/ and /ɔ/ are allowed in pretonic position.

As for the acoustic properties of BP vowels, Faveri's (1991) pioneering study on BP vowel duration provided values for the target vowels within stressed syllable-initial, syllable-medial and syllable-final positions, produced by speakers from Florianópolis. Her study shows that, although the overall durational differences among the target vowels presented very little variation, /i/ generally had the shortest duration in all contexts. Table 2 shows the overall intrinsic duration of BP vowels in stressed-syllable position.

Table 2. Intrinsic duration of BP vowels in stressed syllables (adapted from Faveri, 1991)

BP vowel	Duration (ms)
/i/	83
/e/	116
/ɛ/	119
/a/	107
/ɔ/	126
/o/	104
/u/	103

BP vowels have also been described in terms of formant frequencies in different studies by Callou, Moraes and Leite (1996), Lima (1991) and Pereira (2001). Callou, Moraes and Leite (1996) found that there is “a tendency to centralize high vowels” (p. 27) and to raise /a/. The results in Pereira’s (2001) study suggest that there has been a restructuring of the BP vowel system in the dialect spoken in Florianópolis in that all the vowels were lower than in Lima’s (1991) study. However, Callou, Moraes and Leite’s study had a rather sociolinguistics basis (the data were taken from spontaneous speech), and in all three studies there were only male speakers with different socio-cultural backgrounds, making it difficult to draw conclusions. More recently, Rauber (2006) found minimal differences in vowel duration in BP vowels. The only observed fact was that the lower the vowel, the longer, but this does not mean that BP speakers make use of duration to discriminate BP vowels.

The most crucial information for this investigation provided by all the studies cited in this section is that BP vowels do not differ significantly in terms of duration and that there are only three front vowels in stressed position: one high and two mid vowels. In addition, BP has only one high back vowel in stressed position. As shown in the next section (2.4), these characteristics crucially differentiate BP and AE front and high back vowels, which are the object of the present study.

2.4 The American English vowels

The American English (AE) non-diphthongal vowel system is formed by 11 phonemes – 9 monothongs (/i/, /ɪ/, /ɛ/, /æ/, /ʌ/, /ɑ/, /ɔ/, /ʊ/, /u/) plus 2 semi-diphthongs (/e/ and /o/) – which are distributed as shown in Figure 8.

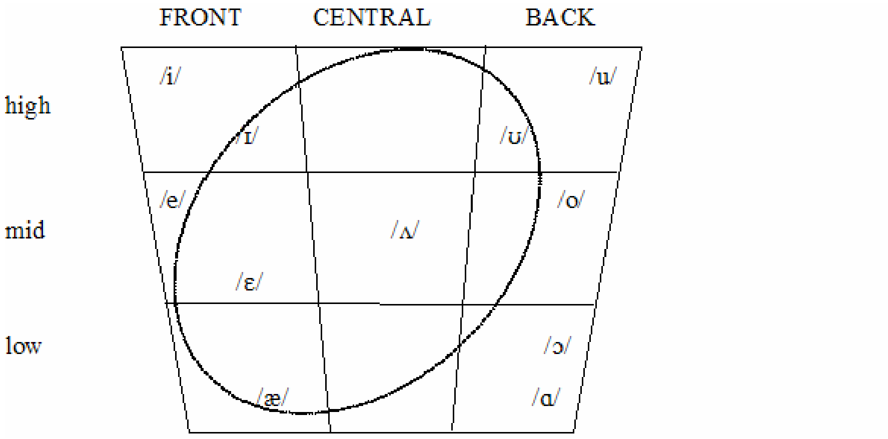


Figure 8. Distribution of AE vowels (adapted from Hammond, 1999, p. 6).

The vowels within the circle in Figure 8 are lax vowels. English vowels are classified in terms of tenseness, a feature which is related to the "force of constriction" of the articulators when vowels are produced (Giegerich, 2005, p. 97). Thus, in terms of production, an English vowel is *tense* when the force of constriction of the articulators is greater in relation to its *lax* counterpart, such as /i/ (tense) compared to /ɪ/ (lax). According to Giegerich, "tense vowels have a tendency (…) to be longer than lax vowels" (2005, p. 98), which is why tense vowels are sometimes called *long*, and lax vowels, *short*. However, the author points out that there is some controversy in the use of such terms (long/tense and short/lax) interchangeably, since in some English dialects (such as the Standard Scottish English) /i/ is realized with the same duration as /ɪ/ (Giegerich, 2005, pp. 99-101). Still, Giegerich (2005, p. 101) suggests that, considering only the Received Pronunciation (RP) and the General American (GA) dialects and ignoring "subtle length differences", English vowels "that are phonemically [+tense] are, on the phonetic level, also [+long]. Conversely [–tense] vowels are [–long]."

Still considering vowel production, the precise location of the AE vowels in terms of vowel quality (F1 and F2) was first reported in a classic study by Peterson and Barney (1952). In that study, 76 native American speakers had their monothongal vowels (excluding /e/ and /o/) recorded within the /hVd/ frame (total of 1520 words); the female speakers were mainly from the Mid-Atlantic region and the males were from different regions within the States[10]. The results were plotted considering the F1 and F2 dimensions, as shown in Figure 9.

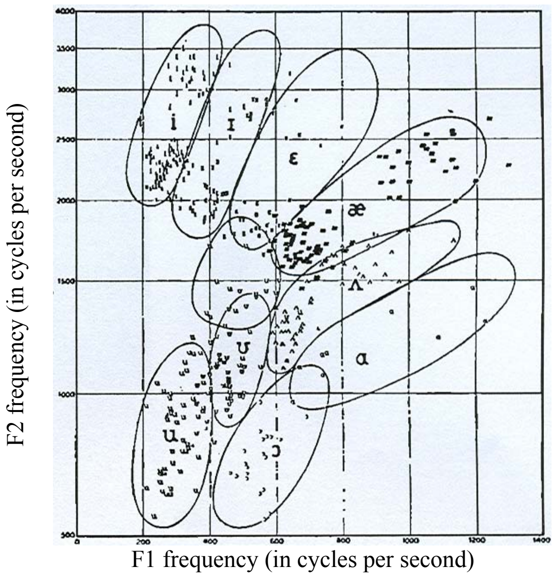


Figure 9. Distribution of ten American English vowels according to the frequencies of F1 and F2 (Peterson & Barney, 1952, p. 182).

More recently, Clopper, Pisoni and Jong (2005) carried out a study similar to that of Peterson and Barney's with the purpose of providing a more updated description of the American English vowels. The vowels produced by 48 participants from six regions of the United States – New England, Mid-Atlantic, North, Midland, South, and West – were analyzed. The regions are shown in Figure 10, and Figure 11 shows the distribution of vowels in the vowel space per dialectal region.

Note that the vowel plot shown in Figure 9 seem to be different from the one in Figure 11 and the ones in Figure 12, in terms of vowel distribution in the plotting space. This difference is because Clopper et al. (2005), but not Peterson and Barney (1952),

[10] According to the authors, the male speakers "represented a much broader regional sampling of the United States" (Peterson & Barney, 1952, p. 177).

plotted their vowels with F1 on the vertical axis with the values of the scales reversed, in order to imitate a traditional articulatory chart, as the one shown in Figures 8. The (reversed) plot by Clopper et al. (2005) is much common nowadays.

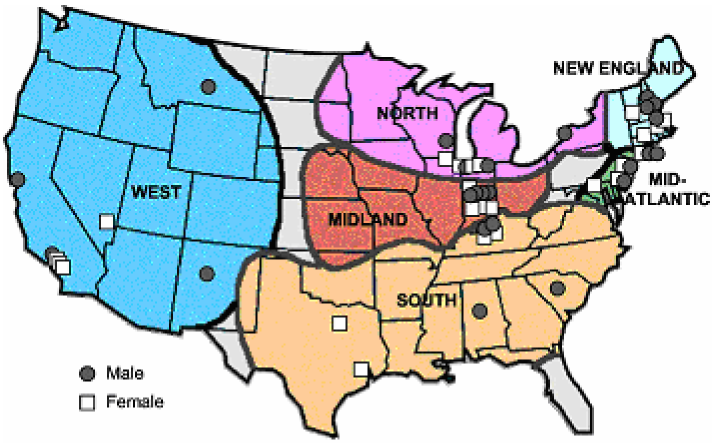


Figure 10. The six dialect regions and speakers' hometowns analyzed by Clopper et al. (Clopper et al., 2005, p. 1662).

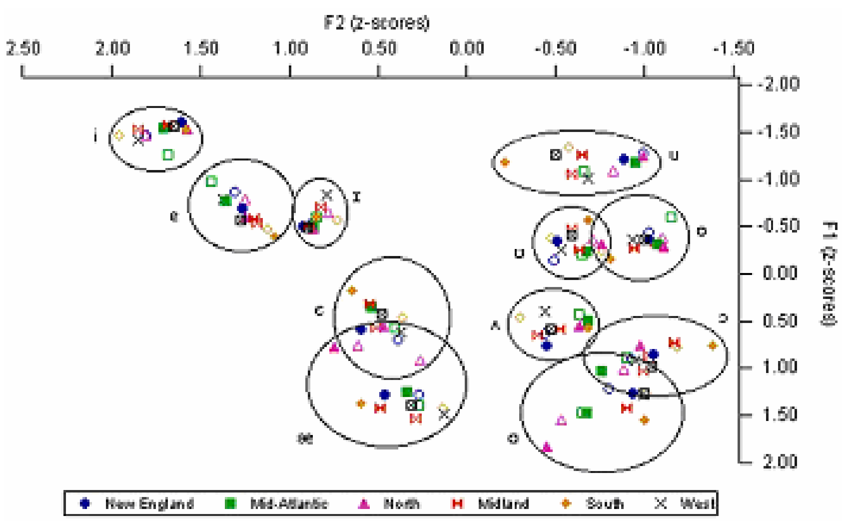


Figure 11. The productions of 11 English vowels by male (filled symbols) and female (open symbols) speakers in each dialectal region (Clopper et al., 2005, p. 1664).

Clopper et al. (2005) found that the location of vowels had changed since Peterson and Barney's study: in some regions the vowels /ɛ/ and /æ/ were overlapping, and in almost all regions[11] /ɑ/ and /ɔ/ were merging, which means that the tendency is to no longer produce these contrasts. The overlapping ellipses in Figure 11 show which vowels are not being contrasted by F1 and F2 in each region. The detailed distribution of each of the vowel tokens per region produced by the female speakers is shown in Figure 12.

[11] Except in the North and South of the United States.

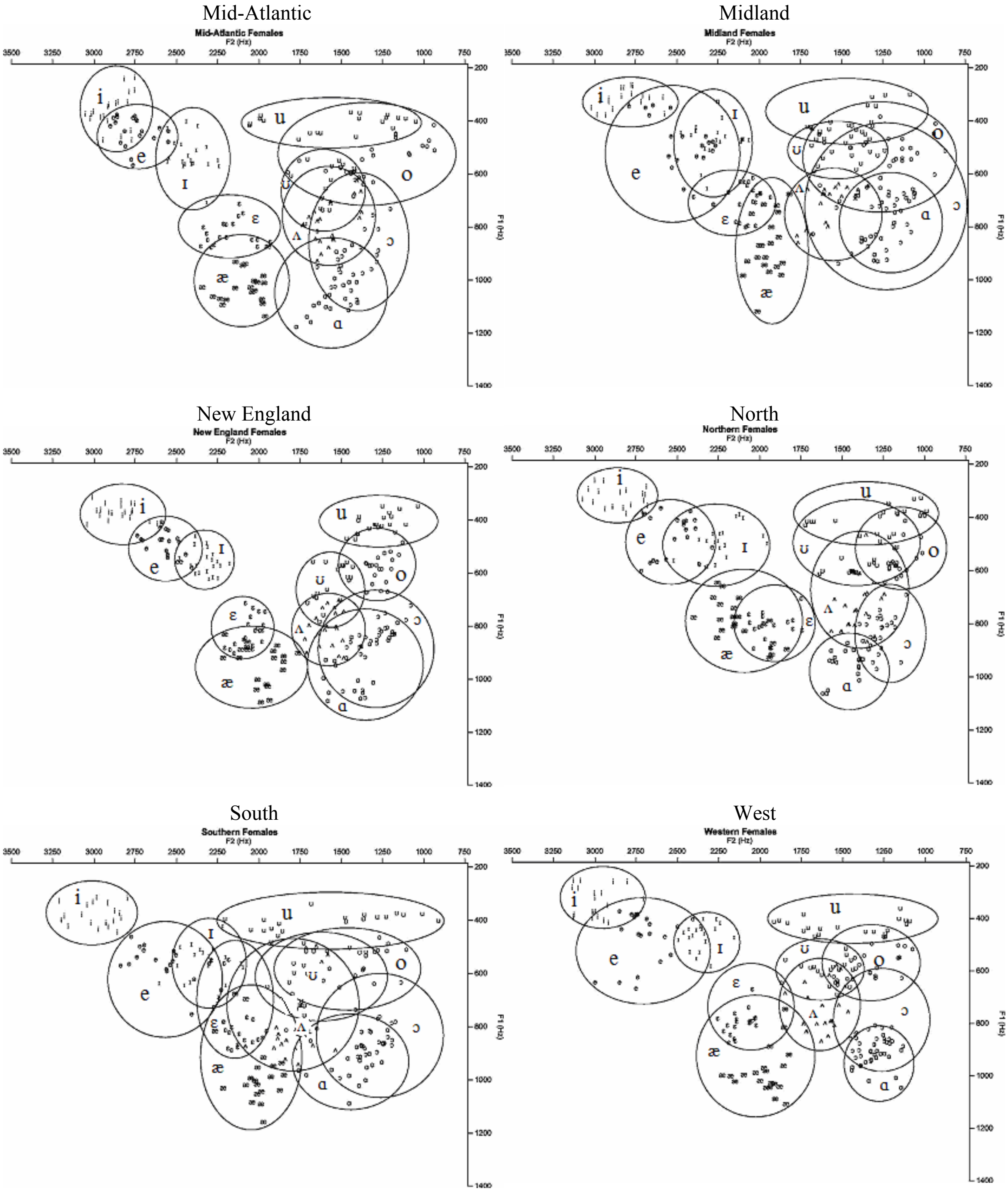


Figure 12. Distribution of AE vowels produced by the female speakers per region. (Clopper et al., 2005, pp. 1668-1673).

Similarly to BP vowels, each AE vowel is classified and distinguished from each other in terms of height, backness and lip roundness – the latter two being redundant features in English as well as in Portuguese. As explained in the beginning of this Section, there is also a third dimension that characterizes English vowels: tenseness.

According to this feature, English vowels can be either tense (long) or lax (short)[12]. Thus, vowels within the circle in Figure 8 (that is, /ɪ, ɛ, æ, ʌ, ʊ/) are classified as short. However, a study that aimed at measuring the intrinsic duration of AE vowels carried out by Peterson and Lehiste (1960) reported that the vowel /æ/ is actually longer than any other English monophthong, as shown in Table 3. Furthermore, although the vowel /ɛ/ is considered short, it is longer than /i/, which is considered long. If compared cross-linguistically, the intrinsic duration of the English lax vowel /æ/ is twice as long as the longest BP vowel (AE /æ/=284ms; BP/ɔ/=126ms); thus, it could potentially be classified as long by BP speakers. However, since the segmentation procedures in the two studies were carried out differently (Faveri [2001] considered only pitch and voicing; Peterson and Lehiste were more criterial, considering pitch, voicing and formant tracking of the target and neighbouring sounds)[13], further investigation is needed in order to make an accurate cross-linguistic comparison.

[12] In this dissertation I will choose the words *long* and *short* to refer to vowel length. As explained in the beginning of Section 2.4, in the GA dialect there is some difference in terms of vowel length, which is easier to measure than the difference between articulatory tenseness.

[13] One would also argue that the speaking rate of people from Florianópolis is generally quite fast in free speech; However, in her study, Faveri (2001) recorded only sentences read by the participants, and the rate of their speech was controlled.

Table 3. Intrinsic duration of AE monophthong (adapted from Peterson & Lehiste, 1960, p. 720)

AE vowel	Duration (ms)
/i/	207
/ɪ/	161
/ɛ/	204
/æ/	284
/ʌ/	181
/ɑ/	265
/ɔ/	250
/ʊ/	163
/u/	235

Another particular aspect recently found in AE vowels is that in some American regions people are not making the distinction between /ɑ/ and /ɔ/, where words such as ‘cot’ and ‘caught’ are pronounced the same, as noted by Ladefoged (2005, p. 46). This can be visualized in Figure 12, in New England and Midland, where these vowels have an extremely large overlap.

In terms of syllable structure, it is crucial to make the distinction between long and short vowels because the phonology of English does not allow short vowels in open syllables (CV). In other words, there are no English words formed by a consonant plus a final short vowel, such as /kæ/ or /bɪ/; such vowels are always followed by a consonant, making it a closed syllable, such as /kæ**t**/ (‘cat’) and /bɪ**g**/ (‘big’). Long vowels, on the other hand, are allowed in closed and in open syllables, like /bi/ (‘bee’), /bit/ (‘beat’), /strɔ/ (‘straw’), and /mɔb/ (‘mob’).

Finally, a visible difference from the AE vowels in Figure 8 and the BP vowels, shown in Section 2.3 (Figure 6), is the number of vowels. As to the front vowels, there are four monophthongs[14] in AE – two high, one mid and one low – whereas BP has only three front vowels – one high and two mid. The number of AE high back vowels is also higher than that of BP, with a proportion of 2 to 1. The two vowel systems differ both in number as well as in quality, and these differences will have consequences that will be discussed in the next section (2.5).

2.5 Cross-linguistic issues

Some of the main causes of difficulties in the perception and production of foreign language vowels are the cross-linguistic differences in terms of: (1) size of the vowel systems (presence or absence of vowels), (2) vowel similarity, and (3) the differential weighting of acoustic cues which characterize vowel contrasts (i. e., spectral quality vs. duration).

In regard to the difference in size of two vowel systems and the presence or absence of particular vowels, the more different two systems are in terms of number of vowels, the more difficulty L2 learners are expected to have in perceiving and producing the L2 vowels (Bradlow, 1996). Taking AE and BP as an example, AE has eleven non-diphthongal vowel phonemes, while BP has only seven oral vowels. This difference in number is accounted for by the five AE vowels with no close correspondents in BP – /ɪ/, /æ/, /ɑ/, /ʊ/ and /ʌ/ – and the one Portuguese vowel without a close English correspondent – /a/.

[14] Since only absolute monophthongs were focused on this research, the semi-diphthong /e/ is not referred to.

Vowel similarity is another contributing factor for cross-linguistic misperception. When two languages have different vowel inventory sizes, learners will tend to assimilate similar L2 sounds into preexisting L1 categories, since L1 and L2 vowel categories will probably overlap. Considering front vowels, BP has only one high vowel whereas AE has two, as shown in Sections 1.3 and 1.4. Since the vowel space occupied by the high front vowel (/i/) in BP is greater than that occupied by the higher front vowel (/i/) in AE – whose space has to be more limited because another high vowel (/ɪ/) has to fit in as well – there will be a potential perceptual overlapping of categories for AE high front vowels perceived by Brazilian learners (see Figure 13); in this case both AE /i/ and /ɪ/ would be perceived as BP /i/. Therefore, Brazilian EFL learners would be expected to have difficulty distinguishing English /ɪ/, /æ/, and /ʊ/ from Portuguese /i/, /ɛ/, and /u/, respectively. The central vowel /ʌ/ has been shown to be treated as a new vowel by Brazilian learners, and associations for the vowel /ɑ/ have been inconsistent (Baptista, 2000, 2002). Acoustic cue enhancement could be a solution to cope with this overlapping problem.

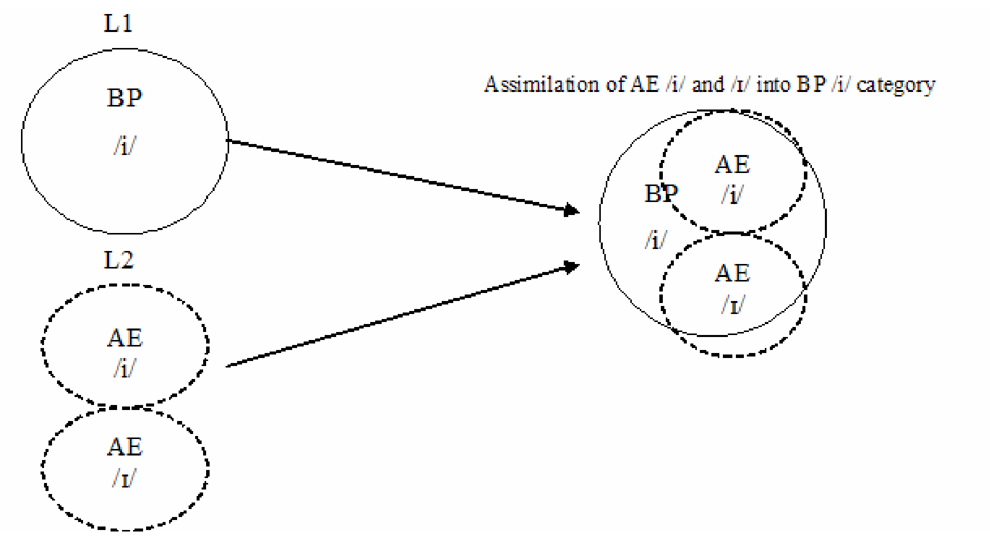


Figure 13. Example of vowel category overlap.

Acoustic cues are physical stimuli patterns in an auditory perceived event. Speech cues are the patterns of speech that characterize particular speech units, and "they are sufficient to cause a person to correctly perceive a given sentence, phrase, word, syllable, or phoneme" (Pickett, 1999, p. 151). For instance, the primary cue to characterize vowel height is F1, nasalization is characterized by a raised F1 frequency and a wider F1 bandwidth, and vowel duration is determined by a longer or a shorter duration.

The weighting of each cue can differ from language to language. For instance, tenseness is one of the cues that characterize English vowels (Ladefoged, 2005); that is, vowels can be either long or short. Differently, in Portuguese the vowels are not characterized by durational cues (Delgado Martins, 1973).

Recent studies have shown that even in a single language regional varieties may not use cues equally. In a study on cue weighting with speakers of two English varieties, Scottish English and Southern British English, Escudero (2001) found that although the speakers of both dialects rely on both spectral properties (F1/F2) and duration to perceive the contrast /i-ɪ/, Scottish English (SE) speakers tend to rely almost exclusively on spectral cues while the Southern British English (SBE) speakers rely predominantly on duration. Interestingly, after testing two groups of Spanish speakers who were exposed to the two English dialects, each group of Spanish speakers followed the same tendency as SE or as SBE, depending on the group they were assigned to, although they do not rely on durational cues in their L1 (i. e., Spanish). Based on these findings, the author concludes that "L1 and L2 perceptual development is decisively influenced by the nature of the input to which the learners are exposed" (p. 259). The results of Escudero's (2001) study suggest that an L2 learner's "perceptual attention to acoustic cues" could be altered (p. 250), in spite of his L1 cue weighting patterns, given the appropriate input[15].

However, under the foreign language classroom condition[16], given that the AE vowel system is larger than that of BP and that some pairs of AE vowels occupy the same relative acoustic space location of a single BP vowel, a BP speaker would be expected to rely primarily on spectral cues to characterize American English vowels because this is his L1 perceptual pattern. Taking the high front vowels as an example, in the same space that Brazilians have only the vowel /i/, Americans have /i/ and /ɪ/, which makes the relative distance between the two L2 sounds quite short by Brazilian standards, reducing their ability to discriminate them.

Considering the acoustic cues AE speakers and BP learners rely on to perceive vowel contrasts, AE speakers rely primarily on spectral quality, followed by vowel duration (Bohn & Flege, 1990), and previous studies (Fox & Maeda, 1999; Wang, 2002; Wang & Munro, 2004; among others) have shown that they tend to rely on their L1 cues to perceive L2 speech, especially in initial learning stages. Still considering cross-language speech perception, there are no studies indicating the tendency followed by Brazilian intermediate EFL learners when they listen to English vowel stimuli. There is a possibility that they will use the same cues of their L1 to listen to L2 speech; however, to this date, only speculations can be made.

[15] In the case of Escudero's participants, they had been living in Scotland for 3 years (personal communication with the author).
[16] In which the amount and sometimes the nature of the material the learners are exposed to is limited and restricted to class time.

Chapter 3

Perception and production in interphonology

3.1 Speech perception models

In this section, I review four of the current models of speech perception – the Native Language Magnet model, the Perception Assimilation Model, the Speech Learning Model, and the L2 Linguistic Perception model – and I highlight some of the most important aspects of each model for the present study.

3.1.1 The Native Language Magnet Model

The Native Language Magnet (NLM) model (Kuhl, 1991, 1992, 1993) is a result of the observation of a series of laboratory experiments involving infants. Considering that speech perception and language-relevant abilities are innate, Kuhl claims that, as early as their first minutes of life, infants *restrict* their L1 perceptual categories as a result of the interaction with the ambient language. This interaction triggers the development of mental representations and, in their first months of life, infants are already able to discriminate between-category contrasts. Later on, infants develop their ability of within-category discrimination as a result of linguistic experience (that is, exposure to new contrasts, new speakers, new words, etc.).

The key concept of the NLM model is that of speech prototypes. Speech prototypes are the best instances of speech sounds, and they function as *perceptual* *magnets*, which "are the result of infants' perception and representation of language input" (Kuhl, 1993, p. 129). These magnets exert influence over similar neighboring sounds, *pulling* them in their direction. Thus, according to the model, the organization of speech sound categories is directly related to prototypes. Furthermore, as remarked by Miller (1994), in terms of brain effort, the concept of storing prototypes is much more economical and less cognitively overloading than storing each and every instance of a speech sound.

As for L2 acquisition, the NLM claims that nonnative language sounds are perceived within the native language frame. Thus, L2 acquisition constraints are caused by prior perceptual experience, and not by lack of brain plasticity. Similar L2 sounds may initially be interpreted as bad exemplars of L1 sounds and mapped onto L1 mental representations. Kuhl further argues that L2 sounds may present different levels of difficulty, with sounds similar to L1 categories being more difficult to discriminate because of their greater magnetic attraction towards the native language category. On the other hand, L2 sounds that are not similar are not affected by this magnet effect, thus making discrimination easier.

Kuhl (2000) also argues that the formation of speech representation is initially monomodal; that is, infants focus only on one kind of perceptual information, which in this case is the auditory information. This monomodal representation later becomes polimodal, and babies start noticing connections between what they hear (auditory-driven perception) and the articulatory gestures they see (visually-driven perception) for each speech sound. This is an important indication of a link between perception and production, since the baby starts making a connection between the visual input of the articulatory gestures and the acoustic input of what he hears. Then, the baby starts *modeling* sounds and gestures (i. e., his perception and production skills, in a constant *feeding* process), and by a *trial-and-error* approach, he adjusts his acoustic boundaries and his articulatory gestures by the feedback he receives from adults.

The interrelation between the NLM effect and "motherese" (a kind of baby talk usually provided by mothers, but also by other family members and friends) that is presented in the following lines is particularly interesting for the present study. The key input that babies receive and that helps them to set psychoacoustic[17] boundaries is provided by motherese. Kuhl et al. (1997) carried out a study in which mothers from the United States, Sweden and Russia had their speech recorded while they were talking to their babies and to other adults. The result was that when mothers talked to their babies they increased their pitches (pitch variability) and they *stretched* considerably their vowel space. As a result, cues for perceiving vowels were enhanced, making vowels longer and more distinct from each other, as well as avoiding overlap among tokens, as represented in Figure 14.

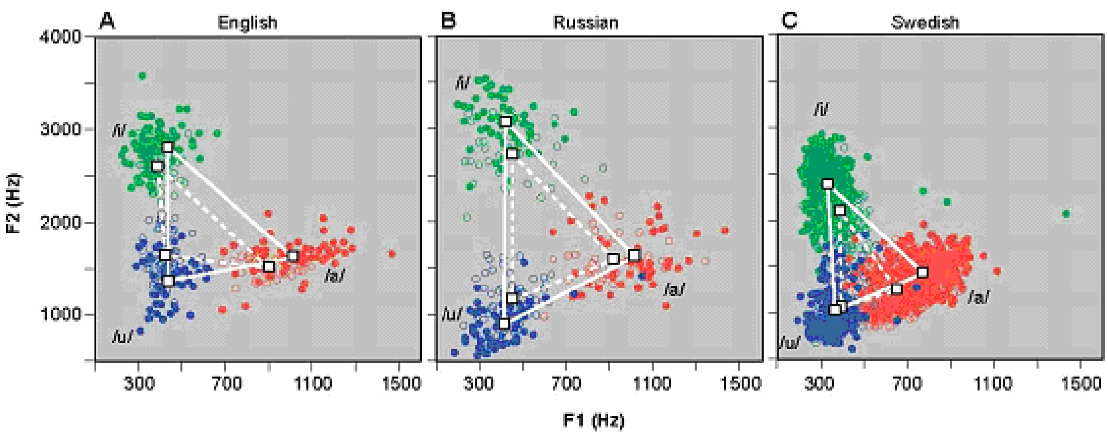


Figure 14. Distribution of the point vowels (/i/, /a/ and /u/) produced by English (A), Russian (B) and Swedish (C) mothers. The dotted lines show the Euclidean Distance[18] between the vowels in speech addressed to adults, and the solid lines show the Euclidean Distance between the vowels in speech addressed to infants (Kuhl, 1997, p. 685).

---

[17] Phonetic representation of language.
[18] According to the National Institute of Standards and Technology, Euclidean distance or Euclidean metric is the straight-line distance between two points.

Thus, an important finding of Kuhl et al.'s study is that the stretching of the vowel space by means of cue enhancement increases the distance between vowels, thus reducing acoustic overlap among categories of sounds. An important pedagogical implication is that, if the L2 learners receive "exaggerated acoustic cues, multiple instances by many talkers, and massed listening experience" (Kuhl, 2000, p. 11855), they may manage to overcome L1 perceptual constraints.

3.1.2 The Perceptual Assimilation Model

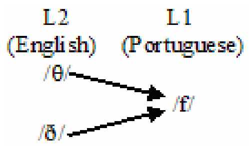
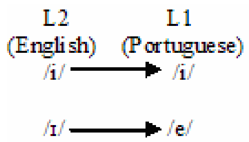
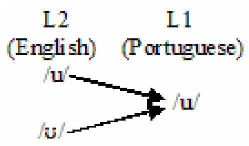
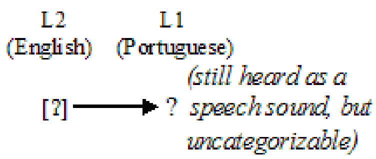
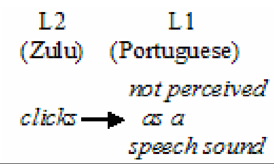
The Perceptual Assimilation Model (PAM) (Best, 1995), an ecological approach to language perception, establishes a link between perception and production by proposing that speech perception aims at identifying how sounds are produced. In this sense, there is a *feeding* relationship between the acoustic information a listener receives and the articulatory response he produces, in which one dimension will provide input and feedback for the other. It is *ecological* in that learning is a result of the interaction of the learner with the environment.

According to this approach, infants understand that they can achieve communicative goals by learning language-specific articulatory gestures. In addition, infants learn to discern which linguistic properties are relevant and which are not. In speech perception, for instance, if on the one hand, they discard any linguistic information that is not crucial, on the other hand they retain the significant aspects of the phonetic and phonological information they need.

Best's model is also based on the definition of gestural constellations. Gestural constellations are stable combinations of multiple speech gestures which are language-specific, although "there is usually a great amount of overlap among languages in the gestures and constellations contained within their individual phonological spaces, at least at the segmental level (…)”. From this perspective, nonnative segments are those whose “gestural elements (…) do not match precisely any native constellations” (Best, 1995, p. 193).

The PAM proposes different levels of relative difficulty in the discrimination of nonnative contrasts. Considering that speech perception is language-specific, and that in the beginning of the learning process L2 sounds are assimilated into L1 categories, the similarities and dissimilarities of the nonnative sounds (compared to the closest L1 gestural constellation) indicate which sounds will potentially be more difficult and which will be easier to perceive. As in Kuhl’s NLM model, similar L2 sounds would be more difficult to discriminate from L1 categories and dissimilar (new) sounds would facilitate discrimination. Some assimilation patterns are described by Best, characterizing the accuracy with which L2 sounds will be discriminated. For instance, optimal perception is achieved when we have *two-category assimilation*, that is, when two nonnative sounds are assimilated into two different categories. On the other hand, *single-category assimilation*, when the learner perceives two different L2 sounds as belonging to a single L1 category, is the poorest assimilation pattern. The assimilation patterns and their predicted discrimination performance are represented in Table 4. As a result of learning development from exposure to L2 input, there would be a reorganization of the assimilation patterns.

Table 4. PAM's assimilation patterns (Best, 1995, p. 195)

Perceptual assimilation pattern	Example	Predicted discrimination
Single-category assimilation (SC)	 	poor
Two-category assimilation (TC)	 	excellent
Category-goodness difference (CG)	 	moderate to very good
Both uncategorizable (UU)	 	poor to very good
Nonassimilable (NA)	 	good to very good

The *single-category assimilation* and the *category-goodness difference*, presented in Table 4, are apparently very similar, since in both of these patterns "both nonnative sounds are assimilated to the same native category". However, while the *single-category assimilation* includes two L2 phonemes that are both "equally acceptable or both equally deviant", the *category-goodness difference* includes a pair of L2 phonemes in which "one is acceptable, the other deviant" (Best, 1995, p. 195).

Although the PAM shares some similarities with the NLM model (for example, the discrimination predictability based on the degree of similarity between two sounds), it seems that there is some difference between the two models regarding the perception/production relationship. Differently from the auditory-driven basis of the NLM model, the PAM suggests that articulatory gestures trigger the development of mental representations. However, more decisive conclusions cannot be drawn as to this articulatory-driven basis of the PAM, since details on this issue are not given by Best (1995).

3.1.3 The Speech Learning Model

The Speech Learning Model (SLM), proposed by Flege (1995), is mainly concerned with ultimate attainment, that is, the final developmental stage of learning, therefore not focusing on the performance of beginners. The model hypothesizes about how L2 sounds are perceived by nonnative speakers based on the idea of phonetic discrimination ability, which is constrained by the speakers' age, since L1 categories are strengthened over the years. Flege also makes predictions concerning the oral performance of nonnative speakers based on their perceptual performance.

According to the SLM, L2 sounds can be perceived as new or similar. An L2 sound perceived as new is interpreted as being so different from any L1 sound that the learners' tendency is to create a new category for that sound. Conversely, when an L2 sound is interpreted as an allophone of an L1 sound, the learner fails to create a new category for this L2 sound, which will be considered perceptually equivalent to the L1 phoneme. This failure in the creation of a new category is caused by a cognitive mechanism called *equivalence classification*, which "permits humans to perceive constant categories in the face of the inherent sensory variability found in the many physical exemplars which may instantiate a category. It is known that children and adults use somewhat different strategies to categorize word, picture, or object arrays" (Flege, 1987, p. 49). In the sense of speech perception, this mechanism drives L2 learners'perception to perceive an L2 sound as being equivalent to an L1 sound. Thus, the smaller the perceived cross-linguistic distance between L1 and L2 sounds, the higher the chances of perceiving the L2 sound as an allophone of an L1 sound, and vice versa.

The SLM suggests that the ability to create new speech sound categories (called *phonetic categories* by Flege) is always active, but it is constrained by the learners' native language (as explained above) and age. Flege argues that age influences L2 phonetic perception in that the exposure to L1 sounds over the years *strengthens* them, constraining the perception of L2 sounds and making L2 perception innacurate. Thus, L1 phonetic categories "will become more powerful attractors of L2 speech sounds" (Flege & MacKay, 2004, p. 6). As a result, learners are not likely to establish new categories for similar L2 sounds (which will cause their speech to stay foreign-accented), although they can still create categories for new sounds.

Since the SLM postulates that oral performance corresponds to the perceptual phonetic category representation, inaccurate perception will cause inaccurate production, and any age-related perceptual constraints will also apply to the production domain. However, Flege also suggests that, if learners are extensively exposed to authentic input in natural settings, they can overcome those age-related constraints. Thus, new L2 sound categories would be formed and near native-like[19] perception could be achieved. Similarly, extensive naturalistic L2 experience can also modify articulatory gestures, which means that, given the appropriate kind and amount of input, adult L2 learners would be able to improve their motoric speech abilities.

[19] The SLM does not predict "mastery" of all L2 sounds (Flege, 1995, p. 243).

The three models of speech perception described in this chapter have some similarities, as well as particularities. For instance, all of them approach speech perception as a language-specific property, since it is a product of the speaker's L1 perceptual learning. Also, the three models have an ecological approach, since they suggest that linguistic development is a result of the learners' interaction with the environment (such interaction triggers the creation/reorganization of sound categories). However, the NLM model and the SLM suggest that perceptual representations lead to speech production, whereas the PAM suggests the opposite, that is, articulatory gestures build mental representations. A summary of the three models is presented in Table 5.

Table 5. Summary of the proposals of three speech perception models (adapted from Escudero, 2005, p. 151)

	NLM (Kuhl, 1991, 1992, 1993)	PAM (Best, 1995)	SLM (Flege, 1995)
Objective	Describe and explain cross-language speech perception	Describe and explain cross-language speech perception	Describe and explain cross-language speech perception; Predict L2 ultimate attainment
Status of perception	Language-specific	Language-specific	Language-specific
Initial learning stage	L1 neural mappings	L1 categories	L1 categories
Linguistic development	Creation of categories	Reorganization of categories	Creation and reorganization of categories
Final learning stage	Not explicit	Not explicit	Depends on the age of learning and on L1 interference
Perception/ production interface	Perception leads to/precedes production	Not explicit	Perception leads to/precedes production

3.2 Two Theories of Speech Production

In this section[20], I describe two speech production theories which were proposed quite some time ago but are still actively used: the Perturbation Theory (Chiba & Kajiyama, 1941) and the Source-Filter Theory (Fant, 1960). Only an overview of these theories is presented in this Section; for a more detailed description of these and other theories of speech production, see Pickett (1999) and Johnson (2003).

3.2.1 Perturbation Theory

According to Johnson (2003), the Perturbation theory, developed by Chiba and Kajiyama (1941), models the effects of vocal tract constrictions over resonances, that is, formant frequencies (see Section 2.2). It is especially useful to predict potential formant frequency values.

Based on Chiba and Kajiyama (1941), Johnson (2003) explains that the articulation of vocal tract components, such as the elevation of the body of the tongue or the rounding of the lips, creates constriction points which will directly interfere in the pressure and velocity of the air molecules that propagate sound. Before relating these constriction points to resonances, it is necessary to introduce two other acoustic concepts: nodes and antinodes.

According to the author, the concept of nodes and antinodes relies on the interaction between air[21] pressure and velocity (Johnson, 2003). When air pressure is at a minimum and velocity is at a maximum, there will be a **node**; when air pressure is at a maximum and velocity is at a minimum, there will be an **antinode**. Each of the

[20] This Section was included for sake of information.
[21] Air coming from the lungs.

resonances produced in the vocal tract is constituted by nodes and antinodes, F1 being formed by one node and one antinode, F2 by 2 nodes and 2 antinodes, F3 by 3 nodes and 3 antinodes, and so on. Thus, looking at the left-hand side of Figure 15, we see that F1 is formed by one antinode (the extreme where the two lines are together, marked with an *A*), where pressure is at a maximum and velocity is at a minimum, and one node (the extreme where the two lines are separated from each other, marked with an *N*).

The consequence of the interaction between vocal tract constrictions and position of nodes and antinodes is that if the constriction point happens near a node, the formant frequency is lowered; if the constriction is near an antinode, the formant frequency is raised. For example: the constriction to produce /i/, in the front part of the vocal tract, is at the same time near a node for F1 and near an antinode for F2 (right-hand Figure 15). Thus, for the vowel /i/ F1 is low and F2 is high.

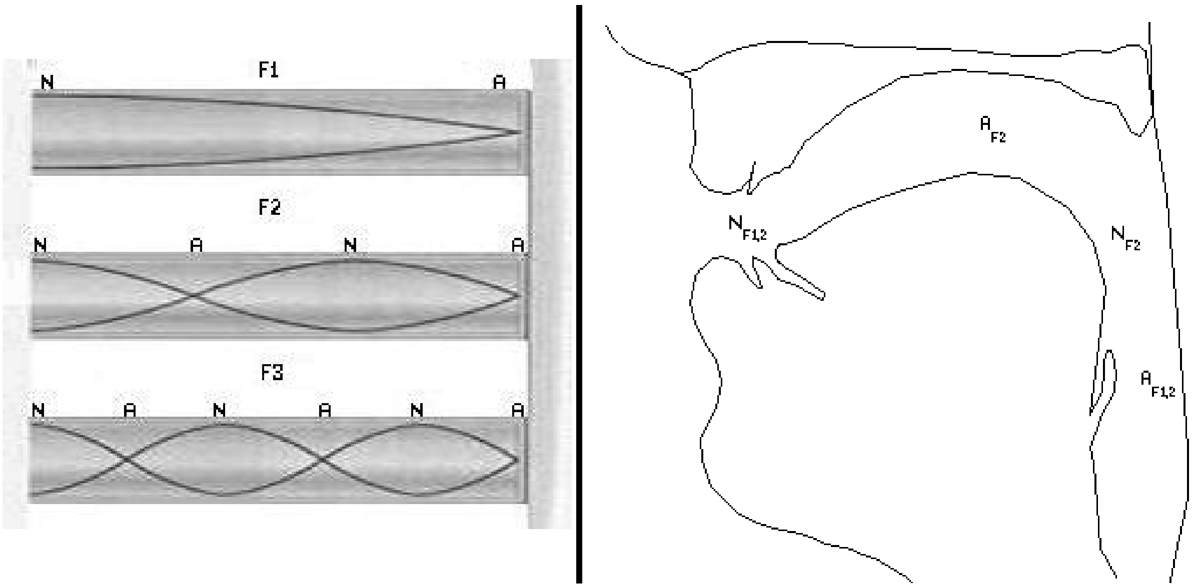


Figure 15. Constitution and location of nodes (N) and antinodes (A). The right-hand figure shows some points of constriction ($N_{F1,2}$, A_{F2}, N_{F2}, $A_{F1,2}$) (www.google.com.br).

3.2.2 Source-Filter Theory

The Source-Filter theory (Fant, 1960) explains how sound waves are generated and modified in order to generate speech sounds. Vowels are generated in the glottis by means of vocal fold vibration. This vibration provides the basic source of sound, which will eventually be modified in the vocal tract. The vocal tract functions as an amplifying filter that will attribute specific characteristics to the original sound wave. The position of the articulators will, therefore, determine which vowel sounds will be produced. This theory of generation and propagation of speech sounds generally gives us a description of how motor procedures interfere in the generation and propagation of speech sounds.

A final consideration about speech production relates the length of our vocal tract and the values of formant frequencies. The vocal tract can be interpreted as a set of tubes that begins in the glottis and ends in the lips; in this case, it will be referred to as vocal tube. By knowing the length of a speaker's vocal tube it is possible to calculate the formant frequency values for the neutral vowel /ə/ by using the formula: $F_{(n)} = (2n-1)C/4L$, where n is the number of nodes, L is the length of the vocal tube (in cm), and C is the speed of sound (34000cm/s). Thus, the F1 (which has one node) for /ə/ of a male speaker with a vocal tube of 17cm, for instance, will be 500Hz ($F_{(1)} = (2.1-1)34000/4.17 = 34000/68 = 500$).

3.3 Studies in cross-language speech perception and production

In this section, after presenting some general issues which relate the acoustic-articulatory dimensions of perception and the processing of speech, the main findings of recent studies on speech perception which have aimed at providing evidence for cross-language speech perception models are reported.

Considering speech perception from an acoustic-articulatory perspective, and according to Boersma (1997, 1998), the acoustic representation of speech sounds functions as a *guide* for articulatory commands. For example, assuming that speaker ***A*** produced a given utterance, the product of the ***A***'s articulated speech – the acoustic signal – is perceptually analyzed by the listener ***B***, who will respond to the speaker ***A***. ***A***'s response will, in turn, serve as a feedback for ***B***: if ***B*** got a possible correct answer for his intended question, he will understand that his utterance was well-produced by himself, and thus well-perceived by ***A***; however, if ***B*** got a nonsense answer from ***A***, he will understand that his production has to be adjusted so that ***A*** can understand his intended message. In this case, ***B*** will try to refine his articulation in order to be understood by ***A***. ***B*** may also listen to his own speech in order to make articulatory adjustments. This procedure of mapping articulatory gestures according to acoustic representations would not be feasible, and thus not necessary, if there was not an intrinsic relation between perception and production.

Accordingly, Kuhl (2000, p. 11854) points out that "vocal learning critically depends on hearing the vocalizations of others and hearing oneself produce sound". The author bases her claim on the fact that innately deaf babies are not able to learn their L1 naturally, since they cannot hear sounds. According to the author, that is also true for singing birds. Thus, there is consistent evidence for the link between perception and production, although there still is some debate on which one comes first.

Wilson, Saygin, Sereno and Iacoboni (2004) also provided visual evidence for the link between perception and production by using functional magnetic resonance imaging (fMRI) equipment. The authors showed that acoustic input activates specific areas in the brain that are responsible for articulatory representations. Although the authors do not make any claims on which ability comes first, the results suggest that perception precedes production.

Baker and Trofimovich (2001) decided to investigate whether accurate perception triggered production, or vice versa, or whether the two abilities evolved in parallel (that is, perception and production would be aligned and synchronic). The authors found that perception precedes production, and that the two abilities "may not necessarily be aligned". The authors also emphasize the role of self-perception[22], which "is essential in learning sounds, enabling speakers to gradually approximate their production ability to their perception ability" (Baker & Trofimovich, 2001, p. 282).

In a study on the perception of English nasals by Japanese and Korean speakers, Aoyama (2003) corroborated the assimilation patterns and their related level of discrimination accuracy proposed by Best's PAM. The author reported that speakers of Korean exhibited different assimilation patterns from the Japanese, but all of them were consistent with PAM's predictions.

In order to find empirical evidence to corroborate the hypotheses of the SLM (and, consequently, of other speech perception models), Flege and colleagues have carried out a series of studies with Italians perceiving and producing English vowels and consonants (Flege, Munro & MacKay, 1995; Munro, Flege & MacKay, 1996; Flege, Munro & MacKay, 1996; Flege, MacKay & Meador, 1999). These studies all found a strong positive correlation between the age of L2 learning and the degree of foreign accent; that is, the later they learned an L2, the more noticeable the foreign accent was.

The same tendency was found with Japanese adults producing English /ɹ/ and /l/. Flege, Takagi and Mann (1995) reported that the production of the nonnative /ɹ/-/l/

[22] Boersma (1997, p. 7) refers to self-perception as *auditory feedback*.

contrast improved as a result of L2 experience, and the degree of detectable foreign accent decreased as Japanese learners gained experience. This study constitutes important evidence against the CPH on the production domain, as well as for Flege's SLM.

In another study with Italians perceiving English vowels, Flege and MacKay (2004) also confirmed the SLM hypothesis that the ability to establish new sound categories stays intact over the life span, since they found that some of the participants who achieved native-like perceptual performance arrived in Canada after the age of six years.

Flege and MacKay (2004) also emphasize that a longer length of residence in a foreign country does not guarantee native-like perceptual performance. Comparing two groups of Italians who reported the same average length of residence (who were living in Canada for 15 to 26 years), but differed in L1 use, the authors observed that early learners who reported a low L1 use performed as good as native speakers. Still based on this comparison, the authors argue that the establishment of the L1 phonetic system does not prevent a native-like perceptual performance of nonnative speakers. Although the participants probably had similar L1 phonetic systems (since they arrived in Canada at approximately the same ages), the ones who reported a high L1 use performed not as well as the ones who reported a low L1 use, whose perceptual performance was native-like.

Flege, Munro and Fox (1994) also carried out a study to assess whether learners' experience[23] with an L2 affects their perceptual performance. Although the authors expected the perceived dissimilarities between L1 and L2 to become more evident as learners gained experience, which would result in an improvement of the perceptual

[23] Flege, Munro and Fox (1994) define *experienced learners* as "subjects who have lived longer in the U.S.(...) and reported using English more in a daily basis." They were also "estimated to have had roughly four times more English-language input" than the "inexperienced" subjects (p. 3628).

discrimination abilities, the authors found no significant difference in the performance of experienced and nonexperienced Spanish speakers perceiving English vowels. The authors attribute this failure in improvement to the rather small difference in amount of L2 experience and to individual differences.

The main objective of Flege, Munro and Fox's study, however, was to find out which factor would potentially influence learners' judgments of perceived dissimilarity between English and Spanish vowels: auditory-acoustic differences, categorical status (that is, whether the sounds were identified as belonging to one or two phonetic categories), or typicality, that is, "the extent to which phones match listeners' phonetic representations in long-term memory" (Flege, Munro & Fox, 1994, p. 3625). The authors found that the greater the distance between the two vowels, the greater the "magnitude of their auditory differences" (Flege, Munro & Fox, 1994, p. 3623), and the more accurate the learners' perception was. This finding demonstrates the importance of auditory acoustic differences for the perception of vowel dissimilarities.

Furthermore, Flege, Munro and Fox (1994) found that the categorical status (also referred to as *differential classification*) of the vowels also affected learners' perception. The authors observed that the learners' differential classification performance increased as they classified vowel pairs as belonging to two different categories.

The authors also concluded that categorical status and acoustic distance between vowels affect the perceived dissimilarity of vowels. Other studies also provided supportive empirical data for the effect of perceived acoustic distance on the discrimination of sounds. This tendency was reported by Lambacher et al. (1997) in a study involving Japanese speakers listening to the English fricatives, and the lack of perceived dissimilarity between /s/ and /θ/ was proved by measuring the perceived spacial distance between the two consonants (using a multidimensional scaling analysis), which indeed was very small (p. 190).

Flege, Munro and Fox (1994) also argue that L2 experience dos not change significantly Spanish/English bilinguals' ability to perceive dissimilarity within language (hearing pairs of English vowels) and between languages (hearing pairs formed by English and Spanish vowels), although recent studies have shown some exceptions. Morrison (2002) reported that one Spanish participant in his study performed native-like perceptual boundary adjustments while perceiving English /i/ and /ɪ/ and, as a result of L2 experience, was able to create a new category for English /ɪ/.

Considering the production of English vowels by BP speakers, Major (1987a) reports that L1 interference causes Brazilian learners to assimilate English /ɛ/ and /æ/ into the Portuguese /ɛ/ category in initial stages of acquisition, and that they eventually create a new category for /æ/ in more advanced stages, as a result of L2 experience and of teacher interference (who makes /æ/ more salient by emphasizing this sound) (Major, 1987a, p. 78). However, according to the author, the negative aspect of this process is that learners tend to overgeneralize, including both /ɛ/ and /æ/ in this new /æ/ category. Major bases all these claims on production data.

In a more recent study, Rauber (2006) found no evidence for the production improvement of the contrast /ɛ/-/æ/ as a result of experience. The author tested experienced Brazilian late learners of English, who had been English teachers for an average of eight years, and found two tendencies: either the participants categorized /ɛ/ and /æ/ as /æ/ (as reported by Major), or the two L2 categories merged.

Rauber (2006) also reports that experienced late learners showed a vowel contrast difficulty scale, according to which /i/-/ɪ/ was the easiest pair to perceive and produce, followed by /u/-/ʊ/. The pair /ɛ/-/æ/ was the most difficult in terms of perception and production (Rauber, 2006, p. 138).

The same perceptual difficulty pattern was reported by Wang and Munro (2004) with Mandarin and Cantonese speakers who received training on the three English vowel contrasts. After training, the perceptual difficulty scale was (from the least to the most difficult contrast): /i/-/ɪ/, /u/-/ʊ/, and /ɛ/-/æ/. However, before training the pattern was /u/-/ʊ/ more difficult than /ɛ/-/æ/, which means that the participants' overall perception of /u/-/ʊ/ improved significantly after intensive and controlled L2 exposure (by means of perceptual training).

Still focusing on speakers of Portuguese listening to English vowels, Flege (1995a, cited in Flege 1995b, p. 249) observed that Portuguese-speaking subjects demonstrated a tendency for discriminative failure, assimilating English /ɛ/ and /ɪ/ into the Portuguese /ɛ/ category. The same tendency was observed with Italian speakers (Flege & MacKay, 2004), whose L1 has the same seven-vowel system as Portuguese. However, there is no register of this assimilation pattern (English /ɛ/ and /ɪ/ as Portuguese /ɛ/) with Brazilian Portuguese speakers.

The studies reported in this Section provide empirical evidence for current speech perception models and shed some light on the role of age in this process. Kuhl and Flege believe that a critical period for learning by itself does not constrain language learning. The authors approach the role of age by a more cognitive perspective, instead of looking only by the physiological side. Thus, they believe that the linguistic limitations are imposed by a *long-date familiarity* with the L1 – not by brain plasticity – which will interfere to a certain extent on L2 acquisition, but which are possible to overcome. The NLM model and the SLM also agree on a connection between perception and production, and that perception comes before production. Best's PAM does not make a clear statement about this issue, although she agrees that there is an interaction between the two skills. More empirical research is needed to shed some light on unanswered questions.

Chapter 4

Perceptual training studies

In this chapter, some previous studies on perceptual training are summarized, and the effect of training, the use of synthesized stimuli and the effect of cue enhancement, generalization of the new knowledge, the perception and production interface, and long-term effects of training are discussed in subsequent sections. All of the studies reviewed in this chapter involve the perceptual training of segments whose errors would degrade intelligibility.

4.1 Effects of training

Training has been the focus of an increasing number of studies since the 1960s, and it has been used to approach both L1 and L2 phonetic contrasts. Many of the L1 training studies served as a means to assess the long-term performance of hearing impaired people or people with speech disorders, such as dyslexia or aphasia. For instance, one of the early studies on training (Lane & Moore, 1962) was a case study which aimed at reestablishing an L1 voicing contrast (/t/-/d/) in an aphasic adult.

As a result of the positive findings in the field of Speech Therapy, training started to be applied to more pedagogical settings as a means to improve L2 phonetic abilities in nonnative speakers, with the purpose of facilitating communication and diminishing foreign accent.

One study that illustrates the use of training to improve L2 production of phonemes is Strange and Dittman (1984). The authors attempted to improve native

Japanese speakers' ability to perceive English /r/ and /l/, which are not contrastive in their L1. The participants were trained on the discrimination of the targets, using synthesized instances of "rock" and "lock". Following Carney et al. (1977), the authors used synthesized speech to manipulate the stimuli by gradually increasing between- and within-category differences. In the discrimination training, the participants heard a pair of words and were simply asked to say if the words sounded the same or different. A control group formed by Americans performed the same tasks as the Japanese and, after 18 training sessions over three months, the Japanese – whose performance was much poorer than the Americans' before training – discriminated the targets as well as the Americans after training.

A secondary objective of Strange and Dittmann's study was to check whether the participants would generalize the knowledge acquired with synthesized stimuli to natural settings; However, the authors could not find any transfer of this kind. They only found transfer to new words of the same nature, that is, synthesized "rake" and "lake". Generalization will be further discussed in Section 4.3.

In another attempt to find generalization of training, Jamieson and Morosan (1986) carried out a study focusing on the learning of English /θ/ and /ð/ by native speakers of Canadian French. The participants received identification training with synthesized stimuli for four days. In identification or labeling tasks, the participants heard a word containing the target segment and they had to give a label to this segment from two or more choices. The authors also applied the perceptual fading technique (Terrace, 1963), which consists of gradually reducing the amount of perceptual contrast between two segments which initially had salient perceptual differences. This can be done, for instance, by adding background noise or manipulating the formants of naturally produced tokens in very small steps – which was the authors' choice – in order to avoid errors and to prevent the participants from feeling frustrated. Results show that, after only 90 minutes of practice, not only did the participants improve their between-category perceptual abilities, but they also transferred this improvement to natural tokens which had not been trained.

The same tendency was found in a study by Yamada et al. (1996) with native speakers of Japanese who were trained on the /r-l/ contrast. The participants received 45 perceptual training sessions over 14 days, after which both their perceptual and production abilities improved. Follow-up tests also showed that the improvement in the participants' perceptual performance was maintained six months after training was over. Long-term effects of training will be addressed again in Section 4.5.

Another study on the perception of /r/ and /l/ by Japanese learners was carried out by Bradlow et al. (1997). After 45 sessions of identification training with natural stimuli over approximately four weeks, the authors found evidence for generalization of the new ability to new speakers and new tokens. However, the most important contribution was in the perception/production domain: The knowledge acquired with perception training was transferred to production, which also improved even without any specific production training. This issue is addressed more thoroughly in Section 4.4.

In another study with Japanese and Korean speakers perceiving /r/ and /l/, Hardison (2002) implemented audiovisual training and compared it with auditory training. Both kinds of training involved identification tasks with immediate feedback plus reinforcement: that is, the participants would hear/see a stimulus, choose an answer, receive feedback, and hear/see the same stimulus again (reinforcement). The results showed that all groups improved their perception significantly after training, but the group that received audiovisual training had a greater improvement than the group that received only auditory training. Production improvement was also assessed and confirmed. The author concluded that, given the appropriate conditions and training design, adults can improve their perceptual abilities for identifying nonnative sounds.

Pisoni et al. (1994) contributed to the perceptual training field by emphasizing the role of selective attention in one's perceptual performance. The fact that adult L2 learners have their L1 categories robustly formed leads them to tune their speech perception abilities to the L1 parameters, causing L2 sounds to be filtered and interpreted within their L1 frame. The authors claim that the decrease of discriminatory ability and its subsequent reestablishment can happen as a result of selective attention, based on experience of what is important. Thus, critical portions of the signal become more distinctive in terms of perception, and portions of the signal that are not crucial become less distinctive.

The results of a study by Rochet (1995) on voicing contrasts of French stops with speakers of Mandarin showed that extensive training (in terms of number of hours and days) is not a prerequisite for the effectiveness of results. After a total of only three hours of training distributed among six sessions, the author showed that the perceptual boundaries of the target stops were significantly closer to the French boundaries. Furthermore, as reported in other studies in this section, the author also reported the transfer of learning to new perception stimuli, as well as to the production level.

As for perceptual training of English vowels, three studies were selected to be presented in this section. Ceñoz Iragui and Garcia Lecumberri (1999) carried out a study with native speakers of Spanish and Basque learning to perceive the vowels of British English. The participants received approximately 14 hours of identification training on the targets. Comparing the results of the participants who performed better with the results of the ones who performed worse in the pretest, the authors found that the degree of improvement of the latter group was much higher than that of the former. They attribute this to a ceiling effect: The group that performed better in the pretest had already done their best and there was little to be improved, whereas the group who performed worse in the pretest had more room for improvement, thus allowing them to have a greater degree of improvement from pretest to posttest. Nevertheless, the authors assume that if those participants who already got good results in the pretest were exposed to more hours of training or to different kinds of training, they would benefit more.

The studies by Wang (2002) and Wang and Munro (2004) focused on the learning of three English vowel contrasts (/i-ɪ/, /ɛ-æ/, /ʊ-u/) by native speakers of Mandarin and Cantonese. The main purpose of these studies was to use computer-based perceptual training to cope with individual differences, especially in terms of amount of training time needed. The authors concluded that computer-based training was an efficient means to provide learners with a more flexible and personalized schedule during the training period. The authors also found evidence for generalization and for long-term retention of training. After training, the participants were able to generalize the new knowledge to new tokens and new speakers, and they retained the perceptual improvement for three months after the training sessions were over. The only shortcoming in these studies was that the authors did not control the individual amounts of training time, preventing them from affirming that the participants who got the best results were the ones who trained longer.

A comparison of percentage of correct vowel categorization by vowel contrast shows that the contrast /i-ɪ/ was correctly categorized more times after training, followed by /ʊ-u/ and by /ɛ-æ/. However, in terms of improvement (that is, the difference from prettest to posttest results), the contrast /ʊ-u/ improved the most, followed by /ɛ-æ/ and by /i-ɪ/. Thus, although the contrast /i-ɪ/ showed the best results in terms of raw number of correct categorization, it was the one that improved the least. This was probably due to a ceiling effect; that is, because learners already performed well in the prettest, the room for improvement of that vowel contrast in the posttest was reduced. Conversely, the contrast /ʊ-u/ was the least correctly categorized in the prettest and, with a 32-point improvement after the posttest, it was the contrast that improved the most, followed by /ɛ-æ/ (16-point improvement) and by /i-ɪ/ (14-point improvement) (Wang, 2002; Wang & Munro, 2004).

A study by Wang et al. (1999), whose object of study was Mandarin tones, shows that the positive effects of training are not restricted to the segmental level. In this study, the authors showed that perceptual training at the suprasegmental level is also effective, and that perceptual improvement can be generalized and retained by learners.

All the studies mentioned in this section illustrate the overall positive effects of perceptual training of L2 sounds. The following section addresses the nature of the training stimuli and how it can interfere in the training process.

4.2 Training with natural or with synthesized stimuli? Using cue enhancement

In this section, the use of cue enhancement as a means to maximize phonetic contrasts in order to cope with the problem of category overlap will be discussed.

In laboratory studies, Kuhl et al. (1997) found that mothers from three different countries (the United States, Russia and Sweden) exhibited the same pattern in the language addressed to their babies, which was different from the everyday language they used to communicate with other people. Kuhl et al. (1997, p. 684) report that "phonetic units in adult-directed speech are often poorly specified", whereas the vowels in infant-directed speech were "acoustically more extreme, resulting in an expanded vowel space" (see Figure 14).

This tendency found by Kuhl et al. (1997, 2000) can be easily applied to the L2 learning context. In this sense, L2 learners could be exposed to a kind of laboratory-developed stimulus with its acoustic properties enhanced. Thus, the American English vowel space could be stretched, making the L2 vowels more distant from one another, preventing category overlap. This increased distance between vowels would, therefore, be expected to facilitate perception. As Logan and Pruitt (1995, p. 358) point out, "exaggeration of acoustic differences between contrasting phones may focus learners' attention on a critical portion of the signal that might not otherwise be salient." Hazan and Simpson (2000) also emphasize the advantages of using "clean" speech as a means to promote intelligibility by amplifying the parts of the signal that contain the crucial acoustic cues.

Sometimes, natural input is not sufficient to clearly show the acoustic differences of some contrasts. Using natural speech stimuli to train diphthongs, for instance, could be very appropriate because their differential acoustic properties are more evident. However, to train the pair /ɛ-æ/ or /ʊ-u/, stimuli treated with cue enhancement would be more appropriate, since it can emphasize the subtle L2 cues which characterize each member of these pairs.

One of the pioneering training studies involving synthesized speech stimuli with cue enhancement of a nonnative contrast was carried out by Strange and Dittmann (1984). The authors used synthesized stimuli in their study in order to provide Japanese learners with more precise instances of /r/ and /l/, since they have difficulty in discriminating these phonemes in American English natural speech.

Jamieson and Morosan (1986) went further and used speech synthesis to apply the perceptual fading technique into their stimuli. As cited in Section 4.1, the authors used this technique in order to teach students to focus on between-category acoustic patterns and to ignore within-category variability. The same technique was used by Wang (2002) and Wang and Munro (2004) in a study on the perceptual training of American English vowels. Mandarin and Cantonese speakers were exposed to two sets of stimuli during the training sessions. The first set of stimuli had duration manipulated. Thus, short and long tokens of /i/, as well as short and long tokens of /ɪ/ were generated. Because /i/ and /ɪ/ are allophones in the learners' L1, Mandarin and Cantonese speakers rely primarily on duration to distinguish these L2 vowels. The purpose of generating long and short instances of /i/ and /ɪ/ was to show learners that in English a short /i/ and a long /i/ are all /i/s, and a short /ɪ/ or a long /ɪ/ all belonged to the same /ɪ/ category. Having learned not to rely primarily on duration to categorize English vowels, the participants were exposed to the second set of stimuli, which was treated with the perceptual fading technique and whose purpose was to teach the participants to focus on between-category differences, as reported above.

Because of the different L1 perceptual cue patterns of the vowels of Japanese and English, Fox and Maeda (1999) carried out a study involving cue enhancement and the categorization of English /i/ and /ɪ/ by Japanese speakers. Whereas Japanese vowels can be distinguished by duration, Americans rely primarily on spectral quality to distinguinsh English vowels. The training stimuli developed by the authors consisted of synthesized natural speech with removal of durational variation of the tokens. This was done to lead learners to rely on L2 primary cues, that is, spectral quality. The results showed that, after 10 sessions, the learners were able to rely on nonnative acoustic cues to categorize L2 sounds.

The studies reviewed in this section show how synthesized stimuli and cue-enhancement can be used in training with the purpose of helping learners to perceive nonnative sounds. The next section presents the findings of previous studies related to generalization of learning to novel stimuli and novel tasks. It also reports briefly how generalization can be affected by the training tasks.

4.3 Generalization of the perceptual learning and the effects of the kind of training task

Logan and Pruitt (1995) define generalization of learning as the ability to transfer the acquired knowledge to multiple dimensions, such as novel productions of new talkers, new productions of the same talker, new tasks, etc. The authors claim that if generalization occurs, "we can be more confident that robust learning has occurred" (p. 371). Generalization enables learners to abstract relevant parameters of the signal, thus ignoring secondary acoustic and phonetic variations which are not distinctive, such as pitch.

Nevertheless, Rochet (1995) points out that the perceptual behavior with the target contrasts should be similar in different phonetic contexts and/or phonological units to claim generalization as a result of training only one category/environment. For instance, in his 1995 study in which the syllable-initial voicing contrast /b-p/ was trained, although the perceptual improvement of synthesized tokens was carried over to new CV tokens and to natural stimuli, Rochet found no transfer to intervocalic targets. According to the author, this specific generalization failure was due to a shift of the primary perceptual cue. Thus, when the targets appear in syllable-initial position, the primary cue is VOT (a cue learners were trained on), whereas in intervocalic position, the primary cue to perceive the targets is "voicing versus lack of voicing of the closure interval" (Rochet, 1995, p. 403). This example supports the importance of a careful selection of materials and training methodology.

As to the generalization gains, Rochet (1995) reports that after training VOT with synthesized tokens of /bu/ and /pu/, the learners performed significantly well on the identification of /b/ and /p/ followed by new vowels (/a/ and /i/), as well as on the identification of new voiceless stops (/t/ and /k/) followed by /u/. However, as to the transfer of learning to natural stimuli, the author only found generalization to the voiceless targets, with no significant change on the perception of natural tokens of voiced stops.

As pointed out in Section 4.1, one of the conclusions of Strange and Dittmann's (1984) study was that the participants' performance, when exposed to natural stimuli in the generalization test, was much poorer than their performance in the tasks with synthesized stimuli. The authors attributed this fact to the kind of task used in training, that is, discrimination training. Jamieson and Morosan (1986) also concluded that the discrimination training was not the best task choice to promote generalization. Accordingly, Logan and Pruitt (1995, p. 359) consider that "identification tasks appear to be more effective than discrimination tasks in promoting generalization to novel stimuli not presented during training."

According to Jamieson and Morosan (1986), discrimination tasks improve only within-category acoustic differences, and "such increased within-category sensitivity could diminish attempts to form a new phonetic category" (p. 207). The authors also claim that identification tasks with feedback seem to be more appropriate to improve listeners' abilities to classify speech sounds into relevant target language categories.

Discrimination training, however, is not worthless. Discrimination training can be very useful both in the very beginning of training – to "show" learners that two sounds are different, even though they cannot say which sound is which, as in category change tasks[24] – as well as in more advanced stages – to train within-category variability.

The generalization of learning was also reported in several training studies on the learning of the /l-r/ contrast by Japanese speakers and on the learning of English vowels by Mandarin and Cantonese speakers. Yamada et al. (1996) reported a high accuracy (approximately 80% of correct answers) in generalization tests with new words spoken by the same speakers and with new words spoken by new speakers. This study was replicated in 1997 by Bradlow et al., who found the same patterns. In another study on the learning of three English vowel contrasts by Mandarin and Cantonese speakers, Wang (2002) and Wang and Munro (2004) also reported generalization of learning with both natural and synthesized stimuli to new tokens and new speakers.

Multiple dimension transfer was also reported by Pruitt et al. (2006), who carried out a study on Hindi consonants. The authors found evidence for generalization of the new knowledge to novel vocalic and to new voicing/aspirating contexts produced by a different speaker. The authors attributed this extensive generalization tendency with adult learners especially to the training material, which has to be carefully selected and highly variable.

Generalization was also reported in a study involving a different modality of perceptual training: Audiovisual training (Hazan et al., 2005). In this study, among other issues, Hazan et al. compared the degree of generalization in two different groups, one that received auditory training and the other received audiovisual training. In spite

[24] In category change tasks, several instances of a sound category are presented in a sequence which also includes instances of a different category, and the listener has to "indicate when a change in category has occurred" (Logan & Pruitt, 1995, p. 356).

of this methodological difference between groups, the authors found that there was transfer of learning to new stimuli in both groups to the same extent.

Finally, Wang et al. (1999) reported high levels of learning transfer on the suprasegmental level in a study with American speakers learning Mandarin tones. The authors found up to 94% of correct answers in the generalization tests with novel talkers and tokens, which correspond to a 25% increase if compared to the posttest. Thus, the results of the generalization tests not only showed a transfer of the new knowledge to new stimuli, but also showed an improvement on the learners' performance.

All of the perceptual training studies involving identification training reviewed so far and which comprised generalization tests showed positive results for the transfer of the acquired knowledge to new settings. In the next section, the relation between perceptual training and eventual gains in production will be discussed.

4.4 The perception and production interface

Some studies (Rochet, 1995; Yamada et al., 1996; Bradlow et al., 1997; Hardison, 2002; Wang et al., 2003; Hazan et al., 2005) have provided empirical evidence of the impact of speech perception training on the production domain. The studies reported in this section show that perception training led to production improvement even without any specific production training.

Rochet (1995) reported production improvement after three hours of perceptual training, and this improvement was parallel to what happened in the generalization of perceptual learning. Therefore, there was production improvement for voiced and voiceless stops, but the results were much better for the production of voiceless stops. Furthermore, there was no production improvement for intervocalic targets (because of the shift of the primary perceptual cue, explained in Section 4.3, p. 56), which was also found in the perceptual generalization results. Rochet attributes this parallel improvement to the fact that both languages in his study share the same phonological patterns for stops (although with different VOT values): "VOT durations are longer for velars than for dentals or labials, and in the presence of a following vowel" (Rochet, 1995, p. 402).

This parallel degree of improvement, however, was not often found in other studies, which have shown improvement in production after perceptual training, but not to the same extent as for perception. Studies by Bradlow et al. (1997), Wang et al. (2003), and Hazan et al. (2005) have shown this tendency.

Bradlow et al. reported a positive correlation between production and perceptual improvement, but they also found that the degree of learning in perception was a little higher than the degree of learning in production after perceptual training, and the authors attribute that to "the wide range on individual variation in learning strategies" (Bradlow et al., 1997, p. 2307). The authors explain that these individual differences are probably related to learners' individual timing, which can cause some learners to acquire the "specific motor commands necessary for improved /r/-/l/ production" faster than others (Bradlow et al., 1997, p. 2308).

Wang et al. also found a positive correlation between perception and production improvement after perceptual training. Furthermore, similarly to Bradlow et al. (1997), the authors found differences in the degree of performance in perception and production. However, these differences were not in the performance of individual subjects producing the target Mandarin tones, but in the production of specific tones which were more difficult than others, although all of them improved perceptually. The authors assumed that this lack of production improvement for certain tones was due to the rather novel properties of the pitch contour for nonnative learners (Wang et al., 2003).

Production improvement to a different degree from perceptual improvement was also found by Hazan et al. (2005). The innovation in this study was that it involved audiovisual perceptual training. The results showed that, although the group that listened only to auditory stimuli and the group that was exposed to audiovisual stimuli improved in perception to the same extent, the group that received audiovisual stimuli showed a greater production improvement than the group that received only auditory training, even though the targets (/l/ and /r/) are a "less visually-distinctive contrast" (Hazan et al., 2005, p. 374). The same tendency was found in a similar study by Hardison (2002).

Besides the evidence that perceptual training leads to production improvement, there is also the advantage that the development of perceptual training materials in terms of feedback is much more feasible than production (pronunciation) materials. To give accurate ratings for the productions of an infinite number of utterances of the same speaker, or even for utterances of a large number of speakers, is almost an impossible task. Teachers cannot give immediate feedback for each and every learner (even with the help of an assistant, that would still be difficult!) and, in the case of computer-based training programs, the rating accuracy of learners' productions would not be very high due to the great variance within and between speakers (Rochet, 1995; Yamada et al., 1996; Hazan et al., 2005). Thus, perception training would be "easier to administer than production training" (Rochet, 1995, p. 396).

4.5 Long-term effects

Some of the studies reported in this chapter also aimed at investigating whether the knowledge acquired by means of perceptual training would be retained some time after training had ended. Wang et al. (1999), Wang (2002) and Wang and Munro (2004), for instance, reported that the degree of perceptual accuracy was maintained over the following three months (Wang, 2002; Wang & Munro, 2004) or even over six months (Wang et al., 1999). Moreover, this retention has been found not only in the phonological level, since the study by Wang et al. (1999) focused on suprasegmentals (Mandarin tones).

The retention of knowledge acquired through perceptual training also goes beyond the domain of speech perception, as shown by Yamada et al. (1996). In this study, learners took a production retention test three months and six months after training was over, and in all tests learners in the experimental group performed significantly better than the ones in the control group, whose performance was worse than in the prettest.

The results of the study by Ceñoz Iragui and Garcia Lecumberri (2002) relate the issue of long-term effects of training to self-perception. One year after the training, the authors carried out a retention test and found something more than just maintenance of perceptual improvement. Indeed, some of the targets not only retained the improvement gained during perceptual training, but even continued to improve to a higher level than in the posttest. However, the perception of other trained vowels decreased to the same level as the results in the pretest, or to a lower level than the results in the posttest, and some of these vowels were among the ones that improved after training. The tentative explanation the authors gave to this fact was that "learners are aware of their own accuracy (…) and are motivated to improve the discrimination of those vowels (…) that they find more difficult" (Ceñoz Iragui & Garcia Lecumberri, 2002, p. 59). Thus, they self-monitored their perception and started giving more attention to the vowels that they found more difficult after the posttest. Conversely, they ignored the "easier" vowels, which may have contributed to a worsening in the perceptual discrimination of those vowels.

Based on the findings of the studies reviewed in this section, it can be concluded that training can provide learning gains that are not generally lost after it is over. This evidence justifies the implementation of training in L2 language classes, since the time spent with training will not be worthless: it is knowledge that will be retained.

Chapter 5

Method

In this chapter the participants, the materials and the procedures in the study are described. Some of the materials and procedures used in the tests are based on Rauber (2006). Also, details on the selection of the participants, on the development of training tools and on the application of tests and training activities are provided.

5.1 Participants

Thirty-six undergraduate students of English at the Universidade Federal de Santa Catarina (UFSC) in Brazil participated in this study: 7 in the control group of Brazilian Portuguese native speakers, with no specific phonetic training, and 29 in the experimental group, which received perceptual training. The experimental group consisted of third- and fourth-semester students of the undergraduate English program and fifth-semester students of the undergraduate Executive Secretary program. There was also a group of two native speakers of American English, which served as a control group for the perception tests. (See Section 4.1.1 for further details on the description of the groups.)

Except for the students in the control groups, all of the participants in this study were enrolled in the "Pronunciation Lab" class, which is taught during the third semester of the undergraduate English program. Note that the data was collected in two different semesters due to a high mortality rate (i.e., many participants did not participate of all tests and had to be excluded) in the first semester. The second data collection compensated for this.

All the Brazilian participants reported to be intermediate speakers of English, more specifically of the General American English variety. None of them had been to an English-speaking country for more than one month or spoke another foreign language fluently. None of them reported having any hearing problems.

An outline of the background information of each of the participants in the study is shown in Table 6 for the experimental group, and in Table 7 for the control groups. Table 6 shows that most of the participants are from and have spent most of their lives in cities in the south of Brazil. The only exceptions are participants 24 and 28, who are originally from São Paulo (southeastern Brazil), but they have also spent most of their lives in southern cities (Porto Alegre and Florianópolis, respectively). All of them are full-time undergraduate students at UFSC and their ages range from 18 to 30 years (mean = 20.8 years).

Table 6. Background information of the participants in the experimental group.

Part.	Gender	Age	Place of birth	Place where spent most of life
E1	F	18	Nova Trento-SC	Nova Trento-SC
E2	F	18	Garopaba-SC	Garopaba-SC
E3	F	18	Xanxerê-SC	Xanxerê-SC
E4	F	20	Erechim-RS	Itá-SC
E5	F	19	Londrina-PR	Florianópolis-SC
E6	F	19	Rio do Sul-SC	Rio do Sul-SC
E7	M	24	Tijucas-SC	Tijucas-SC
E8	M	19	São João Batista-SC	São João Batista-SC
E9	F	20	Pato Branco-PR	Pato Branco-PR

Part.	Gender	Age	Place of birth	Place where spent most of life
E10	F	21	Santo Amaro da Impetatriz-SC	Palhoça-SC
E11	F	28	Presidente Getúlio-SC	São José-SC
E12	F	30	Florianópolis-SC	Florianópolis-SC
E13	F	20	Florianópolis-SC	Biguaçu-SC
E14	F	17	Curitibanos-SC	Curitibanos-SC
E15	F	20	Vacaria-RS	Vacaria-RS
E16	F	19	São Lourenço do Oeste-SC	São Lourenço do Oeste-SC
E17	F	19	Araranguá-SC	Araranguá-SC
E18	M	19	Florianópolis-SC	Florianópolis-SC
E19	F	18	Imbituba-SC	Imbituba-SC
E20	M	25	Florianópolis-SC	Florianópolis-SC
E21	F	21	Florianópolis-SC	Palhoça-SC
E22	F	25	Nova Fátima-PR	Santo Antônio do Paraíso-PR
E23	M	20	Palhoça-SC	Palhoça-SC
E24	F	25	São Paulo-SP	Florianópolis-SC
E25	F	21	Florianópolis-SC	Luzerna-SC
E26	M	24	Florianópolis-SC	Biguaçu-SC
E27	F	20	Florianópolis-SC	Florianópolis-SC
E28	F	20	São Paulo-SP	Porto Alegre-RS
E29	F	19	Florianópolis-SC	Florianópolis-SC

E = Experimental

The information about the participants in the control groups is shown in Table 7. The Brazilian control group consisted of seven fifth-semester students of the undergraduate Executive Secretary program who had not taken any pronunciation (or phonetics) subject by the data collection time. Their ages ranged from 20 to 24 years (mean = 21.8 years) and most of them had spent most of their lives in states of southern Brazil. The American control group was formed by two native speakers of American English, one being from Tennessee and the other from New York.

Table 7. Background information of the participants in the control groups.

	Part.	Gender	Age	Place of birth	Place where spent most of life	Occupation
Brazilians	CP1	F	23	Suzano-SP	Suzano-SP	Undergraduate student
	CP2	F	20	Florianópolis-SC	Florianópolis-SC	Undergraduate student
	CP3	F	20	Palhoça-SC	Palhoça-SC	Undergraduate student
	CP4	F	21	Florianópolis-SC	Florianópolis-SC	Undergraduate student
	CP5	F	21	Estação-RS	Presidente Getúlio Vargas-SC	Undergraduate student
	CP6	F	20	Porto Alegre-RS	Porto Alegre-RS	Undergraduate student
	CP7	F	24	Florianópolis-SC	Florianópolis-SC	Undergraduate student
Americans	CE1	M	39	Kingsport-TN	Kingsport-TN	Graduate student
	CE2	M	33	New York-NY	New York-NY	Economist

CP = Control Portuguese / CE = Control English

The assignment of the participants to one of the four groups in this study (the two experimental and the two control groups previously mentioned) is described in this sub-section. The experimental group was divided into two sub-groups, which I will initially refer to as *Group A* and *Group B*. The participants were assigned to Group A or Group B according to their results in the perception pretest (see Section 4.2.3 for details on the perception pretest) so that each group would be formed with equivalent pretest scores and the level of proficiency regarding perception of the target contrasts would be balanced in both groups. Thus, of the two students who got the highest scores in the pretest, one would be assigned to Group A and the other to Group B, and so on. Table 8 illustrates this distribution. As mentioned above, however, some of the participants had to be excluded because they missed one or both post-tests or one of the training sessions, so the experiment was carried out with another class the following semester in order to increase the number of participants in the two groups. Unfortunately, most of the participants who were excluded were in Group B, and even distributing the students in the following semester in the same manner, it was not possible to make the groups as equivalent as desired, but still the means of the two groups are similar (Group A = 43.6; Group B = 41.6).

Table 8. Distribution of the participants of the experimental group into the two training subgroups.

Group A		Group B	
Participant	Result in the pretest	Participant	Result in the pretest
E10	54	E21	59
E11	53	E22	53
E2	51	E28	51
E4	48	E25	48
E5	49	E24	45
E12	47	E19	44
E9	45	E17	42
E1	43	E23	40
E7	43	E16	38
E13	43	E20	37
E8	42	E29	36
E14	42	E18	35
E6	36	E26	31
E3	34	E27	23
E15	24	-	-

The participants in each of the subgroups received a specific kind of perceptual training: Group A, which from now on will be referred to as *SynS* group, received perceptual training with synthesized stimuli only, and consisted of 15 students; Group B, from now on referred to as *NatS* group, had 14 students who received training involving natural stimuli exclusively.

The control groups will be referred to either as *BP Control* or *AE Control*, depending on the speakers' native language. The BP Control group was formed by the seven undergraduate Executive Secretary students previously mentioned, and the AE Control group consisted of the two native speakers of American English, described in Table 7. Table 9 shows the results of the BP Control group in the perception pretest. The distribution of the groups involved in this study can be visualized in Figure 16.

Table 9. Results of the BP Control group in the perception pretest.

BP Control	
Participant	Result in the pretest
CP1	52
CP5	52
CP3	43
CP6	43
CP2	38
CP7	37
CP4	36

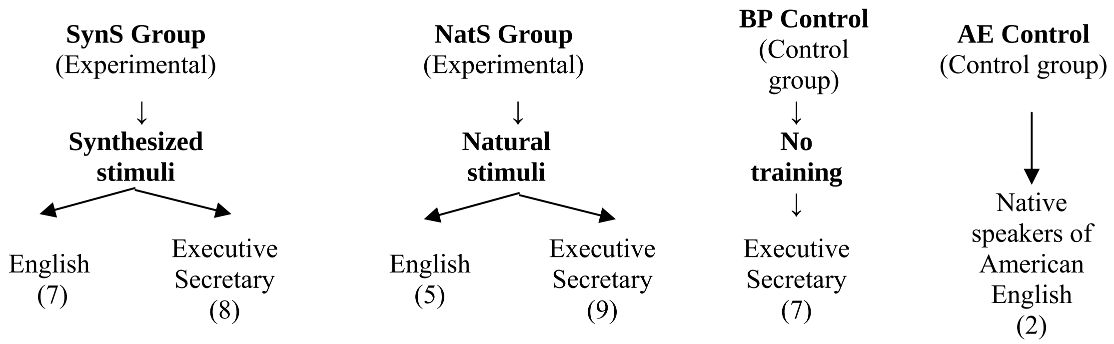


Figure 16. Distribution of the students within each experimental group and control group. Note the number of participants in each group in parentheses.

5.2 Materials

5.2.1 Background Questionnaire

The background questionnaire (Appendix A) consisted of twelve questions about a set of variables that could influence the learner's performance, such as place of birth, other foreign languages spoken (and degree of fluency), the amount of time the learner had been studying English, and/or how long the learner spent in an English speaking country before.

5.2.2 Production test[25]

There was one production test which was given at three different times: a pretest, which was administered before the training began, a posttest, which was administered as soon as the training was over, and a retention test, which was administered one month after the posttest.

[25] This instrument was borrowed from Rauber with permission (2006).

5.2.2.1 Description

In the production tests, the participants read a total of 116 monosyllabic English words containing the vowels /i, ɪ, ɛ, æ, ʊ, u/ inserted in a voiceless consonantal context[26]. As represented in Table 10, each of the targets was read 5 times in isolation and 15 times within a sentence, except for /ʊ/, which was read 4 times in isolation and 12 times within a sentence, for the reasons given below (see Table 11).

Table 10. Number of tokens read in the production test.

Target vowels	Number of tokens read in isolation		Number of tokens read within a sentence		
/i/	5		15		
/ɪ/	5		15		
/ɛ/	5		15		
/æ/	5		15		
/ʊ/	4		12		
/u/	5		15		
Total	29	+	87	=	**116**

In addition, the participants recorded 60 words containing the vowels /ʌ/, /ɑ/ and /ɔ/ (5 in isolation and 15 contextualized for each, just as the target vowels), also within a voiceless context. The main reason for including these vowels was because the data were collected with a group of students enrolled in the subject *Laboratório de Pronúncia* of the undergraduate English program, in which all nine English vowels had to be practiced, even though they were not intended to be included in the data analysis.

[26] There was also one context which included a voiced previous consonant (/bVt/), which was not used in the analysis.

However, after analyzing the perception data, it was decided to analyze the vowel /ʌ/, since it was frequently confused perceptually with the target high back vowels (see Section 6.2.1 for details).

All the words to be recorded appeared first in isolation and then within a sentence, in the format shown in Table 11. Thus, the participants would read: "Beat. Beat and Pete sound like seat. Pete. Beat and Pete sound like seat. Seat. Beat and Pete sound like seat" (see Appendix B).

Table 11. Format of sentences read by the participants in the English production tests. The target words appear in italics.

(bee)	Beat. *Pete*. *Seat*.	Beat and *Pete* sound like *seat*.
(egg)	*Tet*. *Tech*. *Kept*.	*Tet* and *tech* sound like *kept*.
…	…	…

It is important to mention that the target vowels appeared in real words (some of them quite rare), except for one nonsense word (Table 12). For this reason the participants were provided with a very common word containing the target vowel as reference, which was written in parentheses before each set of target words, as shown in Table 11. The presence of the reference words was crucial not only to prevent participants from producing incorrect utterances provoked by orthographic interference, but also to guide their pronunciation when they were confronted with the nonsense word or an unfamiliar word. The nonsense word (*tuke*) was included in order to have almost the same consonantal contexts for all target vowels.

The target vowels in the production tests were presented within voiceless contexts, as in Table 12, because these contexts facilitate the identification of the vowel boundaries and, as a result, the whole process of vowel segmentation. The contexts were /sVt/, /ʃVt/, /pVt/, /pVp/, /tVt/, /tVk/, /kVt/, /kVk/, and /kVp/. Only the target vowel /ʊ/ did not appear in all contexts since it was not possible to find a written representation (real or nonsense) for it in the /tVt/ context that would not be confused with /u/. The target words in the production tests are shown in Table 12.

Table 12. Target words recorded by the participants in the English production tests. The word with an asterisk is a nonsense word.

Context	/i/	/ɪ/	/ɛ/	/æ/	/ʊ/	/u/	/ʌ/
/sVt/, /ʃVt/	seat	sit	set	sat	soot	suit	shut
/pVt/, /pVp/	Pete	pitt	pet	Pat	put	poop	putt
/tVt/	teat	tit	tet	tat	-	toot	tut
/tVk/	teak	tick	tech	tack	took	tuke*	tuck
/kVp/, /kVt/, /kVk/	keep	kit	kept	cat	cook	coot	cut

The purpose of selecting this specific context (CVC) is explained by the phonology of the language. As discussed in Chapter 2, English lax vowels do not appear in open syllables.

5.2.2.2 Production test procedures

The English production pretest was administered on the first day of class. The participants were recorded individually in a quiet room with a Sony Minidisc Recorder and a Sony ECM-MS907 condenser microphone. Before starting the recording session, instructions were given in English and the participants were told that they would not record the words in parentheses in the English production test, which served just as reference and that the three following words would contain the same vowel. The words were presented on sheets of paper and the learners were asked to read each word/sentence at a normal speech rate. On the average, the participants spent 3.5 minutes to record the production test.

5.2.2.3 Production analysis

The recordings of the words and sentences containing the target vowels were digitized at 22kHz, with 16-bit accuracy, MONO, in the *Audacity 1.2.4* software. The production analysis was done following the procedures described in Rauber (2006). Thus, a wideband spectrogram and the waveform of the words containing the target vowels were visualized using the *Praat 4.4.01* software, and two tiers were generated in order to perform the segmentation and labeling actions. In the first tier, named "word", the word containing the target vowel was labeled, whereas the second tier, named "vowel", contained the label of the segmented target vowel. The purpose of the first tier was to facilitate the localization of each word that had a target vowel, making the vowel segmentation process faster. Next, in order to segment the targets, vowel boundaries were selected close to the first and last zero crossings[27], marking the beginning and the end of the target vowel. After the vowel boundaries were marked, the vowels were finally labeled. The actions performed in Praat are illustrated in Figure 17, which exemplifies the segmentation and labeling of the vowel /i/ produced by one of the female participants. After segmentation, the target vowels were visualized on a wideband spectrogram and their acoustic properties (duration, F0, and the first three

[27] Zero crossing is the portion of the waveform that crosses zero amplitude.

formants) were then measured from the central 40% of the target vowels. All these actions were done using the *Praat* software.

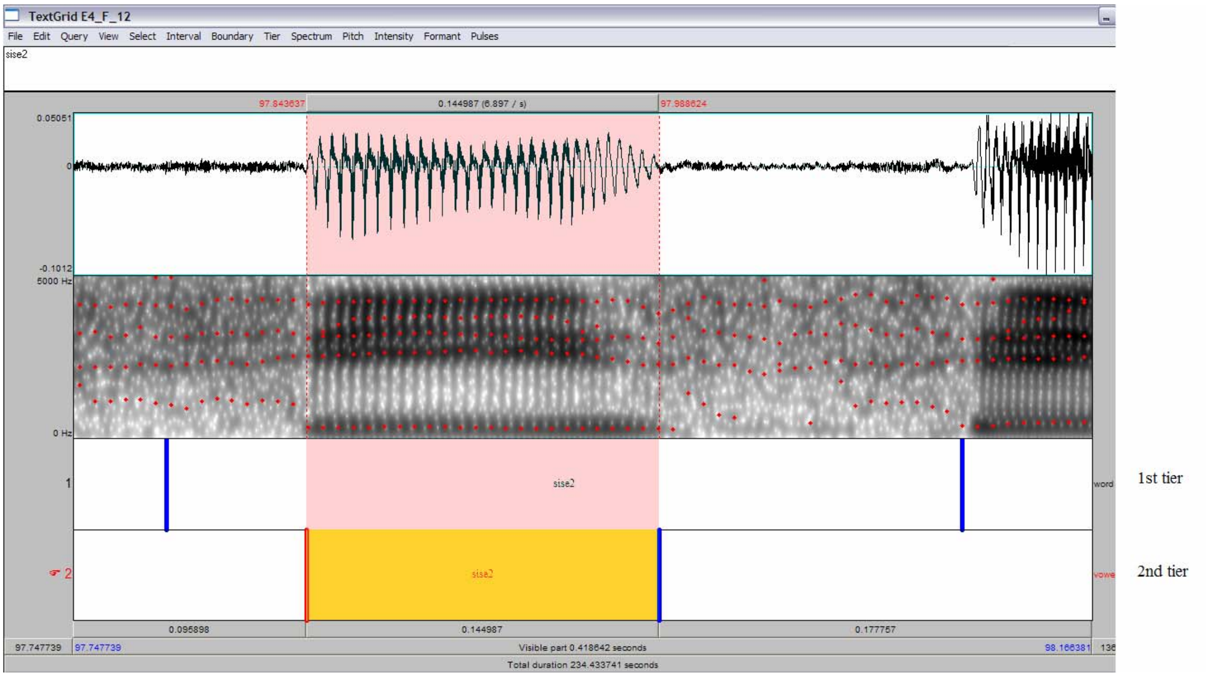


Figure 17. Segmentation and labeling of /i/ in the nonsense word "sisse".

Only the productions of the female participants were analyzed, since they were in greater number than the males and it is not appropriate to join male and female formants for analysis. In order to compare the production of each participant in the prettest and in the posttest, and thus check whether there was improvement after training or not, F0, F1, F2 and F3 were measured, but only F1 and F2 were analyzed in this study. The other acoustic properties were included in the tables only as a reference. The procedures for calculating the pitch floor and pitch ceiling followed Rauber (2006). The pitch floor was set at 120Hz and the pitch ceiling at 400Hz. The formant ceiling was then calculated in order to deal with individual vocal tract differences, varying from 4500Hz to 6500 Hz, calculated in intervals of 10 Hz. Since the F2 is the formant with the highest variability, the F2 instance (per vowel and per speaker) that showed the lowest variability was chosen to be the optimal formant ceiling. After that, the median of each vowel was calculated, as well as the Euclidean Distance (ED) between vowels. All these calculations were done with *Praat* scripts (Appendixes C and D).

Having these values measured, the vowels produced by the learners in the prettest and in the posttest were plotted in *Praat 4.5.14* (see script in Appendix E) and compared to each other. This was done to check whether, after training, there was any improvement in terms of decrease in overlap of the English vowels produced by the nonnative speakers.

5.2.3 Perception test

5.2.3.1 Description

The perception test consisted of a forced-choice labeling task (Appendix I) in which the participants had to identify the American English vowels within 108 CVC words produced by 8 native American English speakers (4 males and 4 females) whose biodata are given in Table 13. The speakers were 8 of the 9 participants of Rauber (2006). The words included nine of the American English vowels (the targets /i, ɪ, ɛ, æ, ʊ, u/), totaling 72 words with the targets plus 36 words with /ʌ, ɑ, ɔ/[28], which served as distractors. Target vowels and distractors appeared twice within the following contexts: /kVt/, /pVt/, /sVt/, /tVk/ and /tVt/.

[28] The extra vowels /ʌ, ɑ, ɔ/ were included only to complete the syllabus, as explained for the production test (see Section 5.2.2.1).

Table 13. Background details of speakers whose utterances were used in the perception test.

Partic.	Gender	Age	Place of birth	Place where spent most of life	Occupation
S1	F	58	Philadelphia-PA	Los Angeles-CA	Professor
S2	F	27	Sacramento-CA	Sacramento-CA	BA student: Medicine
S3	F	22	Sacramento-CA	Sacramento-CA	BA student: International Relations
S4	F	54	Harlingen-TX	Sacramento-CA	Retired medical assistant
S5	M	28	Orange County-CA	Sacramento-CA	BA student: History
S6	M	25	Irvine-CA	Sacramento-CA	Computer programmer
S7	M	18	Sacramento-CA	Sacramento-CA	High school student
S8	M	59	Philadelphia-PA	Sacramento-CA	Psychologist

S = Speaker

The main reasons why only natural stimuli were used in the perception test, even though the learners were trained with different kinds of input (natural x synthesized stimuli) were (i) because that is the kind of stumuli they will be exposed to in natural settings, and (ii) to check the fourth hypothesis of this study, that is, whether knowledge acquired by means of artificial speech with cue enhancement is generalized to natural listening settings.

The target words were each saved in a different file, totaling 108 sound files, which were opened in the Praat software. A Praat script (Appendix J) was created to randomize the sound files and to insert a 6-second interval[29] between each word. The resulting sound file was then recorded onto an audio CD, which was played during the perception test.

[29] This rather long interval was inserted to give the participants enough time to choose one answer from the 9 options they had on the answer sheet.

5.2.3.2 Perception test procedures and analysis

The perception test was administered after the production test. Before the perception test started, the participants received instructions in English on the procedure they should follow: They would hear a word and, on the answer sheet, they should circle the word that contained the same vowel sound as in the word they had just heard. The possible answers corresponding to the vowels /i, ɪ, ɛ, æ, ʌ, ɑ, ɔ, ʊ, u/ on the answer sheet (Appendix I) were, respectively, "sheep, ship, bed, bad, cut, hot, talk, foot, boot"[30]. The perception tests (pre-, post-, and follow-up) were administered in the Language Lab at UFSC. To perform the identification task appropriately, the participants used Sony headsets (H5-95). Also, there was a short break after every 18 words.

The pretest was given on the first day of class, and the posttest as soon as vowel training was over, in the fifth week of class. The results of the perception pretest and posttest were compared and statistical tests[31] were run in order to check whether there was any significant improvement in the participants' perception of the targets or not. The same procedure was adopted to compare the posttest with the follow-up test. The experiment design is summarized in Table 14. Details on the training materials and procedures are given in the following sections.

[30] The words on the answer sheet were not read aloud before the test, since it contained only common words, and it was assumed that the participants were familiar with them.

[31] Independent sample t-tests were used to compare means between groups, and paired t-tests were used to compare means within groups.

Table 14. Summary of the experiment design.

		Week	Content	Homework
Experimental groups	SynS Group	1st	Pretests: Perception and production	-
		2nd	Training: Front vowels	Activities 1 (software)
		3rd	Training: Mid and back vowels	Activities 2 (software)
		4th	Training: All vowels	Activities 3 (software)
		5th	Posttests: Perception and production	-
		9th	Follow-up tests: Perception and production	-
	NatS Group	1st	Pretests: Perception and production	-
		2nd	Training: Front vowels	Activities 1 (software)
		3rd	Training: Mid and back vowels	Activities 2 (software)
		4th	Training: All vowels	Activities 3 (software)
		5th	Posttests: Perception and production	-
		9th	Follow-up tests: Perception and production	-
Control groups	BP Control	1st	Pretests: Perception and production	-
		5th	Posttests: Perception and production	-
	AE Control	1st	Perception test	-

5.2.4 Measuring the relation between production and perception improvement

Pearson's correlation tests were run to check whether the perception and the production results were symmetrical. In order to make perception and production tests results comparable, it was necessary to group the results of the perception test per vowel contrast, that is, /i-ɪ/, /ɛ-æ/, and /ʊ-u/. The raw values in the perception test were number of correct answers per contrast, whereas the raw values in the production test were the EDs (measured in Hertz). The results per vowel contrast in the perception test for the experimental and the control groups are displayed in Appendix F.

To calculate the improvement in production, the EDs of each vowel contrast (which were calculated using the Praat script in Appendix D) in the pretest were subtracted from the EDs in the posttest. The results of each participant per test are shown in Appendix G, and the improvement results are shown in Appendix H. Pearson's correlation tests were carried out and correlation plots were created.

5.2.5 Training tools and procedures

In order to facilitate the explanation about the training tools, this section is divided into 3 subsections. In Section 5.2.4.1, the stimuli used during the training sessions are presented and detailed. In Section 5.2.4.2, the procedures and materials used in class are described. Finally, in Section 5.2.4.3, the materials used by the participants to train vowel perception at home are described.

5.2.4.1 Stimuli

5.2.4.1.1 Natural Stimuli (NatS) training

Two different sets of stimuli were used, depending on the group learners were in. The stimuli used in the natural-stimuli-based training were recorded by 7 native speakers of American English[32] (3 males and 4 females) from different U.S. states: Kentucky, Massachusetts, Michigan and New York. In addition, the utterances of

[32] The native speakers in this Section were borrowed from Rauber (2006) with permission.

speakers 1, 3 and 8 (S1, S3 and S8), which were used in the perception test, were also used in the training material, but the recordings from the other 7 speakers were exclusively used in the training material. The Americans recorded nine monophthongs /i, ɪ, ɛ, æ, ʌ, ɑ, ɔ, ʊ, u/ within the /bVt/ frame. The words were then edited and organized according to the design of each activity.

5.2.4.1.2 Synthesized Stimuli (SynS) training

The stimuli used in the synthesized-stimuli-based training consisted of isolated vowels generated by a Praat script (Appendix K). The values for F1 and F2 are the means of the values found in Peterson and Barney (1952) and in Ohnishi (1991); F3 values were taken from Peterson and Barney (1952). The means of the two studies were used because Peterson and Barney (1952) is a classical study of a variety of AE accent, whereas Ohnishi (1991) is restricted to the Californian accent, but reports more updated formant values. These mean values, shown in Table 15, were used as a basis to establish the enhanced values for the synthesized vowels used in the training stimuli.

Table 15. Peterson and Barney's (1952), Ohnishi's (1991) and mean values for F1, F2 and F3 of vowels produced by male and female native speakers of American English.

		Peterson & Barney (1952)			Ohnishi (1991)			Mean values		
	Gender	F1	F2	F3	F1	F2	F3	F1	F2	F3
/i/	M	270	2290	3310	280	2171	-	275	2230	3310
	F	310	2790	3010	360	2674	-	335	2731	3010
/ɪ/	M	390	1990	2550	429	1869	-	407	1930	2550
	F	430	2480	3070	480	2214	-	455	2347	3070
/ɛ/	M	530	1840	2480	514	1741	-	522	1790	2480
	F	610	2330	2990	634	2143	-	622	2236	2990
/æ/	M	660	1720	2410	629	1686	-	642	1706	2410
	F	860	2050	2850	754	1977	-	808	2014	2850
/ʌ/	M	640	1190	2390	589	1320	-	614	1255	2390
	F	760	1400	2780	691	1560	-	725	1480	2780
/ɑ/	M	730	1090	2440	673	1127	-	701	1108	2440
	F	850	1220	2710	806	1251	-	828	1235	2710
/ɔ/	M	570	840	2410	617	1000	-	593	920	2410
	F	590	920	2710	629	1054	-	609	987	2710
/ʊ/	M	440	1020	2240	429	1203	-	435	1109	2240
	F	470	1160	2680	471	1283	-	469	1218	2680
/u/	M	300	870	2240	343	1097	-	321	983	2240
	F	370	950	2670	400	1114	-	385	1032	2670

Next, following the tests of Escudero and Boersma (2002, 2005), a *continuum* of 7 steps was generated by subtracting the F1 and F2 mean values of /i/ from, respectively, the F1 and F2 mean values of /ɪ/ (shown in Table 16), and dividing the results by 6 to get the interval between the steps, as shown in Figure 18.

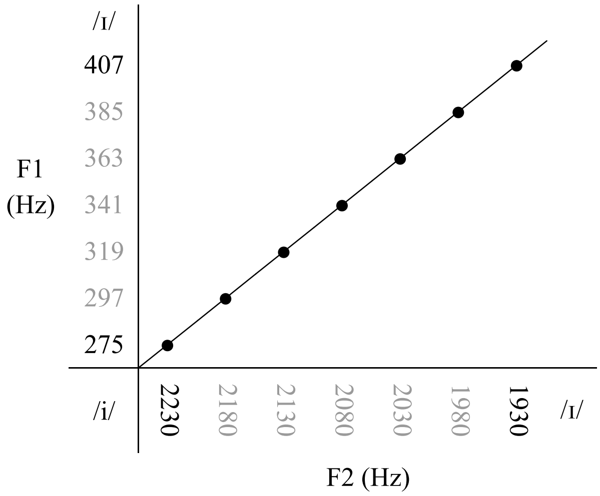


Figure 18. The *continuum* of 7 steps from vowels /i/ to /ɪ/, as produced by male speakers. The difference between each step was 22Hz for F1 and 50Hz for F2.

Table 16. F1 and F2 interval values for males and females per vowel and vowel contrast.

	Gender	/i-ɪ/	/ɛ-æ/	/ʊ-u/	/ʌ/	/ɑ/	/ɔ/
F1 ratio	M	22Hz	20Hz	19Hz	20Hz	20Hz	20Hz
F2 ratio	M	50Hz	14Hz	21Hz	25Hz	25Hz	25Hz
F1 ratio	F	20Hz	31Hz	14Hz	22Hz	22Hz	22Hz
F2 ratio	F	64Hz	37Hz	31Hz	44Hz	44Hz	44Hz

The resulting interval was used to calculate the enhanced values of formant frequencies. The intervals were calculated for each of the three target vowel contrasts as shown in Table 16. The interval used for distractors was kept constant at 22Hz[33] for F1 and 44Hz for F2.

Finally, to calculate the enhanced F1 and F2 values of the vowels, the interval values were either added or subtracted, to increase the distance between the two vowels. For instance, F1 and F2 interval values for /i/ and /ɪ/ were, respectively, 22Hz and 50Hz. To calculate the enhanced F1 for /i/ it was necessary to subtract the interval

[33] These values were the F1 and the F2 interval means of the three target vowel contrasts.

(22Hz) from the F1 value (275Hz); to get the enhanced F2 value for /i/, the interval (50Hz) was added to the F2 value (2230Hz). Thus, the results of enhanced F1 and F2 for /i/ would be, respectively, 253 Hz and 2280 Hz. Conversely, the opposite operations would be done to get the F1 and F2 values for an enhanced /ɪ/. Thus, the enhanced F1 was calculated by adding the interval (22Hz) to the F1 value (407Hz), and the enhanced F2 value was a result of the subtraction of the interval (50Hz) from the F2 value (1930Hz), equaling 429Hz and 1880Hz for the enhanced F1 and F2 of /ɪ/, respectively. The resulting values for enhanced F1 and F2 of all synthesized vowels used in this study for males and females, as well as the information about duration and F0 are shown in Table 17 (for males) and in Table 18 (for females).

Table 17. Enhanced F1 and F2 values for the synthesized stimuli generated in Praat for males. F3 values are the same as in Peterson and Barney (1952).

			Males			
	F1	F2	F3	Duration	F0 initial	F0 final
/i/	253	2280	3310			
/ɪ/	429	1850	2550			
/ɛ/	502	1804	2480		150	
/æ/	662	1692	2410		or	
/ʌ/	594	1280	2390	500ms	180	80
/ɑ/	721	1083	2440		or	
/ɔ/	573	895	2410		200	
/ʊ/	454	1130	2240			
/u/	302	962	2240			

Table 18. Enhanced F1 and F2 values for the synthesized stimuli generated in Praat for females. F3 values are the same as in Peterson and Barney (1952).

Females						
	F1	F2	F3	Duration	F0 initial	F0 final
/i/	315	2795	3010			
/ɪ/	475	2283	3070			
/ɛ/	591	2273	2990		350	
/æ/	839	1977	2850		or	
/ʌ/	703	1524	2780	500ms	380	180
/ɑ/	850	1191	2710		or	
/ɔ/	587	943	2710		400	
/ʊ/	483	1249	2680			
/u/	371	1001	2670			

Enhanced F1 and F2 values were used to cause a stretching of the vowel space, so that vowels would be more distant from each other than usual, preventing the overlap of vowel categories. Furthermore, with a larger vowel space it would be easier for learners to separate each sound into a specific category (Kuhl et al., 1997).

In order to generate the vowel sounds produced by a male speaker and by a female speaker, pitch values were also manipulated. In addition, three different initial pitch values were used in order to prevent the participants from relying in the pitch, instead of in the spectral quality. Pitch variation was also good for minimizing boredom during the task. Thus, as shown in Table 17, initial pitch values were 200Hz, 180Hz and 150Hz, and the final pitch value was fixed in 80Hz[34] for male participants. For females (see Table 18), initial pitch values were 400Hz, 380Hz and 350Hz, and 180Hz for final pitch. All nine AE vowels had tokens with the three initial pitches.

[34] 80Hz was the standard value in the Praat script.

The duration of the vowels was kept constant at 500ms (0.5s) in order to train learners to rely primarily on the spectral properties of the vowels, rather than on duration. Americans rely primarily on spectral quality cues to discriminate the vowels in their L1 (Fox & Maeda, 1999), even though AE vowels differ both in duration and in spectral quality. Furthermore, although previous research found no evidence that Brazilian advanced speakers of English always use L1 cues to identify L2 vowels (Rauber, 2006), speech perception models suggest that perception is a language-specific property and the learners' L1 will *guide* L2 speech perception.

As for the difference in linguistic context of the targets between groups (vowel within a word for the NatS group and isolated vowel for the SynS group), research has found that consonants may interfere in vowel duration, which would either facilitate or complicate vowel identification (Garcia Lecumberri & Ceñoz Iragui, 1997). Peterson and Lehiste (1960) found that although preceding consonants do not affect the duration of vowels, vowel duration is significantly affected by the following consonants. The extent to which vowel duration is affected depends on the voicing and the natural class of the following consonant. Thus, voiced following consonants make vowels longer than their voiceless counterparts. Considering their natural classes, voiced fricatives lengthen the syllable nucleus most, followed by nasals and voiceless plosives, as represented in Figure 19. The consonants that affect vowel duration the least are the voiceless plosives.

[+long] vowel ←-------------------------------→ [-long] vowel

voiced fricatives > voiced plosives, nasals > voiceless plosives

Figure 19. Durational scale of the syllable nucleus as a function of the following consonant (adapted from Peterson & Lehiste, 1960, p.702). “>” means “lengthens more than”.

Since the lack of a consonantal context is also a way to enhance the stimuli, isolated vowels were selected for the SynS group training stimuli. Conversely, it was decided to use words for the natural stimuli tokens because naturally produced vowels extracted from words are too short and, thus, more difficult to perceive and potentially inappropriate for training materials.

Moreover, although 500ms is a rather long duration for vowels (it is actually much longer than real vowels), it was decided to keep it for the synthesized vowels in order to facilitate the perception of the different spectral properties of each vowel during the training phase, which would hopefully help learners to improve their ability of categorizing L2 vowels successfully.

5.2.4.2 In class training

In class training was divided into two phases: theoretical and practical. In the first phase, which took 40 minutes, the learners were introduced to some basic articulatory vowel properties, such as vowel height and backness, and the relation between vowel articulation and their representation in vowel charts. In the second phase of the in class training, which took 50 minutes, the learners went to the Language Lab, where they performed listening activities in which they listened to specific stimuli according to the group they were in (natural words with the target vowels or just synthesized vowels) – and had to choose one option in their answer sheet (Appendix L). In the first week of training they practiced only front vowels, in the second week they practiced back vowels, and the third week of training was dedicated to practicing all vowels together. All activities performed in lab (the second phase of the in class training) were divided into two blocks ("Part 1" and "Part 2") and they consisted of identification tasks. In the first block, the learners were asked to listen to a vowel and say whether it was vowel "X" or not. For instance, when they were practicing the front vowels, they had to say if the vowel they heard was /i/ or another vowel. In the second block, learners had two vowel choices. For instance, they heard the vowel /i/ and they had to say if it was an /i/ or an /ɪ/. The teacher provided immediate feedback after each trial. In each class, the learners were assigned homework to be done and sent to the teacher via e-mail, which is described in the next section.

5.2.4.3 Take-home training

In order to provide learners with a larger amount of training, software consisting of four activities was developed and saved on a CD, which was handed out to the participants as homework to be done and sent back to the teacher.

The software consisted of 2 identification (or labeling) tasks and 2 discrimination tasks, in order to provide learners with both within-category and between-category variability (for details see Section 4.3). The design of all activities was exactly the same, differing only in the kind of stimuli, according to the learner's group: activities with synthesized stimuli for the learners in the SynS Group, and activities with natural stimuli for learners in the NatS Group. Moreover, each week the vowels focused on in the software activities were different according to the vowels trained in class. For instance, in the first week of training, the homework consisted of front vowels – synthesized for the SynS Group and natural for the NatS Group. Again, the exercises were exactly the same for both groups, and only the nature of the sound files (natural or synthesized) was different. Screenshots of each software activity are reproduced in Figures 20a (Activities 1 and 2) and 20b (Activities 3 and 4).

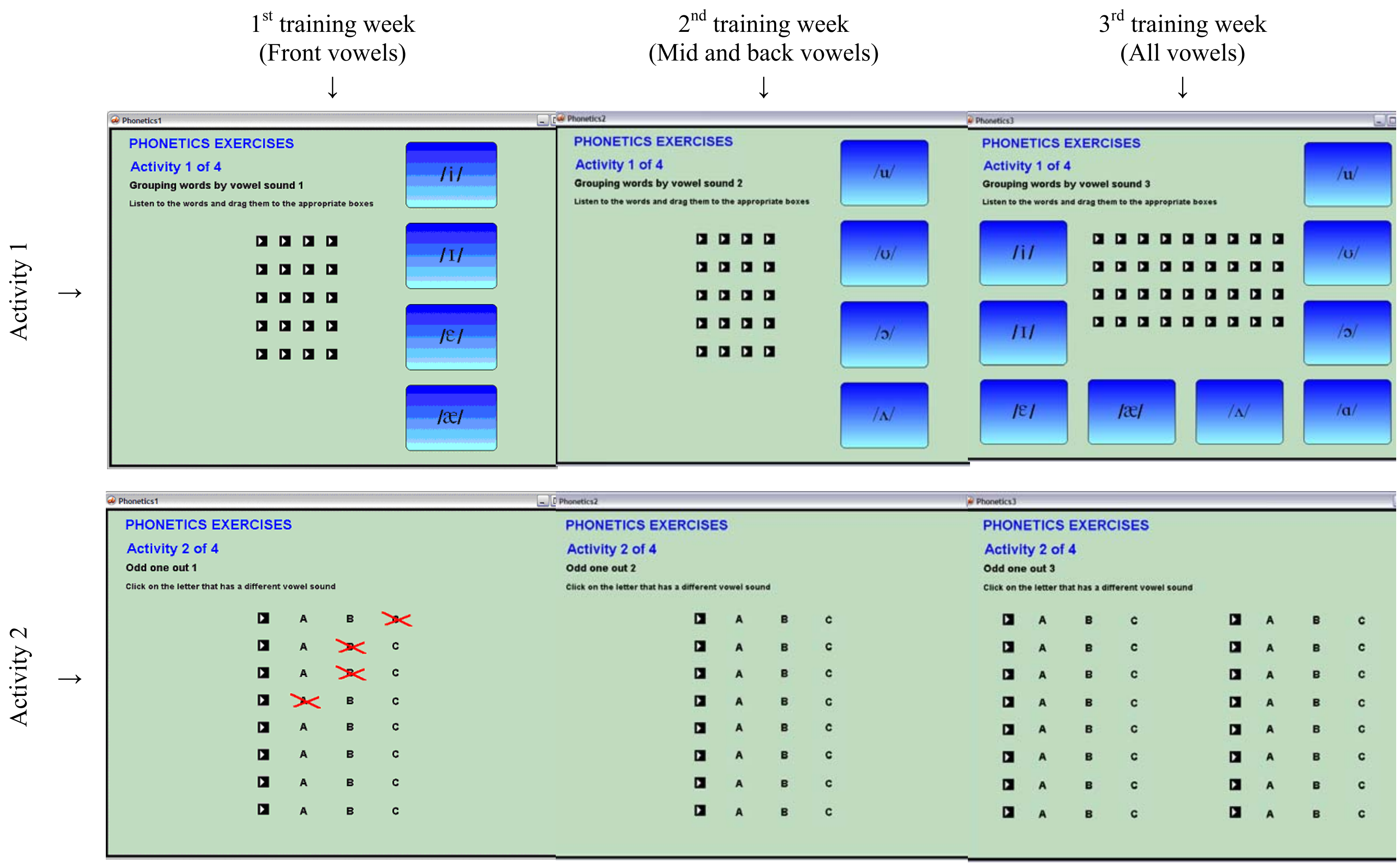


Figure 20. Software activities to be done as homework.

Figure 21. Software activities to be done as homework.

After finishing each activity, the learners were asked to save the results and send them to the teacher by e-mail. The file that was generated after the learner finished the activities was cryptographed (encoded), so that his/her results could not be manually changed. Decoding software was used to transform the learners' results into intelligible data, which provided the teacher with feedback on each learner's correct and incorrect answers, the day the activity was done, and the amount of time the learner spent doing that activity. Figure 21 displays a sample screen with encoded and decoded results of Activity 1.

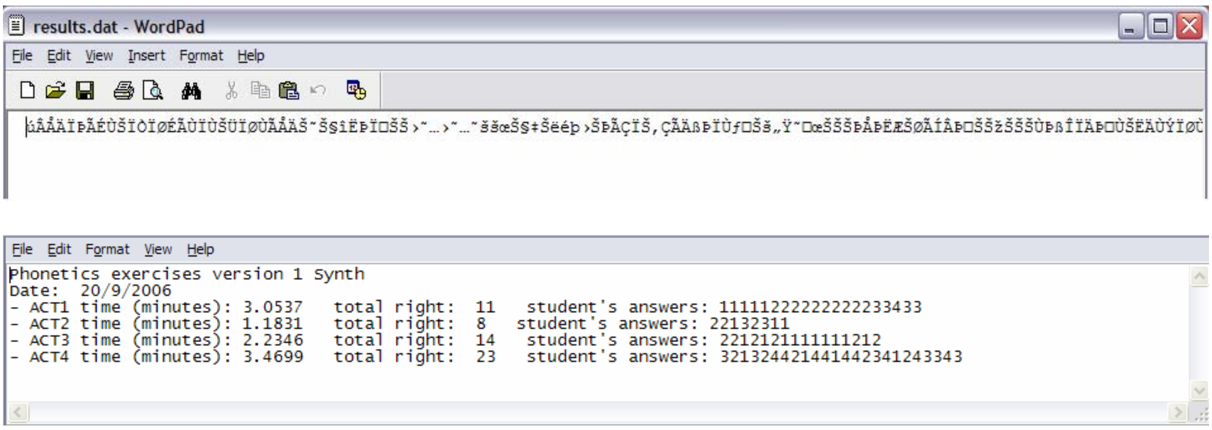


Figure 22. Cryptographed (top) and decoded (bottom) results. Each number represents one of the English vowels.

As soon as she received the results file and decoded it, the teacher corrected the exercises and wrote and sent a feedback report (Figure 22) to each learner. The learners had a deadline to send their result files and they were told that they could redo the activities in the software as many times as they wanted.

Phonetics 1

Activity 1

Positive: the answers for the vowels /i/ and /ɪ/ were all correct; one vowel /æ/ was correctly identified.
Negative: all vowels /ɛ/ were identified as /ɪ/, and four (out of five) vowels /æ/ were identified as /ɛ/.
Suggestion: You should practice more the mid (/ɛ/) and the low (/æ/) front vowels.

	Activity 2		Activity 3		Activity 4	
	Your answer	Correct answer	Your answer	Correct answer	Your answer	Correct answer
1.	B	B	/ɪ/	/ɪ/	3	3
2.	B	B	/ɪ/	/ɪ/	2	2
3.	A	A	/i/	/i/	1	3
4.	C	C	/ɪ/	/ɪ/	3	3
5.	B	B	/i/	/i/	2	2
6.	C	C	/ɪ/	/ɪ/	Same	Same
7.	A	A	/i/	/i/	Same	Same
8.	A	A	/i/	/i/	2	2
9.	-	-	/ɛ/	/ɛ/	1	1
10.	-	-	/ɛ/	/ɛ/	Same	Same
11.	-	-	/ɛ/	/æ/	Same	Same
12.	-	-	/ɛ/	/æ/	1	1
13.	-	-	/ɛ/	/ɛ/	Same	Same
14.	-	-	/æ/	/æ/	Same	Same
15.	-	-	/ɛ/	/ɛ/	2	2
16.	-	-	/æ/	/æ/	3	3
17.	-	-	-	-	Same	Same
18.	-	-	-	-	1	1
19.	-	-	-	-	2	2
20.	-	-	-	-	Same	Same
21.	-	-	-	-	3	3
22.	-	-	-	-	3	3
23.	-	-	-	-	Same	Same
24.	-	-	-	-	3	3

Number of correct answers in each activity:

Activity 1	Activity 2	Activity 3	Activity 4
11	8	14	23

Figure 23. Feedback report sent to the learners.

Chapter 6

Results and discussion

In this chapter, the results of the experiments are reported and discussed according to the order of the objectives they are related to. A general discussion of the results, which will make connections with models of speech perception and other related issues, is also provided at the end of this chapter.

6.1 Pretraining status and effects of training

Before reporting the results of the experiments which are related to the effects of training, it is important to illustrate and discuss the initial perceptual status of the participants in the control group and in the experimental groups before training started. In order to make predictions about the effect of training, it was necessary for the participants in each group to exhibit similar perceptual performance.

As explained in Chapter 5 (Section 5.2.3), the participants performed a forced-choice perceptual labeling task (pretest) in which they listened to a stimulus and chose one of the nine options in the answer sheet. Figure 23 shows that the two groups (control and experimental) had similar results in the perception pretest.

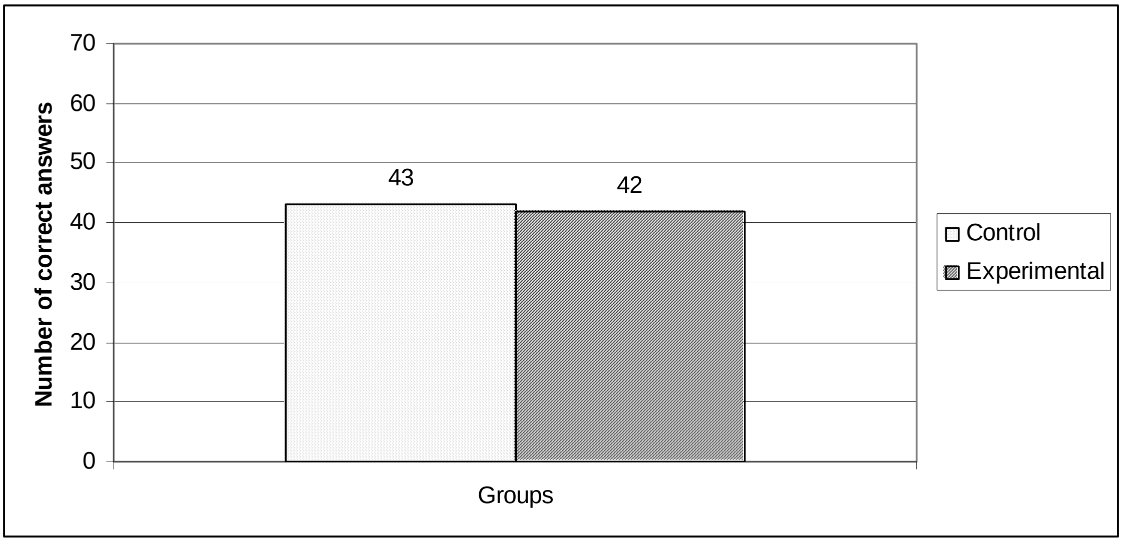


Figure 24. Results of the pretest for the control group and experimental groups (means).

As illustrated in Figure 23, from a total of 70 targets, the control group identified a mean of 43 vowels correctly, whereas the experimental groups reached a mean of 42 correct identifications. Independent-samples t-tests showed that there was no significant difference between groups ($t(34) = .091$, $p = .928$), which confirms that the performance of the control and experimental groups was similar before training.

Figure 24 shows the performance of the control and experimental groups after training. Statistical tests indicate that there was a significant difference in the results of the experimental groups from pretest to posttest ($t(28) = -8.333$, $p = .0001$). Conversely, the performance of the control group in the posttest did not differ significantly from the pretest[35] ($t(6) = -2.203$, $p = .07$). These results suggest that training had positive effects and contributed to the improvement of the participants' perceptual ability.

[35] Since the pretest/posttest differences for the control group were small and non-significant, they will not be discussed further.

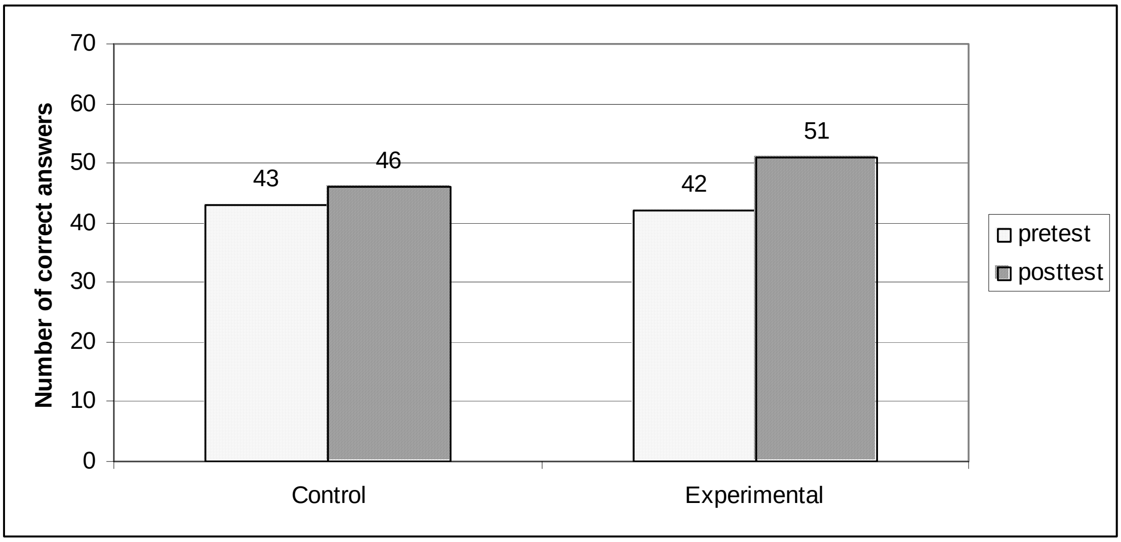


Figure 25. Results of the posttest for the control group and experimental groups (means).

6.2 Identification of the L2 vowels

To check if the identification of the targets improved after training, first the participants' errors (misidentifications) in the pretest were observed to understand what the misidentification tendency was, and to check whether it was consistent or not with the hypotheses of this study. Then, the results of the posttest were analyzed to check whether the number of errors remained constant or was reduced in order to draw conclusions regarding the relationship between training and misidentification rate[36]. The follow subsections report the pretest and the posttest results.

6.2.1 Misidentification in the pretest

In this section, the misidentification results in the pretest are presented for all six target English vowels. In general, there was no significant difference in terms of

[36] Since not all of the target vowels had the same number of tokens, the misidentification patterns of each target vowel are presented in percentages in Sections 6.2.1 and 6.2.2.

misidentification rate between the control group and the experimental groups in the pretest, as shown in Figure 25. A Mann-Whitney test[37] was used to compare the means of the two groups and to confirm the similarity in their performance (Z = .000, p > .05).

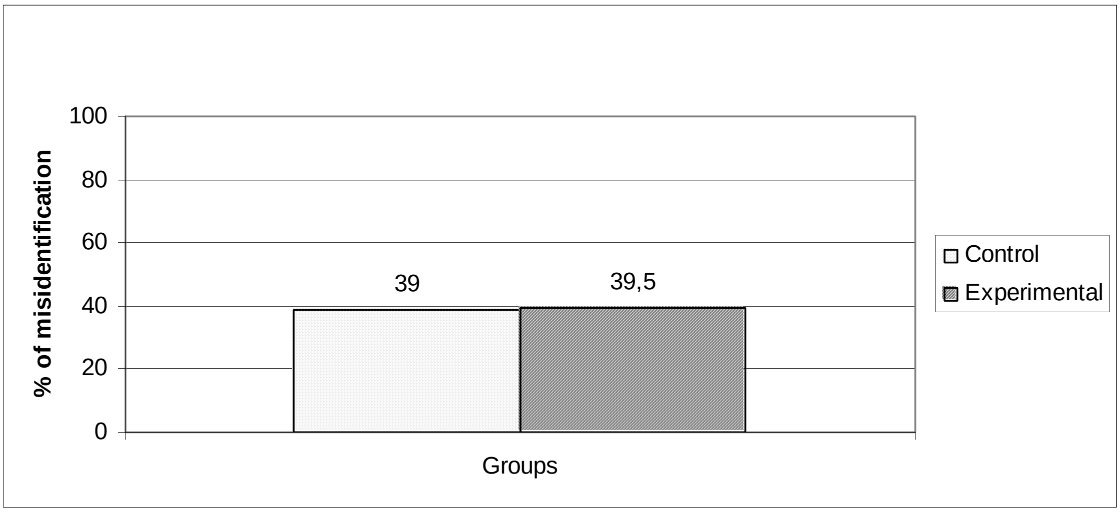


Figure 26. Misidentification rate in the pretest for the control group and experimental groups.

Looking at the errors of the two groups per target vowel, it was observed that there were some misidentification tendencies that were not predicted. Thus, although the English vowels /i/, /ɛ/ and /u/ do not occupy exactly the same locations of Portuguese /i/, /ɛ/ and /u/, it was expected that BP speakers would have no difficulty categorizing them as /i/, /ɛ/ and /u/. However, the results of both groups revealed that English /i/ was sometimes identified as /ɪ/, the vowel /ɛ/ was sometimes identified as /ɪ/ and /æ/, and the vowel /u/ was identified as /ʌ/ and /ʊ/, as shown in Figures 26, 27 and 28, respectively. Furthermore, Figures 26 and 27 show that English /i/ was sometimes misidentified as /æ/, /ʌ/, /ɔ/ and /u/, English /ɛ/ as /ʌ/, /ɑ/ and /ɔ/, and English /u/ as /i/,

[37] Mann-Whitney is a non-parametric statistical test used to compare means of two independent samples, and it is similar to an independent sample t-test. Mann-Whitney tests were applied to the data shown reported in Section 6.2.1.

but due to the extremely rather low misidentification rates, these occurrences may be attributed to lack of attention.

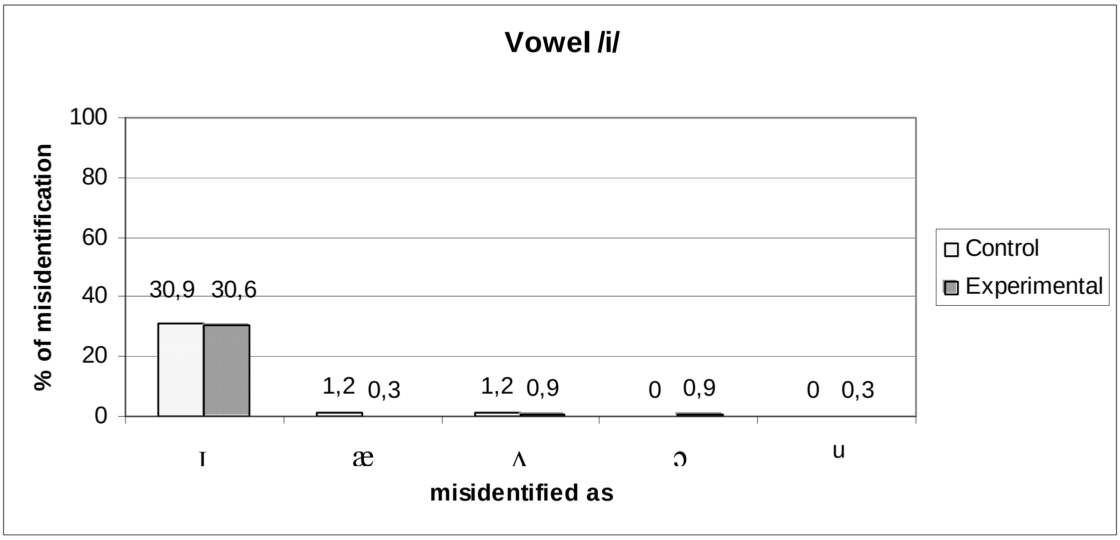


Figure 27. Misidentification rate of /i/ in the pretest for the control group and experimental groups.

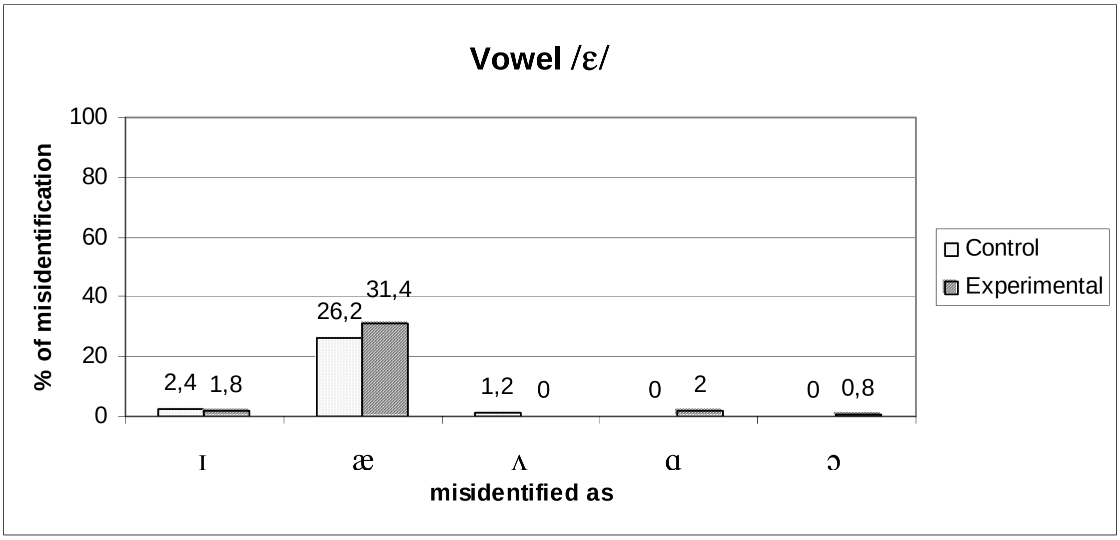


Figure 28. Misidentification rate of /ɛ/ in the pretest for the control group and experimental groups.

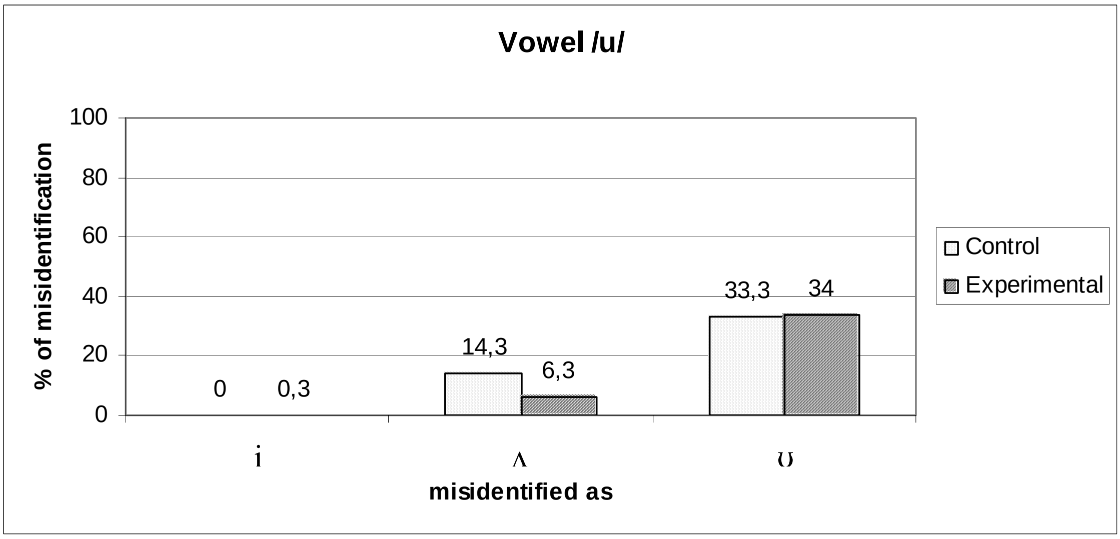


Figure 29. Misidentification rate of /u/ in the pretest for the control group and experimental groups.

As to the target English vowels that do not have a close correlate in BP, that is, /ɪ/, /æ/, and /ʊ/, Figure 29 shows that /ɪ/ was mostly misidentified as /i/, as predicted, but also as /ɛ/, /ʌ/, /ʊ/ and /u/. No significant difference was found between groups for /ɪ/ perceived as /i/. Also, surprisingly, /ɪ/ was misidentified less frequently than /i/, as can be seen from the comparison of Figures 26 and 29.

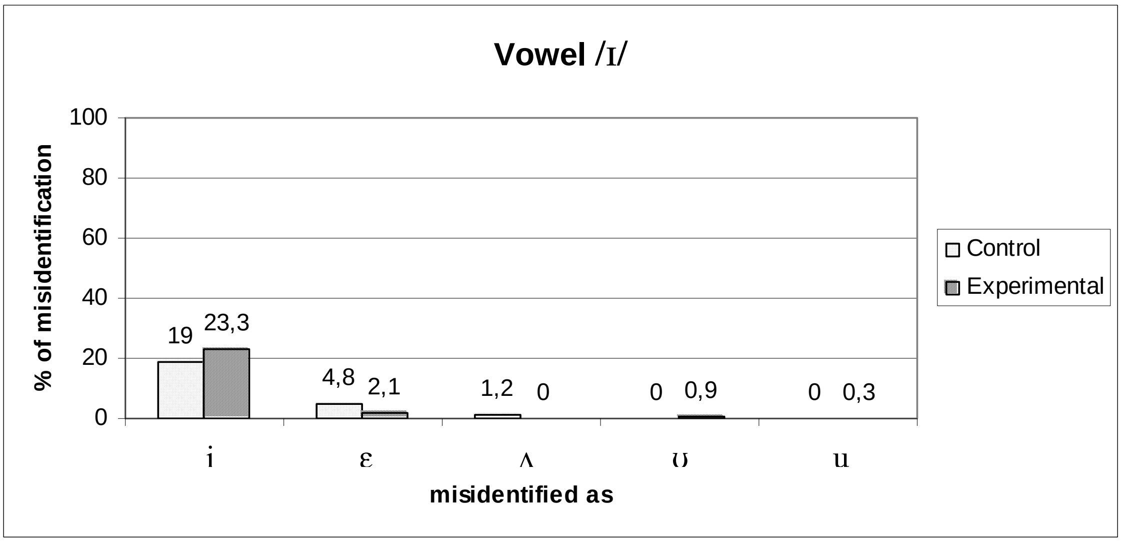


Figure 30. Misidentification rate of /ɪ/ in the pretest for the control group and experimental groups.

Figure 30 also shows that /æ/ was misidentified as /ɛ/ most of the time, but it was also perceived as /ɑ/, /ɔ/, /ʊ/ and /u/ (by the experimental groups) and /ʌ/ (by both groups), although these latter confusions were probably due to lack of attention. Finally, Figure 31 shows that the back vowel /ʊ/ was misidentified mostly as /ʌ/ and /u/ by both groups. The misidentification of /ʊ/ as /æ/ and /ɪ/ by the experimental group was also potentially due to lack of attention.

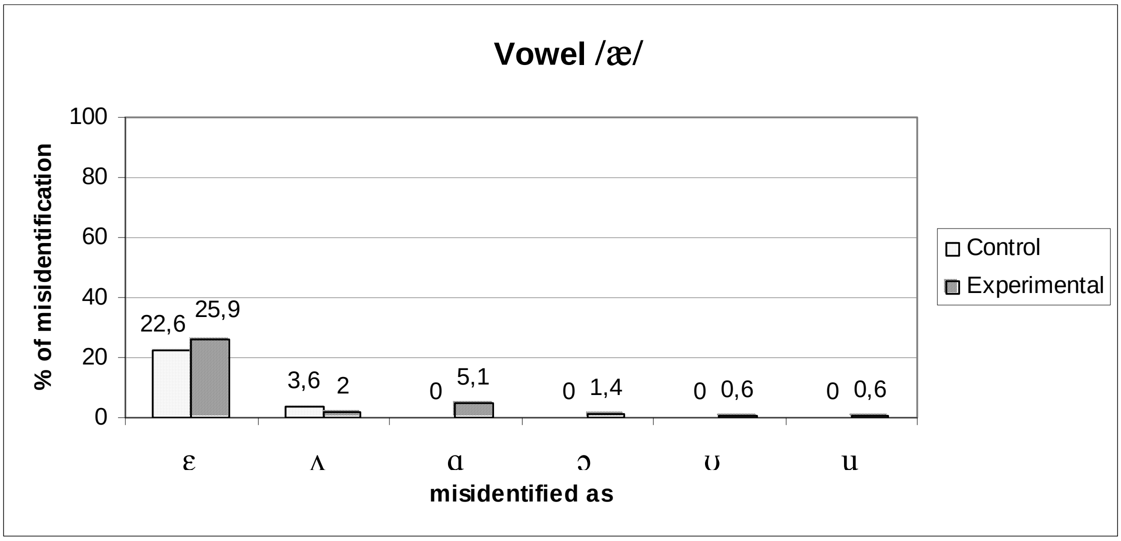


Figure 31. Misidentification rate of /æ/ in the pretest for the control group and experimental groups.

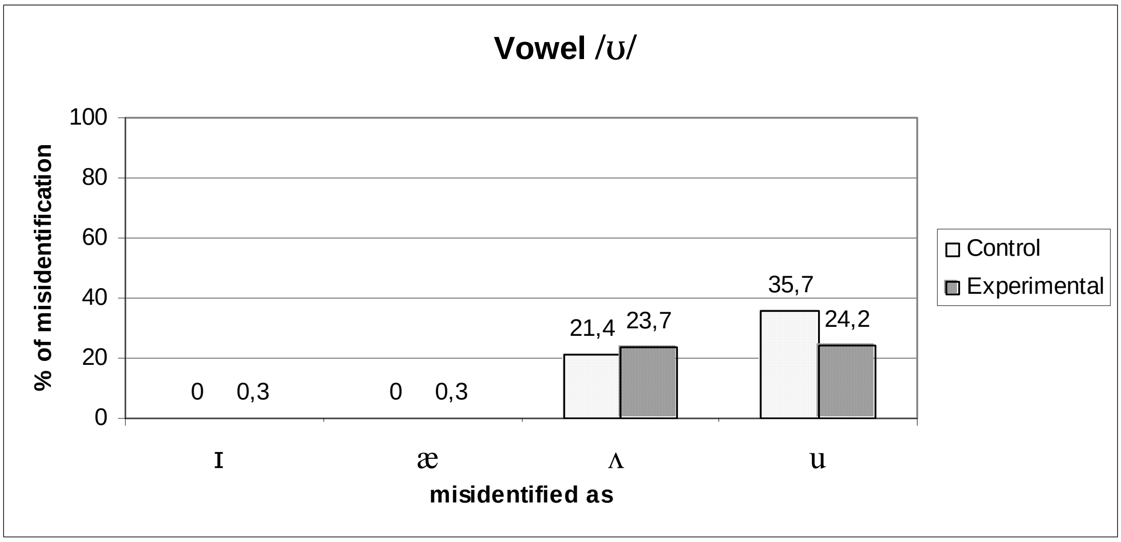


Figure 32. Misidentification rate of /ʊ/ in the pretest for the control group and experimental groups.

The results reported in this section confirmed the assimilation patterns hypothesized in the Introduction of this study, which suggested that (i) /ɪ/ would be perceived as /i/, (ii) /æ/ as /ɛ/, and (iii) /ʊ/ as /u/. However, the participants also showed other identification patterns that were not expected before training. First, they misidentified /ɪ/ as /ɛ/, /ʊ/ as /ʌ/, and /æ/ as /ɑ/. As reported in Section 3.3, the misidentification of /ɪ/ as /ɛ/ by Portuguese speakers was already reported by Flege (1995a). More recently, Flege and MacKay (2004) found the same tendency (of identifying English /ɛ/ and /ɪ/ as /ɛ/) with native speakers of Italian, whose vowel system contains the same oral vowels as does the BP system.

As to English /ʊ/, the identification pattern of this vowel as /ʌ/was not predicted in previous studies. However, considering that all of the native speakers whose utterances were recorded for the tests lived in California, and that, according to Clopper et al.'s (2005) distribution of regional dialects, California is within the West region, it would not be surprising that some of the /ʊ-ʌ/ contrasts were not being actually "contrasted" by the native speakers. As shown in Figure 12 (Chapter 2), in the US Western dialectal region, approximately half of the vowel space for /ʊ/ and /ʌ/ is overlapping. Besides, even the native speakers in the control group, who are not from the same dialectal region, sometimes confused /ʊ/ and /ʌ/ in the perception test (see Appendix N).

The misidentification of /æ/ as /ɑ/ was probably due to how EFL teachers in Brazil approach the teaching of /æ/. When teaching the articulation of /æ/, teachers usually say that it sounds like a very open BP /ɛ/, almost a BP /a/. Therefore, whenever the participants heard a vowel that sounded like an /a/ (i. e. an AE /ɑ/), they may have perceived it as an /æ/.

Another unexpected categorization pattern found in the pretest involved English vowels that have a similar correlate in BP, that is, /i/, /ɛ/ and /u/. Although most of the tokens were correctly identified, sometimes (i) /i/ was perceived as /ɪ/, (ii) /ɛ/ as /æ/, and (iii) /u/ as /ʊ/. Such errors can be attributed either to interference of orthography, to lack of attention, or to chance. In the pretest, after hearing a word, the participants had to choose one of the words in the answer sheet, which for the vowels /i/ and /ɪ/ were *sheep* and *ship*, respectively. The orthographic form of these words might have interfered since the participants would associate /i/ with the word *ship*[38], which has the letter *i* written in it. Chance would explain the errors involving the back vowels /ʊ/ and /u/, and the mid and low front vowels /ɛ/ and /æ/. Since the participants had the choices *foot* and *boot* on the answer sheet for the vowels /ʊ/ and /u/, respectively, and since the orthography of these words does not give any clue as to the correct option (both words are written with *oo*), the participants may have chosen one of the two words at random. Furthermore, the participants were asked to choose only one word for each stimulus they heard, some of them circled both *foot* and *boot* whenever they were in doubt about their answer. The same happened with the vowels /ɛ/ and /æ/. In this case, although the words corresponding to each vowel sound in the answer sheet were not written in the same way (*bed* and *bad*), still the participants were not sure to which written form /ɛ/ and /æ/ corresponded, and they sometimes circled both words in their answer sheet whenever they heard one of these two vowels. Another possibility is that they just did

[38] In BP, the letter *i* is always pronounced [i].

not hear any difference between /ɛ/ and /æ/, and they assumed that there was no difference in pronunciation between *bed* and *bad*, so that they circled both answers because both vowels sounded the same.

Finally, the vowel /ʌ/ was sometimes chosen for any of the six targets. However, this error can also be attributed to lack of attention since the rate of misidentification of the vowels as /ʌ/ was quite low. The same explanation applies to the very few times that /æ/ was perceived as /ɪ/, and /ʊ/ as /æ/. Since the test was rather long (with a total of 108 tokens), and the answer sheet contained many labels for the answers (consisting of written words), tiredness and lack of attention may have contributed to such misidentifications.

6.2.2. Misidentification in the posttest

A Wilcoxon test[39] confirmed that, after training, the experimental groups showed a considerable decrease in the misidentification rate that nearly reached statistical significance ($Z = -1.73$, $p = .08$), as illustrated in Figure 32.

[39] The Wilcoxon test is a non-parametric test used to compare two means that belong to the same group. It is similar to a paired-sample t-test. Wilcoxon tests were applied in the data shown in Sections 6.2.2.

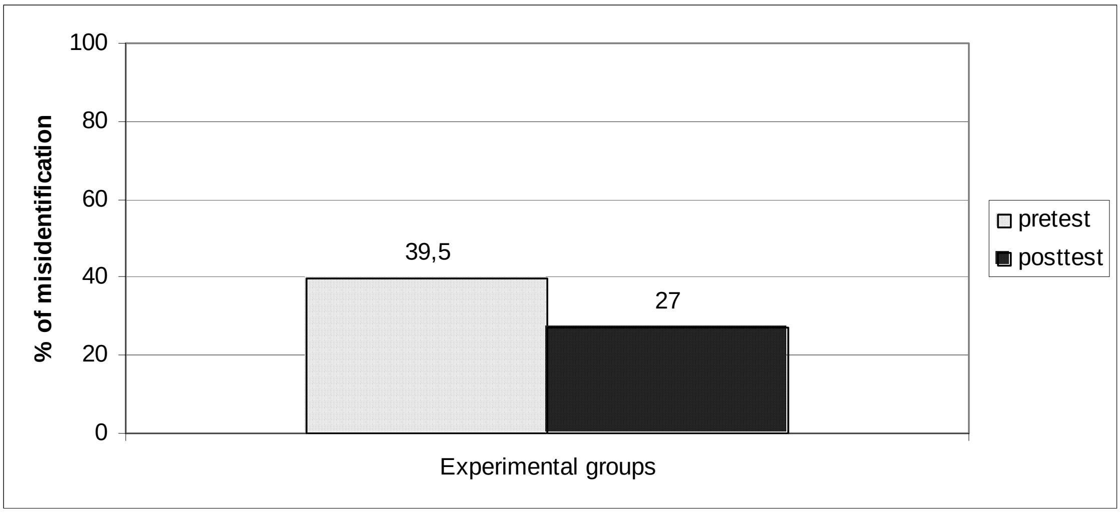


Figure 33. Misidentification rate of the experimental groups from pretest to posttest.

Looking at the results of the experimental groups for individual vowels, Figure 33 shows that there was a significant decrease in the misidentification of /i/ ($Z = -3.79$, $p = .001$). The confusions of /i/ with /ɪ/ decreased after training, and the confusions with /æ/, /ʌ/, /ɔ/ and /u/ did not occur in the posttest.

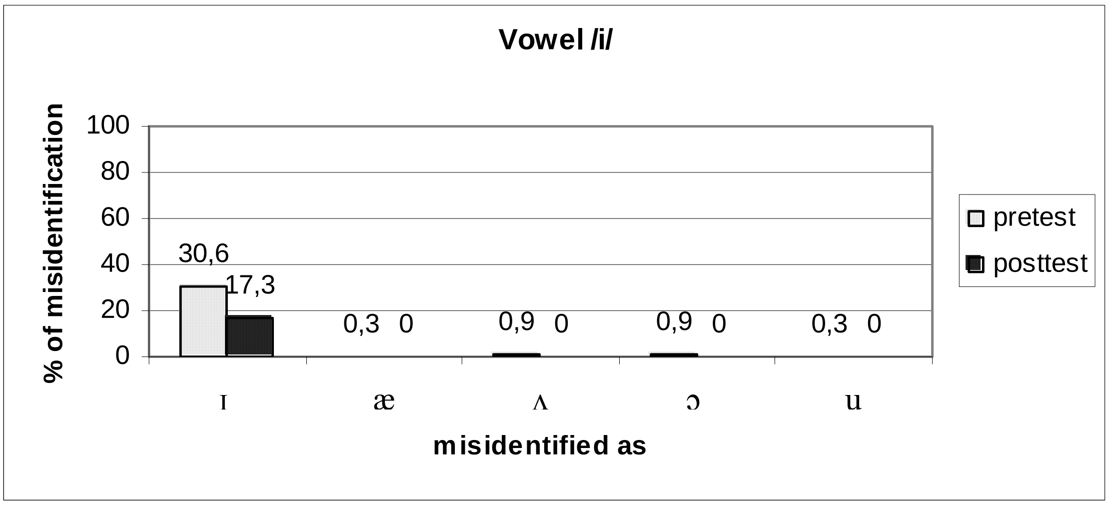


Figure 34. Misidentification rate of /i/ in the experimental groups from pretest to posttest.

Statistical tests confirmed that the results for the vowel /ɪ/ were maintained after training ($Z = -1.72$, $p > .05$), that is, the participants' improvement was little and did not reach statistical significance from pre- to posttest. As shown in Figure 34, there was just a small decrease in the misidentifications as /i/, as well as a small decrease of confusion with /ɛ/. The confusion between /ɪ/ and /ʊ/ and /u/ found in the pretest, however, did not happen in the posttest. As for the vowel /ɛ/, statistical tests showed that the perception of the experimental groups improved significantly after training (Z = -2.44, p = .01). Figure 35 shows that the confusions of /ɛ/ with /ɪ/, /æ/, /ʌ/, /ɑ/ and /ɔ/ tended to decrease after training.

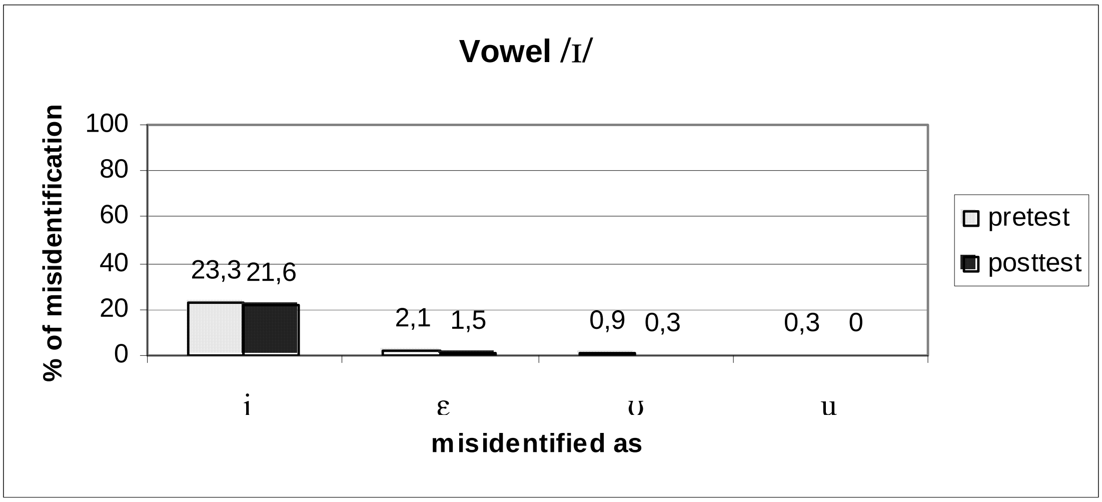


Figure 35. Misidentification rate of /ɪ/ in the experimental groups from pretest to posttest.

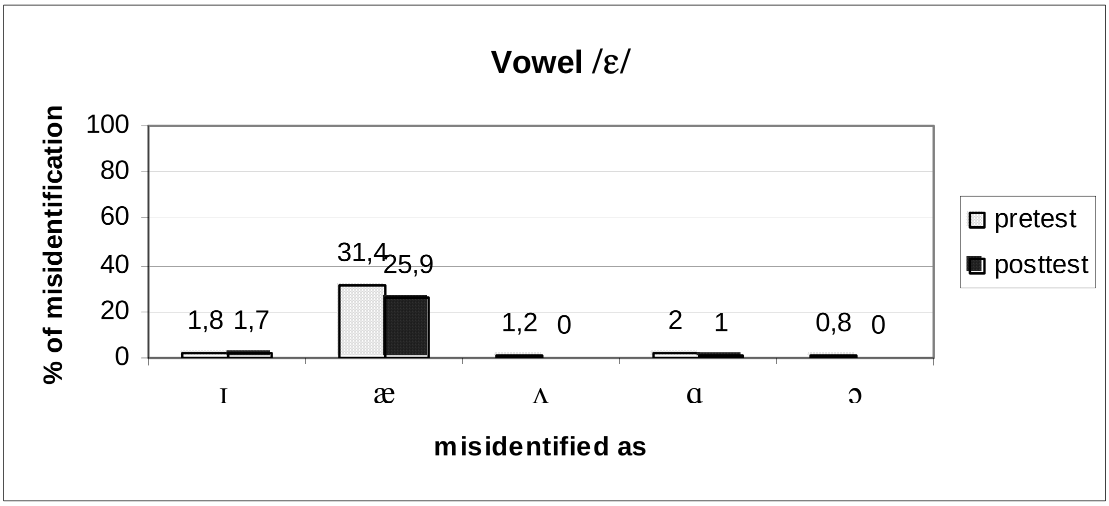


Figure 36. Misidentification rate of /ɛ/ in the experimental groups from pretest to posttest.

The results for the vowel /æ/ show a significant decrease in its misidentification rate (Z = -2.84, p = .004). As shown in Figure 36, the tendency of misidentifying /æ/ as /ɛ/, /ʌ/, /ɑ/ and /ɔ/ decreased after training. In addition, /æ/ was never misidentified as /ʊ/ and /u/ in the posttest

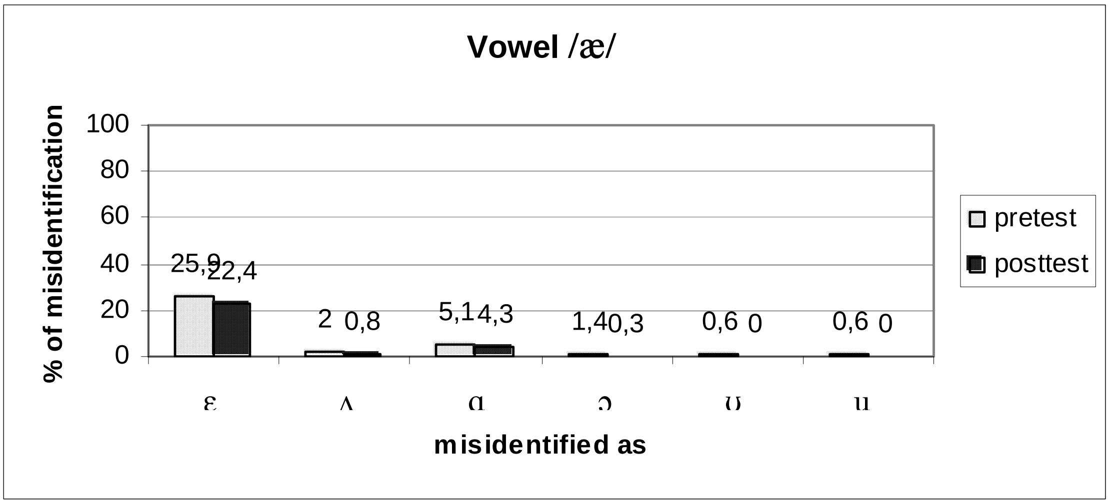


Figure 37. Misidentification rate of /æ/ in the experimental groups from pretest to posttest.

The best results in terms of decrease in number of misidentifications were achieved with the back vowels. Besides showing that /ʊ/ was no longer perceived as /ɪ/ and /æ/ in the posttest, Figure 37 shows that there was a considerable decrease in the misidentifications of /ʊ/ as /ʌ/ and as /u/ (Z = -4.21, p = .001). The same tendency was found for /u/, whose misidentification rate decreased significantly after training (Z = -3.73, p = .001), as shown is Figure 38.

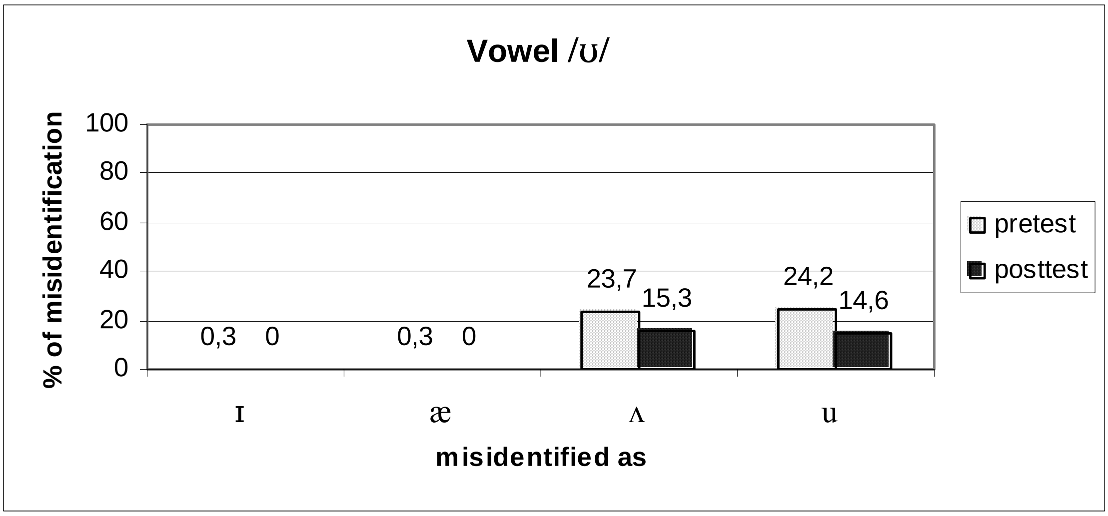


Figure 38. Misidentification rate of /ʊ/ in the experimental groups from pretest to posttest.

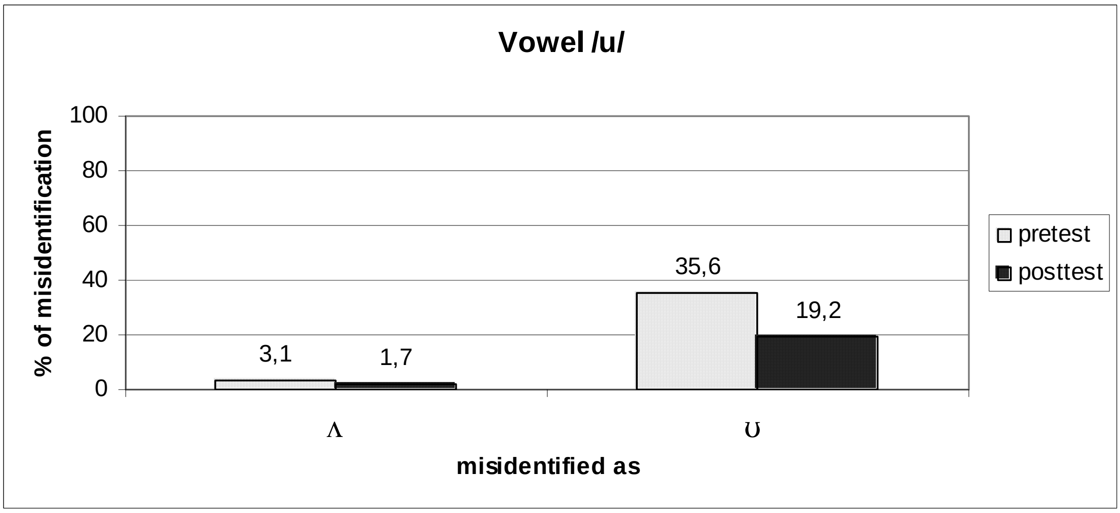


Figure 39. Misidentification rate of /u/ in the experimental groups from pretest to posttest.

Figure 39 shows that although the participants in both experimental groups significantly improved their perception of the targets after training (SynS: $Z = -3.41$, $p = .001$; NatS: $Z = -3.02$, $p = .003$), their performance is still quite different from that of the native speakers. However, the tendency shown by the results of the experimental groups indicates that the participants are moving in the right direction. Therefore, with more training the participants would be likely to achieve more native-like results.

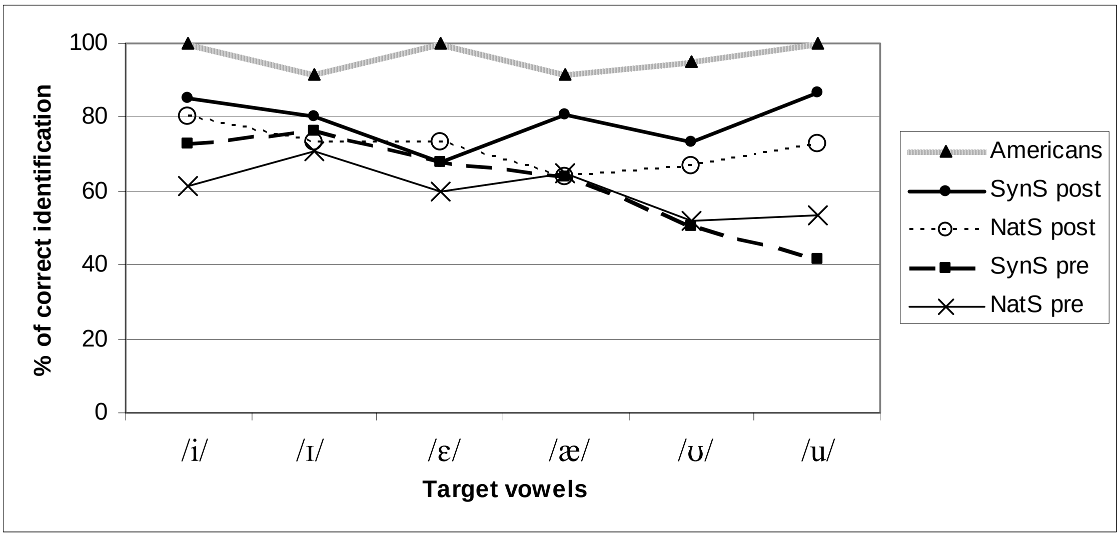


Figure 40. Identification accuracy of each group in the perception tests.

As to each target vowel, the results reported in this section indicate that, after training, the misidentification rate of the vowels /i/, /ɛ/, /æ/, /ʊ/, and /u/ decreased significantly, which suggests that training contributed to a more accurate perception. The unlikely misidentifications of /i/ as /ɔ/, /i/ and /ʊ/ as /æ/, /i/ and /ɛ/ as /ʌ/, /i/, /ɪ/ and /æ/ as /u/, /æ/ as /ʊ/, /ʊ/ as /ɪ/, and /u/ as /i/, which were observed in the pretest, were not found in the posttest, which support the suggestion that such confusions were due to lack of attention. Finally, although some of the misidentification rates did not decrease significantly for some of the targets, they did not increase either, which can be considered a positive finding. The results of the next section (6.2.3) will provide extra evidence about the positive effects of training on the correct identification of the targets.

6.2.3 Improvement in the identification of L2 sounds

The results so far have been reported considering the experimental groups as a whole. Figure 43 shows the improvement of the targets from pretest to posttest in each experimental group and for each vowel. As shown in Figure 40, except for the vowel /æ/ in the NatS group, all of the targets in both groups improved to some extent. The behavior of each of the targets from pretest to posttest is analyzed in this section.

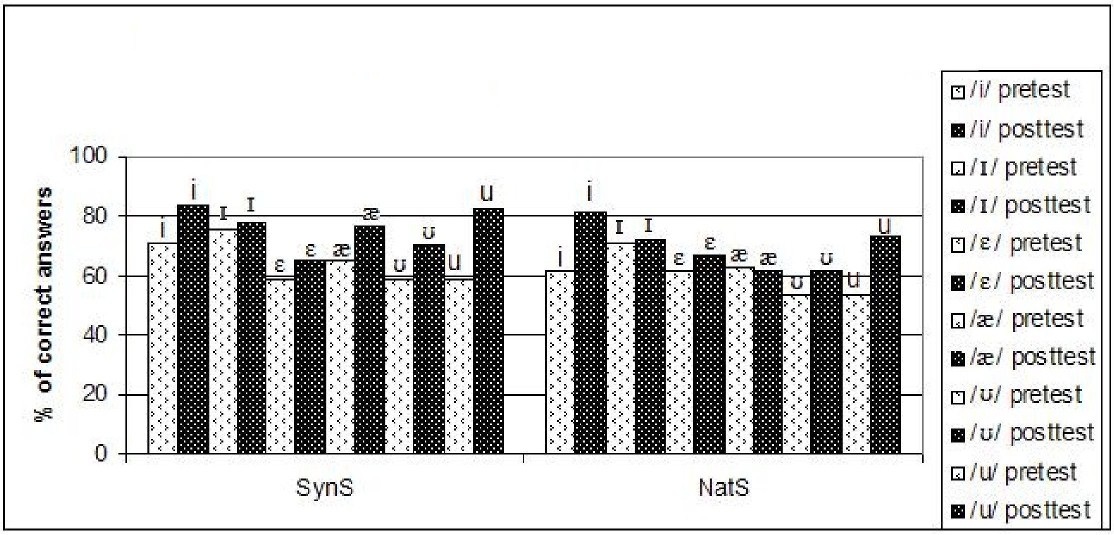


Figure 41. Improvement of the targets after training in each experimental group.

As for the high front vowels, the results for /i/ (see Figure 41) show that, although the participants in both groups improved their perception of this vowel, paired-sample t-tests[40] confirmed that only the ones in the NatS group improved it significantly ($t(13) = -2.293$, $p = .04$). Figure 42 shows that the perceptual improvement of /ɪ/ for neither group was significant.

[40] Paired-sample t-tests compare two means that belong to the same group.

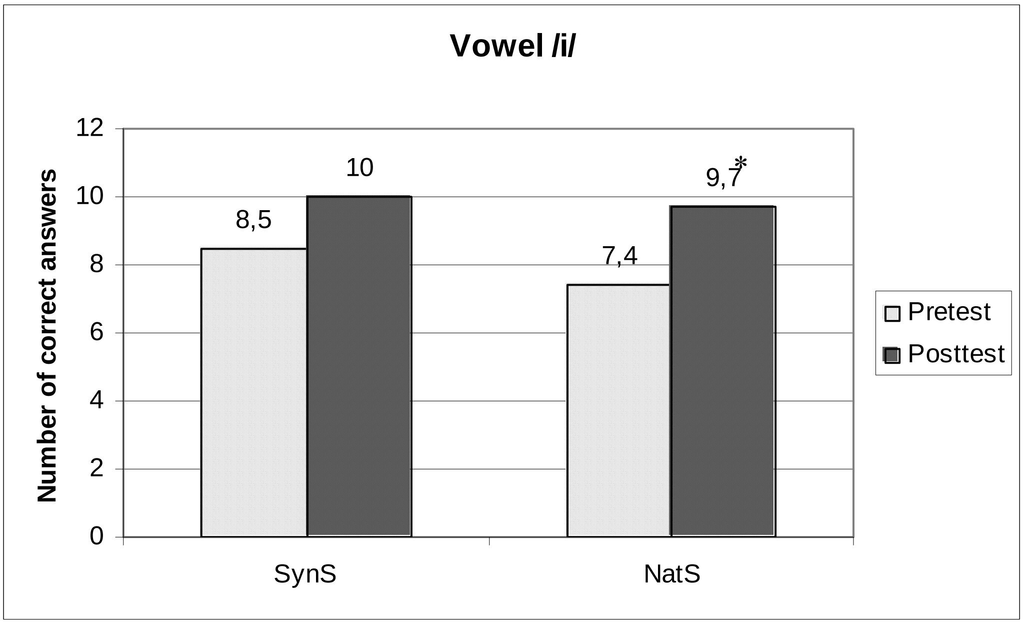


Figure 42. Number of correct identifications of /i/ before and after training in each experimental group. The asterisk indicates significant improvement.

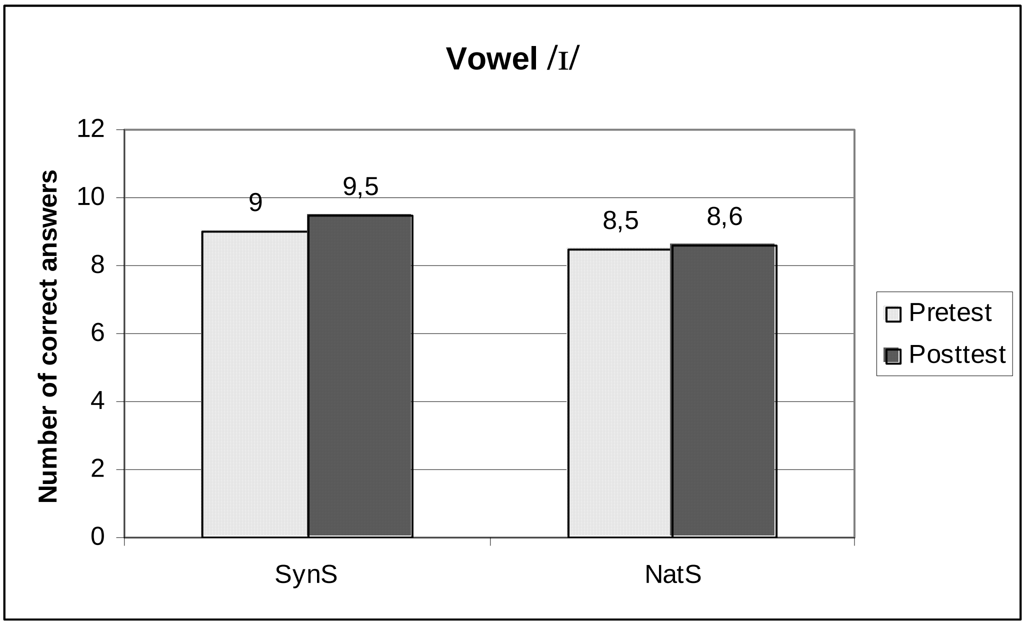


Figure 43. Number of correct identifications of /ɪ/ before and after training in each experimental group.

Considering the mid and low front vowels, statistical tests confirmed that the slight improvement from pretest to posttest of the vowel /ɛ/, shown in Figure 43, was significant for the SynS group ($t(14) = -3.72$, $p = .02$). Similarly, the improvement of the vowel /æ/ in the SynS group was significant ($t(14) = -2.98$, $p = .03$) after training, as shown in Figure 44. Conversely, the results of /æ/ for the NatS group show that they performed slightly worse in the posttest. Thus, if we consider the experimental group as a whole, the good results in terms of perceptual improvement rate of the vowels /ɛ/ and /æ/ after training is basically due to the performance of the SynS group.

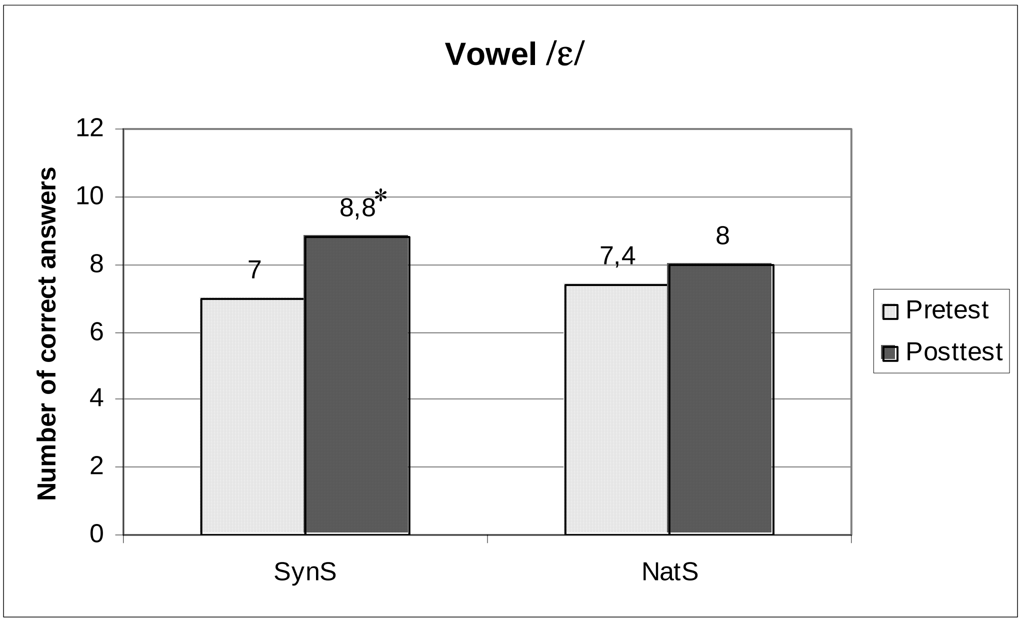


Figure 44. Number of correct identifications of /ɛ/ before and after training in each experimental group. The asterisk indicates significant improvement.

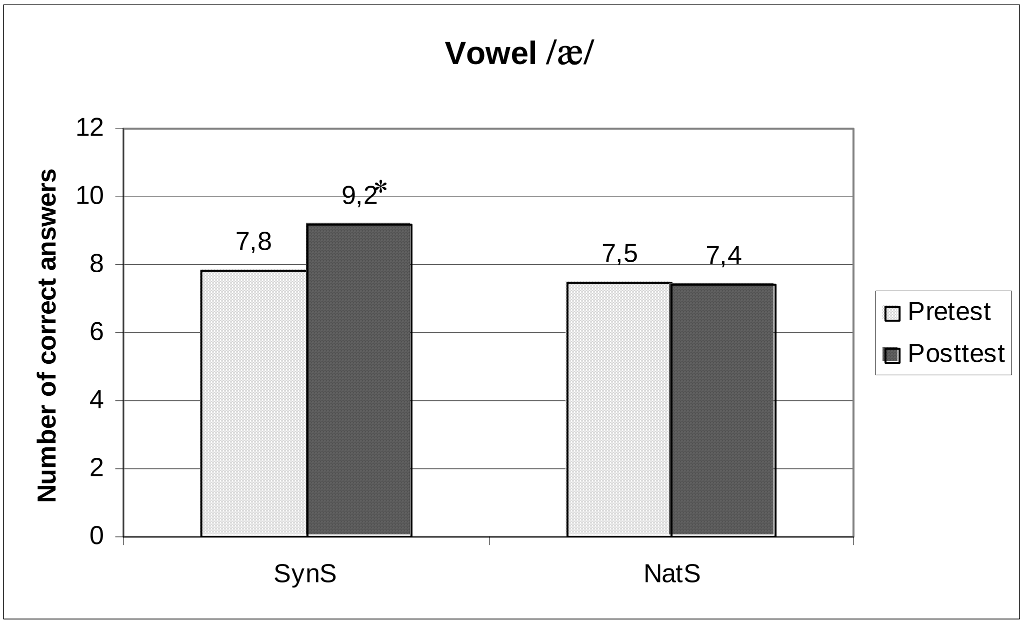


Figure 45. Number of correct identifications of /æ/ before and after training in each experimental group. The asterisk indicates significant improvement.

The best results were found with the high back vowels. Figure 45 shows a significant improvement in the perception of /ʊ/ for both experimental groups (SynS group: t(14) = -3.398, p = .005; NatS group: t(13) = -2.929, p = .013), especially for the SynS group, which performed slightly better than the NatS group. The same pattern was observed for the vowel /u/, as shown in Figure 46: Both groups improved their perception of the target vowel significantly after training (SynS group: t(14) = -3.985, p = .003; NatS group: t(13) = -2.621, p = .022), but the performance of the SynS group was slightly better than that of the NatS group.

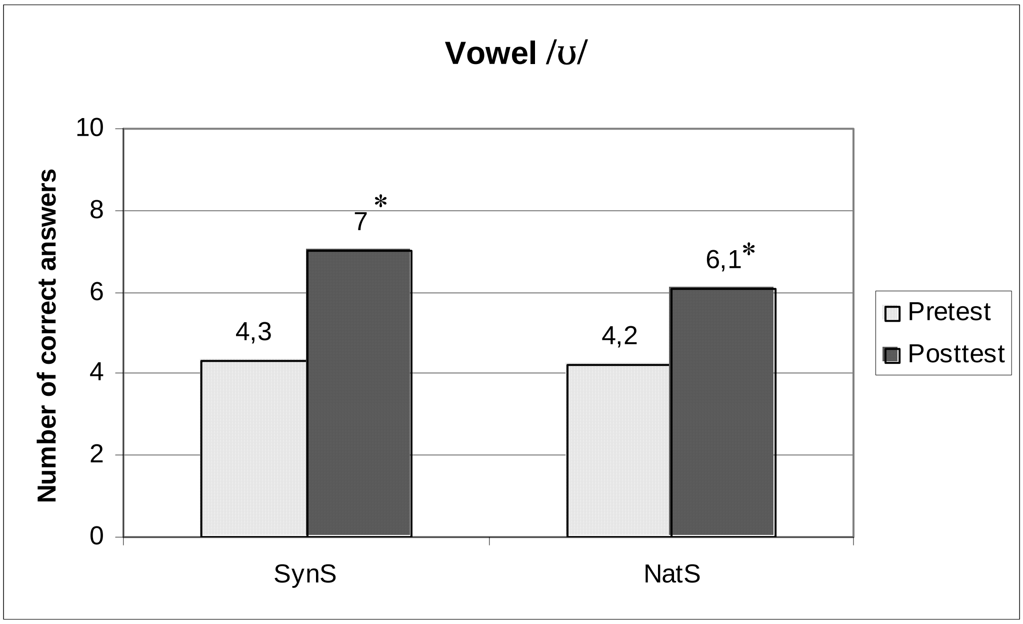


Figure 46. Number of correct identifications of /ʊ/ before and after training in each experimental group. The asterisk indicates significant improvement.

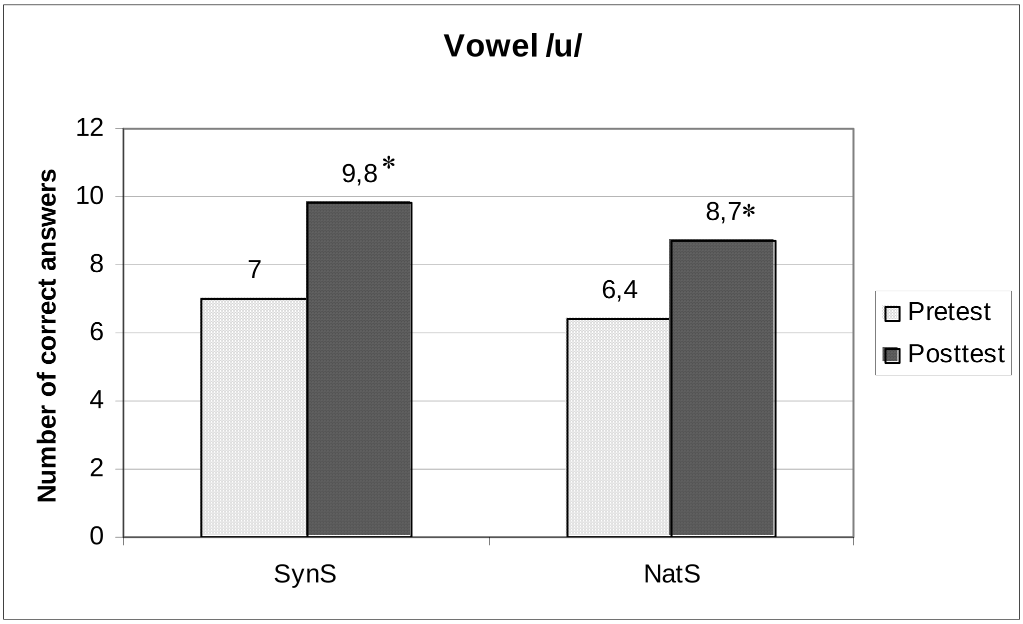


Figure 47. Number of correct identifications of /u/ before and after training in each experimental group. The asterisk indicates significant improvement.

The results reported in this section show that, although in terms of raw values they were the most frequently identified vowels, /i/ and /ɪ/ were the vowels that improved the least in the posttest, followed by /ɛ/ and /æ/; /ʊ/ and /u/ were the vowels that improved the most after training. Considering raw values, the number of correct identifications of /i/ and /ɪ/ in the pretest in each experimental group was already high (SynS: /i/ = 8.5, /ɪ/ = 9; NatS: /i/ = 7.4, /ɪ/ = 8.5, out of 12 tokens). Consequently, there was not much room for improvement for this pair, resulting in the lowest improvement rate of all L2 vowel pairs.

Conversely, the rate of correct identifications in the pretest for the back vowels (especially for /ʊ/) was the lowest (SynS: /ʊ/ = 4.3, /u/ = 7; NatS: /ʊ/ = 4.2, /u/ = 6.4, from a total of 10 tokens of /ʊ/ and 12 of /u/). As a result, /ʊ/ and /u/ had a large room for improvement, which allowed an increase in the correct identification rate in the posttest.

The performance of the two experimental groups perceiving the mid and low front vowels was intermediate to /i/ and /ɪ/ and to /ʊ/ and /u/. In the pretest both groups perceived /ɛ/ and /æ/ better than /ʊ/ and /u/, but worse than /i/ and /ɪ/. Although /ɛ/ and /æ/ had a lot of room for improvement in the posttest, this pair did not improve considerably due to the fact that /ɛ/ and /æ/ improved only in the SynS group. Thus, this pair improved slightly more than /i/ and /ɪ/, but less than /ʊ/ and /u/.

6.3 Training with natural stimuli versus training with synthesized stimuli

Comparing the results of the pretest with those of the posttest in each group, a Wilcoxon test confirmed that there was a significant difference in the performance of the participants in the two experimental groups (SynS group: $Z = -3.41$, $p = .001$; NatS group: $Z = -3.02$, $p = .003$). This difference indicates that both groups performed much better after training. The control group, however, did not show any significant difference in its performance from pretest to posttest ($Z = -1.78$, $p > .05$), although a slight improvement was found. These within-group results, shown in Figure 47, suggest that training indeed has positive effects on the improvement of the perception of L2 sounds.

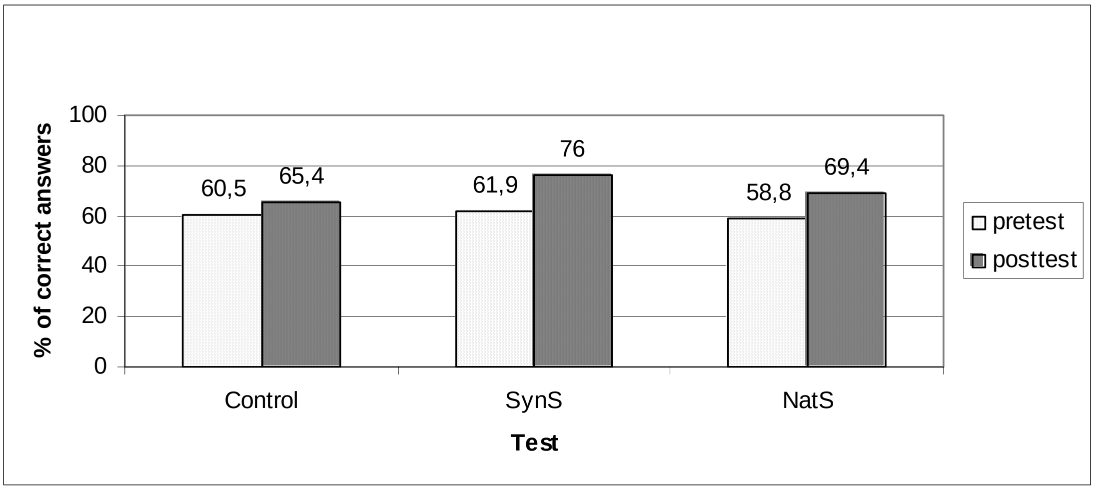


Figure 48. Overall pretest and posttest results per group (within-group performance).

In order to provide a between-group comparison that would indicate which training stimuli are more effective, the pretests and the posttests of the three groups were compared, as illustrated in Figure 48.

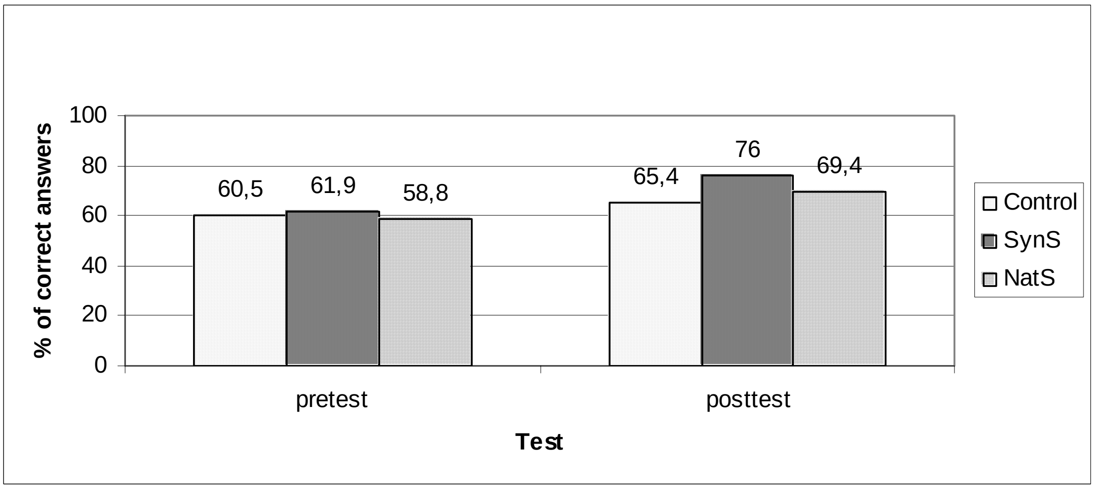


Figure 49. Comparison of overall pretest and of posttest results of the three groups (between-group performance).

Comparing the pretests of the three goups, a Kruskal-Wallis test[41] confirmed that there was no significant difference between groups before training ($X^2 = -.812$, $p > .05$).

[41] Kruskal-Wallis is a non-parametric test used to compare three or more means for different groups. It is equivalent to a One-Way ANOVA.

Similarly, no significant difference was found between groups after training ($X^2 = -.114$, $p > .05$). Still, Figure 48 shows an improvement tendency from pretest to posttest for the three groups, but the highest improvement rate was achieved by the SynS group (14 percentage points), followed by the NatS group (10.5 p.p.) and the Control group (5 p.p.).

Although they did not control for this variable, previous studies (Wang 2002; Wang & Munro, 2004) suggested that the amount of extra-class training time would influence in the performance of the training groups, and thus should be controlled. As for the present study, the slight difference in performance of the experimental groups after training might be explained by an eventual difference in amount of *at home* training time. However, Figure 49 shows that there was not a big difference in amount of at-home training time between the SynS and the NatS groups. Therefore, no conclusions can be drawn as for the effect of extra-class training time on the performance of the two experimental groups.

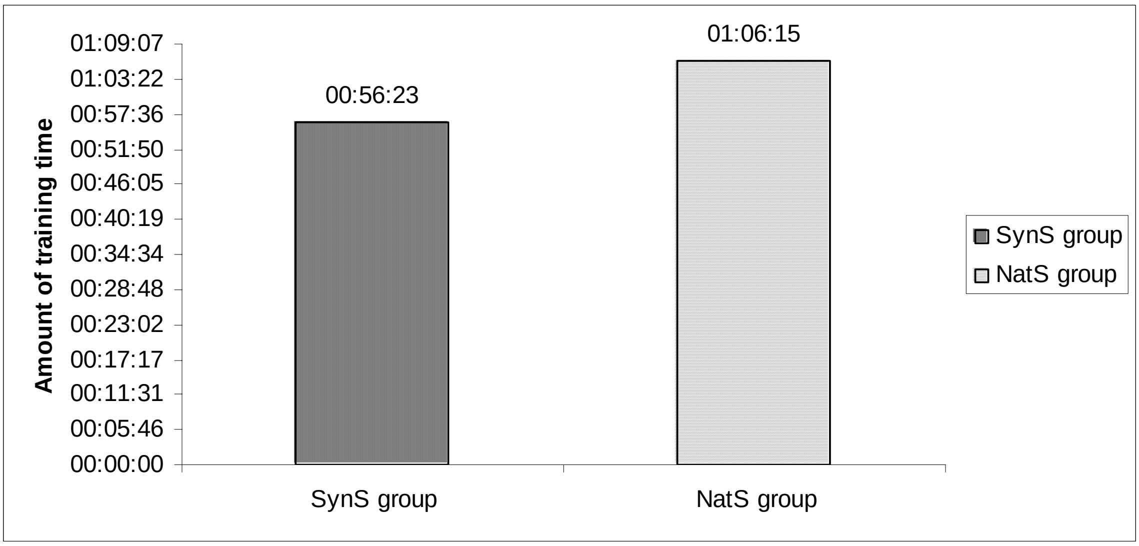


Figure 50. Amount of extra-class (*at home*) training time per experimental group.

Individual performance after training also does not seem to relate to with individual amount of training time. Only the results of Participant 29 corresponded in terms of amount of training time and improvement rate: She was the participant in the NatS group who had the poorest performance and who spent the least time training at home. The individual results and the amount of *at home* training time are shown in Appendix M.

6.4 Generalization

Generalization effects were assessed by comparing the performance of the SynS group in the pretest and posttest. Since the participants in this group were trained exclusively with synthesized stimuli, and were tested with natural stimuli, an eventual perceptual improvement in the posttest would constitute evidence for the transfer of the knowledge acquired through synthesized stimuli to natural speech settings. As shown in Figure 47 (Section 6.3), this hypothesis was confirmed by means of a Wilcoxon test, which showed that there was a significant difference from pretest to posttest.

Secondary evidence of generalization was found in the NatS group. Although the participants in this group were trained and tested with natural stimuli, most of the native speakers whose utterances were recorded for the test stimuli were different from the native speakers who recorded the training stimuli, as reported in Chapter 5 (Section 5.2.4.1.1). Thus, an improvement from the pretest to the posttest of the participant in the NatS group would indicate that they generalized the new knowledge to new speakers. Figure 47 also showed that there was a significant improvement from pretest to posttest, confirming the generalization predictions to the NatS group.

6.5 Production results

In this Section, the results of the production tests are presented and discussed, and in Section 6.7, the production results are approached in the light of current perception models.

As explained in Chapter 5, the production data of the female participants were segmented, labeled and measured, and the results of each of the tests were compared to check whether there was production improvement or not, as a result of perception training. Thus, in order to make this comparison, the first step was to measure the F1 and F2 of the six targets, which was done with a *Praat* script (see Appendix C). The same procedure was followed with the vowel /ʌ/, since it was also confused with /u/ and /ʊ/ in the perception tests, and it would be interesting to observe its behavior in the production domain. The measurement results of the L2 vowels in the pretest and in the posttest of the control group are shown in Tables 19 and 20, respectively. The pretest, posttest and follow-up test results of the experimental groups are shown in Tables 21, 22 and 23, respectively. The F0 and duration of the L2 vowels are also displayed for reference.

Table 19. Mean, median and SD of duration, F0, F1, F2 and F3 values of the English vowels produced by the BP Control group in the pretest (all females).

		/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/ʊ/	/u/
	Mean	239	236	275	290	256	256	253
Duration	Median	234	234	269	282	247	249	248
	SD	56	51	57	52	57	48	49
	Mean	218	217	188	195	211	207	231
F0	Median	217	219	197	203	207	210	230
	SD	54	59	69	56	48	67	50
	Mean	333	372	661	659	506	407	356
F1	Median	322	363	667	670	505	394	354
	SD	59	65	77	83	65	78	49
	Mean	2530	2456	2199	2169	1831	1247	1062
F2	Median	2561	2449	2190	2157	1861	1295	1032
	SD	284	237	189	202	164	357	293
	Mean	3171	3125	2933	2948	2895	2855	2783
F3	Median	3108	3125	3006	3061	2865	2781	2695
	SD	322	310	348	366	322	358	338

Table 20. Mean, median and SD of duration, F0, F1, F2 and F3 values of the English vowels produced by the BP Control group in the posttest (all females).

		/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/ʊ/	/u/
	Mean	244	231	272	278	253	255	248
Duration	Median	237	227	276	272	251	255	248
	SD	52	49	45	49	48	48	54
	Mean	215	218	195	194	205	210	229
F0	Median	221	218	197	198	209	209	224
	SD	71	69	51	62	64	56	49
	Mean	352	369	654	663	509	404	382
F1	Median	350	371	657	669	511	391	377
	SD	55	54	71	77	61	83	59
	Mean	2604	2513	2243	2195	1856	1173	1113
F2	Median	2580	2501	2242	2216	1860	1067	1054
	SD	271	175	149	264	147	337	301
	Mean	3178	3158	2837	2857	2868	2830	2857
F3	Median	3207	3171	2941	3010	2900	2743	2818
	SD	345	286	392	384	363	280	312

Table 21. Mean, median and SD of duration, F0, F1, F2 and F3 values of the English vowels produced by the female participants in the Experimental groups in the pretest.

		/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/ʊ/	/u/
	Mean	206	205	236	241	229	215	221
Duration	Median	201	202	234	236	230	212	222
	SD	61	61	66	66	64	61	62
	Mean	221	216	193	192	208	213	227
F0	Median	215	217	199	200	208	212	224
	SD	52	55	58	59	50	48	48
	Mean	405	421	714	725	575	440	430
F1	Median	402	416	718	726	574	439	429
	SD	64	71	85	88	65	64	63
	Mean	2556	2480	2061	1978	1730	1160	1204
F2	Median	2587	2492	2063	2074	1712	1125	1187
	SD	235	220	142	331	149	216	264
	Mean	3074	3033	2646	2605	2772	2729	2731
F3	Median	3075	3078	2778	2738	2817	2738	2740
	SD	261	261	384	402	301	240	256

Table 22. Mean, median and SD of duration, F0, F1, F2 and F3 values of the English vowels produced by the female participants in the Experimental groups in the posttest.

		/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/ʊ/	/u/
	Mean	250	247	282	294	267	258	255
Duration	Median	254	252	282	295	273	264	258
	SD	58	53	63	55	55	48	59
	Mean	212	198	183	185	204	202	221
F0	Median	209	207	195	196	207	205	215
	SD	59	68	66	61	54	51	52
	Mean	392	441	712	747	570	459	417
F1	Median	387	426	709	744	555	446	410
	SD	61	115	76	89	80	73	59
	Mean	2542	2358	2021	2000	1771	1248	1197
F2	Median	2570	2398	2054	2058	1760	1226	1178
	SD	246	348	228	250	134	266	273
	Mean	3070	2969	2637	2588	2790	2723	2713
F3	Median	3083	3006	2756	2701	2924	2726	2716
	SD	287	282	386	409	267	204	239

Table 23. Mean, median and SD of duration, F0, F1, F2 and F3 values of the English vowels produced by the female participants in the Experimental groups in the follow-up test.

		/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/ʊ/	/u/
	Mean	221	226	248	265	240	237	225
Duration	Median	215	230	244	258	239	238	227
	SD	60	57	57	61	55	50	57
	Mean	207	203	190	187	196	201	215
F0	Median	206	205	195	195	200	200	212
	SD	57	56	56	59	57	48	53
	Mean	381	431	701	738	547	450	409
F1	Median	371	427	700	724	544	445	404
	SD	89	83	79	95	65	83	56
	Mean	2536	2386	2071	2001	1796	1297	1282
F2	Median	2554	2369	2069	2038	1789	1274	1258
	SD	246	217	136	258	153	295	314
	Mean	3083	2948	2696	2699	2779	2736	2726
F3	Median	3071	2984	2807	2810	2835	2720	2704
	SD	294	309	378	402	292	206	217

After the measurement of the F1 and F2 values, another *Praat* script (see Appendix D) was used to calculate the Euclidean Distance (ED) between the means of the L2 vowel pairs of each participant, whose values are presented in Appendix N. A third *Praat* script (Appendix E) was used to generate the L2 vowel plots, which are shown in this Section. After all the target acoustic properties were measured, comparisons between and within groups were made.

Similarly to the perception results, the production results of the control group and of the experimental groups in the pretest were compared to check whether the initial learning stage of the groups would approximate or not. Figure 50 shows that the ED between vowel pairs in the control group was slightly larger than in the experimental groups. However, statistical tests showed that this difference was not significant. The distribution of the vowels in the pretest in the control groups and in the experimental groups can be visualized in Figure 51.

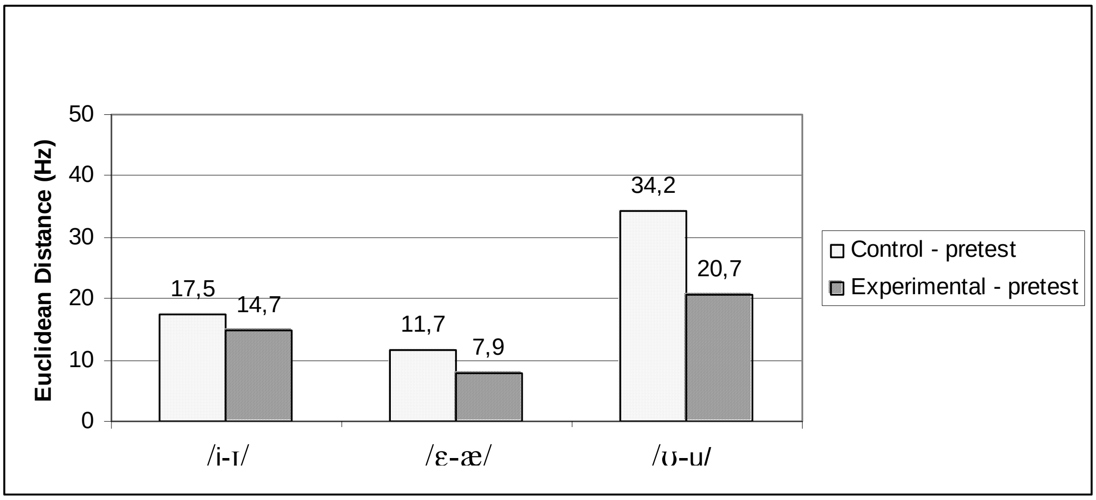


Figure 51. Mean ED of each L2 vowel pair in the production pretest of the control group and of the experimental groups.

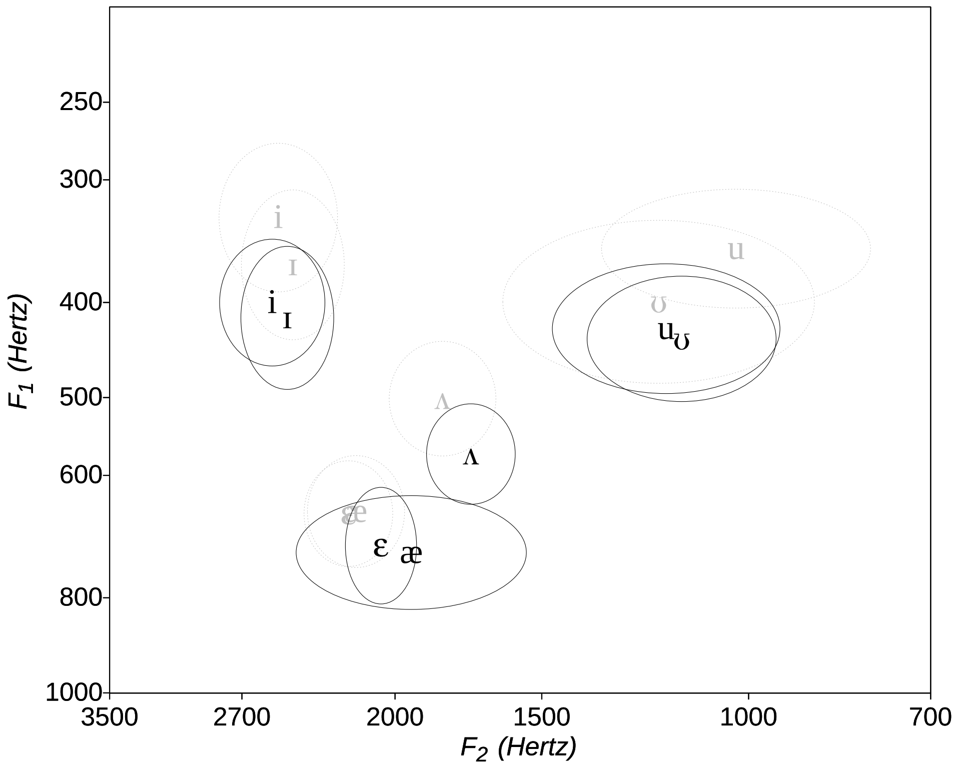


Figure 52. Means and SD (in ellipses) of normalized L2 vowels produced by the female participants in the pretest in the Control group (in grey) and in the Experimental groups (in black).

The results of the control group are probably due to chance, and that might be supported by the comparison of the results of this group in the pretest and in the posttest. Figure 52 shows that the performance of the control group decreased in the posttest and this decrease was statistically significant for the pairs /ɛ-æ/ and /ʊ-u/ (/ɛ-æ/: $t(6) = 2.719$, $p = .036$; /ʊ-u/: $t(6) = 2.696$, $p = .037$). These results indicate that the vowels within each pair produced by the control group in the posttest exhibited a higher degree of overlap than in the pretest, as shown in Figure 53.

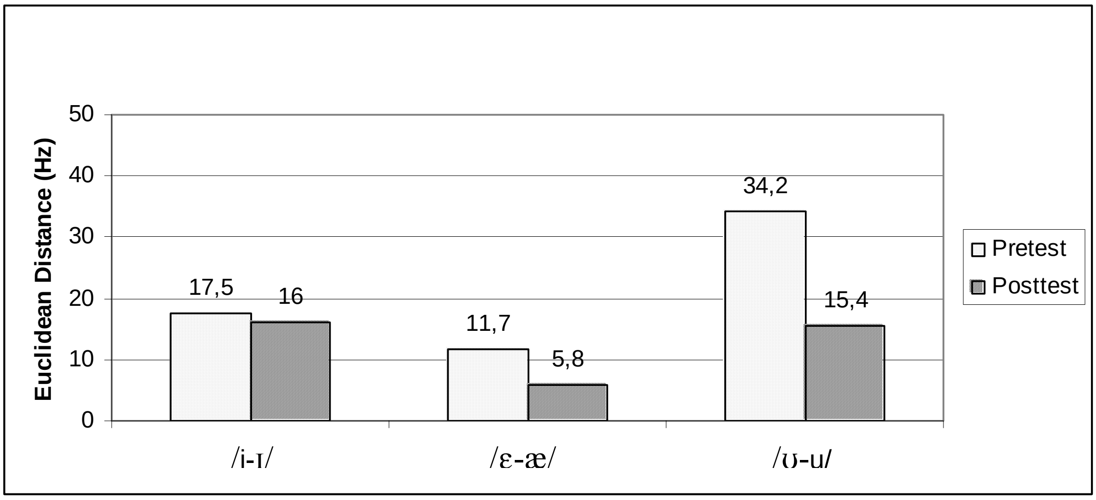


Figure 53. Mean ED of each L2 vowel pair in the pretest and in the posttest of the Control group.

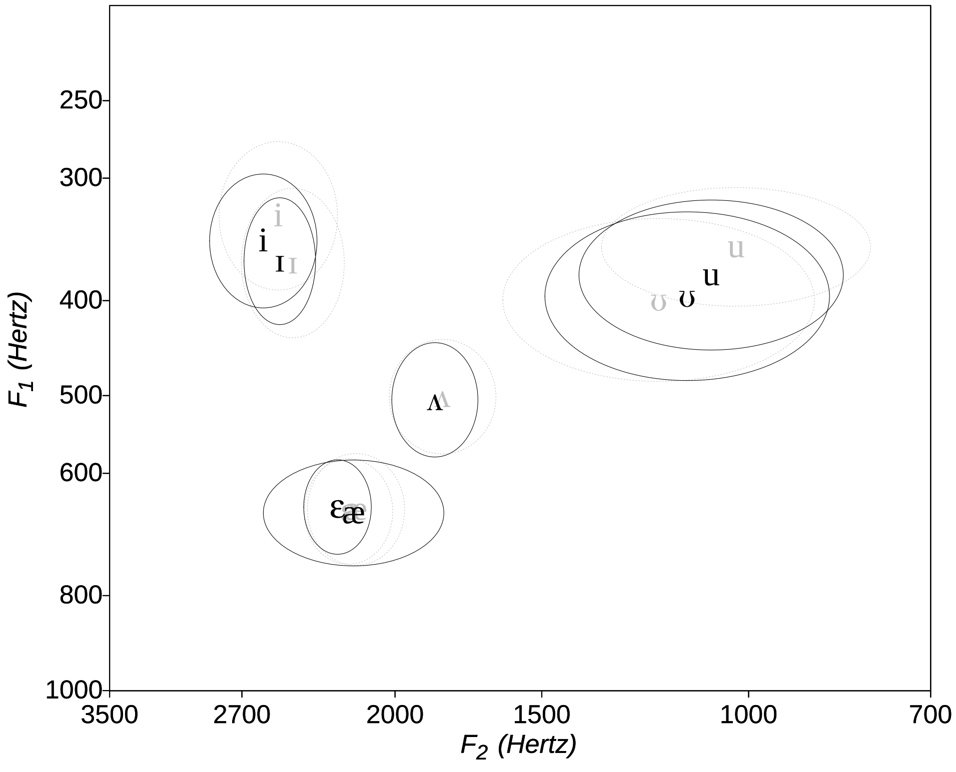


Figure 54. Normalized L2 vowels produced by the female participants in the pretest (grey line) and in the posttest (black line) in the Control group.

Visualizing the plotted results of the experimental groups in the pretest and in the posttest (see Figure 54), it is possible to observe that in the pretest (grey line) the vowel pairs /i-ɪ/ and /ʊ-u/ are occupying approximately the same space. After training, /ɪ/ is more dispersed than in the pretest and its degree of overlap with /i/ is also slightly lower.

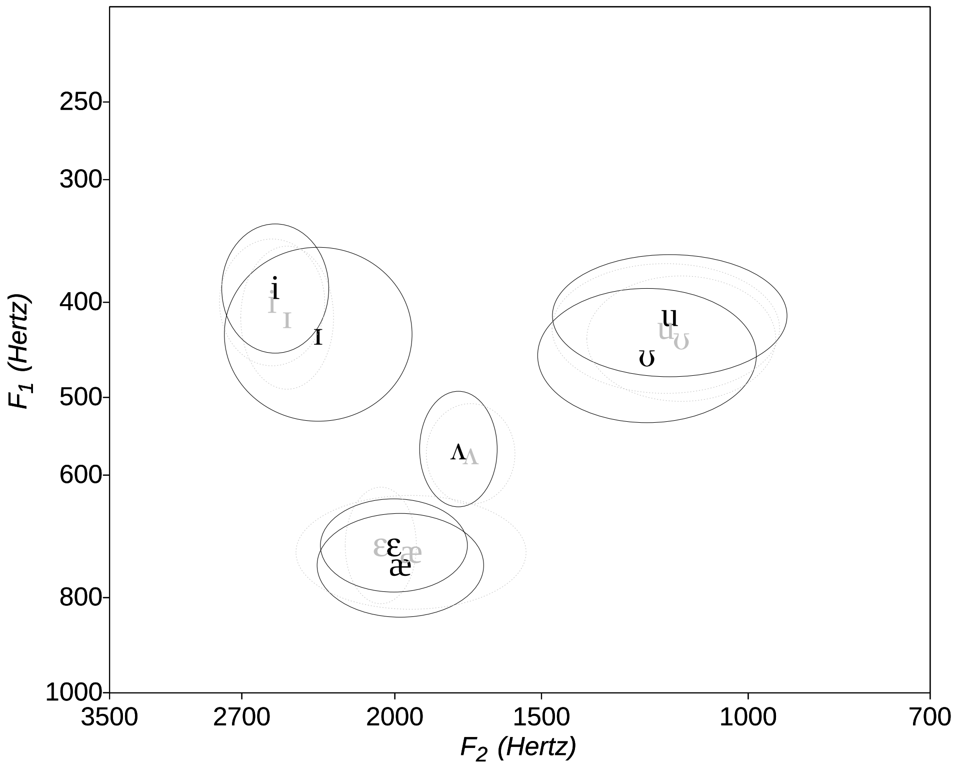


Figure 55. Normalized L2 vowels produced by the female participants in the pretest (grey line) and in the posttest (black line) in the Experimental groups.

These results suggest two things: First, related to dispersion, there has been more variation of the production of /ɪ/ in terms of height and backness, which suggests that the participants actually perceived some difference and tried to adjust to it, although it was small in extent. Second, they slightly increased the ED between /i-ɪ/. It is possible to better visualize such increase in Figure 55. Statistical tests showed that the performance of the participants was considerably better in the posttest for the vowels /i-ɪ/ ($t(22) = -3.33$, $p < .006$).

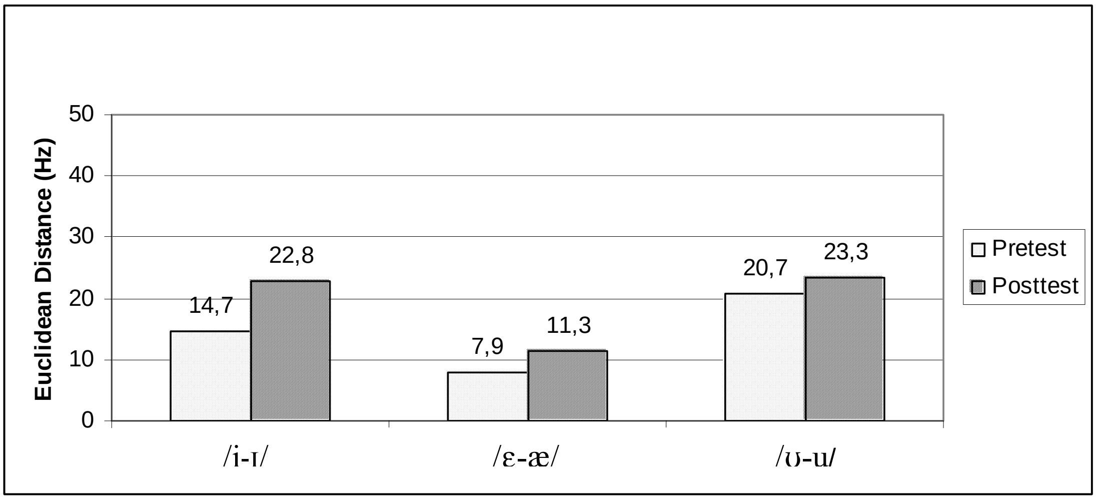


Figure 56. Mean ED of each L2 vowel pair in the pretest and in the posttest of the Experimental group.

As to the pairs /ɛ-æ/ and /ʊ-u/, no statistical significance was found in terms of increase of ED after training. Figure 54 shows that /ʊ/ was slightly lower and more fronted than /u/ after training, and that /ɛ/ was somewhat higher than /æ/, but these differences were not significant.

However, if the results of each experimental group are analyzed separately, it is possible to see that the poor results of the NatS group are masking the good results of the SynS group. Therefore, as shown in Figures 56 and 57, respectively, the participants in the NatS group did not improve their production of /ɛ-æ/ and /ʊ-u/, and the improvement of /i-ɪ/ was not statistically significant. On the other hand, the oral performance of the participants in the SynS group in the posttest was statistically better than that in the pretest for all vowel pairs (/i-ɪ/: $t(12) = 2.405$, $p = .03$; /ɛ-æ/: $t(12) = -2.544$, $p = .02$; /ʊ-u/: $t(12) = -2.145$, $p = .05$).

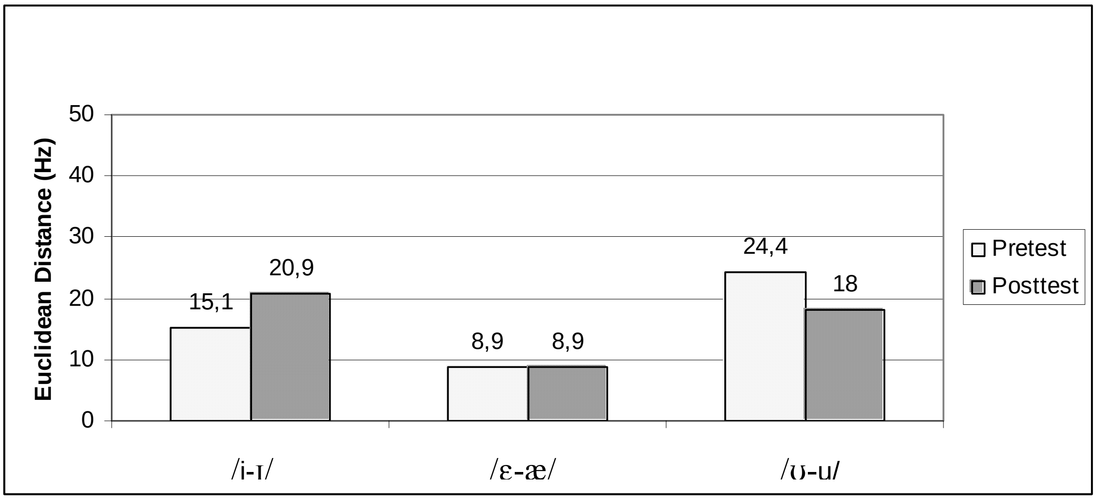


Figure 57. Mean ED of each L2 vowel pair in the pretest and in the posttest of the NatS group.

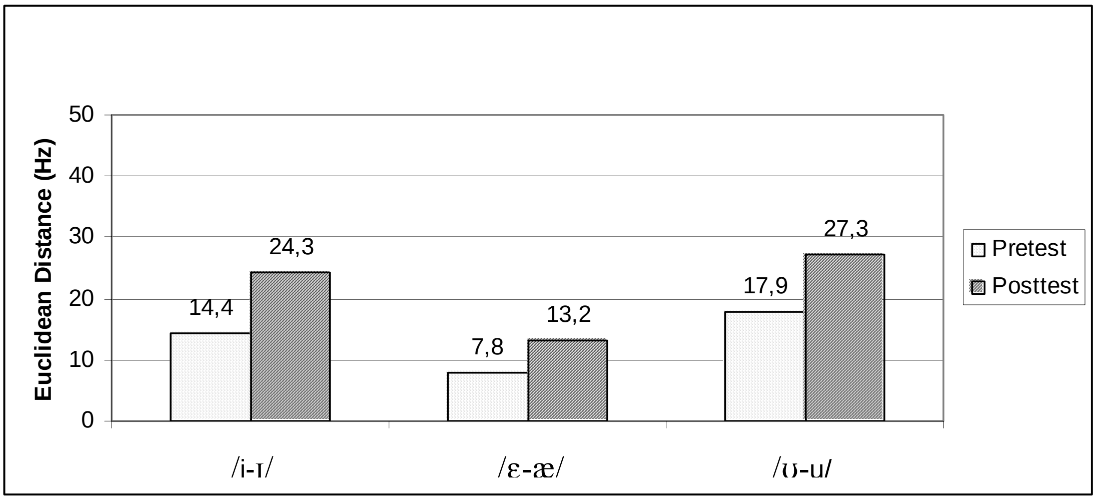


Figure 58. Mean ED of each L2 vowel pair in the pretest and in the posttest of the SynS group.

The better results of the SynS group, compared to the NatS group, can also be visualized in Figures 58 and 59, respectively. As for /i-ɪ/, Figure 58 shows that the SynS group produced /ɪ/ in a more back and lowered position, whereas the NatS group tendency (see Figure 59) was only to make it lower. The degree of overlap between the back vowels also diminished in the SynS group, which also produced /ʊ/ more lowered and more fronted in the posttest. The production of /ɛ-æ/ also improved in the SynS group in that both vowels had similar F1 values before training, but after training /ɛ/ was slightly higher than /æ/.

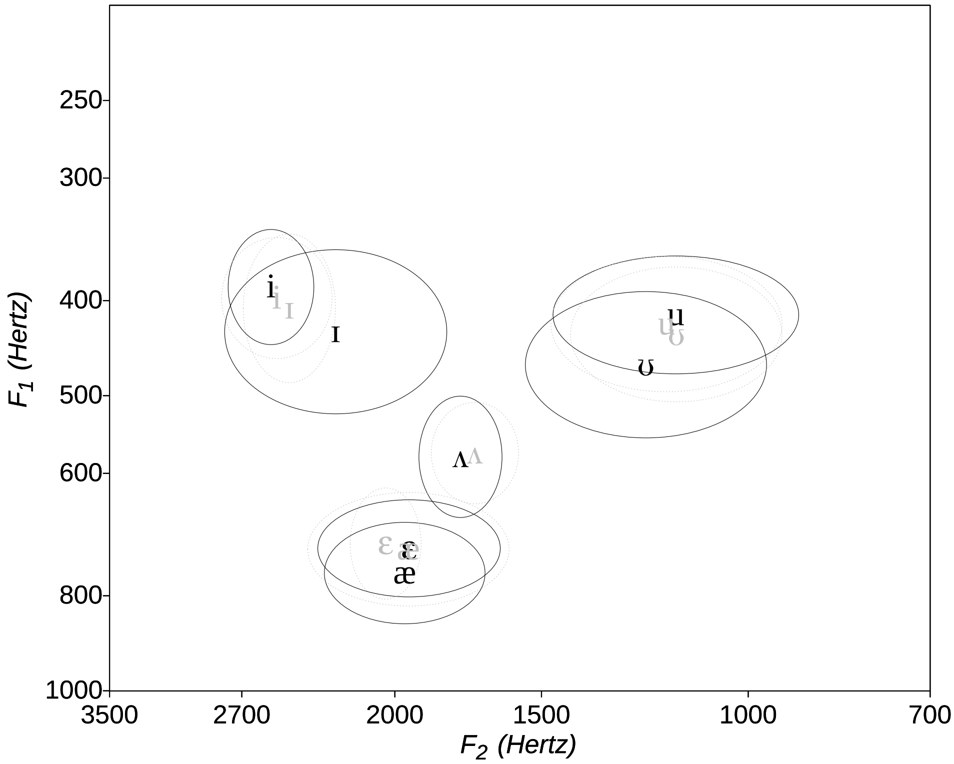


Figure 59. Normalized L2 vowels produced by the female participants in the pretest (grey line) and in the posttest (black line) in the SynS group.

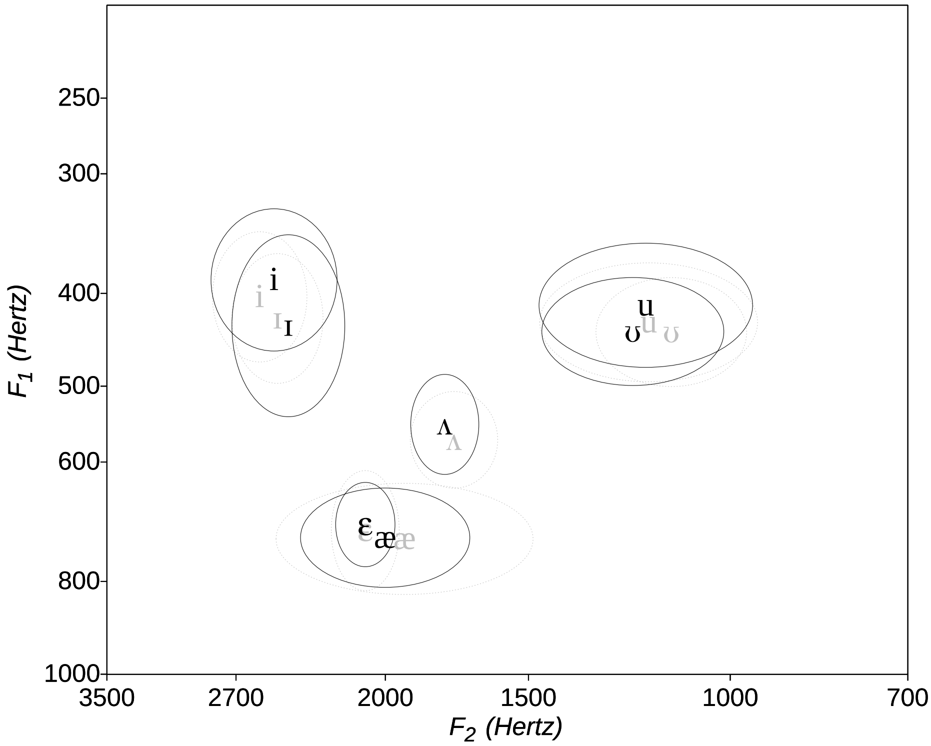


Figure 60. Normalized L2 vowels produced by the female participants in the pretest (grey line) and in the posttest (black line) in the NatS group.

Differently from what was found in perception, the vowel /ʌ/ was not confused with /ʊ/ or /u/ in any of the groups. Looking at the distribution of the vowels in Clopper et al.'s (2005) study (see Figure 12), some overlap between /ʊ/ and /ʌ/ was found because /ʌ/ is more back and /ʊ/ is more fronted. However, the participants in the present study produced more central /ʌ/ (though slightly fronted, rather than back), and a higher /ʊ/ (which was produced at the same height as /u/), and that prevented an eventual overlap between /ʌ/ and /ʊ-u/, as shown in Figures 53 and 54.

The production results of the experimental groups presented in this Section apparently correlate weakly with the perception results[42]. Considering a possible

[42] The correlations were run with ED values in Hz and raw number of correct identifications per vowel pair.

correlation between perception and production by participant performance, Pearson's correlation tests confirmed that there was no significant correlation between perception and production results (see Appendix O).

Considering a possible parallel improvement between perception and production by vowel pair in the experimental group, the perception results showed that /ʊ-u/ improved more than /ɛ-æ/, which improved more than /i-ɪ/. These results also showed that the perception of /i-ɪ/ was already very good before training. The production results showed that the participants' production improved more for /i-ɪ/, followed by /ɛ-æ/ and /ʊ-u/, respectively.

6.6 Long-term effects of training

In order to check whether training had long-term effects, the results of the perception posttest were compared to those of the perception follow-up test. Friedman tests[43] showed that the means for the three tests in the SynS group are significantly different (X^2 = 24.1, p = .001), and Wicoxon tests confirmed that there was significant improvement from pretest to posttest (p = .001) and from pretest to follow-up test (p = .001), but no significant difference from posttest to follow-up test (p > .05).

As for the NatS group, similar results were obtained. Friedman tests confirmed that the means were significantly different for the pretest and post- and follow-up tests (X^2 = 15.42, p = .001). Wilcoxon tests confirmed that there was significant improvement from pretest to posttest (p = .003) and from pretest to follow-up test (p = .005), but no significant difference from posttest to follow-up test (p > .05).

[43] Friedman is a non-parametric test used to compare three or more means for the same group. It is similar to a Repeated-Measures ANOVA.

Figure 60 illustrates the results. Although there was no significant improvement from posttest to follow-up test in the experimental groups, the results also show that there was not any worsening of performance. These results indicate that, although no further learning occurred one month after training was over, there was no loss of the new acquired ability either, which is a positive finding.

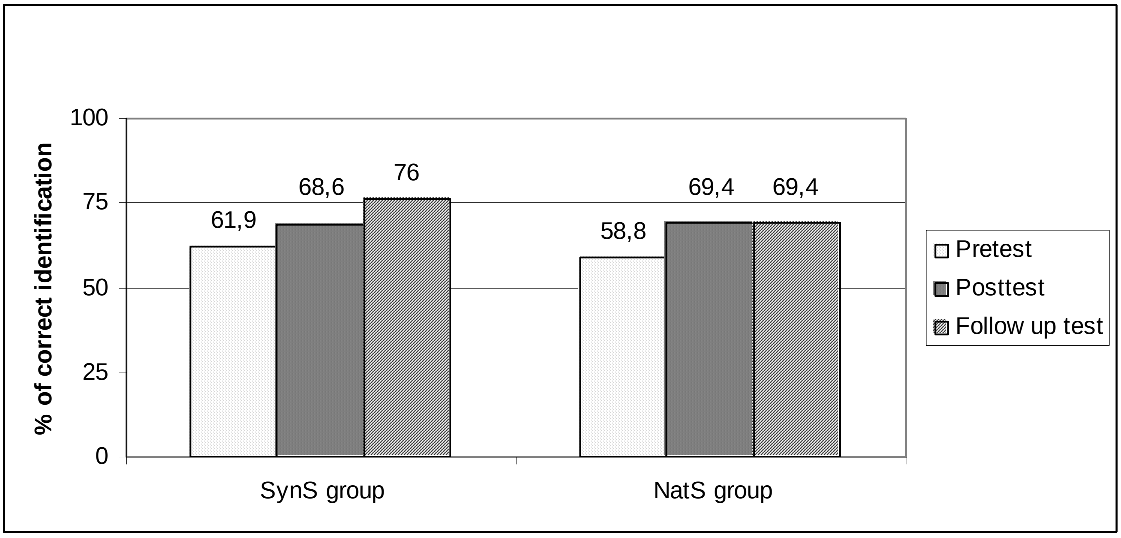


Figure 61. Results of perception pretest, posttest and follow-up test in the experimental groups.

Similar results were found in the production domain. Although Figure 61 shows that there was a slight increase of the ED from posttest to follow-up test for the contrast /u-ʊ/, paired t-tests confirmed that this difference was not significant. This lack of significant improvement can also be visualized in Figure 62, which shows a high degree of overlap between the target vowels in each pair.

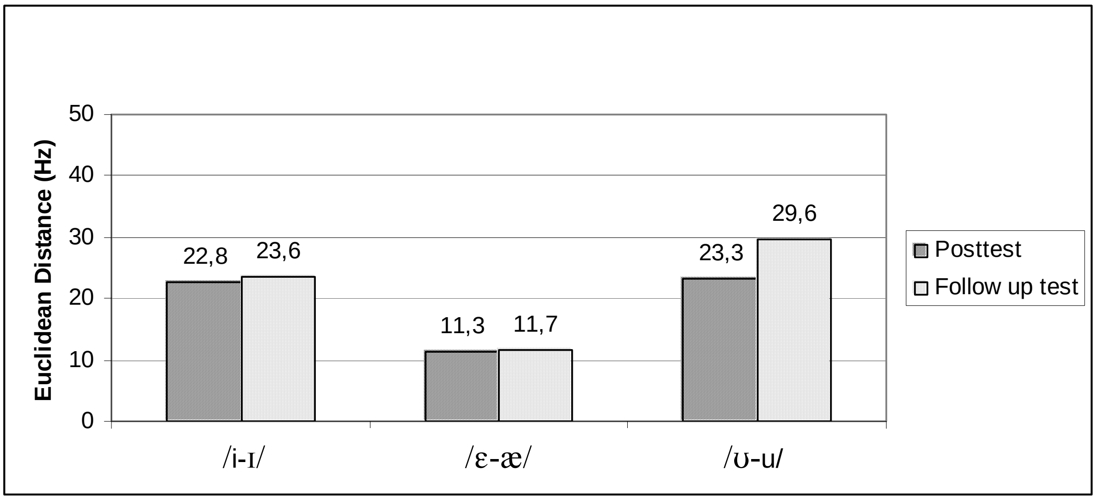


Figure 62. Mean ED of each L2 vowel pair in the production posttest and follow-up test of the experimental group.

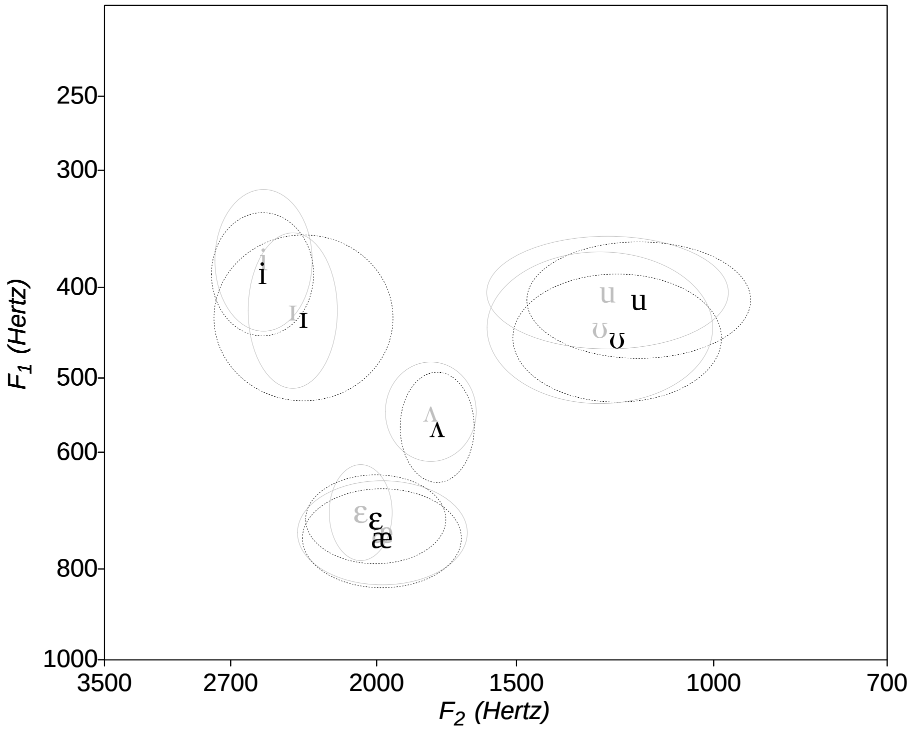


Figure 63. Means and SD (in ellipses) of normalized L2 vowels produced by the female participants in the posttest (in black) and in the follow-up test (in grey).

Statistical tests also confirmed that the production improvement found from the pretest to posttest for the contrast /i-ɪ/ was maintained from pretest to follow-up test (pre → post: $t(22) = -3.23$, $p = .006$; pre → follow-up: $t(22) = -3.213$, $p = .005$), which indicates that, although no considerable improvement was found, no loss of the acquired knowledge was registered either.

These results also address the question of whether testing long-term effects of training only one month after training finishes is already sufficient to see any further improvement. All previous studies (Yamada et al., 1996; Ceñoz Iragui & Garcia Lecumberri, 2002; Wang et al., 1999; Wang, 2002; Wang & Munro, 2004) reported further improvement from posttest to follow-up test, but the follow-up tests were carried out at least three months after training was over. Thus, it is possible that the participants in this study did not improve significantly from posttest to follow-up test because of the rather short time interval between tests, especially if we note the tendency shown by the SynS group in Figure 61: Some improvement from posttest to follow-up test is visible in the SynS group.

6.7 General discussion

In this chapter, the results related to each objective set in the Introduction are reported and briefly discussed in each section. The main objective of this dissertation, related to the effect of training on the learning of English vowels by BP speakers, was approached in Section 6.1. The results in that section showed that, initially, the control and the experimental groups exhibited a similar pretraining status in terms of perceptual performance. This equivalence of initial learning state between groups is important, since it allows for more reliable comparisons and conclusions as to the effects of training. Another finding shown in the section was that the perceptual performance of the experimental groups improved significantly after training, whereas there was no significant difference in the performance of the control group, which suggests positive effects of training on the perceptual improvement of the participants.

In sum, the results presented in Section 6.1 show the initial and final learning stages of the participants in the control and in the experimental groups, providing evidence for the effectiveness of perceptual training. However, in order to relate these findings with the linguistic models presented in Chapter 2, it is necessary to take into account the secondary hypotheses of this dissertation, which are related to the difference of L1 and L2 vowel inventories.

The results presented in Section 6.2 show the participants' incorrect identification of the English vowels in the pretest, their incorrect answers in the posttest, and the overall rate of correct identification of the experimental groups before and after training. The main reason why the misidentifications in the pretest were analyzed was to check whether the identification pattern of the participants matched those predicted in Hypothesis 1. Thus, differently from what was predicted in the Introduction, Table 24 shows that before training the participants not only perceived /ɪ/ as /i/, /æ/ as /ɛ/, and /ʊ/ as /u/, but also perceived (i) target /i/ as /ɪ/, (ii) /ɛ/ as /ɪ/ and /æ/, (iii) /u/ as /ʌ/ and /ʊ/, (iv) /ɪ/ as /ɛ/, and (v) /æ/ as /ɑ/. Other patterns were also found (i. e. /i/ as /ɔ/, /i/ and /ʊ/ as /æ/, /i/ and /ɛ/ as /ʌ/, /i/, /ɪ/ and /æ/ as /u/, /æ/ as /ʊ/, /ʊ/ as /ɪ/, and /u/ as /i/), but they were probably due to lack of attention.

Table 24. Principal predicted and unpredicted misidentification of the targets before training.

Target vowel	Perceived as… (predicted)	…but also as (not predicted)
/i/	/i/	/ɪ/
/ɪ/	/i/	/ɛ/
/ɛ/	/ɛ/	/ɪ/, /æ/
/æ/	/ɛ/	/ɑ/
/ʊ/	/u/	/ʌ/
/u/	/u/	/ʊ/, /ʌ/

As already discussed in Section 6.2.1, some of these misidentification patterns can potentially be due to the native speakers' pronunciation of the targets, since, according to Clopper et al.'s (2005) dialectal mappings, some of the target vowels are overlapping in some U.S. regions.

Another possibility is that there was confusion of the auditory stimulus with the written form of the words. As discussed in Section 6.2.1, the sound /i/ is represented by the letter *i* in BP, and since this does not apply to English, the participants may sometimes have chosen the answer *ship* when they perceived the stimulus as containing /i/. Assmann et al. (1982) have already investigated this problem of orthographic interference in vowel identification tasks, and they concluded that, indeed, "when listeners are required to respond to a set of vowels using written response labels, orthographic and labeling difficulties provide a potential source of error" (p. 985). In association with this orthographic mismatch between L1 and L2, there was also the problem of fatigue from the test, as informally reported by some of the participants.

The misidentification tendencies found in the pretest and reported in Section 3.5.2 partially support the predictions of current speech perception models. The models presented in Chapter 3 suggest that in initial learning stages speech perception is L1 based, that is, it is driven by L1 neural mappings. Therefore, when the participants heard L2 /i/ or /ɪ/, they would interpret both of them as L1 /i/, L2 /i/ being a better instance of the L1 vowel than L2 /ɪ/. The same would be true for L2 /ɛ/ and /æ/: both perceived as L1 /ɛ/, and L2 /ʊ/ and /u/ perceived as L1 /u/. The unpredicted confusion between /ɪ/ and /ɛ/ was explained by Flege and MacKay (2004), who suggest that the two vowels have similar acoustic specifications. The authors claim that these two vowels "differ relatively little in terms of midpoint formant frequencies and duration and do not show a differing pattern of formant movement" (Flege & MacKay, 2004, p. 13). Furthermore, the authors explain that the confusion between /ɪ/ and /ɛ/ is also made by native speakers. In fact, one of the native speakers in the control group of the present study misidentified /ɪ/ as /ɛ/ once, as shown in Appendix N. According to Flege and MacKay (2004), these errors can be attributed to something other than L1 interference, which in this case may be the similar acoustic properties of the two vowels.

According to the assimilation patterns proposed by the PAM, and which are represented in Table 4, the participants' perceptual tendency before training as to the vowels /ɛ/ identified as /ɪ/ and /æ/, and /u/ identified as /ʌ/ and /ʊ/ was to perform *single-category assimilation*, according to which both misidentified vowels are considered bad (*deviant*) instances of the target and the predicted discriminatory performance of the participant is expected to be poor. As a matter of fact, /ɛ/ and /u/ are among the vowels which had the lowest correct identification rate in the pretest.

Another assimilation pattern proposed by the PAM which was found for most of the targets in the pretest results was the *category-goodness difference*. As explained in Chapter 3, this assimilation pattern postulates that two L2 phonemes are assimilated into the same L1 category, but comparing these two L2 sounds to the L1 phoneme, one is a better instance than the other. For instance, English /i/ and /ɪ/ were both perceived as Portuguese /i/, and English /i/ is probably considered a better instance of Portuguese /i/ than English /ɪ/. In this case, discriminatory performance of the learners is expected to range from moderate to very good.

After training was over, there was a considerable decrease in the misidentification rate in the experimental groups. The results presented in Section 6.2.2 show that this decrease was significant for the vowels /æ/, /ʊ/, and /u/. There was also a slight (but not significant) decrease in perceptual errors for /i/ and /ɛ/. The improvement for the vowel /ɪ/ was the subtlest.

Considering the results of Sections 6.2.2 and 6.2.3, the main findings were that training led to perceptual improvement of the target vowels and that each vowel contrast improved to a different extent. Therefore, /ʊ-u/ improved more than /ɛ-æ/, which improved more than /i-ɪ/. As discussed in Section 6.2.3, this difference in improvement rate for each vowel contrast can potentially be a result of the participants' self-perception and selective attention, and of the available room for improvement of the targets. Thus, after the pretest, the participants probably realized that most of their errors involved the vowels /ʊ/ and /u/; consequently, the contrast /ʊ-u/ was considered the most difficult for them. The participants, then, most likely started to focus their attention on that contrast during the treatment period (training sessions). Parallel to that, since the number of correct identifications of the back vowels in the pretest was the lowest, /ʊ/ and /u/ also had the largest room for improvement. Therefore, the participants had a high probability of improving their perception of the target back vowels. Moreover, the resulting perceptual improvement after training constitutes evidence for the PAM and the SLM, since it indicates that the participants' sound categories underwent some kind of reorganization, moving towards near-native-like patterns.

Besides, although the participants performed exactly the same task in the posttest, it was not as difficult and as boring as it was in the pretest, since they *knew* what they would have to pay attention to (in terms of vowel quality), and they were also aware that the vowels in each contrast did not sound the same (and, consequently, the options in the answer sheet represented words with different vowel sounds). Still, the possible interference of orthography cannot be discarded, and further tests are necessary in this respect.

In order to overcome this problem, Assmann et al. (1982) suggested that phonetic symbols should be used on the answer sheet, and learners should be trained on their use. However, that would be problematic in a pretest-posttest study design with similar objectives as the present one, since the participants are not expected to know much about phonetic symbols and acoustic specification of the targets before training. Thus, following the idea of Assmann et al., the pretest should have a different format from the posttest (the former containing words or drawings, and the latter containing the phonetic symbols), and the results would not be very reliable.

Another suggestion given by Assmann et al. (1982) was to elicit the answers from the learners by means of oral repetition. That would not be reliable, either, because nonnative speakers obviously do not have a native-like pronunciation of the L2 vowels, and that would also interfere negatively in the computation of their pretraining results.

The results presented in Section 6.3 showed that both training methods (training with synthesized stimuli and training with natural stimuli) led to perceptual improvement. Based on the findings of previous research, it was hypothesized that learners trained with enhanced stimuli would benefit more than learners trained with natural stimuli in terms of perceptual improvement. Although the difference in improvement rate was not significant between training groups, the SynS group performance in the posttest was slightly better than the performance of the NatS group.

One could argue that the use of isolated vowels in the synthesized stimuli prevented the participants in the SynS group from having much better results, since the tests involved CVC stimuli, and that isolated vowels do not contribute to accurate perception because they lack the acoustic dynamic information contained in the neighboring consonants in coarticulated syllables (Gottfried & Strange, 1980). In this sense, the NatS group would benefit more and thus have a higher rate of improvement. However, this possibility was discarded since the results indicate that there was no significant difference between the NatS and the SynS groups. Further counterevidence was provided by Assmann et al. (1982), who showed that “vowels may be well-identified even in the absence of context” (p. 985).

The results related to the amount of *at home* training time were also presented in Section 6.3. They showed that there was no significant difference between groups as to the time spent on training. Similarly, no correlation was found within groups. Looking at the performance of each participant, only one participant followed the expected tendency: the greater time spent on training, the better the performance, and *vice versa*. However, this was an isolated case, and none of the other individual performances could have been predicted by *at home* training time. This variable was not controlled for in previous studies, although it was suggested by Wang (2002) and Wang and Munro

(2004). Therefore, in order to understand more thoroughly the relationship between amount of training time and gains in perceptual performance, further investigation is necessary.

As reported in Section 6.4, generalization of the new knowledge to new stimuli and new speakers was found in the two training groups. The group that trained exclusively with synthesized stimuli performed considerably better in the posttest, in which they responded to natural stimuli (that is, stimuli produced by real people, and not created in a computer) consisting of CVC words. Therefore, the new knowledge acquired with training was transferred to a new kind of stimuli (natural) and to a new syllable structure (CVC).

Similar positive results were also found for the participants in the group that trained exclusively with natural stimuli: They performed significantly better in the posttest, which consisted of utterances of new speakers. However, further research is necessary to investigate generalization to new words and new contexts.

As for the production results presented in Section 6.5, it was possible to observe that the experimental groups improved their oral performance of the pair /i-ɪ/ considerably, but no production gains were found for /ɛ-æ/ and /ʊ-u/. Furthermore, it was observed that the lack of improvement of these two pairs was due to poor performance of the NatS group, since the participants in the SynS group improved their production of the three L2 vowel pairs. Thus, this particular finding in the production domain is another indication of the effectiveness of training with the synthesized, enhanced stimuli over training with natural stimuli.

Considering the performance of the experimental group as a whole, only /i-ɪ/ improved significantly, especially because /ɪ/ was lower and more back after training. This back and downwards movement indicates that learners were trying to adjust their production to the perceptual targets they were exposed to. The plots in Figure 12 show that /i/ is more fronted and higher than /ɪ/ in all U.S. dialects. This interpretation would apply not only to the NatS group (which trained with natural stimuli) but also to the SynS group, since the F1 and F2 values used to generate the synthesized stimuli were based on those of Peterson and Barney (1952) and of Ohnishi (1991), both of whose /ɪ/ is lower and more back than /i/.

These production results support the claims of the NLM model and of the SLM, which suggest that perception precedes production. Although in the perception domain most of the L2 targets improved significantly, in the production domain only one pair improved for both experimental groups. It is interesting to remark that the only pair that did not improve in perception was the only one that improved in production, that is, /i-ɪ/. This fact, however, is explained by the very good perceptual identification rate in the pretest, which was maintained after training. It was more difficult to improve significantly something that was already very good before training, which was the case of this pair. Thus, since /i-ɪ/ was already very good in the perception domain, it was the pair that improved the most in the production domain. Following this line of thought, there was still “a lot of work” to be done with /ɛ-æ/ and /ʊ-u/ in the perception domain before any improvement could be seen in the production domain, and that is why no considerable gains were found for those pairs in the production posttest.

The fact that /ʊ-u/ was the pair that improved the most in perception but not in production is explained by the fossilization of speech gestures. Considering that the learners had produced /ʊ/ and /u/ as /u/ since they started learning English, and taking into account that their teachers probably had not called their attention to the difference between these two vowels (either because they did not know or could not produce them themselves), it is likely that they would have difficulty articulating /ʊ/, even after learning that /u/ and /ʊ/ are not the same, and after being able to perceive these vowels accurately. This fact would support Major's (1987a) claims about the degree to which teachers can interfere positively or negatively on the learners achievements.

A finding reported in Section 6.5, to which the above explanation applies, is the lack of parallel improvement between perception and production by vowel pair. The perception results showed that the order of improvement was (from best to worst) /ʊ-u/, /ɛ-æ/ and /i-ɪ/, while the production results showed that the participants' production was better for /i-ɪ/, followed by /ɛ-æ/ and /ʊ-u/. As explained above, /i-ɪ/ improved the least because its perception was already very good. Therefore, it would be very unlikely to find a significant positive correlation by vowel pair between perception and production improvement and very logical that /i-ɪ/ was the pair which improved the least in perception and the most in production. Considering the production results by participant, Pearson's correlation tests showed that they weakly correlate with the perception results. Thus, similarly to the findings of previous studies (Bradlow et al., 1997; Wang et al, 2003), the lack of correlation between perception and production indicates that the improvement in these two domains occurs at different rates, suggesting that perception precedes production.

Considering the long-term effects of training, previous studies reported in Chapter 2 found not only the retention of the perceptual learning acquired during the training period, but also further improvement in perception and in production after the training sessions were over. As to the present study, statistical tests revealed no significant improvement from posttest to follow-up test in the experimental groups in either domain (Section 5.6). However, maintenance of the improvement achieved by the participants from the pretest to the posttest was shown in the results of the follow-up tests. Therefore, the positive finding as to the long-term effects of training was that the acquired knowledge in perception and in production is not *lost* after training is over.

As for the lack of significant improvement in the follow-up test, a factor which may have contributed to that was the different design of the present study, compared to previous studies, in terms of time interval between the posttest and the follow-up test. Most of the previous studies had a three-month to one-year interval from posttest to follow-up test, whereas in this study the participants had only a one-month interval. With a longer time interval, the participants in previous studies were able to improve not only their perception, but also their production. In the present study, although not significant, the participants in the SynS group showed a tendency toward continued perceptual improvement; however, further testing would be necessary to find out whether this tendency would be confirmed or not. As to production, long-term maintenance of improvement was found only for the pair /i-ɪ/ in the experimental groups, although a non-significant tendency for improvement was observed for /ʊ-u/ after training was over.

Chapter 7

Conclusion

The main objective of this study was to test the effects of perceptual training on the learning of English vowels by BP speakers. This objective was achieved and the main hypothesis supported by the results that show the perceptual improvement of the participants in the two training groups. Furthermore, the results of this study confirmed some of the specific hypotheses presented in the Introduction.

As to the first hypothesis, according to which the BP learners would assimilate the L2 sounds into their L1 categories before training, and after training their perceptual ability would be improved, the results show that it was only *partially confirmed*. As a matter of fact, the BP learners presented some assimilation patterns of L2 sounds, as predicted by L2 speech perception models, and their ability to identify L2 sounds improved after training. However, only the mid/low front vowel pair and the high back vowel pair improved significantly after training. Nevertheless, the nonsignificant improvement of the high front vowels after training was possibly due to the limited room for improvement. As to the hierarchy of perceptual improvement of the targets, the following tendency was found: /ʊ-u/ improved more than /ɛ-æ/, which in turn improved more than /i-ɪ/.

Differently from previous studies involving perceptual training, two different training groups were tested and had their results compared, one that was trained with natural CVC stimuli and one trained with synthesized (cue-enhanced) V stimuli. Since both groups improved significantly after training, the second hypothesis, which predicted that training with synthesized stimuli would provide learners with more perceptual gains than training with natural stimuli, was *not confirmed.* However, the SynS group did obtain a somewhat greater improvement than the NatS group, indicating a possible tendency toward greater perceptual gains.

The hypothesis predicting generalization of learning was *confirmed*. Not only did the SynS group transfer their perceptual learning gains to a new kind of stimuli (natural) and to a new syllable structure, but also the NatS group transferred the learning to syllables produced by new speakers. Thus, regarding generalization, the results of the present study corroborate those of previous research (Rochet, 1995; Yamada et al., 1996; Bradlow et al., 1997; Wang et al., 1999; Wang, 2002; Wang & Munro, 2004).

The fourth hypothesis, which predicted that perceptual training would lead to production improvement, was only *partially supported*, since only /i-ɪ/ improved in the experimental groups after training. However, if the results of the two experimental groups are analyzed separately, the performance of the SynS group was much better than the performance of the NatS group. Thus, the participants in the former improved their production of all three L2 vowel pairs (/i-ɪ/, /ɛ-æ/, /ʊ-u/), whereas the participants in the latter did not improve significantly in any of the pairs. This finding is another piece of evidence for the effectiveness of training with synthesized and enhanced stimuli.

The results that partially support the fourth hypothesis are also consistent with the postulates of the NLM model and the SLM. These models implicitly suggest that perception precedes production. Thus, any eventual production gains are the result of previous perception gains. In this study, the learners' perception of /i-ɪ/ was already very good at the time of the pretest and improved slightly after training, resulting in the improvement of their production of this pair. However, this improvement does not seem to occur in parallel; that is, changes in the two domains do not happen to the same extent.

Finally, it was hypothesized that the perceptual training gains would be maintained some time after training was over, and the results of both training groups *confirmed* this hypothesis. Besides this maintenance of the perceptual improvement, previous studies have reported further perceptual gains from posttest to follow-up test. Possibly due to the restricted time interval between tests, this finding was not found in the present study (although the SynS group did show a tendency toward further improvement). Similarly, the long-term production results show that the improvement found after training was retained one month after the end of the training, although no further gains of significance were registered. A summary of the results per hypothesis is shown in Table 25.

Table 25. Summary of the results of the present study per hypothesis (H).

Hypotheses	Result
H1. Brazilian learners without training would identify the L2 vowels /ɪ, æ, ʊ/ as L1 /i, ɛ, u/, respectively; after the training sessions, their ability to identify L2 sounds would improve;	Partially confirmed.
H2. Training involving synthesized stimuli would be more effective than training with natural stimuli;	Not confirmed.
H3. Improvement in synthesized speech would be transferred to natural listening settings;	Confirmed.
H4. Perceptual training would lead to production improvement;	Partially confirmed.
H5. The learners' improvement would be maintained some time after perception training is over.	Confirmed.

The amount of extra-class training time was also controlled. However, no effect of this variable was found, since the two experimental groups did not differ significantly in this respect. Furthermore, no positive correlation was found between amount of training time and perceptual improvement.

The findings of the present study lead to some pedagogical and methodological implications. Regarding the former, first it was concluded that perceptual training improved both perception and production skills, even in the absence of any production practice. However, the improvement in the two dimensions did not occur in parallel, and the improvement in perception was greater than the improvement in production, corroborating the postulates of current speech perception models (NLM and SLM).

Second, it was found that the gains in perception were generalized to new talkers (for the NatS group) and to a new kind of stimuli (for the SynS group). Furthermore, these perceptual gains were maintained some time after training was over, and no L2 perceptual learning loss was found in the present study.

Third, although the improvement did not achieve statistical significance, the results of the SynS group suggest that training with enhanced stimuli is more effective than training with natural stimuli, since the rate of improvement in both skills (perception and production) was higher for the SynS group. Therefore, materials developers should consider using synthesized stimuli. Much of the effectiveness of the synthesized training in helping learners to identify L2 sounds is that subtle and crucial cues of the signal are enhanced, drawing learners' attention to them (and the less important features attenuated). Thus, the results of the present study suggest that enhanced stimuli help learners to develop selective attention to the main phonetic cues of certain sounds in a given L2, although more research on this issue is needed.

Finally, language teachers should consider increasing the use of perceptual training in their classes. Nowadays, pronunciation courses foreground training the oral modality, giving less importance to listening practice. As pointed out by Rochet (1995), perception training "is easier to administer" (p. 396), since it can be done without the presence of the teacher (it can be done at home), and learners can be provided with immediate feedback.

The methodological implication of the present study is that extra-class learning and immediate feedback can also be provided by computer-assisted perceptual training, which has another advantage compared to production training. The results concerning perceptual training corroborate what was pointed out by Hazan et al. (2005), who claim that "computer-based perceptual training programmes are more reliable than computer-based pronunciation training programmes" since the latter needs "to provide accurate automatic ratings of the learner's productions" (p. 376). Therefore, perceptual training would be more feasible both for presential and nonpresential courses.

This study had several limitations, which I will leave as suggestions for future research. First, I suggest testing more training groups: Besides having one group training with naturally produced CVC words and one training with synthesized V targets, I would include two other training groups, one training with naturally produced V targets and one training with synthesized CVC words. Thus, more accurate comparisons between training groups would be carried out.

Another limitation was that, due to logistical considerations of the present study, the follow-up test was done only one month after the posttest, which might have prevented learners from achieving better long-term results. Therefore, I suggest that learners are tested at least three months after training is over, similarly to previous studies, in order to check the long-term effects of training.

Limitations on the generalization of the new knowledge were also found since the participants were not tested on the perception of new contexts (such as the target vowels within voiced consonantal contexts). Further investigation in this respect is still needed.

Finally, further research is needed on the use of computer-based perception training, since no correlation was found in this study as to the amount of training time and perceptual improvement. It would be interesting to investigate the rate of perceptual improvement in groups that use software to train during different amounts of time.

In sum, the perceptual improvement of the participants who received training reported in this study constitutes evidence for the reorganization of sound categories predicted by the PAM and the SLM. Furthermore, since the participants started learning English after the critical period and had never lived in an English speaking country, their perceptual improvement was a result of intensive exposure to very controlled material with immediate feedback.

References

Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. *Second Language Research, 19*, 251-265.

Ashby, M., Maidment, J. (2005). *Introducing Phonetic Science*. Cambridge Introductions to Language and Lisguistics. Cambridge: Cambridge University Press.

Assmann, P. F., Nearey, T. M. & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual, and acoustic aspects. *Journal of the Acoustical Society of America, 71*, 975-989.

Baker, W. & Trofimovich, P. (2001). Does perception precede production? Evidence from Korean-English bilinguals. In M. Brend, A. K. Melby & A. R. Lommel (Eds.) *The XXVII LACUS Forum 2001* (pp. 273-284). Lake Bluff, IL: Linguistic Association of Canada and the United States.

Baptista, B. O. (2000). *The acquisition of English vowels by Brazilian Portuguese speakers*. ARES. Florianópolis:UFSC.

Baptista, B. O. (2002). Adult phonetic learning of a second language vowel system. In A. James & J. Leather (Eds.), *New Sounds 2000: Proceedings of the fourth international symposium on the acquisition of second language speech* (pp. 32-41). University of Amsterdam. Klagenfurt, Austria: University of Klagenfurt.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In: Strange, W. (Ed.) *Speech perception and linguistic experience: issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Bisol, L. (1999). *Introdução a estudos de fonologia do português brasileiro.* 2nd ed. Porto Alegre: EDIPUCRS.

Bohn, O.-S. & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*, 303–328.

Boersma, P. (1997). *The elements of functional phonology*. Available at http://roa.rutgers.edu.

Boersma, P. (1998). *Functional Phonology*. Doctoral Dissertation, University of Amsterdam. The Hague: Holland Academic Graphics.

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.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A. & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, *101*, 2299-2310.

Bradlow, A. R., Yamada, R. A., Pisoni, D. B. & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: long-term retention of learning in perception and production. *Perception & Psychophysics*, *61*, 977-985.

Callou, D. & Leite, Y. (1990). *Iniciação à fonética e à fonologia*. 4th ed. Rio de Janeiro: Jorge Zahar Editor.

Callou, D., Moraes, J. A. & Leite, Y. (1996). O vocalismo do português do Brasil, *Letras de Hoje, 31*, 27-40.

Câmara Jr., J. M. (1970). *Estrutura da língua portuguesa.* Petrópolis: Vozes.

Carney, A. E., Widin, G. P., Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *54*, 421-428.

Ceñoz Iragui, J. & Garcia Lecumberri, L. (1999). The effect of training on the discrimination of English vowels. *International Review of Applied Linguistics in Language Teaching*, *37*, 261-275.

Ceñoz Iragui, J. & Garcia Lecumberri, L. (2002). The acquisition of English vowels: a longitudinal study. In A. James & J. Leather (Eds.), *New Sounds 2000: Proceedings of the fourth international symposium on the acquisition of second language speech* (pp. 55-60). Klagenfurt, Austria: University of Klagenfurt.

Chiba, T. & Kajiyama, M. (1941). *The vowels: its nature and structure*. Tokyo: Kaiseikan.

Clopper, C. G., Pisoni, D. B. & Jong, K. (2005). Acoustic characteristics of the vowel systems of six regional varieties of American English. *Journal of the Acoustical Society of America*, *118*, 1661-1676.

Cristófaro Silva, T. (2001). *Fonética e fonologia do português.* 4th ed. São Paulo: Contexto.

Delgado Martins, M. R. (1973). Análise acústica das vogais tônicas em português, Centro de Estudos Filológicos (Lisbon), *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. *Proceedings of the 25th Annual Boston University Conference on Language Development* (pp. 250-261). Somerville, MA: Cascadilla Press.

Escudero, P. (2005). *Linguistic perception and second language acquisition*: *Explaining the attainment of optimal phonological categorization*. Utrecht: LOT.

Escudero, P. & Boersma, P. (2002). The subset problem in L2 perceptual development: Multiple-category assimilation by Dutch learners of Spanish. *Proceedings of the 26th*

Annual Boston University Conference on Language Development, vol. I (pp. 250-261). Somerville, MA: Cascadilla Press.

Escudero, P. & Boersma, P.. (2005). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*, 551-585.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mounton.

Faveri, C. B. (1991). Análise da duração das vogais orais do português de Florianópolis-Santa Catarina, Unpublished MA thesis, Universidade Federal de Santa Catarina.

Flege, J. E. (1987) 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. (1995a). Oddity discrimination of English vowels by non-natives. Unpublished manuscript. University of Alabama at Birmingham.

Flege, J. E. (1995b). 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. & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*, 1-34.

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, 106*, 1-15.

Flege, J. E., Munro, M. J. & Fox, R. A. (1994). Auditory and categorical effects on cross-language perception. *Journal of the Acoustical Society of America, 95*, 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.

Flege, J. E., Munro, M. J. & MacKay, I. R. A. (1996). Factors affecting the production of word-initial consonants in a second language. In. R. Bayley & D. R. Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 47-73). Amsterdam: John Benjamins.

Flege, J. E., Takagi, N. & Mann, V. (1995). Japanese adults can learn to produce English /ɹ/ and /l/ accurately. *Language and Speech*, *38*, 25-55.

Fox, M. M. & Maeda, K. (1999). Categorization of American English vowels by Japanese speakers. *Proceedings of the International Conference of Phonetic Sciences* (pp. 1437-1440). San Francisco.

Frieda, E. M., Walley, A. C., Flege, J.E. & Sloane, M.E. (2000). Adults' perception and production of the English vowel /i/. *Journal of Speech, Language, and Hearing Research*, *43*, 129-143.

Garcia Lecumberri, L. & Ceñoz Iragui, J. (1997). Identification by L2 learners of English vowels in different phonetic contexts. In J. Leather & A. James (Eds.) *New Sounds 97: Proceedings of the third international symposium on the acquisition of second language speech* (pp. 196-205), Klagenfurt, Austria: University of Klagenfurt.

Giegerich, H. J. (2005). *English phonology: an introduction*. 8th ed. Cambridge Textbooks in Linguistics. Cambridge: Cambridge University Press.

Gottfried, T. L. & Strange, W. (1980). Identification of coarticulated vowels. *Journal of the Acoustical Society of America, 68*, 1626-1635.

Hammond, M. (1999). *The phonology of English: a prosodic optimality-theoretic approach*. Oxford Linguistics. Oxford: Oxford University Press.

Hardison, D. (2002). Transfer of second-language perceptual training to production improvement: focus on /r/ and /l/. *New Sounds 2000: Proceedings of the fourth international symposium on the acquisition of second language speech* (pp. 166-173), Klagenfurt, Austria: University of Klagenfurt.

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24*, 429-522.

Harrington, J. & Cassidy, S. (1999). *Techniques in Speech Acoustics*. Dordrecht, the Netherlands: Kluwer Academic Publishers.

Hawkey, D. J. C., Amitay, S. & Moore, D. R. (2004). Early and rapid perceptual learning. *Nature Neuroscience, 7*, 1055-1056.

Hayward, K. (2000). *Experimental Phonetics*. Essex: Pearson Education.

Hazan, V., Sennema, A., Iba, M. & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication, 47*, 360-378.

Hazan, V. & Simpson, A. (2000). The effect of cue-enhancement on consonant intelligibility in noise: speaker and listener effects. *Language and Speech*, *43*, 273-294.

Jamieson, D.G. & Morosan, D.E. (1986). Training non-native speech contrasts in adults: acquisition of the English /θ/-/ð/ contrast by francophones. *Perception & Psychophysics, 40*, 205-215.

Jenkins, J. J. (1987). A selective history of issues in vowel perception. *Journal of Memory and Language, 26*, 542-549.

Johnson, K. (2003). *Acoustic and auditory phonetics,* 2nd ed. Oxford: Blackwell Publishing.

Kuhl, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93-107.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics, 21,* 125-139.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97*, 11850-11857.

Kuhl, P. K, Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U. & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*, 684-686.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N. & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606-608.

Ladefoged, P. (2003). *Phonetic data analysis: an introduction to fieldwork and instrumental techniques*. Oxford: Blackwell Publishing.

Ladefoged, P. (2005). *Vowels and consonants: an introduction to the sounds of languages*, 2nd ed. Oxford: Blackwell Publishing.

Lane, H. L. & Moore, D. J. (1962). Reconditioning a consonant discrimination in an aphasic: An experimental case history. *Journal of Speech and Hearing Disorders, 27*, 232-243.

Lima, R. (1991). *Análise acústica das vogais orais do português de Florianópolis-Santa Catarina*, Unpublished MA thesis, Universidade Federal de Santa Catarina.

Logan, J. S. & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic* *experience: issues in cross-language research* (pp. 351-377). Timonium, MD: York Press.

Major, R. C. (1987a). A model for interlanguage phonology. In G. Ioup & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 101-124). Cambridge, MA: Newbury House Publishers.

Major, R. C. (1987b). Phonological similarity, markedness, and rate of L2 acquisition. *Studies in Second Language Acquisition, 9*, 63-82.

Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition, 50*, 271-285.

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, Canada: Simon Fraser University Linguistics Graduate Student Association.

Munro, M. J., Flege, J. E. & MacKay, I. R. A. (1996). The effects of second language learning on the production of English vowels. *Applied Linguistics, 17*, 313-334.

Nobre-Oliveira, D. (2005). *Perceptual training and the learning of English vowels by Brazilian EFL learners.* Unpublished paper, Universidade Federal de Santa Catarina.

Ohnishi, M. (1991). *A spectrographic investigation of the vowels of Californian English (Southwest General American).* Paper presented in the 1991 Convention of the Phonetic Society of Japan. Osaka, Japan.

Ortega-Llebaria, M., Faulkner, A. & Hazan, V. (2001). Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. *Speech, Hearing and Language: Work in progress, 13,* 39-51.

Ortega-Llebaria, M. & Hazan, V. (1999). Enhancing acoustic cues to aid L2 speech perception. *Proceedings of the International Congress of Phonetic Sciences,* San Francisco, 1-7 August 1999, 1, 117-120.

Pereira, A. L. D. (2001). *Caracterização acústica do sistema vocálico tônico oral florianopolitano: alguns indícios de mudança*, Unpublished MA thesis, Universidade Federal de Santa Catarina.

Peterson, G. E. (1952). The information-bearing elements of speech. *Journal of the Acoustical Society of America, 24*, 629-637.

Peterson, G. E. (1961). Parameters of vowel quality. *Journal of Speech and Hearing Research, 4*, 10-29.

Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24* (2), 175-184.

Peterson, G. E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America*, *32*, 693-703.

Pickett, J. M. (1999). *The Acoustics of Speech Communication: Fundamentals, Speech Perception Theory, and Technology*. Needham Heights, MA: Allyn & Bacon.

Pisoni, D. B., Lively, S. E. & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.), *The Transition from Speech Sounds to Spoken Words* (pp. 121-166). Cambridge, MA: MIT Press.

Pruitt, J. S., Jenkins, J. J. & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *Journal of the Acoustical Society of America*, *119*, 1684-1696.

Rauber, A. S. (2006). *Perception and production of English vowels by Brazilian EFL speakers*. Unpublished doctoral dissertation. Universidade Federal de Santa Catarina.

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.

Rost, M. (1990). *Listening in language learning*. New York: Longman.

Silveira, R. (2004). *The influence of pronunciation instruction on the perception and production of English word-final consonants.* Unpublished doctoral dissertation. Universidade Federal de Santa Catarina.

Stevens, K. (1997). Articulatory-acoustic-auditory relationships. In W. J. Hardcastle and J. Laver, *The Handbook of Phonetic Sciences* (pp.462-506). Oxford: Blackwell Publishers.

Strange, W. (1999). Perception of vowels: Dynamic constancy. In J. M. Pickett, *The acoustics of speech communication: Fundamentals, speech perception theory, and technology* (pp.153-165). Needham Heights, MA: Allyn & Bacon.

Strange, W. & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36* (2), 131-145.

Terrace, H. S. (1963). Discrimination learning with and without "errors". *Journal of Experimental Analysis of Behavior, 6,* 1-27.

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: long-term retention and effects on production*. Unpublished doctoral dissertation. Simon Fraser University.

Wang, X. & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *34*, 539-552.

Wang, Y., Spence, M. M., Jongman, A. & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649-3658.

Wilson, S. M., Saygin, A. P., Sereno, M. I. & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience, 7*, 701-702.

Yamada, R. A., Tohkura, Y., Bradlow, A. R. & Pisoni, D. B. (1996). Does training in speech perception modify speech production? Proceedings of the fourth International Conference on Spoken Language Processing, v. 2 (pp.606-609). Wyndham Franklin Plaza Hotel, Pennsylvania. Philadelphia: Institute of Electrical and Electronics Engineers.

APPENDIXES

APPENDIX A

BACKGROUND QUESTIONNAIRE

Date: _____/_____/_____

Name: ____

E-mail: ____

Age: ________________ Place and date of birth: ____________________

1) Write the names of the cities and countries you've been to or lived at for more than 1 month since you were born:

City and country:___________________, how long you stayed there: __________________

City and country:___________________, how long you stayed there: __________________

City and country:___________________, how long you stayed there: __________________

2) Where were your parents born? Write the city and state.

a) Mother: ______________________ b) Father: ________________________

3) Are you studying any foreign languages (other than English)? (If not, go to question 6.)

Write the language and the level (beginner, intermediate or advanced):

Language: ________________, Level: _____________

Language: ________________, Level: _____________

Language: ________________, Level: _____________

4) Where do you study these languages? (extracurricular course, private classes, etc.)

Language: ________________, Place: ____

Language: ________________, Place: ____

Language: ________________, Place: ____

5) How many hours per week do you study these languages?

Language: ________________, Hours per week: _____________

Language: ________________, Hours per week: _____________

Language: ________________, Hours per week: _____________

6) Have you ever studied other languages before? _____________ (If not, go to question 10.)

Which language? ____

7) How old were you when you started studying other languages?

Language: ________________, Age: _____________

Language: ________________, Age: _____________

Language: ________________, Age: _____________

8) How long did you study other languages?

Language: ________________, Years: _____________

Language: ________________, Years: _____________

Language: ________________, Years: _____________

9) If you studied in a language school, up to which level did you study? _______________

10) Circle the number corresponding to your listening comprehension level of the language(s) you study or studied. (0 means you don't understand anything; 7 means you understand everything.)

Language: ________________,	1	2	3	4	5	6	7
Language: ________________,	1	2	3	4	5	6	7
Language: ________________,	1	2	3	4	5	6	7

11) Circle the number corresponding to how much you can speak in the language(s) you study or studied. (0 means you can't speak anything; 7 means you speak fluently, native-like.)

Language: ________________,	1	2	3	4	5	6	7
Language: ________________,	1	2	3	4	5	6	7
Language: ________________,	1	2	3	4	5	6	7

12) How much time do you speak other language (when you're not in your language class)?

Language: ________________, Hours and minutes per week: _____________

Language: ________________, Hours and minutes per week: _____________

Language: ________________, Hours and minutes per week: _____________

APPENDIX B

PRODUCTION TEST SAMPLE AND CORRESPONDING VOWEL PHONEME

/i/:	(bee)	Beat. Pete. Seat.	Beat and Pete sound like seat.
/ɑ/:	(dog)	Bot. Pot. Sot.	Bot and pot sound like sot.
/ɛ/:	(egg)	Tet. Tech. Kept.	Tet and tech sound like kept.
/ʊ/:	(book)	Book. Put. Soot.	Book and put sound like soot.
/æ/:	(cat)	Tat. Tack. Cat.	Tat and tack sound like cat.
/ɑ/:	(dog)	Tot. Tock. Cot.	Tot and tock sound like cot.
/ɔ/:	(orange)	Bought. Ought. Sought.	Bought and ought sound like sought.
/æ/:	(cat)	Bat. Pat. Sat.	Bat and pat sound like sat.
/ʊ/:	(book)	Book. Took. Cook.	Book and took sound like cook.
/ɛ/:	(egg)	Bet. Pet. Set.	Bet and pet sound like set.
/ɪ/:	(ship)	Bit. Pitt. Sit.	Bit and Pitt sound like sit.
/ʌ/:	(cup)	But. Putt. Shut.	But and putt sound like shut.
/ɪ/:	(ship)	Tit. Tick. Kit.	Tit and tick sound like kit.
/u/:	(boot)	Boot. Poop. Suit.	Boot and poop sound like suit.
/ɔ/:	(orange)	Taught. Talk. Caught.	Taught and talk sound like caught.
/ʌ/:	(cup)	Tut. Tuck. Cut.	Tut and tuck sound like cut.
/i/:	(bee)	Teat. Teak. Keep.	Teat and teak sound like keep.
/u/:	(boot)	Toot. Tuke. Coot.	Toot and tuke sound like coot.

APPENDIX C

SCRIPT TO CALCULATE THE MEAN AND THE MEDIAN OF THE FORMANTS AND DURATION

```
# Ricardo Bion

clearinfo

for formant from 1 to 5

select all
table1 = selected("Table")
Pool... vowel "" "F4 F5 F1 F2 F3" "" ""
nrows = Get number of rows
table2 = selected("Table")

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

select all
minus table1
Remove

printline 'newline$'
endfor
```

```
#Ricardo Bion

#clearinfo

for formant from 1 to 5

select all
table1 = selected("Table")
Pool... vowel "" "" "F4 F5 F1 F2 F3" ""
nrows = Get number of rows
table2 = selected("Table")

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

select all
minus table1
Remove

printline 'newline$'
endfor
```

```
# This script will calculate the durations of all labeled segments in
a TextGrid object.
# The results will be save in a text file, each line containing the
label text and the
# duration of the corresponding segment..
# A TextGrid object needs to be selected in the Object list.
#
# This script is distributed under the GNU General Public License.
# Copyright 12.3.2002 Mietta Lennes

# ask the user for the tier number
form Calculate durations of labeled segments
	comment Which tier of the TextGrid object would you like to
analyse?
	integer Tier 1
	comment Where do you want to save the results?
	text textfile durations.txt
endform

# check how many intervals there are in the selected tier:
numberOfIntervals = Get number of intervals... tier

# loop through all the intervals
for interval from 1 to numberOfIntervals
	label$ = Get label of interval... tier interval
	# if the interval has some text as a label, then calculate the
duration.
	if label$ <> ""
		start = Get starting point... tier interval
		end = Get end point... tier interval
		duration = end - start
		# append the label and the duration to the end of the
text file, separated with a tab:
		resultline$ = "'label$' 'duration''newline$'"
		fileappend "'textfile$'" 'resultline$'
	endif
endfor
```

APPENDIX D

SCRIPT TO CALCULATE THE EUCLIDEAN DISTANCE IN THE PRODUCTION TEST

```
# Ricardo Bion, August 20, 2006
# this scripts creates perceptual vowel plots

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

   finename$ = Get stimulus... trial_c
   response$ = Get response... trial_c
   p$ = finename$ - ".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$'%
iI= sqrt(((f13 - f18)^2)+((f23 - f28)^2))
percent = 100*iI/235
printline i-I'tab$''iI: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
endfor

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 maaaaybe 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$ = "\ef"
elsif resp$ = "O"
   label$ = "\ct"
elsif resp$ = "a"
   label$ = "\ae"
elsif resp$ = "e"
   label$ = "e"
elsif resp$ = "i"
   label$ = "i"
elsif resp$ = "o"
   label$ = "o"
elsif resp$ = "u"
   label$ = "u"
elsif resp$ = "A"
   label$ = "\as"
elsif resp$ = "I"
   label$ = "\ic"
elsif resp$ = "U"
   label$ = "\hs"
elsif resp$ = "v"
   label$ = "\vt"
else
label$ = resp$
endif
endproc

procedure formantvalues
```

```
if finename$ = "1_1_1.wav"
  f1= 239.99999999999997
  f2= 580.0000000000001
elsif 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 E

PLOT GENERATOR SCRIPT

```
# Ricardo Bion
# November, 2006

clearinfo

form PARTICIPANT
comment put 0 for all participants
integer plot_participant: 0
integer max_F2: 3500
integer min_F2: 700
integer max_F1: 1000
integer min_F1: 200
endform



# L2 vowels
#Read Table from table file... C:\a\Male_L2 speakers_English\All 5 L2
male speakers.txt




############################################################
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... 2700 yes yes no
One logarithmic mark bottom... 3500 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

#One logarithmic mark right... 909 yes yes yes
#One logarithmic mark top... 1100 yes yes yes
#One logarithmic mark right... 273 yes yes yes
```

```
#One logarithmic mark top... 2883 yes yes yes

Draw inner box


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

#####################################################

select all

tempt = selected("Table")

Copy... temp
Formula... F1 log10(self)
Formula... F2 log10(self)


if plot_participant > 0
Extract rows where column (number)... speaker "equal to"
'plot_participant'
endif

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


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


for formant from 1 to 2
f'formant'_em_Hz = Get value... i F'formant'
endfor

select table1
Extract rows where column (text)... vowel "is equal to" 'label$'

for formant from 1 to 2
sd_F'formant'_em_Hz  = Get standard deviation... F'formant'
endfor

call plot

endfor

select all
minus tempt
Remove


procedure plot
```

```
f1 = f1_em_Hz
f2 = f2_em_Hz
stdev_f2 = sd_F2_em_Hz
stdev_f1 = sd_F1_em_Hz

'color_of_the_vowel$'
Text special... 'f2' Centre 'f1' Half Times 24 0 'label$'
Plain line
Line width... 1

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'
endproc
```

APPENDIX F

RESULTS PER VOWEL CONTRAST IN THE PERCEPTION TEST FOR THE EXPERIMENTAL AND THE CONTROL GROUPS

Results of the experimental groups in the perception pretest and improvement rate per vowel contrast.

		Perception pretest	Perception posttest	Improvement
/i-ɪ/	total = 24 (100%)	16.7 (69.6%)	18.9 (78.8%)	2.2 (9.2 p.p.)*
/ɛ-æ/	total = 24 (100%)	14.7 (61.4%)	16 (66.8%)	1.3 (5.4 p.p.)
/ʊ-u/	total = 22 (100%)	11.1 (50%)	15.7 (71.2%)	4.6 (21.2 p.p.)

* p.p. = percentage points

Results of the BP control group in the perception pretest and improvement rate per vowel contrast.

		Perception pretest	Perception posttest	Improvement
/i-ɪ/	total = 24 (100%)	16.8 (70.2%)	17.4 (72.6%)	0.6 (2.4 p.p.)*
/ɛ-æ/	total = 24 (100%)	16.4 (68.4%)	17.7 (73.8%)	1.3 (5.4 p.p)
/ʊ-u/	total = 22 (100%)	9.7 (44.1%)	11.1 (50.6%)	1.4 (6.5 p.p.)

* p.p. = percentage points

APPENDIX G

MEAN INDIVIDUAL EUCLIDEAN DISTANCE OF L2 PAIRS PER TEST

BP CONTROL GROUP (PRETEST):

Participant	/i-ɪ/	/ɛ-æ/	/u-ʊ/
CP1	35	19	52
CP2	15	8	7
CP3	9	8	18
CP4	4	9	10
CP5	40	9	30
CP6	10	9	70
CP7	10	20	53

BP CONTROL GROUP (POSTTEST):

Participant	/i-ɪ/	/ɛ-æ/	/u-ʊ/
CP1	26	8	42
CP2	13	4	6
CP3	7	4	5
CP4	13	10	5
CP5	21	5	12
CP6	8	6	15
CP7	24	4	23

EXPERIMENTAL GROUPS (PRETEST):

Participant	Training Group	/i-ɪ/	/ɛ-æ/	/u-ʊ/
E1	SynS	6	3	23
E2	SynS	17	14	17
E3	SynS	8	8	18
E4	SynS	24	3	10
E5	SynS	1	5	22
E6	SynS	46	7	20
E9	SynS	15	7	3
E10	SynS	15	13	12
E11	SynS	17	7	12
E12	SynS	8	4	47
E13	SynS	5	10	6
E14	SynS	15	13	15
E15	SynS	11	6	28
E16	NatS	3	3	23
E17	NatS	14	17	3
E19	NatS	55	5	17
E21	NatS	3	6	54
E22	NatS	8	14	24
E24	NatS	14	5	13
E25	NatS	7	11	17
E27	NatS	16	5	23
E28	NatS	21	9	65
E29	NatS	10	14	5

EXPERIMENTAL GROUPS (POSTTEST):

Part	Training Group	/i-ɪ/	/ɛ-æ/	/u-ʊ/
E1	SynS	4	17	26
E2	SynS	16	15	53
E3	SynS	32	8	43
E4	SynS	15	9	24
E5	SynS	8	14	21
E6	SynS	58	2	33
E9	SynS	52	7	24
E10	SynS	34	34	18
E11	SynS	14	15	22
E12	SynS	43	14	16
E13	SynS	11	5	7
E14	SynS	11	13	24
E15	SynS	19	19	45
E16	NatS	14	22	35
E17	NatS	8	4	21
E19	NatS	61	15	26
E21	NatS	3	6	18
E22	NatS	26	9	18
E24	NatS	19	7	10
E25	NatS	5	3	16
E27	NatS	35	6	22
E28	NatS	18	5	11
E29	NatS	20	11	3

EXPERIMENTAL GROUPS (FOLLOW-UP):

Part	Training Group	/i-ɪ/	/ɛ-æ/	/u-ʊ/
E1	SynS	3	11	50
E2	SynS	2	14	43
E3	SynS	42	14	39
E4	SynS	52	2	14
E5	SynS	7	8	17
E6	SynS	60	8	48
E9	SynS	47	6	28
E10	SynS	30	40	40
E11	SynS	25	6	38
E12	SynS	24	20	3
E13	SynS	20	2	11
E14	SynS	21	14	57
E15	SynS	8	14	23
E16	NatS	11	10	11
E17	NatS	2	1	22
E19	NatS	52	8	45
E21	NatS	12	23	13
E22	NatS	22	9	33
E24	NatS	18	11	9
E25	NatS	31	8	27
E27	NatS	33	12	51
E28	NatS	11	15	28
E29	NatS	11	15	31

APPENDIX I

PERCEPTION TEST – ANSWER SHEET

UFSC/CCE/LLE
Name:____ Do you have any hearing problems: Yes No

PERCEPTION TEST

Instructions: Circle the word that has the same vowel sound as the word you hear.

Example: You will hear: "feet"

01 (sheep) ship – bed – bad – cut – hot – talk – foot – boot

You should circle the word "sheep" because it contains the same vowel sound as "feet".

1. sheep – ship – bed – bad – cut – hot – talk – foot – boot	19. sheep – ship – bed – bad – cut – hot – talk – foot – boot	37. sheep – ship – bed – bad – cut – hot – talk – foot – boot
2. sheep – ship – bed – bad – cut – hot – talk – foot – boot	20. sheep – ship – bed – bad – cut – hot – talk – foot – boot	38. sheep – ship – bed – bad – cut – hot – talk – foot – boot
3. sheep – ship – bed – bad – cut – hot – talk – foot – boot	21. sheep – ship – bed – bad – cut – hot – talk – foot – boot	39. sheep – ship – bed – bad – cut – hot – talk – foot – boot
4. sheep – ship – bed – bad – cut – hot – talk – foot – boot	22. sheep – ship – bed – bad – cut – hot – talk – foot – boot	40. sheep – ship – bed – bad – cut – hot – talk – foot – boot
5. sheep – ship – bed – bad – cut – hot – talk – foot – boot	23. sheep – ship – bed – bad – cut – hot – talk – foot – boot	41. sheep – ship – bed – bad – cut – hot – talk – foot – boot
6. sheep – ship – bed – bad – cut – hot – talk – foot – boot	24. sheep – ship – bed – bad – cut – hot – talk – foot – boot	42. sheep – ship – bed – bad – cut – hot – talk – foot – boot
7. sheep – ship – bed – bad – cut – hot – talk – foot – boot	25. sheep – ship – bed – bad – cut – hot – talk – foot – boot	43. sheep – ship – bed – bad – cut – hot – talk – foot – boot
8. sheep – ship – bed – bad – cut – hot – talk – foot – boot	26. sheep – ship – bed – bad – cut – hot – talk – foot – boot	44. sheep – ship – bed – bad – cut – hot – talk – foot – boot
9. sheep – ship – bed – bad – cut – hot – talk – foot – boot	27. sheep – ship – bed – bad – cut – hot – talk – foot – boot	45. sheep – ship – bed – bad – cut – hot – talk – foot – boot
10. sheep – ship – bed – bad – cut – hot – talk – foot – boot	28. sheep – ship – bed – bad – cut – hot – talk – foot – boot	46. sheep – ship – bed – bad – cut – hot – talk – foot – boot
11. sheep – ship – bed – bad – cut – hot – talk – foot – boot	29. sheep – ship – bed – bad – cut – hot – talk – foot – boot	47. sheep – ship – bed – bad – cut – hot – talk – foot – boot
12. sheep – ship – bed – bad – cut – hot – talk – foot – boot	30. sheep – ship – bed – bad – cut – hot – talk – foot – boot	48. sheep – ship – bed – bad – cut – hot – talk – foot – boot
13. sheep – ship – bed – bad – cut – hot – talk – foot – boot	31. sheep – ship – bed – bad – cut – hot – talk – foot – boot	49. sheep – ship – bed – bad – cut – hot – talk – foot – boot
14. sheep – ship – bed – bad – cut – hot – talk – foot – boot	32. sheep – ship – bed – bad – cut – hot – talk – foot – boot	50. sheep – ship – bed – bad – cut – hot – talk – foot – boot
15. sheep – ship – bed – bad – cut – hot – talk – foot – boot	33. sheep – ship – bed – bad – cut – hot – talk – foot – boot	51. sheep – ship – bed – bad – cut – hot – talk – foot – boot
16. sheep – ship – bed – bad – cut – hot – talk – foot – boot	34. sheep – ship – bed – bad – cut – hot – talk – foot – boot	52. sheep – ship – bed – bad – cut – hot – talk – foot – boot
17. sheep – ship – bed – bad – cut – hot – talk – foot – boot	35. sheep – ship – bed – bad – cut – hot – talk – foot – boot	53. sheep – ship – bed – bad – cut – hot – talk – foot – boot
18. sheep – ship – bed – bad – cut – hot – talk – foot – boot	36. sheep – ship – bed – bad – cut – hot – talk – foot – boot	54. sheep – ship – bed – bad – cut – hot – talk – foot – boot

55. sheep – ship – bed – bad – cut – hot – talk – foot – boot
56. sheep – ship – bed – bad – cut – hot – talk – foot – boot
57. sheep – ship – bed – bad – cut – hot – talk – foot – boot
58. sheep – ship – bed – bad – cut – hot – talk – foot – boot
59. sheep – ship – bed – bad – cut – hot – talk – foot – boot
60. sheep – ship – bed – bad – cut – hot – talk – foot – boot
61. sheep – ship – bed – bad – cut – hot – talk – foot – boot
62. sheep – ship – bed – bad – cut – hot – talk – foot – boot
63. sheep – ship – bed – bad – cut – hot – talk – foot – boot
64. sheep – ship – bed – bad – cut – hot – talk – foot – boot
65. sheep – ship – bed – bad – cut – hot – talk – foot – boot
66. sheep – ship – bed – bad – cut – hot – talk – foot – boot
67. sheep – ship – bed – bad – cut – hot – talk – foot – boot
68. sheep – ship – bed – bad – cut – hot – talk – foot – boot
69. sheep – ship – bed – bad – cut – hot – talk – foot – boot
70. sheep – ship – bed – bad – cut – hot – talk – foot – boot
71. sheep – ship – bed – bad – cut – hot – talk – foot – boot
72. sheep – ship – bed – bad – cut – hot – talk – foot – boot
73. sheep – ship – bed – bad – cut – hot – talk – foot – boot
74. sheep – ship – bed – bad – cut – hot – talk – foot – boot
75. sheep – ship – bed – bad – cut – hot – talk – foot – boot
76. sheep – ship – bed – bad – cut – hot – talk – foot – boot
77. sheep – ship – bed – bad – cut – hot – talk – foot – boot
78. sheep – ship – bed – bad – cut – hot – talk – foot – boot
79. sheep – ship – bed – bad – cut – hot – talk – foot – boot
80. sheep – ship – bed – bad – cut – hot – talk – foot – boot
81. sheep – ship – bed – bad – cut – hot – talk – foot – boot
82. sheep – ship – bed – bad – cut – hot – talk – foot – boot
83. sheep – ship – bed – bad – cut – hot – talk – foot – boot
84. sheep – ship – bed – bad – cut – hot – talk – foot – boot
85. sheep – ship – bed – bad – cut – hot – talk – foot – boot
86. sheep – ship – bed – bad – cut – hot – talk – foot – boot
87. sheep – ship – bed – bad – cut – hot – talk – foot – boot
88. sheep – ship – bed – bad – cut – hot – talk – foot – boot
89. sheep – ship – bed – bad – cut – hot – talk – foot – boot
90. sheep – ship – bed – bad – cut – hot – talk – foot – boot
91. sheep – ship – bed – bad – cut – hot – talk – foot – boot
92. sheep – ship – bed – bad – cut – hot – talk – foot – boot
93. sheep – ship – bed – bad – cut – hot – talk – foot – boot
94. sheep – ship – bed – bad – cut – hot – talk – foot – boot
95. sheep – ship – bed – bad – cut – hot – talk – foot – boot
96. sheep – ship – bed – bad – cut – hot – talk – foot – boot
97. sheep – ship – bed – bad – cut – hot – talk – foot – boot
98. sheep – ship – bed – bad – cut – hot – talk – foot – boot
99. sheep – ship – bed – bad – cut – hot – talk – foot – boot
100. sheep – ship – bed – bad – cut – hot – talk – foot – boot
101. sheep – ship – bed – bad – cut – hot – talk – foot – boot
102. sheep – ship – bed – bad – cut – hot – talk – foot – boot
103. sheep – ship – bed – bad – cut – hot – talk – foot – boot
104. sheep – ship – bed – bad – cut – hot – talk – foot – boot
105. sheep – ship – bed – bad – cut – hot – talk – foot – boot
106. sheep – ship – bed – bad – cut – hot – talk – foot – boot
107. sheep – ship – bed – bad – cut – hot – talk – foot – boot
108. sheep – ship – bed – bad – cut – hot – talk – foot – boot

APPENDIX J

SCRIPT TO RANDOMIZE SOUND FILES

```
"ooTextFile"
"ExperimentMFC 2"
fileNameHead = "C:\TestSoundFiles\"
fileNameTail = ""
carrierBefore = ""
carrierAfter = ""
initialSilenceDuration = 1.0 seconds
interStimulusInterval = 4.0
numberOfDifferentStimuli = 108
"ae_bat_F9.wav"
"ae_bat_M6.wav"
"ae_cat_F9.wav"
"ae_cat_M6.wav"
"ae_pat_F9.wav"
"ae_pat_M6.wav"
"ae_sat_F9.wav"
"ae_sat_M6.wav"
"ae_tack_F9.wav"
"ae_tack_M6.wav"
"ae_tat_F9.wav"
"ae_tat_M6.wav"
"a_bot_F7.wav"
"a_bot_M1.wav"
"a_bought_F7.wav"
"a_bought_M1.wav"
"a_caught_F7.wav"
"a_caught_M1.wav"
"a_cot_F7.wav"
"a_cot_M1.wav"
"a_pot_F7.wav"
"a_pot_M1.wav"
"a_sot_F7.wav"
"a_sot_M1.wav"
"a_sought1_F7.wav"
"a_sought1_M1.wav"
"a_sought2_F7.wav"
"a_sought2_M1.wav"
"a_talk_F7.wav"
"a_talk_M1.wav"
"a_taught_F7.wav"
"a_taught_M1.wav"
"a_tock_F7.wav"
"a_tock_M1.wav"
"a_tot_F7.wav"
"a_tot_M1.wav"
"E_bet_F9.wav"
"E_bet_M6.wav"
"E_kept_F9.wav"
"E_kept_M6.wav"
```

```
"E_pet_F9.wav"
"E_pet_M6.wav"
"E_set_F9.wav"
"E_set_M6.wav"
"E_tech_F9.wav"
"E_tech_M6.wav"
"E_tet_F9.wav"
"E_tet_M6.wav"
"IC_bit_F9.wav"
"IC_bit_M1.wav"
"IC_kit_F9.wav"
"IC_kit_M1.wav"
"IC_pitt_F9.wav"
"IC_pitt_M1.wav"
"IC_sit_F9.wav"
"IC_sit_M1.wav"
"IC_tick_F9.wav"
"IC_tick_M1.wav"
"IC_tit_F9.wav"
"IC_tit_M1.wav"
"i_beat_F9.wav"
"i_beat_M6.wav"
"i_keep_F9.wav"
"i_keep_M6.wav"
"i_pete_F9.wav"
"i_pete_M6.wav"
"i_seat_F9.wav"
"i_seat_M6.wav"
"i_teak_F9.wav"
"i_teak_M6.wav"
"i_teat_F9.wav"
"i_teat_M6.wav"
"UC_book1_F9.wav"
"UC_book1_M6.wav"
"UC_book2_F9.wav"
"UC_book2_M6.wav"
"UC_cook_F9.wav"
"UC_cook_M6.wav"
"UC_put_F9.wav"
"UC_put_M6.wav"
"UC_soot_F9.wav"
"UC_soot_M6.wav"
"UC_took_F9.wav"
"UC_took_M6.wav"
"u_boot_F9.wav"
"u_boot_M6.wav"
"u_coot_F9.wav"
"u_coot_M6.wav"
"u_poop_F9.wav"
"u_poop_M6.wav"
"u_suit_F9.wav"
"u_suit_M6.wav"
"u_toot_F9.wav"
"u_toot_M6.wav"
"u_tuke_F9.wav"
"u_tuke_M6.wav"
"vt_but_F9.wav"
"vt_but_M6.wav"
"vt_cut_F9.wav"
"vt_cut_M6.wav"
"vt_putt_F9.wav"
```

```
"vt_putt_M6.wav"
"vt_shut_F9.wav"
"vt_shut_M6.wav"
"vt_tuck_F9.wav"
"vt_tuck_M6.wav"
"vt_tut_F9.wav"
"vt_tut_M6.wav"

numberOfReplicationsPerStimulus = 1
breakAfterEvery = 0
randomize = <PermuteBalancedNoDoublets>
numberOfResponseCategories = 9
      0.25 0.35 0.8 0.9 "sheep" "i"
      0.25 0.35 0.6 0.7 "ship" "ic"
      0.25 0.35 0.4 0.5 "bed" "ep"
      0.25 0.35 0.2 0.3 "bad" "ae"
      0.45 0.55 0.6 0.7 "cut" "vt"
      0.65 0.75 0.2 0.3 "hot" "as"
      0.65 0.75 0.6 0.7 "foot" "hs"
      0.65 0.75 0.4 0.5 "talk" "ct"
      0.65 0.75 0.8 0.9 "boot" "u"
```

APPENDIX K

SCRIPT TO GENERATE SYNTHESIZED STIMULI

```
# Artificial vowel generator by repetitive multiplication of spectrum
pulse train with 2nd order 'formant' spectra
form Artificial Vowel Generator (multiplicative filtering)
      positive duration_(sec) 0.5
      positive initial_F0_(Hz) 150
      positive final_F0_(Hz) 80
      positive F1_(Hz) 820
      positive F2_(Hz) 1300
      positive F3_(Hz) 2300
      real formantfrequency/bandwidth 12
      word name vowel
endform

d = 'duration'
sweep = 'final_F0' - 'initial_F0'
slope = sweep / d
q = 'formantfrequency/bandwidth'

# create initial pulse period
per1 = 1/'initial_F0'
Create Sound... pulse 0 1/22050 44100 1
Create Sound... zero 0 'per1'-1/22050 44100 0
plus Sound pulse
Concatenate
Rename... pulsetrain

# make pulse train
curtime = 0
while curtime < d + 1/'initial_F0'
     curperiod = 1/('initial_F0' + curtime * slope) - 1/22050
     Create Sound... null 0 'curperiod' 44100 0
     plus Sound pulse
     Concatenate
     Rename... period
     plus Sound pulsetrain
     Concatenate
     select Sound null
     plus Sound pulsetrain
     plus Sound period
     Remove
     select Sound chain
     Rename... pulsetrain

     curtime = curtime + curperiod + 1/22050
endwhile

select Sound pulsetrain
Rename... 'name$'
```

```
b = f1 / q
Filter (one formant)... 'f1' 'b'

b = f2 / q
Filter (one formant)... 'f2' 'b'

b = f3 / q
Filter (one formant)... 'f3' 'b'
```

APPENDIX L

ANSWER SHEET FOR TRAINING IN CLASS

CLASS 1

UFSC/CCE/LLE
Name:____

Perception Practice 1 – High front vowels

Instructions: Circle the best option according to the vowel sound in the word you hear.

> *Example:* You will hear "feet"
> 01. (/i/) – other
> You should circle "/i/" because it sounds the same as the vowel in the word "feet".

Part 1:

1.	/i/	-	other	11.	/i/	-	other
2.	/i/	-	other	12.	/i/	-	other
3.	/i/	-	other	13.	/i/	-	other
4.	/i/	-	other	14.	/i/	-	other
5.	/i/	-	other	15.	/i/	-	other
6.	/i/	-	other	16.	/i/	-	other
7.	/i/	-	other	17.	/i/	-	other
8.	/i/	-	other	18.	/i/	-	other
9.	/i/	-	other	19.	/i/	-	other
10.	/i/	-	other	20.	/i/	-	other

Part 2:

21.	/i/	-	/ɪ/	36.	/i/	-	/ɪ/
22.	/i/	-	/ɪ/	37.	/i/	-	/ɪ/
23.	/i/	-	/ɪ/	38.	/i/	-	/ɪ/
24.	/i/	-	/ɪ/	39.	/i/	-	/ɪ/
25.	/i/	-	/ɪ/	40.	/i/	-	/ɪ/
26.	/i/	-	/ɪ/	41.	/i/	-	/ɪ/
27.	/i/	-	/ɪ/	42.	/i/	-	/ɪ/
28.	/i/	-	/ɪ/	43.	/i/	-	/ɪ/
29.	/i/	-	/ɪ/	44.	/i/	-	/ɪ/
30.	/i/	-	/ɪ/	45.	/i/	-	/ɪ/
31.	/i/	-	/ɪ/	46.	/i/	-	/ɪ/
32.	/i/	-	/ɪ/	47.	/i/	-	/ɪ/
33.	/i/	-	/ɪ/	48.	/i/	-	/ɪ/
34.	/i/	-	/ɪ/	49.	/i/	-	/ɪ/
35.	/i/	-	/ɪ/	50.	/i/	-	/ɪ/

CLASS 2

UFSC/CCE/LLE
Name:____
NS

Perception Practice 2 – Mid and low front vowels

Instructions: Circle the best option according to the vowel sound in the word you hear.

> *Example:* You will hear "fed"
> 01. (/ɛ/) – other
> You should circle "/ɛ/" because it sounds the same as the vowel in the word "fed".

Part 1:

1.	/ɛ/	-	other	11.	/ɛ/	-	other
2.	/ɛ/	-	other	12.	/ɛ/	-	other
3.	/ɛ/	-	other	13.	/ɛ/	-	other
4.	/ɛ/	-	other	14.	/ɛ/	-	other
5.	/ɛ/	-	other	15.	/ɛ/	-	other
6.	/ɛ/	-	other	16.	/ɛ/	-	other
7.	/ɛ/	-	other	17.	/ɛ/	-	other
8.	/ɛ/	-	other	18.	/ɛ/	-	other
9.	/ɛ/	-	other	19.	/ɛ/	-	other
10.	/ɛ/	-	other	20.	/ɛ/	-	other

Part 2:

21.	/ɛ/	-	/æ/	36.	/ɛ/	-	/æ/
22.	/ɛ/	-	/æ/	37.	/ɛ/	-	/æ/
23.	/ɛ/	-	/æ/	38.	/ɛ/	-	/æ/
24.	/ɛ/	-	/æ/	39.	/ɛ/	-	/æ/
25.	/ɛ/	-	/æ/	40.	/ɛ/	-	/æ/
26.	/ɛ/	-	/æ/	41.	/ɛ/	-	/æ/
27.	/ɛ/	-	/æ/	42.	/ɛ/	-	/æ/
28.	/ɛ/	-	/æ/	43.	/ɛ/	-	/æ/
29.	/ɛ/	-	/æ/	44.	/ɛ/	-	/æ/
30.	/ɛ/	-	/æ/	45.	/ɛ/	-	/æ/
31.	/ɛ/	-	/æ/	46.	/ɛ/	-	/æ/
32.	/ɛ/	-	/æ/	47.	/ɛ/	-	/æ/
33.	/ɛ/	-	/æ/	48.	/ɛ/	-	/æ/
34.	/ɛ/	-	/æ/	49.	/ɛ/	-	/æ/
35.	/ɛ/	-	/æ/	50.	/ɛ/	-	/æ/

CLASS 3

UFSC/CCE/LLE
Name:____
NS

Perception Practice 3 – Front vowels

Instructions: Circle the best option according to the vowel sound in the word you hear.

Example: You will hear: "feet"

01. (/i/) - /ɪ/ - /ɛ/ - /æ/

You should circle option "/i/" because it sounds the same as the vowel in the word "feet".

1. /i/ - /ɪ/ - /ɛ/ - /æ/
2. /i/ - /ɪ/ - /ɛ/ - /æ/
3. /i/ - /ɪ/ - /ɛ/ - /æ/
4. /i/ - /ɪ/ - /ɛ/ - /æ/
5. /i/ - /ɪ/ - /ɛ/ - /æ/
6. /i/ - /ɪ/ - /ɛ/ - /æ/
7. /i/ - /ɪ/ - /ɛ/ - /æ/
8. /i/ - /ɪ/ - /ɛ/ - /æ/
9. /i/ - /ɪ/ - /ɛ/ - /æ/
10. /i/ - /ɪ/ - /ɛ/ - /æ/
11. /i/ - /ɪ/ - /ɛ/ - /æ/
12. /i/ - /ɪ/ - /ɛ/ - /æ/
13. /i/ - /ɪ/ - /ɛ/ - /æ/
14. /i/ - /ɪ/ - /ɛ/ - /æ/
15. /i/ - /ɪ/ - /ɛ/ - /æ/
16. /i/ - /ɪ/ - /ɛ/ - /æ/
17. /i/ - /ɪ/ - /ɛ/ - /æ/
18. /i/ - /ɪ/ - /ɛ/ - /æ/
19. /i/ - /ɪ/ - /ɛ/ - /æ/
20. /i/ - /ɪ/ - /ɛ/ - /æ/
21. /i/ - /ɪ/ - /ɛ/ - /æ/
22. /i/ - /ɪ/ - /ɛ/ - /æ/
23. /i/ - /ɪ/ - /ɛ/ - /æ/
24. /i/ - /ɪ/ - /ɛ/ - /æ/

APPENDIX M

INDIVIDUAL RESULTS AND AT HOME TRAINING TIME

Participant	Training group	pretest (total=72)	posttest (total=72)	improvement (posttest - pretest)	amount of training time
E1	SynS	43	48	5	0:33:38
E2	SynS	51	55	4	0:45:19
E3	SynS	33	48	15	0:20:15
E4	SynS	42	56	14	1:14:28
E5	SynS	49	58	9	1:34:29
E6	SynS	34	42	8	1:07:16
E7	SynS	53	54	1	1:07:09
E8	SynS	51	60	9	0:29:11
E9	SynS	45	59	14	0:56:45
E10	SynS	54	62	8	1:31:33
E11	SynS	45	54	9	0:38:31
E12	SynS	47	52	5	1:08:50
E13	SynS	43	54	11	1:13:32
E14	SynS	42	53	11	0:46:40
E15	SynS	24	46	22	0:38:14
E16	NatS	42	53	11	0:48:46
E17	NatS	35	52	17	0:59:35
E18	NatS	44	49	5	1:13:23
E19	NatS	38	52	14	1:15:17
E20	NatS	59	67	8	0:47:47
E21	NatS	53	61	8	0:44:54
E22	NatS	48	59	11	0:51:47
E23	NatS	45	52	7	1:10:08
E24	NatS	48	57	9	1:56:03
E25	NatS	26	28	2	0:47:33
E26	NatS	31	39	8	0:52:58
E27	NatS	40	48	8	1:47:09
E28	NatS	36	38	2	1:38:46
E29	NatS	37	30	-7	0:33:22

APPENDIX N

NUMBER OF MISIDENTIFICATIONS (RAW NUMBERS) IN THE PERCEPTION TESTS OF THE AMERICAN ENGLISH CONTROL GROUP

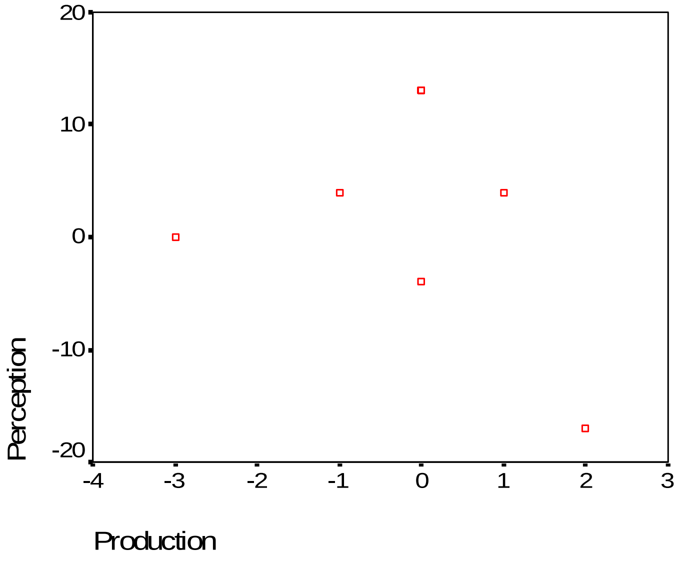
AE CONTROL GROUP:

	Targets					
perceived as	/i/ (24)	/ɪ/ (24)	/ɛ/ (24)	/æ/ (24)	/ʊ/ (20)	/u/ (24)
/i/						
/ɪ/						
/ɛ/		1				
/æ/						
/ʌ/					2	
/ʊ/						
/u/						

APPENDIX O

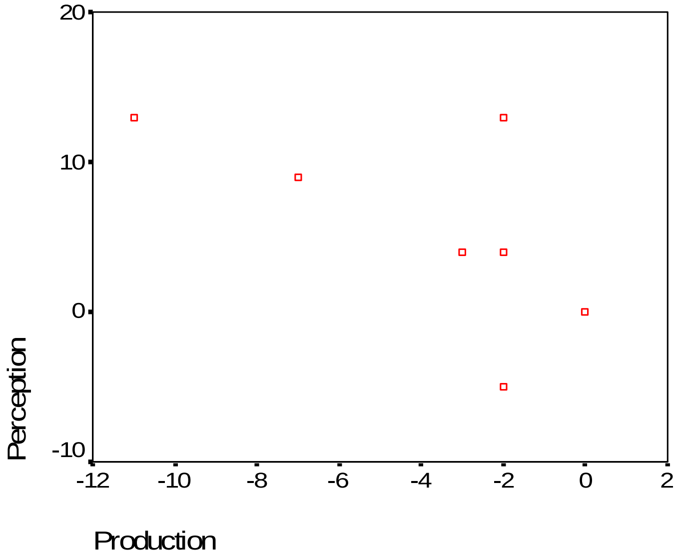
PERCEPTION vs. PRODUCTION CORRELATION GRAPHS

Control group
Pretest vs. Posttest
/i-ɪ/
Pearson = -.327
N=7, p > .474



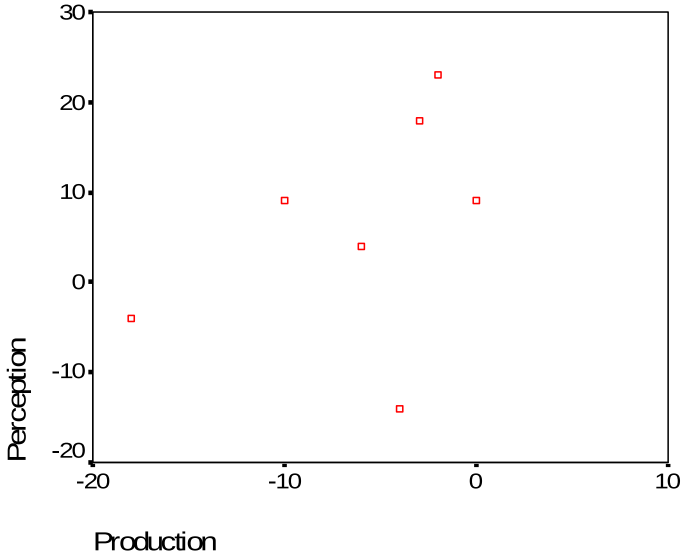
Control group
Pretest vs. Posttest
/ɛ-æ/

Pearson = -.623
N=7, p > .135



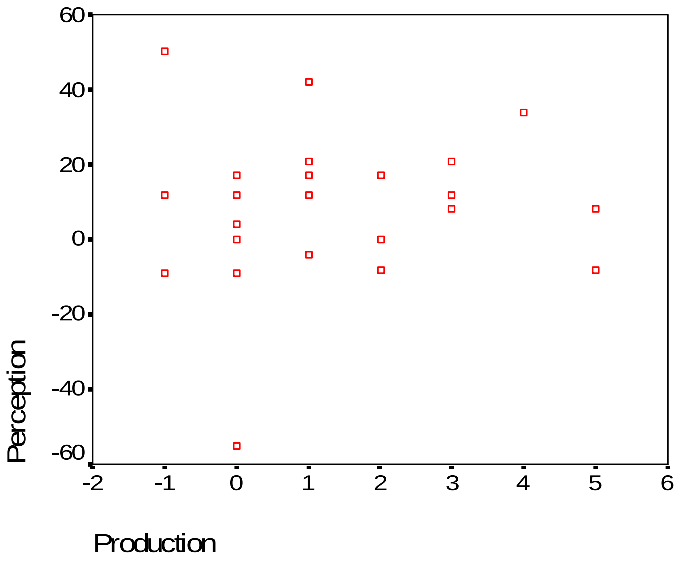
Control group
Pretest vs. Posttest
/u-ʊ/

Pearson = .411
N=7, p > .359



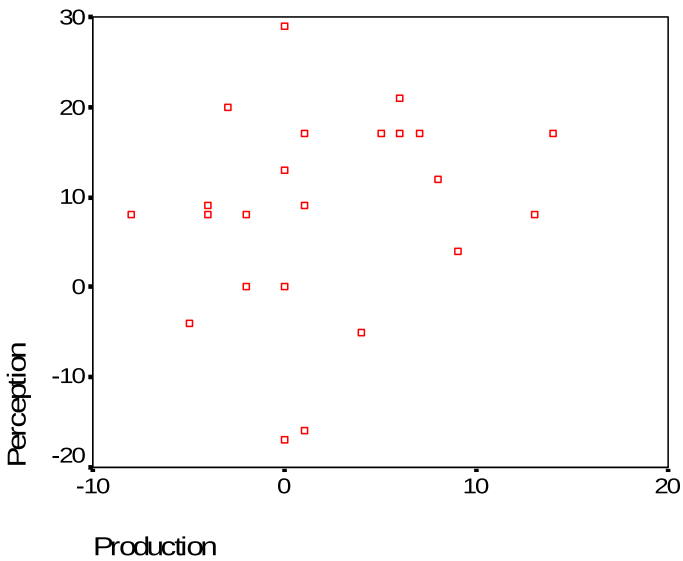
Experimental groups
Pretest vs. Posttest
/i-ɪ/

Pearson = .061
N=23, p > .782

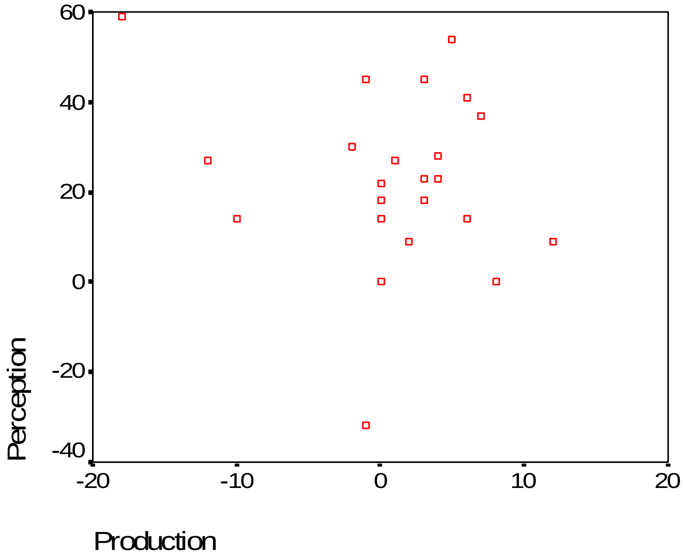


Experimental groups
Pretest vs. Posttest
/ɛ-æ/

Pearson = .221
N=23, p > .312



Experimental groups
Pretest vs. Posttest
/u-ʊ/
Pearson = -.201
N=23, p > .359



APPENDIX P

NUMBER OF MISIDENTIFICATIONS (PERCENTAGES) IN THE PERCEPTION TESTS OF THE EXPERIMENTAL GROUP

	/i/ pre	/i/ post	/ɪ/ pre	/ɪ/ post	/ɛ/ pre	/ɛ/ post	/æ/ pre	/æ/ post	/ʊ/ pre	/ʊ/ post	/u/ pre	/u/ post
E1	42	17	25	17	42	0	8	0	50	10	50	25
E2	17	0	25	0	66	42	8	8	50	10	0	0
E3	58	8	50	33	33	17	42	8	40	40	67	59
E4	8	8	25	17	42	33	8	17	100	10	50	17
E5	25	17	0	17	33	8	42	17	50	20	33	8
E6	58	25	8	8	33	42	41	33	80	30	33	16
E7	25	8	8	0	33	58	58	17	20	20	0	8
E8	8	17	0	8	42	41	0	8	70	0	50	0
E9	0	0	42	25	33	25	17	0	80	20	41	17
E10	0	0	25	8	33	17	25	17	30	10	8	17
E11	25	8	50	42	16	16	33	17	20	10	0	8
E12	50	0	25	0	33	0	8	0	10	10	33	8
E13	58	0	33	50	33	16	17	16	30	0	58	0
E14	33	0	58	0	33	16	16	17	70	30	75	25
E15	58	17	33	8	58	42	50	50	70	40	66	25
E16	42	8	8	8	42	33	42	0	40	20	42	17
E17	17	0	8	0	42	50	42	8	30	30	75	58
E18	75	25	17	42	17	33	33	0	50	40	25	8
E19	8	8	16	16	33	17	58	25	40	30	33	17
E20	0	0	0	0	8	8	8	25	30	0	8	0
E21	0	0	0	0	0	8	50	33	60	20	33	8
E22	25	25	25	0	33	17	8	0	50	20	25	33
E23	25	8	0	8	33	50	25	0	60	8	58	17
E24	8	0	25	8	8	16	0	25	50	50	75	17
E25	67	42	34	42	42	33	33	0	40	40	75	25
E26	8	16	66	33	42	50	25	25	60	40	33	25
E27	58	0	25	33	41	33	58	17	40	30	33	25
E28	33	10	0	75	50	25	33	50	50	30	75	0
E29	42	50	42	33	42	33	25	25	80	50	50	83